Самарский государственный аэрокосмический
университет имени академика С.П. Королева
(национальный исследовательский университет)

ПИТ-2010

Перспективные информационные технологии для авиации и космоса



Труды
Международной конференции с элементами научной школы для молодежи

СГАУ

САМАРА

Министерство образования и науки Российской Федерации

Государственное образовательное учреждение высшего профессионального образования «Самарский государственный аэрокосмический университет имени академика С.П. Королева» (национальный исследовательский университет)

ПЕРСПЕКТИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ АВИАЦИИ И КОСМОСА (ПИТ-2010)

ТРУДЫ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ
С ЭЛЕМЕНТАМИ НАУЧНОЙ ШКОЛЫ ДЛЯ МОЛОДЕЖИ

29 сентября – 1 октября 2010 г.

Самара
2010

УДК 004:001.8(06) + Ч 480.253я 54

Перспективные информационные технологии для авиации и космоса (ПИТ-2010). Труды Международной конференции с элементами научной школы для молодежи. – Самара, 2010. – 977 с.

Труды конференция содержат отобранные Программным комитетом материалы работы секций, посвященные обсуждению проблем фундаментальных и прикладных исследований, компьютерному моделированию, разработке и практическому внедрению компонентов информационно-телекоммуникационных систем в авиации и космонавтике, научно-практической деятельности исследователей в этом направлении, обмену опытом учебно-образовательной деятельности при подготовке ИТ-специалистов с использованием инновационных образовательных и инструментальных технологий.

Издание осуществлено при финансовой поддержке Министерства образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы

ISBN 978-5-7883-0851-7

© Самарский государственный аэрокосмический университет, 2010

Секция 1

АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ

Подсекция 1
«Информационные технологии в научных исследованиях»

SCAN: Software package of spectral-correlation analysis

Sergey A. Prokhorov, Ilona M. Kulikovskikh
Samara State Aerospace University, Samara

Introduction

Today there is a great number of present-day mathematical systems of statistical data processing including both standard functions of numerical data processing and means of obtaining analytical expressions of functional characteristics.

Navigation in the area of present-day mathematical systems is difficult. While practically everybody is familiar with such systems as Mathcad, Matlab, R specialized statistical systems are often unknown for the wide circle of researchers. However the latter are most suitable for solving narrowly specialized problems; besides they are developed for solving concrete problems, for example, in the area of applied physics subject to approved techniques of execution of the complete research cycle, therefore the loss of universality is accompanied by an improvement of the quality of data processing, including the increase of its efficiency.

Statistical general-purpose packages are notable for the absence of straightforward orientation to a specific subject area, broad range of statistical methods, friendly user interface. Specialized packages usually realize the methods used in the concrete subject area; they include a sufficiently complete set of conventional methods, and also unique methods and algorithms created by developers. Their use is advisable when systematic solving problems of a narrow subject area is required. The **SCAN** belongs to such specialized packages.

The SCAN

The problem statement

Many statistical systems make it possible to perform correlation-spectral analysis, they maintain discrete Fourier transform, such systems as Matlab having a powerful enough tool to perform fast Fourier transform. However the problem of definition of the spectral power density is solved numerically in the general case and, as a result, the processing speed problem remains relevant for vast data arrays.

Various subproblems of approximative analysis of the correlation-spectral characteristics of stochastic processes can be solved using general-purpose and specialized systems, however, it is impossible to solve such a problem in corpore. The universality of the known systems requires additional modification, realization of original algorithms for solving the complex scientific and research problems. Therefore it is necessary either to add package for the known mathematical systems, or to develop their own software package with the help of any high level language.

The **SCAN** represents the own developed software package of approximative correlation-spectral analysis of stochastic processes, combining the analytical and numerical methods.

Analytical solutions possess a number of advantages including the opportunity of research of the influence of physical parameters, initial and final conditions on the character of the solution. The results of analytical solutions promote the development of adequate mathematical models, they are more informative, robust, possess the opportunity of calculating a value at any point with a given accuracy, without resorting to calculation of values at other points. The disadvantage of such solutions is that in practice processed data are accepted as series, especially it concerns stochastic processes, therefore obtaining analytical expressions involves computational and material outlays. Numerical decisions are universal and are applied when an analytical solution is impossible, while high efficiency of modern computer techniques compensate for their low speed. However the emergence of various instabilities, the complexity using the calculation results, the accumulation of rounding-off errors reduce the value of numerical expressions substantially.

The combining of the mentioned methods is an obvious and effective step in the analysis of large data arrays including stochastic processes. It allows increasing the efficiency and integrating the advantages of both methods eliminating part of the drawbacks. There are approaches based on the analytical description of digital data with their subsequent processing. Thus the following requirements are imposed on a similar solution: providing the given accuracy of the description with a simpler analytical expression, adaptibility of the analytical description to the features of each signal, the unified nature of description structure of a signal, the opportunity of realization of the method in the absence of a priori information on the signal.

Approximative approach is presented primary based on the numerical analysis of informational data arrays, stochastic processes, consisting in the definition of functional characteristics with their further analytical processing. Thus it is possible to avoid essential mistakes when choosing an expression model, a method and an algorithm of approximation, to receive a priori information on the processes under investigation and finally to remove or to reduce the importance of observing the above mentioned restrictions.

Methods of approximative analysis

The essence of the proposed approximative approach underlying the software package **SCAN** is a follows. Let's present an orthogonal model of a correlation function the expansion into a Fourier series [1]

$$K_a(\tau) = \sum_{k=0}^{\infty} \beta_k \psi_k(\tau, \gamma).$$

However it is impossible to realize infinity when calculating on a computer, consequently, it is necessary to be limited to a final number of series terms and instead of the function $K_a(\tau)$ to use its estimation $\hat{K}_a(\tau)$ with number of members m

$$\hat{K}_a(\tau) = \sum_{k=0}^{m} \beta_k \psi_k(\tau, \gamma)$$

where

$$\beta_k = \frac{1}{\|\psi_k\|^2} \int_0^\infty K_x(\tau)\psi_k(\tau,\gamma)\mu(\tau,\gamma)d\tau \quad (1)$$

are coefficients of the orthogonal decomposition, $F_x(\tau)$ is an analyzed correlation function, and $\psi_k(\tau,\gamma)$ are orthogonal functions with the scale parameter γ and the weight function $\mu(\tau,\gamma)$.

The problem of the definition of the Fourier coefficients (1) is solved subject to minimum of an approximation error square

$$\Delta = \int_0^\infty \left[K_x(\tau) - \sum_{k=0}^{m} \beta_k \psi_k(\tau,\gamma) \right]^2 \mu(\tau,\gamma)d\tau = min. \quad (2)$$

Having defined the coefficients (1), it is necessary to solve the problem of defining scale parameter γ subject to (2). There are several algorithms of scale parameter definition [1, 2]. Having defined m from the equation (2) we will recalculate the decomposition coefficients β_k subject to the basic property of the analyzed functional characteristic. For example, the basic property of correlation function $\hat{K}_a(\tau)$ is

$$\hat{K}_a(0) = \sum_{k=0}^{m} \beta_k \psi_k(0,\gamma) \neq \sigma_x^2$$

where σ_x^2 is a dispersion of the analysed stochastic process.

Let's determine coefficients b_k in order to realize the basic property [1, 2]

$$\hat{K}_a(0) = \sum_{k=0}^{m} b_k \psi_k(0,\gamma) = \sigma_x^2 \quad (3)$$

where b_k is defined from the following equation

$$\Delta_1 = \int_0^\infty \left[K_x(\tau) - \sum_{k=0}^{m} b_k \psi_k(\tau,\gamma) \right]^2 \mu(\tau,\gamma)d\tau + \lambda \sum_{k=0}^{m} b_k \psi_k(0,\gamma) = min.$$

Let's determine the partial derivatives and set them equal to zero. Hence

$$b_n = \beta_n + \frac{\left[\sigma_x^2 - \sum_{k=0}^{m} \beta_k \psi_k(0,\gamma) \right]}{\sum_{k=0}^{m} \frac{1}{\|\psi_k\|^2}} \frac{\psi_n(0,\gamma)}{\|\psi_n\|^2} = \beta_n + C_n.$$

Then we will estimate the spectral power density by the approximating expression of the correlation function (3)

$$\hat{S}_a(\omega) = \frac{1}{\pi} \sum_{k=0}^{m} b_k \mathrm{Re} W_k(\mathrm{j}\omega)$$

where $\mathrm{Re} W_k(\mathrm{j}\omega)$ is the real part of a orthogonal function Fourier transform.

The parameters of the orthogonal model for other characteristics are defined similarly. It is necessary to note that the orthogonal bases of Jacobi and Sonin-Laguerre represented by ten various orthogonal functions are realized in the **SCAN**.

The SCAN functions

The software package **SCAN** is developed with the help of a high level language Borland Delphi v.7.0. It includes the following functions:

- generation of a stochastic process with the required type of a correlation function;

- setting of recommended parameters during the generation of a stochastic process (e.g. interval discretization, number of ordinates);
- filtration of a generated stochastic process and construction of a correlation function and a phase portrait;
- item construction of a cross correlation function;
- approximation of a cross (auto -) correlation function by means of orthogonal functions and obtaining generalized correlation characteristics;
- construction of an orthogonal model of spectral power density, its components and spectral function by approximating analytical expression of a cross (auto -) correlation function and obtaing generalized spectral characteristics;
- approximation of spectral power density components (e.g. the real part, the imaginary part, modulus) by means of orthogonal functions and obtaining generalized spectral characteristics;
- construction of an orthogonal model of a spectral function by approximating analytical expression of a spectral power density modulus and obtaing generalized spectral characteristics;
- construction of an orthogonal model of a cross (auto -) correlation function by approximating analytical expression of spectral power density of the real and imiginary parts and obtaing generalized correlation characteristics;
- construction of a specialized signal (e.g. delta-shaped, rectangular, low-frequency signal etc.);
- investigation of the Jacobi and Sonin-Laguerre orthogonal polynomial and functions and their properties;
- simulation for the checking of the adequacy of the algorithms developed;
- data input/output from files of required format;
- export of processed data and results of simulation into the MS Office (Word, Excel).

Application of the SCAN

Let's present an example of processing magnetic fields data (level 2 data) on ACE Satellite [3] with the help of the **SCAN**. The ACE Satellite serves as Earth's sentry for Solar generated particles and magnetic waves. It is the primary source for the measurement of solar wind speed. The ACE Satellite is positioned at a Lagrangian L1 point. This places ACE about 1,5 million km from Earth and 148,5 million km from the Sun. The space craft is approximately 1/100 of the distance from the Earth to the Sun. The satellite travels in an elliptical orbit in gravitational equilibrium between the Sun and Earth. This orbit gives it an excellent view of the Sun and a range of deep space beyond.

The magnetic field experiment on ACE provides continuous measurements of the local magnetic field in the interplanetary medium. These measurements are essential in the interpretation of simultaneous ACE observations of energetic and thermal particles distributions. The experiment consists of a pair of twin, boom-mounted, triaxial fluxgate sensors which are located 165 inches (= 4,19 meters) from the center of the spacecraft on opposing solar panels. The electronics and digital processing unit (DPU) is mounted on the top deck of the spacecraft. The two triaxial sensors provide a balanced, fully redundant vector instrument and permit some enhanced assessment of the spacecraft's magnetic field [4].

The data set was created at NSSDC after the downloading from the ACE Science Center of 4-min IMF data (from MAG instrument, PI: N.F. Ness) [5].

In Fig. 1 the form view with the correlation functions change of magnetic fields data since 1998 to 2010 in January is presented.

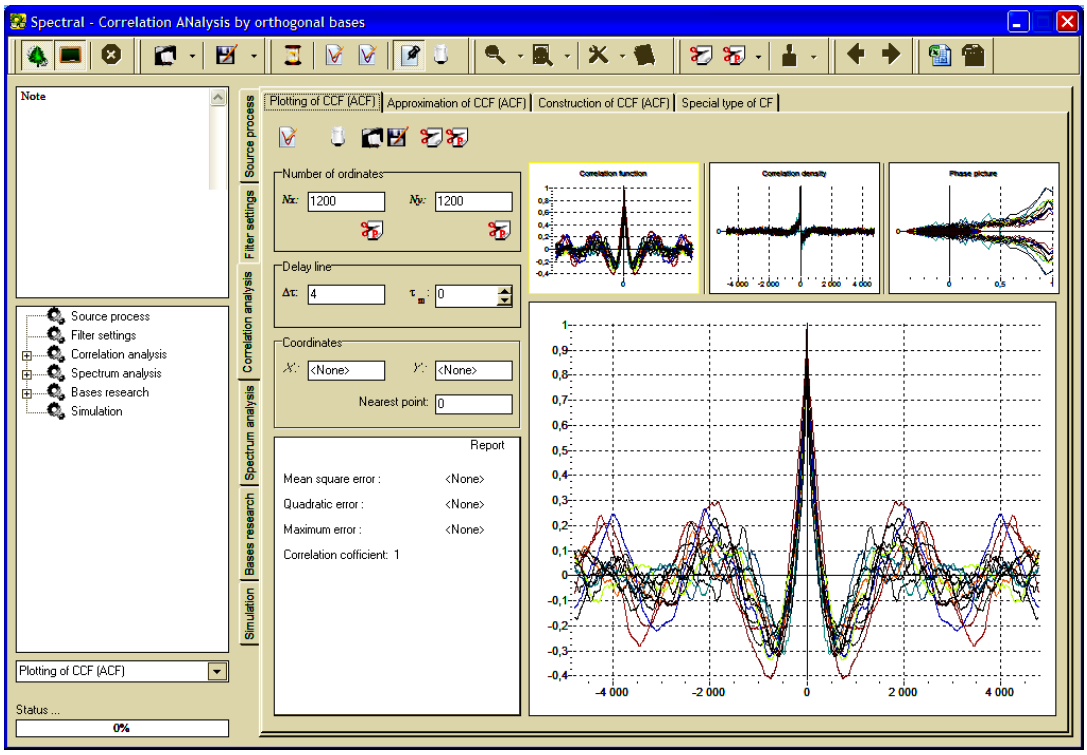


Fig. 1. Correlation functions change of magnetic fields data since 1998 to 2010 in January

In Fig. 2 the form view with are spectrum power densities on constructed correlation functions.

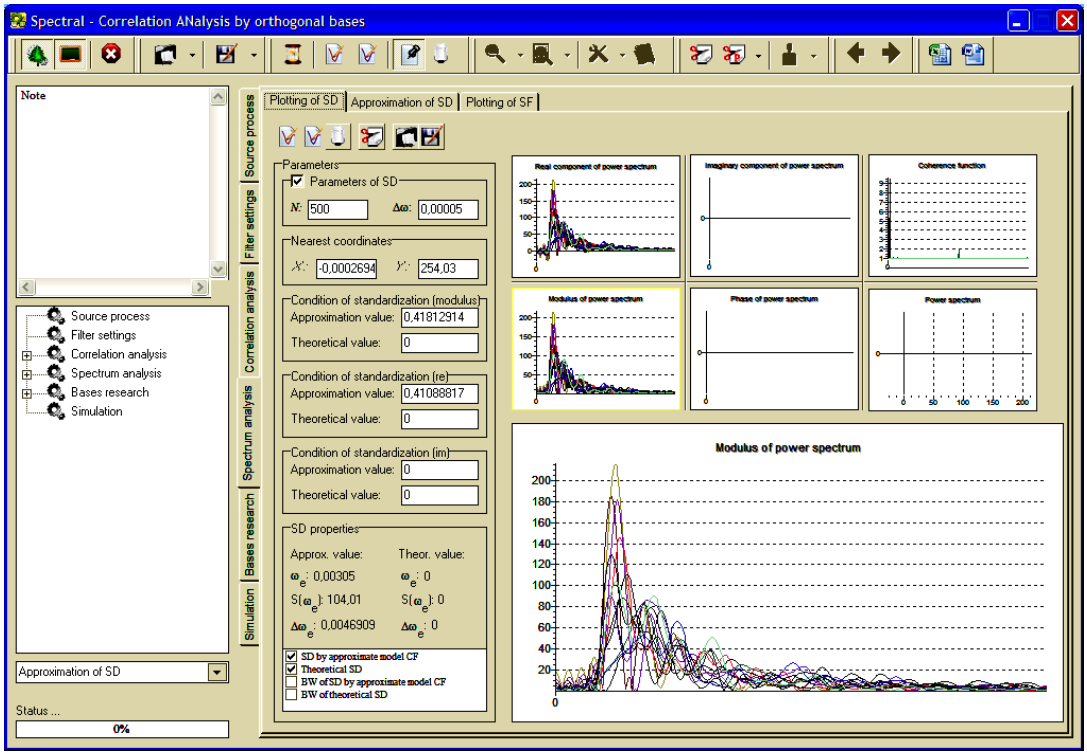


Fig. 2. Spectrum power densities change of magnetic fields data since 1998 to 2010 in January

Conclusions

The proposed software package **SCAN** includes subpackages connected with the estimation of various correlation and spectral characteristics such as correlation windows, a correlation function and a cross correlation function, a phase portrait used when solving

identification problems, spectral power density and its components, a spectral function and generalized spectral characteristics and others. Besides the software package provides a research subpackages including a subpackage of simulation, a subpackage of research into the properties of the orthogonal functions used when constructing models of correlation and spectral characteristics.

Finally the **SCAN** is used when training experts on the speciality 23.01.02 “Automated systems of information processing and control”.

References

1. Applied analysis of stochastic processes [in Russian]. Edited by Sergey A. Prokhorov. Samara Scientific Center of Russian Academy of Sciences, 2007, 582 p.
2. Sergey A. Prokhorov, Ilona M. Kulikovskikh The orthogonal models of correlation-spectral analysis. Laboratory work [in Russian]. Samara Scientific Center of Russian Academy of Sciences, 2008, 301 p.
3. Charles W. Smith, Jacques L'Heureux and Norman F. Ness, Mario H. Acuna, Leonard F. Burlaga and John Scheifele The ACE Magnetic Fields Experiment, 1998, 23 p.
4. www.kc4cop.bizland.com/ace_satellite.htm.
5. www.srl.caltech.edu/ACE/ASC/level2/index.html.

Пропаганда открытых кодов

Е.В. Авдеев
Тольяттинский филиал Самарского государственного аэрокосмического университета имени академика С.П. Королева (национальный исследовательский университет), Тольятти

CAE (англ. Computer-aided engineering) — общее название для программ или программных пакетов, предназначенных для инженерных расчётов, анализа и симуляции физических процессов.

В работе выполнен выбор среди бесплатных CAE пакетов с целью использования открытых реализаций моделирования мультифизических процессов.

OEM версию Ansys Muliphisics можно скачать всего за $30, цена же корпоративной версии на год может достигать нескольких тысяч евро.

Salome — это открытая интегральная программная платформа для выполнения численных расчётов. Это конечно-элементный пре-процессор, но он является ядром вычислительной системы, вокруг которого в настоящее время обхединяется множество CAE решателей. Кроме того, Salome позволяет разрабатывать собственные программные решения.

Продукты марки Salome распространяются на условиях GNU Lesser General Public License.

Salome используется как база для проекта NURESIM, который предназначен для полномасштабного моделирования ядерных реакторов.

Причинами малого использования Salome являются: малый объём документации на русском языке, более сложный интерфейс пользователя (по сравнению с платными аналогами).

Задачей работы является пропаганда использования открытых реализаций CAE пакетов, создание условий для упрощения и доступности использования.

В ходе работы произведена установка пакета SALOME на GNU/Debian и приведено несколько примеров использования.

Информационная технология сравнительной оценки доказательств

К.А. Арутюнян, А.А. Ильичев
Самарский государственный архитектурно-строительный университет, Самара

В процессе функционирования современного общества все чаще возникают ситуации, когда имеется некольно мнений по определенному вопросы, каждое из которых абсолютно правильно подкреплено своими доводами и доказательствами. В такой ситуации становится понятной необходимость создания инструмента, который смог бы сопоставить и оценить доводы, выдвигаемые сторонниками той или иной точки зрения. Наиболее универсальным инструментарием, способным справиться с поставленной задачей обладает лишь математика. Очевидно, что внедрение математического понятийного аппарата в какие-либо прикладные области невозможно без использования информационных технологий. Частным случаем данной проблемы является внедрение информационных технологий в область юриспруденции, конкретнее — в область судебного процесса, с целью оценки доводов и доказательств, представляемых стороной защиты и стороной обвинения.

Хотелось бы подчеркнуть некоторые важные аспекты прикладной области. На сегодняшний день математика все в большей степени становится необходимым атрибутом юридической науки. Это объясняется тем, что правовые системы, явления и процессы, а это, прежде всего, механизмы правотворчества, правового регулирования, законности, борьбы с преступностью, наряду с качественными свойствами, такими как структурностью, целостностью, устойчивостью обладают и количественной мерой (количеством норм, связей, интенсивностью потоков информации, степенью развития, целенаправленности и т.д.). Кроме того, внедрение математики в сферу права также можно объяснить тем, что в юридических науках в связи с правовой информатизацией общества, созданием информационных комплексов и систем в области права и решением на компьютерах юридических задач возникло значительное число проблем, связанных с оптимизацией функционирования правовых систем, юридических органов и процессов. Эти проблемы не могут быть решены без привлечения разнообразных математических методов, так как сущность оптимизации в этом случае состоит в разработке формализованных способов достижения целей функционирования систем с наименьшими затратами материальных средств, времени в решении информационных, логических и математических задач.

Среди целей данной научной работы является использование математического метода многокритериального принятия решений ПРИНН для весовой оценки аргументов стороны защиты и стороны государственного обвинения- прокуратуры в судебных прениях при рассмотрении уголовного дела. Для примера было рассмотрено реальное уголовное дело, опубликованное в Бюллетене Верховного Суда за сентябрь 2009 года. Далее было проведено статистическое и аналитическое исследование на предмет совпадений решений судей с точкой зрения математического метода принятия решений, в результате которого были получены ошеломительные результаты.

Следует отметить необходимость модернизации метода ПРИНН, возникающую в процессе приспособления метода для оценки агументации сторон защиты и обвинения в судебном процессе. В настоящее время назрела задача его модернизации путем замены базовой линейной свертки на свертку Гермейера, что позволит устранить общий для большинства методов принятия многокритериальных решений недостаток, связанный с некорректностью линейной свертки. В числе целей данной работы был вывод аналитических соотношений, позволяющих рассчитывать в методе ПРИНН Н-обобщенные потери, как при линейной свертке, так и при свертке Гермейера, и на этой

основе провести сопоставление результатов. Поскольку переход к свертке Гермейера потребует в методе ПРИНН численного интегрирования, также, цель работы состояла в том, чтобы на основе сравнения с полученным точным решением оценить сходимость метода Монте-Карло при вычислении специфических интегралов метода ПРИНН.

В данной работе перед авторами был поставлен ряд следующих задач:

1. Разработать аналитические формулы для расчета Н-обобщенных потерь в методе ПРИНН для типовых способов учета неопределенности при линейной базовой свертке и свертке Гермейера и провести численное сопоставление результатов. Эта цель представляет научный интерес, поскольку на ее основе может быть выполнена модификация метода ПРИНН путем отказа от некорректной линейной свертки и перехода к корректной свертке Гермейера.

2. Исследовать сходимость метода Монте-Карло при вычислении специфических интегралов метода ПРИНН. Эта цель представляет практический интерес, поскольку помогает решить вопрос о выборе численного метода интегрирования в модифицированном методе ПРИНН, в котором, ввиду сложности свертки Гермейера, невозможно аналитическое вычисление интегралов.

Достижение указанных целей потребовало решения следующих задач:

1. Изучить основы теории принятия решений и, в частности, метод ПРИНН
2. Вывести аналитические выражения для Н-обобщенных потерь.
3. Проверить их правильность сопоставительными расчетами на всем диапазоне изменения аргументов.
4. Разработать программный комплекс, реализующий указанные расчеты.
5. Запрограммировать метод Монте-Карло, провести множественные расчеты и сопоставить результаты численного интегрирования этим методом с точными результатами, полученными аналитически.

Также, в рамках данной работы осуществлен вывод аналитических зависимостей, предназначенных для использования в методе ПРИНН для многокритериальной оценки. Выведены аналитические зависимости для расчета Н-обобщенных потерь при пяти способах учета неопределенностей:

- среднем;
- оптимистическом;
- релейном;
- нивелирующем;
- осторожном.

Рассмотрены различные базовые свертки критериев, линейной и Гермейера, и два различных типов отношений между критериями: критерии равнозначны или один из них более значим, чем другой.

Исследована корреляция результатов при переходе от одной базовой свертки к другой. Исследование показало, что результаты тесно коррелируют, в пользу свертки Гермейера.

Разработана Программная Система, обеспечивающая вычисление Н-обобщенных потерь при всех указанных выше исходных установках.

Проведено исследование целесообразности использования метода Монте-Карло, при вычислении многомерных интегралов в методе ПРИНН, что необходимо, больше чем два критериев. Сравнение результатов численного решения по методу Монте-Карло, с точными расчетами при двух критериях, показало, что этот метод может использоваться, причем, внимание должно быть обращено на качество датчика случайных чисел.

Разработанная информационная технология базируется на смешанных методах преобразования информации: как компьютерных, так и на бумажных носителях. В рамках

реализации технологии создана математическая модель судебного заседания, которая оперирует четырьмя ключевыми таблицами - две таблицы используются для вычисления достоверности доказательств в уголовном дела, а в конечном счете и самих обстоятельств дела. Еще две таблицы используются для установления статьям Уголовного Кодекса Российской Федерации соответствия квалифицирующих признаков преступления и соответствия между квалифицирующими признаками преступления и обстоятельств уголовного дела.

Для расчета достоверности обстоятельств дела создается таблица с одинаковым числом колонок и столбцов. В данную таблицу заносятся обстоятельства дела, причем они транспонируются. Далее производится расстановка флагов влияния обстоятельств друг на друга, по системе: 1-, 0- не влияет. На пересечении одних и тех же обстоятельств, соответственно, будет значение 0, поскольку обстоятельство не может влиять само на себя. Таким образом, если одно обстоятельство влияет на другое, соответственно на их пересечении будет значение 1. Заполнив всю матрицу отношений обстоятельств, можно перейти к расчету достоверности.

На основе формально-логических суждений была выведена формула для расчета достоверности, при условии, что нет прямых доказательств, поскольку их достоверность по умолчанию равна максимально возможной. Философию формулы

$$W = \frac{k * p}{\sum_{i=1}^{n} \varphi_i}$$

можно объяснить следующим образом. Коэффициент k- степень влияния обстоятельства на любые обстоятельства. Поскольку, нельзя ставить экспертную оценку со свидетельскими показаниями на одну планку по достоверности, поэтому введен данный коэффициент. Он позволяет наиболее правильно согласовать эти значения с матрицей отношений обстоятельств. Φ_i – флаг влияния на обстоятельство. При расчете производится суммирование всех флагов влияния обстоятельств, которые влияют на обстоятельство. Коэффициент p- достоверность обстоятельства на предыдущем этапе, на первом этапе, все обстоятельства имеют максимальную степень достоверности, и в процессе согласования достоверности тех обстоятельств на которые влияют другие будут снижаться пропорционально степени влияния влияющих на него обстоятельств. К примеру, у нас есть обстоятельство, обладающее степенью влияния 0,8 и на него влияют три обстоятельства. В итоге получится, что на первой итерации достоверность этого обстоятельства была равна 1, на второй – 0,4, на третьей – 0,16. Трех итераций достаточно для того чтобы получить достаточную разность между достоверностями обстоятельств, чтобы их успешно упорядочить в порядке убывания достоверности.

После расчета достоверности обстоятельств производится расчет доказанности квалифицирующих признаков. Формируется таблица квалифицирующие признаки – обстоятельства дела. Той же системой флагов отмечается какие обстоятельства дела содержат в себе какие-либо квалифицирующие признаки, после чего производится расчет доказанности этих признаков, поскольку имеются значения достоверностей обстоятельств, полученные на предыдущем этапе. На основе логических суждений было получено, что целесообразно вычислять доказанность признаков путем сложения произведений векторов W_i и C_i, где W_i – вектор достоверностей обстоятельств, а вектор C_i - вектор флагов наличия квалифицирующих признаков в обстоятельствах. Таким образом, рассчитывается доказанность квалифицирующих признаков. Остается лишь произвести связь квалифицирующих признаков и статей УК РФ. Это представляется очень простой задачей, если использовать тот же подход, что и в двух предыдущих случаях.

В итоге, на основе данных, полученных на первом уровне модели, производится расчет при помощи метода ПРИНН. Формируется таблица, среди альтернатив которой перечисляются статьи УК РФ, а среди критериев – квалифицирующие признаки. Для метода ПРИНН этих данных достаточно чтобы произвести расчет. Производится оптимизация, и в результате, получаются эффективности альтернатив, в нашем частном случае – это вероятности использования в приговоре суда данной альтернативы- нормы УК РФ.

На основе полученных данных можно утверждать, что математический метод ПРИНН, абсолютно полностью справляется с поставленной задачей. Разработанный математический аппарат и программные средства позволяют наиболее правдоподобно моделировать деятельность судьи с учетом мельчайших философских подробностей предметной области. Также, при исследовании на большом объеме данных, возможно получение информации о новых правовых явлениях и процессах.

Действительно, математический метод и математическая логика позволяют выявить не только количественные показатели, но и структурированность, параметризацию исследуемого с их помощью явления, но и использовать фундаментальные свойства математики в анализе закономерностей, зависимостей, в том числе путем построения функций и системы координат, учёте динамики развития. Это позволяет проводить группирование эмпирической базы, прогнозирование результатов, основанных на выявлении функциональных закономерностей правовой системы.

Таким образом, проведенное исследование можно назвать результативным.

К задаче оптимального выбора элементов псевдослучайной последовательности

М.С. Воронов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Современная информатика широко использует псевдослучайные числа в самых разных направлениях – от метода Монте-Карло и имитационного моделирования до криптографии. При этом от качества последовательности псевдослучайных чисел напрямую зависит качество получаемых результатов. А от результатов зачастую (особенно в криптографии) зависит очень многое, это обстоятельство иллюстрирует известный афоризм Роберта Кавью: «Генерация случайных чисел слишком важна, чтобы оставлять её на волю случая». В данной статье рассматривается новый способ получения последовательности псевдослучайных чисел без циклов.

По определению, числа $x_1, x_2, \dots, x_n \in [0,1]$ называются псевдослучайными, если $\ln(f) = \frac{1}{n}\sum_{i=1}^{n} f(x_i) - \int_0^1 f(x)dx \to 0 \ (n \to 0)$ для любой непрерывной на отрезке [0,1] функции f.

Вычислим норму функционала ln f на линейном многообразии C[0,1] с полунормой $\|f\|_p^{(1)} = (\int_0^1 |f^{(1)}(x)|^p dx)^{\frac{1}{p}}, \ 0 < p < \infty$, подставив $f(x) = \int_x^1 \varphi(t)dt$. Получим $\|ln\|_{p'} = (\int_0^1 \left|x - \frac{1}{n}\sum_{i=1}^{n} \theta(x - x_i)\right|^{p'} dx)^{\frac{1}{p'}}, p' = \frac{p}{p-1}$.

Предположим, что у нас уже сгенерирована последовательность $x_1, x_2, \dots, x_{n-1} \in [0,1]$ из n-1 псевдослучайного числа. Число x_n найдём с помощью минимизации нормы $\|ln\|_{p'}$ по переменной x_n и условия $\frac{\partial}{\partial x_n}\|ln\|_{p'}^{p'} = 0$. Оно будет равно $x_n = \frac{1}{2n} + \frac{\sum_{i=1}^{n-1}\theta(x_n - x_i)}{n}$.

Из данного равенства следует важный вывод о том, что число x_n содержится среди чисел вида $\frac{2k-1}{2n}, k = 1,2, \dots, n$, т.к. в этом случае значение $\|ln\|_{p'}^{p'}$ является наименьшим. При p=2 можно получить явную формулу: $\|ln\|_2^2 = \int_0^1 (x - \frac{1}{n}\sum_{i=1}^{n} \theta(x - x_i))^2 dx = \frac{1}{3} + \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n^2}\sum_{1 \le i < j \le n} |x_i - x_j|$, откуда x_n можно определить минимизацией функции $(x - \frac{1}{2})^2 - \frac{1}{n}(|x - x_1| + \dots + |x - x_{n-1}|)$ на наборе $\frac{2k-1}{2n}, k = 1,2, \dots, n$.

Таким образом, в ходе данной работы была рассмотрена задача оптимального выбора элементов последовательности псевдослучайных чисел, предложен новый способ её решения, на основе которого был разработан генератор псевдослучайных чисел.

Перспективы применения программного комплекса аппроксимативного корреляционно-спектрального анализа для обработки результатов испытаний камер сгорания ГТД

Я.В. Газетова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

При решении различных прикладных задач нередко возникает необходимость из небольшого объема данных выявлять максимум необходимой информации, снижать временные и ресурсные затраты на их обработку. Данным требованиям удовлетворяет программный комплекс аппроксимативного корреляционно-спектрального анализа в ортогональном базисе Бесселя. Суть метода, лежащего в основе работы комплекса, заключается в получении моделей корреляционных и спектральных характеристик, необходимых для определения диагностических параметров различных динамических систем в промышленности, в данном случае, камер сгорания ГТД. Алгоритмы, положенные в его основу, являются уникальными и постоянно совершенствуются, что повышает конкурентоспособность данного проекта. Предлагаемый программный комплекс содержит ряд подсистем, связанных с определением корреляционных и спектральных характеристик, по которым специалисты предметной области смогут анализировать исходные данные.

Описание предметной области

Функцией камеры сгорания (КС) как модуля газотурбинного двигателя является подвод тепла к рабочему телу путем сжигания топлива.

В данной работе рассматривается топливная система камеры сгорания ГТД семейства «НК» с выносными жаровыми трубами состоит из пяти контуров, один для подачи топлива в дежурную зону и, одновременно, во все зоны, и четыре для подачи топлива в основную зону камеры.

Запуск двигателя производится по отработанной схеме подачи топлива в дежурный контур камеры. После завершения запуска включается в работу АСУ КС. При увеличении режима работы двигателя расход топлива в дежурный контур регулируется таким образом, чтобы температура пламени в дежурной зоне увеличивалась и до определенного режима поддерживалась постоянной. Затем в заданном диапазоне температура пламени в дежурной зоне линейно уменьшается и вновь поддерживается постоянной на этом уровне.

После завершения запуска одновременно включается подача топлива в первый основной контур и открывается перепуск из зоны горения в выходную зону. Расход топлива повышается до достижения оптимальной температуры [2].

Применение термопары для измерения поля температур камеры сгорания ГТД

Описание термопары

Под термопарой принято понимать два проводника из разнородных металлов (термоэлектроды), соединенных на одном конце и образующих часть устройства, использующего термоэлектрический эффект для измерения температуры. Тип термопары в данном случае – ТХА (хромель-алюмель). На испытательных стендах невыгодно тянуть длинные и достаточно дорогие термопарные провода от точки измерения температур до измерительных цепей. В этом случае для удлинения термопары и соответственно удаления холодного спая от места нагрева (двигателя) широко применяют так называемые компенсационные провода [2].

Схема подключения термопар к измерительному прибору приведена на рис. 1.



Рис. 1. Цепь термопары с компенсационными проводниками

Характеристики термопары

Погрешность измерения температуры

Погрешность измерения температуры термопарой ТХА включает в себя следующие составляющие:

погрешность термопары ТХА (2 класс допуска).

Термопары хромель-алюмель предназначены для измерения температуры при длительном применении от 0 до 1000°С (как в данном случае) и при кратковременном — до 1300°С.

В таблице 13.2 учебника [2] предел допускаемых отклонений термоЭДС от номинальных статических характеристик преобразования в диапазоне измерений от 333°С до 1300°С составляет 0,0075t. Для значения 1000°С:

$$\delta_T = 0,0075 \cdot 1000 = 7,5^0 C,$$

а относительная погрешность составляет 0,75%.

погрешность компенсационного провода и проводящих линий – при правильном выполнении соединения, в соответствии с технологией, этой погрешностью можно пренебречь.

погрешность модуля преобразования LC-227 – в соответствии с паспортными данными LC-227, данная погрешность $\delta_П$ составляет 0,2%.

Таким образом, суммарная погрешность определения температуры

$$\delta_\Sigma = \delta_T^2 + \delta_П^2 = \sqrt{(0,75)^2 + (0,2)^2} = 0,77\%.$$

Для значения 1000°С δ_Σ составляет $9,7^0 C$.

Инерционность термопары

Инерционность - это время, за которое температура рабочего спая станет равна температуре среды, с момента помещения в неё термопары. Инерционность в данном случае определяется массой горячего спая (диаметр горячего спая при «склеивании» двух проводников диаметром 0,7мм равен 1 мм).

Все типы хромель-алюмелевых термопар характеризуются большой инерционностью, что вносит дополнительные погрешности в измерение температуры.

Для устранения влияния инерционности термопары применяются меры:

уменьшение диаметра горячего спая;

установка выдержки 18-20с при переходе в соседнюю позицию.

Время измерения температуры

Время измерения поля температуры камеры сгорания ГТД определяется временем перемещения из одной позиции измерения в другую ($\approx 2c$), временем выдержки ($\approx 18-20c$) и временем многократного опроса измерительных каналов, т.е. непосредственно съема данных. Заметим, в каждой позиции снимается среднее значение, т.е. динамика не отслеживается.

Распределение временных затрат в процессе измерения температуры показано на рис. 2.

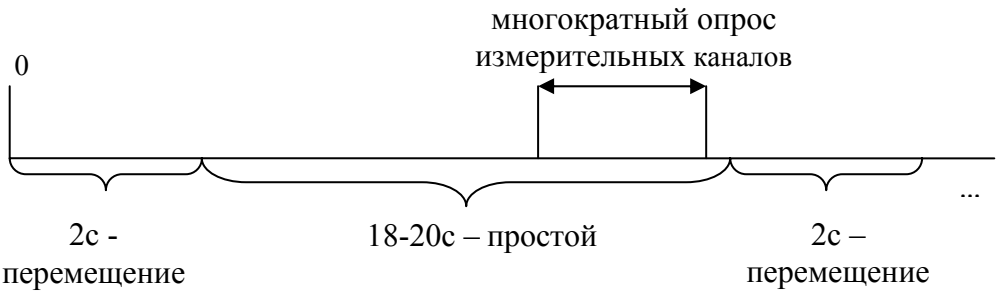


Рис. 2. Диаграмма распределения временных затрат в процессе измерения температуры

Задачи, решенные с помощью программного комплекса

В качестве примера, рассмотрим задачи, полученные в результате испытаний КС ГТД, которые были решены с использованием разработанного программного комплекса:

1) В тракте газотурбинного двигателя проводятся измерения динамических давлений с целью оценки уровня колебательных процессов, связанных с неустойчивостью в компрессоре и с нестабильностью горения в камере сгорания вплоть до пульсационного горения. С этой целью в контрольных точках по тракту двигателя устанавливаются зонды пульсаций давления. Типовые измерения проводятся в двух сечениях двигателя: за компрессором высокого давления (ВД) и в камере сгорания (КС).

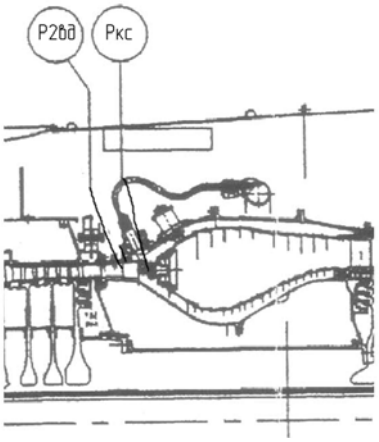


Рис.3. Точки замеров пульсаций давления

При этом теоретически присутствуют неравномерности потоков за счет влияния процессов в компрессоре, неравномерности потоков газов за счет процессов горения, а также влияние динамических процессов, связанных с горением, и процессов в турбине на процессы за компрессором ВД и в камере сгорания.

При помощи разработанного программного комплекса стало возможным определение величин взаимовлияния процессов, их корреляции путем нахождения коэффициентов корреляции по построенным взаимным корреляционным функциям (ВКФ) предварительно отцентрированных исходных процессов.

Также возможно проведение амплитудно-частотного анализа процессов за ВД и в КС с использованием разработанного программного комплекса.

Таким образом, проводя обработку данных по предложенной методике, возможно выявление дефектов камер сгорания, сравнение камер сгорания различных изделий с целью выявления камеры сгорания лучшего качества.

2) При испытании камер сгорания снимается характеристика, которая называется «поле температур» за КС. Снятие «поля температур» происходит с помощью специальной многоточечной гребенки – 6 точек в 140 позициях. Гребенка с термопарами вращается по окружности, останавливаясь в 140 позициях и измеряя температуру [1].

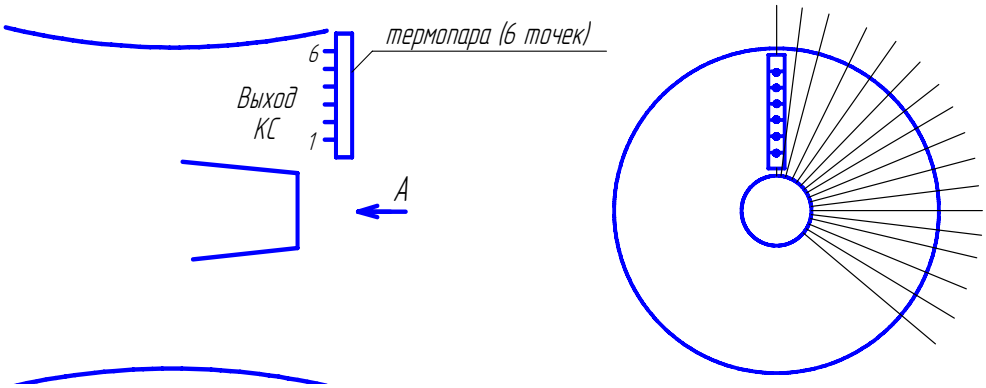


Рис.4. Вид по стрелке *А* – положение гребенки с термопарами (всего 140 позиций)

В результате испытаний получаем таблицу, в строках указывается номер пояса (всего 6), в столбцах – номер позиции (всего 140).

С использованием разработанного программного комплекса были решены следующие задачи, связанные с обработкой данных, полученных при измерении полей температур контуров камеры сгорания ГТД:

– выявление регулярности путем выделения тренда исходного процесса с последующим определением флуктуирующей составляющей и моментных характеристик. Решение данной задачи позволяет сказать, насколько результирующие данные эксперимента соответствуют желаемым результатам, а также определить «холодные места» - места, куда не поступает топливо (не работает горелка);

– определение взаимосвязи отдельных частей камеры сгорания путем нахождения коэффициентов корреляции по построенным взаимным корреляционным функциям с последующим построением по ним матриц корреляции, а именно выявление взаимовлияния между поясами, влияние 1-й точки пояса на 2-ую, 3-ю и остальные. Решение данной задачи позволяет сказать, насколько поля температур взаимодействуют друг на друга и «смешиваются»;

– проведение гармонического анализа с использованием автоматизированной системы аппроксимативного корреляционно-спектрального анализа в ортогональном базисе Бесселя. Решение данной задачи позволяет определить, какая из составных частей камеры сгорания – горелка, стойка или охлаждающее отверстие вносит нерегулярность в процесс измерения температуры. Известно, что камера сгорания данного типа в развертке состоит из ряда форсунок, охлаждающих отверстий, стоек и др., расположенных с определенной частотой следования по окружности. В случае если число, обратное найденной частоте, кратно количеству форсунок, стоек или охлаждающих отверстий, то предполагается, что именно данная составляющая камеры сгорания вносит нерегулярность в процесс измерения температуры. Полученные результаты позволят оптимально выбрать количество элементов КС и их расположение.

Также представляется возможным при помощи данного подхода обработать каждый пояс и весь массив целиком с выявлением частот появления пиков и в позиции.

Результаты, полученные при решении перечисленных задач, могут быть использованы при исследовании работы камеры сгорания ГТД данного типа. Проведение дальнейших исследований в данном направлении представляется актуальным, т.к. это позволит восстанавливать характеристики двигателя по экспериментальным данным, проводить статистический анализ полученных в ходе эксперимента результатов, изучать рабочие процессы двигателя и влияние различных условий на эти процессы, а значит, создавать в будущем высокоэффективные и конкурентоспособные двигатели нового поколения [1].

Аналоги и потребители программного продукта

Аналогами данного проекта являются системы типа «Data Mining», хорошо известные западным потребителям. Однако в силу громоздкости и тривиальности алгоритмов, положенных в основу их работы, приспособленных к решению общих, а не конкретных задач, подобные системы не так часто используются среди российских потребителей.

Основным потенциальным потребителем предлагаемого в проекте пакета программ является СНТК им. Кузнецова. Представители данной организации заинтересованы в продолжение работы над данным проектом и получении законченного продукта. Также данный пакет программ может быть внедрен и в других научно-технических комплексах, заводах и предприятиях, т.к. задачи обработки данных посредством предлагаемого в проекте пакета программ встречаются при анализе сигналов в различных областях науки и техники.

Заметим, что данный пакет программ используется и в учебном процессе при подготовке специалистов по специальности 23.01.02 «Автоматизированные системы обработки информации и управления» (наличие акта внедрения).

Литература

1. Прохоров, С.А. Применение автоматизированной системы аппроксимативного корреляционно-спектрального анализа в ортогональном базисе Бесселя при обработке данных измерения полей температур контуров камеры сгорания ГТД /С.А. Прохоров, Я.В. Газетова// Надежность и качество – 2010: труды Международного симпозиума: в 2 т./ под ред. Н.К. Юркова. – Пенза: Изд-во ПГУ, 2010. – 1 т. – С. 49-50.

2. Испытания авиационных двигателей: Учебник для вузов /под общ. ред. В.А. Григорьева и А.С. Гишварова. – М.: Машиностроение, 2009. – 504 с.:ил.

Сравнительный анализ методов вычисления спектральной плотности мощности сигналов

В.В. Графкин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Основными целями, преследуемыми исследователями в области анализа временных рядов, являются определение природы ряда и прогнозирование его отсчетов по имеющимся [1, 2, 3, 4, 7, 8]. Эти цели требуют от модели ряда идентификации и, более или менее, формального описания [7, 8, 9], после которых можно интерпретировать рассматриваемые данные. В данной работе уделяется внимание таким характеристикам временных рядов, как структурная функция (СФ) и спектральная плотность мощности (СПМ). Более распространенной и изученной является СПМ, а одним из самых популярных методов ее определения является преобразование Фурье (ПФ) [1, 2, 8].

В случае определения спектра $g_X(\omega)$ с помощью прямого ПФ неявно предполагается, что сигнал $X(t)$ задан на бесконечном интервале. И, если бы на практике интервал наблюдения действительно был таковым, ПФ давало бы точную оценку спектра [10]:

$$g_X(\omega) = \int_{-\infty}^{\infty} X(t) e^{-j\omega t} dt \tag{1}$$

В случае же конечности интервала

$$g_n = \frac{1}{N} \sum_{k=0}^{N-1} x_k e^{-j\frac{2\pi nk}{N}} \tag{2}$$

где n – номер гармоники; k – индекс отсчета сигнала (от 0 до N-1). При этом временной ряд, подвергаемый прямому ПФ, является всего лишь частью сигнала, который существует в действительности, то есть преобразованию подвергается не сигнал, существующий на самом деле, а сигнал, составленный из повторенной бесконечное число раз упомянутой части. Данное обстоятельство приводим к стыкам на краях частей, избавляются от стыков применением сглаживающих окон [1, 2, 3, 4, 10], что в свою очередь требует существенного количества априорной информации об анализируемом сигнале, с дефицитом которой обычно сталкиваются исследователи на практике (в табл. 1 представлена информация о наиболее часто используемых сглаживающих окнах, более подробную информацию о которых можно найти в работе [10]).

Определение спектра в этом случае проводится после перемножения исходного сигнала на весовую функцию W_k, то есть выражение (2) примет вид:

$$g_n = \frac{1}{N} \sum_{k=0}^{N-1} x_k W_k e^{-j\frac{2\pi nk}{N}} \tag{3}$$

где W_k – весовой множитель.

Выбор окна – немаловажная проблема, так как его неверный выбор может привести к результату, абсолютно расходящемуся с действительностью [1, 2, 4]. Получается, что при кажущейся на первый взгляд простоте прямого ПФ, существует множество нюансов [1, 2, 4, 8, 10], не учитывая которые анализ сигнала будет некорректным.

Другой метод определения СПМ, рассматриваемый в данной работе, основан на определении СПМ не напрямую по сигналу, а по его СФ, точнее по аппроксимации его СФ [5, 6, 7]. Важной особенностью данного метода является получение аналитической (ортогональной) модели СПМ, а значит и все остальные характеристики сигнала, связанные функционально с СПМ и СФ, могут определяться аналитически, то есть вычисления будут производиться значительно проще и быстрее, нежели численными методами.

Таблица 1. Сглаживающие окна

Вид окна	Аналитическое выражение окна, W_k	Характеристики окна
Прямоугольное	1	Очень узкий основной лепесток, хорошее частотное разрешение, плохое амплитудное разрешение, хорошее подавление белого шума
Блэкмана	$0.42 - 0.5\cos(2\pi k/N) + 0.08\sin(4\pi k/N)$	Широкий основной лепесток, заметный спад боковых лепестков, хорошее амплитудное разрешение
Хэннинга	$0.5(1-\cos(2\pi k/N))$	Достаточно широкий входной лепесток, быстрый спад боковых лепестков
Хэмминга	$0.54 - 0.46\cos(2\pi k/N)$	Умеренно широкий основной лепесток, постоянный уровень боковых лепестков
С плоской вершиной	$0.281 - 0.521\cos(2\pi k/N) + 0.198\sin(4\pi k/N)$	Размытые пики, хорошее разрешение по амплитуде, плохое разрешение по частоте

Наличие ортогональной модели СФ дает ряд преимуществ:

- отсутствие необходимости хранить все отсчеты, как процесса, так и его характеристик (СФ, СПМ);
- задавать шаг дискретизации функциональных характеристик;
- получать аналитические выражения характеристик, функционально связанных с моделью;
- производить вычисления имея минимум временных затрат, так как при этом не используются численные методы, к тому же, таким образом уменьшать количество составляющих погрешности вычислений;

Особенно ценна именно последняя особенность данного метода. Примером ее наглядного представления может служить определение одной из мер частотного анализа – спектральной функции (другое ее название – спектральная мера [4]):

$$G_X(\omega) = \int_0^{\omega} g_X(\omega) d\omega \qquad (4)$$

где $g_X(\omega)$ – спектральная плотность мощности (СПМ), которая может быть выражена через ортогональную модель структурной функции [7]:

$$g_X(\omega) = \frac{1}{2\pi\omega}\sum_{k=0}^{m}\beta_k\int_0^{\infty}\sin(\omega\tau)\frac{\partial\psi_k(\tau,\alpha)}{\partial\tau}d\tau \qquad (5)$$

где β_k – коэффициенты разложения структурной функции; $\psi_k(\tau,\alpha)$ – семейство базисных функций; m – число членов разложения; α – параметр масштаба базисных функций.

Далее подстановкой (5) в (4), в зависимости от используемого ортогонального базиса [7], спектральная функция определяется по выражениям, представленным в табл. 2. Переход от спектральной плотности мощности к спектральной функции позволяет получить ряд преимуществ: определение мощности процесса в заданном диапазоне частот $G(\omega_1,\omega_2) = G_x(\omega_2) - G_x(\omega_1)$, а также аналитическое определение эквивалентной ширины спектра. Вычисление спектральной функции случайных процессов по аппроксимации структурной функции характеризуется применимостью данного метода к случайным процессам со стационарными приращениями.

Таблица 2. Спектральные функции

Базис	Аналитическое выражение $G_X(\omega)$
Лагерра	$\frac{1}{2\pi}\sum_{k=0}^{m}\beta_k\left[\sum_{s=1}^{k}(-1)^{s+1}\frac{\sin(2s\varphi)}{s}-\varphi\right]$, где $\varphi = \operatorname{arctg}\left(\frac{2\omega}{\alpha}\right)$
Лежандра	$\frac{1}{2\pi}\sum_{k=0}^{m}\beta_k\sum_{s=0}^{k}(-1)^{s+1}C_{k+s}^{s}C_k^s\operatorname{arctg}\left[\frac{\omega}{(2s+1)\alpha}\right]$
Дирихле	$\frac{1}{2\pi}\sum_{k=0}^{m}\beta_k\sum_{s=0}^{k}(-1)^{k-s+1}C_{k+s+1}^{s+1}C_k^s\operatorname{arctg}\left[\frac{\omega}{(s+1)\alpha}\right]$

Из табл. 2 видно, что определение спектральной функции в случае ортогональной модели Лагерра структурной функции обладает значительным преимуществом – отсутствие в нем необходимости вычисления факториалов. Это свойство, во-первых, позволяет вычислять ортогональную модель структурной функции очень высоких порядков (сотни членов разложения ряда, тогда как для Лежандра Дирихле это число обычно ограничивается двадцатью), а, во-вторых, производится меньшее количество операций, что приводит к увеличению скорости вычисления спектральной функции. Примеры определения спектральных функций в различных базисах для одного процесса приведены на рис. 1.

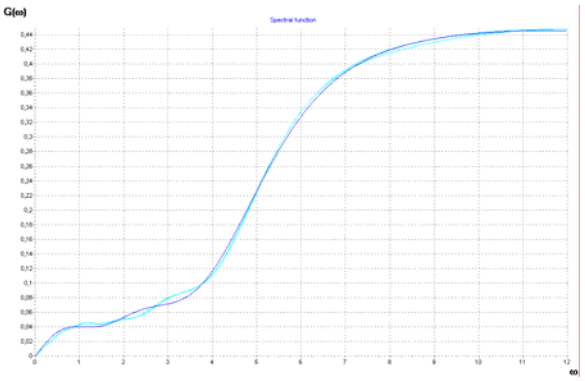


Рис. 1. Спектральные функции

Таким образом, при своей неоспоримой эффективности, спектральный анализ требует аккуратной настройки множества параметров в зависимости от конкретной задачи и специфики анализируемых сигналов, а для правильной интерпретации результатов нужен еще и некоторый практический опыт. В таблице 3 приведены основные сравнительные характеристики рассмотренных методов вычисления СПМ: БПФ (быстрое преобразование Фурье – метод вычисления прямого ПФ [1, 2, 4, 10]) и

косвенного метода, основанного на определении ортогональной модели СФ и дальнейшем вычислении СПМ уже по полученной модели.

Таблица 3. Характеристики методов вычисления СПМ

Особенность вычисления	БПФ	по ОМ СФ
Возможность изменения шага дискретизации по частоте	Требуется модифицирование самого временного ряда	Доступно
Аналитическая модель по результатам вычислений	Недоступно	Доступно
Основные этапы вычисления	Непосредственное вычисление СПМ по временному ряду Применение сглаживающих фильтров	Вычисление ортогональной модели СФ, вычисленной по временному Вычисление СПМ по ортогональной модели СФ

Литература

1. Дженкинс Г., Ваттс Д., Спектральный анализ и его приложения, Том 1, М.:МИР, 319 с. – 1971.
2. Дженкинс Г., Ваттс Д., Спектральный анализ и его приложения, Том 2, М.:МИР, 287 с. – 1972.
3. Витязев В.В., Спектрально-корреляционный анализ равномерных временных рядов. – Спб.: Изд-во С.-Петерб. ун-та, 2001. – 48 с.
4. Бриллинджер Д. Временные ряды. Обработка данных и теория. М.: МИР, 536 – 1980.
5. Прохоров С.А., Графкин В.В. Аппроксимация структурных функций случайных процессов // Математическое моделирование и краевые задачи: М33. Труды Третьей Всероссийской научной конференции. Ч. 4: Математические модели в информационных технологиях. – Самара: СамГТУ, 2006. – с. 82-86.
6. Прохоров С.А., Графкин В.В. Определение спектральной плотности мощности случайного процесса по аппроксимации структурной функции // Радиотехника и связь: материалы третьей Междунар. науч.-техн. конф. Саратов: СГТУ, 2006. – с. 13-17.
7. Прикладной анализ случайных процессов. Под ред. С.А. Прохорова. – Самара: СНЦ РАН, 2007. - 582с.
8. Андерсон Т., Статистический анализ временных рядов. М.:МИР, 757 с. – 1976.
9. Тюрин Ю.Н., Макаров А.А., Статистический анализ данных на компьютере / Под ред. В.Э. Фигурнова. М.: ИНФРА-М. 1998. – 528с.
10. Дьяконов В.П., Фурье- и вейвлет-преобразования в компьютерной математике // Материалы международной научно-практической конференции «Информационные технологии в образовании и фундаментальных науках (ИТО-Поволжье – 2007)». Казань: ТГГПУ, 2007. – с. 346-354.

Извлечение нечетких правил из числовых данных и структуризация базы знаний для задачи классификации

А.Н. Даниленко, А.Н. Коварцев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Качество функционирования системы принятия решений в значительной степени зависит от содержимого её базы знаний. На сегодняшний день существуют две основные группы методов получения знаний: прямые (интервью, изучение литературы и др.) и косвенные (анализ обучающего множества примеров, наблюдения за экспертом и др.).

Многочисленные проведённые исследования показали, что при принятии решений в условиях неопределённости, то есть при работе со слабоструктурированной, неполной или нечеткой информации, большую предпочтительность имеют методы второй группы.

Основная цель анализа данных – извлечение из них информации, то есть получение в качестве конечного результата лаконичного, наглядного и полезного описания каждого объекта.

Таблица, составленная не из модельных, а реальных данных содержит много характерных трудностей. Признаки данных отличаются друг от друга допустимыми наборами значений. Они могут принимать любое вещественное значение, являться выражением какой-либо меры и служить меткой варианта ответа на вопрос или отражать степень проявления определенного качества (например, степень уверенности в ответе).

Зачастую в таблице результатов количество строк и столбцов измеряется сотнями. Очевидно, что восприятие такого массива данных весьма затруднено. Следовательно, и анализ полученных результатов становится практически невозможным. Графики и диаграммы способны наглядно показать отношения лишь между двумя – тремя признаками, не учитывая остальные количественные показатели. В связи с этим получается, что чем больше признаков содержит таблица, тем, с одной стороны, полнее описываются объекты таблицы, а, с другой стороны, тем труднее извлекать из таблицы необходимую информацию для принятия решения.

В настоящее время существуют различные методы решения задачи кластеризации, однако изначально количество кластеров в них задается пользователем, исходя из примерного представления о характере будущего решения. Практически неизвестны методы, в которых помимо решения задачи кластеризации проводилась бы оценка значимости признаков.

Для задач с небольшим количеством примеров (не более 20) и 2 – 3 признаками классификации по данным методам можно провести вычисления вручную. Однако для задач с большой размерностью данные методы не пригодны.

В связи с вышеперечисленными проблемами встает вопрос об автоматической кластеризации данных по примерам обучающей выборки с выдачей результата в виде совокупности продукционных правил вида «ЕСЛИ – ТО».

В данной работе рассматривается задача принятия решения о профессиональной пригодности кандидата. Вследствие достаточно большого числа параметров и признаков требуется привлечение инструментального (программного) средства, реализующего те или иные алгоритмы кластеризации. Таким образом решается задача соотнесения соискателя к одному из четырех классов: кандидат полностью подходит, в основном отвечает требованиям специальности, частично соответствует, не

соответствует. Входными векторами сети являются результаты психологических тестирований по 12 тестам, имеющих более 50 факторов.

При решении задачи классификации в условиях неопределенности используются нечеткие правила. Однако вследствие того, что пространство каждой входной переменной, в свою очередь, делится на несколько интервалов, это приводит к ограничению возможного количества входных переменных.

Для преодоления этой проблемы при решении задачи извлечения нечетких правил из числовых данных изначально полагаем, что нечеткие «ЕСЛИ – ТО»-правила уже представлены в лингвистической форме или в виде четкой кластеризации корректных входных – выходных наборов данных.

Для решения задачи аппроксимации функции Абе и Лэн предложили метод извлечения нечетких правил непосредственно из числовых данных. Адаптируем этот метод для задачи классификации.

Пусть неизвестная функция характеризуется одномерным выходом y и m-мерным входным вектором x. Область, на которой определена переменная y, разделяется на n интервалов:

$$[y_0, y_1], (y_1, y_2], \ldots ([y_{n-1}, y_n],$$

где $y_0 = M_1, y_n = M_2$. Назовем i-й интервал $(y_{i-1}, y_i]$ выходным интервалом i.

Используя заданные выходные данные, для которых выходы находятся в выходным интервале i, рекурсивно определяется область входных значений, соответствующая входному интервалу i.

А именно, прежде всего, находятся в области активизации, которые определяют входную область, соответствующую выходному интервалу i, посредством вычисления минимального и максимального значений входных данных для каждого выходного интервала.

Если область активизации для выходного интервала i перекрывается с областью активизации для выходного интервала j, то область перекрытия определяется как область запрещения.

Если входные данные для выходных интервалов i и/или j находятся внутри области запрещения, то определяются одна или две дополнительные области активизации. В дальнейшем если области активизации определены также и перекрываются, то далее определяется дальнейшая область запрещения. Данный процесс повторяется до тех пор, пока проблема наложения областей не будет решена. Данный процесс показан на рис. 1.

Нечеткие правила определяются на основе областей активизации либо на основе областей активизации и соответствующих областей запрещения (если они сгенерированы).

Рассмотрим алгоритм, вычисляющий степени принадлежности входных данных к соответствующим выходным интервалам с последующим приведением к четкому значению. Так для входного вектора x степень принадлежности к выходным интервалам от 1-го до n-й определяется данным алгоритмом и затем на основе полученных степеней принадлежности вычисляется четкое значение выхода y.

Данный алгоритм состоит из трех этапов. При этом между функциями, реализующими различные выходные интервалы переменной y на этих этапах, отсутствует какая-либо связь.

В начале работы алгоритма реализуются нечеткие правила «ЕСЛИ–ТО», вычисляющие степени принадлежности переменных входного вектора x.

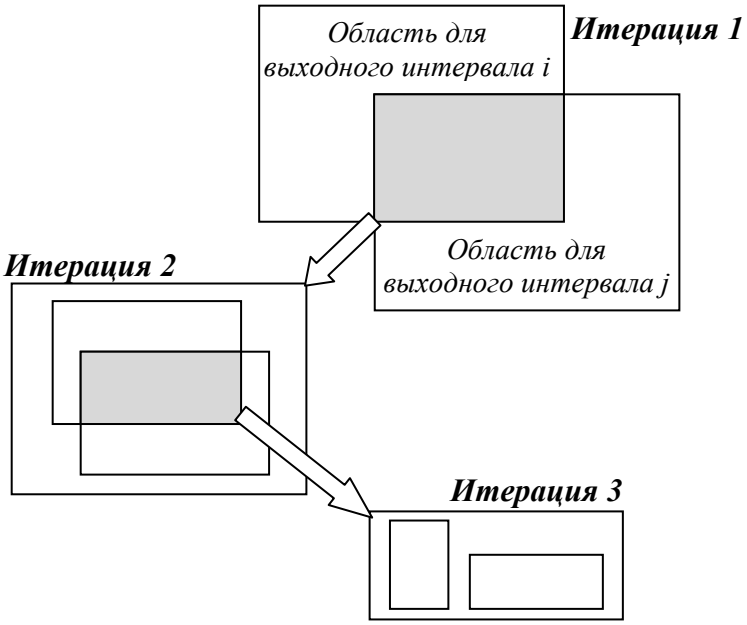


Рис. 1. Рекурсивное определение областей активизации и запрещения

На следующем этапе вычисляются максимальные значения выходов элементов первого этапа, которые представляют собой степени принадлежности, полученные в результате ликвидации перекрытий между парой выходных интервалов. Число элементов данного этапа для выходного интервала i определяется по числу выходных интервалов, чьи входные пространства перекрываются со входным пространством данного выходного интервала i. Поэтому, если нет перекрытия между входными пространствами выходного интервала i и какого-либо другого выходного интервала, то алгоритм для выходного интервала i уменьшается до одного этапа.

На третьем этапе для выходного интервала i определяет минимальное значение среди максимальных значений, определенных на предыдущем этапе и соответствующих перекрытию между двумя выходными интервалами. Поэтому если выходной интервал i перекрывается только с одним выходным интервалом, то алгоритм для выходного интервала i уменьшается до двух этапов. Вычисление минимумам на последнем этапе позволяет ликвидировать перекрытия среди более чем двух выходных интервалов. Таким образом, в процессе образования входных областей необходимо избавиться от перекрытия между двумя интервалами одновременно.

Данный подход позволяет сформировать начальную базу правил и оптимизировать ее в ходе работы с данными. В автоматически сформулированной базе продукционных правил контролируется ее полнота и непротиворечивость, что приводит к улучшению механизмов нечеткого логического вывода. Структуризация базы знаний позволит повысить качество и эффективность систем принятия решения.

Фреймовая база знаний информационной компьютерной среды

В.П. Дерябкин, А.И. Белоусов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Уровень и качество решения задач и принятия решений человеком во многом определяется его информационной компьютерной средой (ИКС). В [1-3] обсуждалась проблема построения параметрически настраиваемой ИКС для решения задач синтеза автоматизированных систем. При этом под ИКС понимается поддерживаемое программно-техническим комплексом (ПТК) информационное окружение, позволяющее человеку с тем или иным уровнем качества и комфортности выполнять установленные функции. ИКС в различных режимах может играть роль инструментальной среды разработки и развёртывания автоматизированной системы, а также среды непосредственного исполнения функций системы в соответствии с её назначением.

Показателем интеллектуальности среды с точки зрения представления знаний считается способность среды использовать в нужный момент необходимые (релевантные) знания.

Важнейшей частью ИКС является база знаний, структура и принципы построения которой излагаются далее.

Общность хранения абстрактных и материальных объектов

Хранение объектов любой природы**,** как материальных, так и абстрактных (информационных), присутствует практически во всех современных бизнес-технологиях. Сохраняется тенденция к построению универсальных систем и подсистем хранения, особенно в связи с внедрением территориально-распределённых сетей хранения и совершенствованием методов транспортировки и доступа к хранимым объектам.

При хранении материальных объектов (складировании) осуществляется переход к автоматизированным хранилищам класса WMS (Warehouse Management System) с автоматическим доступом и размещением хранимых объектов [4]. Системы хранения данных и знаний развиваются в направлении реализации концепции ILM (Information Lifecycle Management) – концепции управления данными на протяжении времени жизни информации [5]. Время жизни данных и знаний может достигать 100-150 лет и более, что намного больше времени жизни технологий представления и доступа. Возникает проблема обратной совместимости данных при смене поколений систем хранения в течение жизненного цикла информации. Аналогичные проблемы возникают и при хранении материальных объектов (например, книг в библиотеках), хотя здесь технологии размещения и доступа развиваются не столь быстро. Несмотря на несопоставимую разницу в скоростях доступа и транспортировки материальных и абстрактных объектов, способы адресации и доступа к хранимым объектам могут иметь много общего. Кроме того, информационная модель хранимых материальных объектов практически всегда присутствует в любой автоматизированной системе хранения.

Распределённые системы хранения вызвали развитие методов прикладной логистики и использование новых типов архитектур: NAS (Network Attached Storage) – на основе локальных сетей предприятий, и, далее, SAN (Storage Area Network) – на основе корпоративных и глобальных сетей размещения хранимых объектов. При этом возрастает роль стандартов открытых систем, обеспечения взаимодействия данной

системы с другими системами хранения. Новый технологический стандарт высокоскоростной транспортировки данных канал Fibre-Channel существенно влияет на прогресс SAN-архитектур и методов резервного копирования для повышения надёжности систем хранения.

Усложнение систем хранения и увеличение их объёма требует нового подхода к построению концептуальных моделей схем хранения. На базе SAN-архитектуры развиваются методы виртуализации ресурсов хранения – отображения информации об объектах, хранящихся на любом количестве разнообразных устройств и адресов в виде единого централизованного хранилища ("виртуального пула") корпоративных данных. Таким образом, обеспечивается единый доступ к разнообразно размещённым объектам, что повышает эффективность функционирования системы хранения в целом. Виртуализация ресурсов хранения облегчается, если существует единая модель хранимых объектов и логики доступа к этим объектам с целью их извлечения и организации систем учёта движения и их изменения в процессе жизненного цикла.

Фреймы как элементарные единицы хранения знаний

В качестве метода представления знаний, позволяющего повысить уровень интеллектуальности и унифицировать структуру единицы хранимой информации в ИКС, более эффективно использовать средства визуализации и современные объектно-ориентированные СУБД в процессе решения функциональных задач, предлагается использовать фреймовый подход. Он был предложен ещё в середине 70-х годов прошлого века Мински [6] и привёл к появлению множества различных языков описания и манипулирования данными и знаниями, а также различных инструментальных средств и приложений. Концепция фреймов способствовала распространению методов объектно-ориентированного программирования и объектного описания хранимых данных.

Характерной особенностью ИКС является постоянно изменяемая и расширяемая база знаний. Связано это как с совершенствованием методической и технической базы предприятий и повсеместным внедрением корпоративных компьютерных сетей, так и с распространением и популяризацией всемирной компьютерной сети Интернет. Представляется актуальной идея разработки систем на базе унифицированного представления знаний, которое будет, с одной стороны, обеспечивать прозрачную интеграцию с объектной и компонентной моделями построения программных систем, с реляционными и объектными базами данных, а так же допускать распределение фрагментов знаний по сети с последующим распределенным выводом для решения некоторой прикладной задачи. Следует отметить, что классическая фреймовая модель знаний должна быть дополнена и уточнена с целью удобного построения иерархий не только по направлению обобщений (отношение is_a), но и по направлению агрегаций (отношение part_of), что является характерным при проектировании автоматизированных систем. Кроме того, модель должна предусматривать хорошую визуализацию элементов хранения и их иерархий, как показал, например, опыт использования известной инструментальной среды G2 [7].

Дополнительным преимуществом фреймов как модели представления знаний является способность отражать концептуальную основу организации памяти человека, а также ее гибкость и наглядность. Следуя классическому подходу [8], будем различать фреймы-прототипы и фреймы-экземпляры. Фрейм-прототип представляет собой абстрактный образ какой-либо сущности, а фрейм-экземпляр наследует все свойства прототипа и конкретизирует их. При описании экземпляра фрейма указывается ссылка на фрейм-прототип и задаются значения его слотов. С каждым фреймом ассоциируется несколько видов информации, слоты могут быть заполнены по умолчанию.

Графические образы фреймов позволяют использовать визуальные методы проектирования и представления данных в разработке приложения.

Формально фрейм имеет уникальное имя и рассматривается как структура вида

F = < s1, s2, s3, s4>,

где s1 – множество слотов, содержащих общесистемную информацию фрейма;

s2 – множество слотов ссылочного типа, позволяющих работать с фреймовыми иерархиями;

s3 – множество слотов, определяющих индивидуальные нестандартные свойства (характеристики) фрейма;

s4 – множество предопределённых слотов, содержащих дополнительную информацию фрейма.

Введём понятие *хранимая единица знаний* в виде фрейма-прототипа, имеющего имя, стереотип и слоты:

- синонимы;
- адрес хранения;
- категории (обобщения, суперфреймы) ближайшего верхнего уровня;
- подфреймы ближайшего нижнего уровня;
- состав (агрегация) ближайшего нижнего уровня:
- входимости в составы (другие агрегации) ближайшего верхнего уровня;
- мультимедийная информация;
- характеристики (свойства);
- ассоциации (логические связи);
- соединения (физические связи);
- комментарий.

К множеству s1 относятся слоты синонимов, адресов хранения; s2 – указатели на суперфреймы, подфреймы, слоты состава, входимости в состав, ассоциации и соединения; s3 – характеристики (свойства); s4 – мультимедийная информация и комментарий.

Характеристики могут задаваться в одном слоте множеством наборов «имя-тип-размерность-значение» или отдельными слотами. Как и в классической модели [8], с каждым слотом может быть связан набор процедур типа IF-ADDED, IF-DELETED, IF-NEEDED. Указанные процедуры могут содержать условную часть – проверку их применимости в сложившейся ситуации к другим слотам и реализуют требуемые функции приложения. Таким образом, предлагаемая модель является одним из вариантов формализации классической фреймовой модели и оставляет возможность использования некоторого подмножества существующих языков фреймового представления знаний.

Особенности реализации иерархических фреймовых структур

Каждая единица хранения может относиться к определённому классу в различных системах классификации и обладать набором характеристик – как принадлежащих классу в целом, так и индивидуальных. Наличие различных систем классификации принципиально важно, так как способ использования одной и той же единицы хранения может быть разным и интерес при доступе могут представлять различные характеристики. Единица хранения размещается по определённому адресу, в общем случае являющимся составным, содержащим в своём составе подадреса, отражающие структуру мест и устройств хранения и доступа. Аналогично строятся фреймы, представляющие знания об адресных пространствах, классах единиц хранения, характеристиках классов единиц хранения. Деление по стереотипам отражает роль фреймов, например, системные фреймы управления и функциональные фреймы приложения. Однако, все фреймы базы знаний имеют общий приведённый выше

прототип, что позволяет добиться довольно строгой унификации. Само интеллектуальное приложение задаётся фреймом аналогичной структуры, все подсистемы и модули приложения и среды исполнения также являются фреймами. Обобщённая концептуальная модель базы знаний представляет собой сеть фреймов типовой структуры с набором иерархических связей и может использоваться для виртуализации ресурсов хранения в любой сложной автоматизированной системе хранения. Указанная концепция была успешно реализована средствами системы программирования Microsoft Visual Studio 8 (язык C#) в виде инструментальной интеллектуальной системы ИКС и фреймовой системы оптимального планирования автомобильных перевозок однородного груза. Более подробные сведения о реализации приводятся в отдельном докладе А.И. Белоусова.

Выводы

Предложенная модель фреймового представления знаний, являющаяся вариантом классической фреймовой модели, рекомендуется для унифицированного представления знаний в информационных компьютерных средах (ИКС) пользователей, разрабатывающих и эксплуатирующих автономные и распределённые приложения, в том числе интеллектуальные.

Модель позволяет использовать современные визуальные объектные и COM-технологии при разработке приложений, оставляя возможность использования существующих реляционных и объектных СУБД.

Эффективность модели подтверждена реализацией фреймовой системы планирования и оптимизации маршрутов грузоперевозок.

Расширения модели в сторону использования продукций и переход к унифицированным продукционно-фреймовым моделям [3, 9] , являются предметом дальнейших исследований с точки зрения эффективности по быстродействию работающих приложений.

Литература

1. Deriabkin V.P. Synthesis of Intellectual Information Environment: technique and development tools/V.P. Deriabkin // The First International Conference on Distance Education in Russia ICDED'94. Proceedings, Moscow, 1994.
2. Дерябкин, В.П. База знаний системы синтеза и параметрической настройки проблемно-ориентированной информационной компьютерной среды / В.П. Дерябкин // Перспективные информационные технологии в научных исследованиях, проектировании и обучении «ПИТ-2006», Т.1. – Самара, СГАУ, 2006. – С.65-69.
3. Дерябкин, В.П. Среда визуальной разработки интеллектуальных систем продукционно-фреймового типа / В.П. Дерябкин // Математическое моделирование информационных процессов и систем в науке, технике и образовании. – Самара, СГАСУ, 2010. – С. 48-51.
4. Система управления складом (WMS) на платформе Microsoft Dynamics NAV (Navision)// www.elfor-soft.ru.
5. Лобанов А.К. Методы построения систем хранения данных // www.citforum.ru.
6. Minsky M. A Framework for Representing Knowledge/M.A. Minsky. - Cambridge: MIT Press, 1974.
7. Gensym Announces Release of Gensym G2 8.4 R2// www.gensym.com.
8. Джарратано, Д. Экспертные системы: принципы разработки и программирования, 4-е издание / Д. Джарратано, Г. Райли . – М.: ООО «И.Д. Вильямс», 2007. -1152 с.
9. Сошников Д.В. Инструментарий JULIA для построения распределённых интеллектуальных систем на основе продукционно-фреймового представления знаний/ Д.В. Сошников // Электронный журнал «Труды МАИ» - М.: МАИ, 2002, №7.

Математическая модель и решение обобщенной задачи о покрытии

Б.А. Есипов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Совершенствование и автоматизация организационно-технических систем приводит к необходимости математического моделирования сложных задач оптимизации. В частности, при постановке задачи наиболее эффективного распределения ресурсов по отдельным объектам возникает так называемая задача об оптимальном покрытии. Например, мы хотим обеспечить всю совокупность функций при производстве на отдельных объектах: цехах, участках за счет установки оборудования с частично перекрывающимися функциями. Нужно, чтобы все заданные функции на каждом объекте были обеспечены и при этом затраты на оборудование были бы минимальны. Аналогичные задачи возникают при кадровом обеспечении удаленных воинских частей, кораблей, больниц, выбора инструментов для ремонтов или при проектировании систем на основе стандартных наборов микропроцессоров и электронных схем. Ниже мы рассмотрим математическую модель подобной задачи на «классическом» примере.

Имеется m туристических групп, в каждой из которых известен набор языков общения и n переводчиков, каждый из которых может обеспечить определенный набор языков общения из общего числа K языков. Необходимо распределить переводчиков по туристическим группам так, чтобы обеспечить общение во всех группах и при этом, чтобы количество занятых переводчиков было бы минимально.

Обозначим i – номер туристической группы ($i=1,m$), а j - номер переводчика ($j=1,n$).

Пусть k – номер языка общения ($k=1,K$). Тогда можно ввести величины a_{jk} , которые принимают значение 1, если j - й переводчик знает k - й язык, и значение 0 в противном случае. Обозначим также величины b_{ik} , которые принимают значение 1, если в i - й группе необходим язык k, и значение 0 – в противном случае. В качестве управляемых переменных, определяющих распределение переводчиков по туристическим группам рассмотрим величины x_{ij}, равные 1, если в i - ю группу направлен j - й переводчик и равные 0 – в противном случае. Сформулируем ограничения данной задачи.

$$x_{ij} \in \{0,1\}. \tag{1}$$

Каждый переводчик может работать только в одной группе или не работать вовсе. Поэтому

$$\sum_{i=1}^{m} x_{ij} \le 1, \;\; j = 1,n\,. \tag{2}$$

В каждой группе должны быть обеспечены переводом все необходимые языки. Это требование можно математически выразить так

$$\sum_{j=1}^{n} a_{jk} x_{ij} \ge b_{ik}\,, \;\; i = 1,m; \;\; k = 1,K\,. \tag{3}$$

Ограничения (3) допускают решения, когда в одной группе могут быть несколько переводчиков, знающих один и тот же язык (поэтому знак не =, а $\ge$). Это расширяет множество вариантов и, следовательно, может улучшить значение целевой функции.

Эффективность распределения можно выразить в виде целевой функции

$$f = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \to \min . \qquad (4)$$

где величины c_{ij} - затраты на использование j - го переводчика в i -ой группе. Если c_{ij} одинаковы, то необходимо минимизировать общее число занятых переводчиков (все c_{ij} =1).

Очевидно, что задача (1) – (4) может иметь допустимые решения, только если $n \geq m$, а также

$$\sum_{j=1}^{n} b_{jk} \geq \sum_{i=1}^{m} a_{ik}, \ \ k = 1, K . \qquad (5)$$

Сформулированная математическая модель и задача (1) – (4) принадлежит к классу задач линейного булевского программирования, в которой $m \cdot n$ булевских переменных и $m \cdot K+n$ ограничений неравенств. Легко заметить, что те ограничения, в которых b_{ik} =0 можно не записывать и не проверять, т.к. они выполняются для любых решений. Поэтому на самом деле ограничений (3) столько, сколько ненулевых элементов b_{ik} в конкретной задаче, т.е. $p=\sum\sum b_{ik}$. .Для решения рассмотренной задачи при больших значениях m,n и K разработана специальная программа, использующая в основе «аддитивный алгоритм». Перед проверкой ограничений в программе проверяется условие превышения f достигнутого рекорда. Кроме разработанной программы проведены эксперименты по решению этой задачи с использованием стандартных программ, работающих на основе метода ветвей и границ (LINDO, EXCEL), которые подтвердили правильность модели и решения. Временные затраты на решения при числе переменных порядка 40 и числе ограничений 20 мало различаются. При размерностях порядка 100 переменных и 40 ограничений специальная программа дает преимущество.

Рассмотренная модель может быть уточнена и расширена для частных практических случаев, которые отличаются способами получения исходных данных для решения рассмотренной задачи оптимизации, а именно: величин a_{jk} , b_{ik} и c_{ij}.

Автоматизированная система выполнения математических расчетов в сети Интернет с элементами социальной сети

А.В. Камынин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

За годы существования Интернет состав web-приложений, выполняемые ими функции, принципы и архитектура их построения претерпели весьма значительные изменения - от простейших средств хранения HTML-страниц до современных решений, ориентированных на поддержку работы развитых корпоративными информационных систем. На сегодняшний день многие мировые лидеры в производстве программного обеспечения стремятся занять эту нишу, выпуская версии всем известных приложений, доступных в виде web-приложений (web-сервисов). Яркими примерами этого могут служить такие web-приложения, как Google Docs от компании Google, SalesForce CRM от компании SalesForce, Microsoft Office Live от компании Microsoft. Несмотря на то, что web-приложения для браузера имеет ограничения, связанные с безопасностью и производительностью, их разработка более сложна и запутана по сравнению с разработкой стандартных приложений, усилия обычно оправдываются, потому что:

- не требуется установка приложения; обновление и распространение приложения – быстрый и автоматизированный процесс;
- обновление версий автоматическое;
- пользователи могут использовать приложение на любом компьютере, имеющем соединение с Интернет в независимости от установленной операционной системы;
- при работе web-приложения компьютер пользователя гораздо меньше подвержен вирусному заражению, чем при запуске exe-файлов.

Стоит отметить, что ниша сложных математических вычислений, требующих больших вычислительных мощностей от компьютера остаётся слабо тронутой в среде Интернет. Из-за отсутствия онлайн-аналогов пользователи больше отдают предпочтение настольному математическому программному обеспечению. Для такого ПО свойственно:

- высокие требования к аппаратной части ПК;
- индивидуальная работа. Над одним файлом в один момент времени может работать только один пользователь;
- несовместимость файлов разных версий ПО.

Однако применение Интернет в этой среде могло бы устранить эти недостатки и сделать возможным недоступную ранее совместную работу пользователей над одним массивом данных, поддерживая общение между ними в реальном времени. Важно отметить, что пользователи могут находиться на больших расстояниях друг от друга и поддерживать связь исключительно посредством Интернет.

Цель и особенности

Целью разработки автоматизированной системы (далее «система») является повышение эффективности индивидуальной и коллективной работы над математическими моделями в научной деятельности и учебном процессе за счёт применения web-технологий.

Система представляет собой сайт, доступный в сети Интернет или другими словами Web-сервис.

Web-сервис является адаптированной Интернет версией CAD системы, разработанной для удалённого использования посредством сети Интернет, и ориентирован на непрофессиональных или полупрофессиональных пользователей, не использующих широкие возможности дорогостоящих CAD-систем, таких как MathCAD, Maple и других, а использующих данный сервис для решения локальных задач либо для обучения. Под локальными задачами подразумеваются: решение уравнений либо систем уравнений, построение графиков, вычисление сложных выражений. В отличие от упомянутых выше аналогов, одной из основных задач web-сервиса является предоставление пользователю возможности максимально быстро и без лишних затрат приступить к его использованию. Это достигается посредством следующих особенностей web-сервиса:

1. Использование модели SaaS (Software as a service - «Программное обеспечение как услуга», или Software on Demand «Программное обеспечение по требованию»). Это модель предоставления программного обеспечения, которая подразумевает разработку и управление web-сервисом самостоятельно, независимо от пользователей, предоставляя им доступ к сервису через Интернет. Основное преимущество модели SaaS для потребителя (пользователя web-сервиса) состоит в отсутствии затрат, связанных с установкой, обновлением и поддержкой работоспособности оборудования и программного обеспечения, работающего на нём. В данном случае это означает, что все рабочие данные и настройки пользователей хранятся на удалённых серверах, на которых также происходит их обработка.
2. Web-сервис является кроссплатформенным. Это означает, что его работа не зависит от операционной системы пользователя. Для взаимодействия с web-сервисом требуется только наличие браузера и подключения к сети Интернет.

В web-сервисе приняты следующие понятия:

- Рабочий лист (worksheet) — набор формул, графиков, результатов вычислений и комментариев к ним, объединенных общим смыслом. Рабочий лист хранится на сервере. Аналогом рабочего листа является файл.
- Документ (document) — текстовая информация с описанием чего-либо, хранящаяся на сервере.

Работа с системой состоит из нескольких этапов. Для начала работы пользователю необходимо пройти регистрация, чтобы создать себе учётную запись, которую он впоследствии будет использовать для хранения рабочих данных. После подтверждения регистрации по e-mail, пользователю становятся доступны функции web-сервиса, а именно: создание и редактирование рабочих листов, создание и редактирование документов, публикация рабочих листов и документов для доступа к ним других пользователей, причём доступно может быть не только их чтение, но и изменение.

Возможность публикации информации в системе для доступа к ней других лиц является очень важной особенностью системы. Это позволяет коллективно разрабатывать и обсуждать математические модели в рамках web-сервиса, не прибегая к использованию сторонних средств (таких как средства моментального обмена сообщениями, электронная почта и т.п.). После публикации данных их автором, те пользователи, которые имеют к ней доступ, могут одновременно, работая параллельно за разными ПК, вносить в неё правки. Таким образом, достигается коллективная работа над поставленными задачами.

В условиях постоянного совершенствования возможностей компьютерной техники, а особенно в условиях распространения глобальной сети Интернет, создание и внедрение компьютерных обучающих систем приобретает все большее значение. Автоматизация процесса обучения требует использования современных методик

представления и обработки информации, использования последних достижений в области информационных и телекоммуникационных технологий, построения в рамках обучающей системы адекватной модели не только обучаемого, но и обучающего. Использование web-сервиса в сфере обучения может значительно упростить процесс коммуникации преподавателя со студентами и процесс контроля знаний посредством возложения большинства обязанностей на web-сервис.

Рассмотрим пример работы с web-сервисом в сфере обучения. Здесь можно выделить два типа пользователей: преподаватель и студент. Преподаватель формулирует задание и помещает его в web-сервис в виде документа, к которому открывает доступ на чтение для своих студентов, имеющих учётную запись в web-сервисе. Данные действия могут проводиться удалённо через Интернет. При получении студентом документа с заданием в системе, он может приступить к его выполнению в удобное для себя время.

Как уже упоминалось, доступ к системе осуществляется удалённо и все данные хранятся на удалённых серверах, а это означает, что отпадает необходимость переносить рабочие файлы, и процесс решения одной и той же задачи может легко осуществляться как из учебного заведения, так и с домашнего компьютера. Кроме того, существует возможность вернуться к предыдущей версии рабочих данных, если сделанные изменения окажутся неверными, так как система хранит промежуточные версии рабочих листов и документов. После завершения работы над заданием, студент открывает доступ преподавателю на результирующий рабочий лист, и преподаватель может приступить к проверке. Так как у рабочего листа существуют ревизии (версии), то преподаватель может посмотреть ход работы студенты и сделать вывод о степени знаний студента. Разграничивая учётные записи студента и преподавателя в системе можно установить ограничения таким образом, чтобы студенты не могли обмениваться данными, касающимися одного варианта задания, исключив возможность списывания. После проверки, общаясь внутри системы, преподаватель уведомляет студента о результатах.

Доступность web-сервиса в сети Интернет предоставляет широкие возможности для публикации научных достижений и результатов исследования. Сервис, по сути, является виртуальным сообществом людей, интересующихся математическим моделированием, а значит, публикация материала на смежную тему в нём может достичь высокого уровня отклика из-за присутствия на сайте целевой аудитории. Организация сайта в форме виртуального сообщества преследует следующие цели:

1) объединение пользователей в группы по интересам;
2) каталогизация пользователями информации по различным характеристикам с использованием тегирования. Это позволяет пользователю самому определить удобные и понятные ему категории [1];
3) простой обмен информацией между пользователями;
4) совместная работа;
5) возможность высказывания мнения пользователями, рецензирования.

Предоставление доступа к системе как к сервису подразумевает возможность одновременного использования системы большим числом пользователей. При условии возложения на сервер обязанностей по выполнению вычислений, web-сервис будет являться требовательным к ресурсам сервера. Оптимальная работа в такой ситуации будет обеспечиваться использование облачных вычислений (cloud computing). Это технология обработки данных, в которой компьютерные ресурсы и мощности предоставляются по требованию в необходимом объеме [2]. Достигается это за счёт виртуализации. В итоге, для поддержки приложения в рабочем состоянии существует ряд рабочих серверов, непосредственно выполняющих запросы пользователей и ряд

балансирующих серверов, осуществляющих равномерное распределение этих запросов между всеми рабочими серверами. При этом увеличение процессорной мощности, объёма оперативной памяти и дискового пространства происходит незаметно для web-приложения при необходимости.

Техническая реализация

Разработка web-сервиса включает в себя разработку клиентской (видимой непосредственно для пользователя) и серверной (отвечающей за обработку и хранение данных) частей. Клиентская часть состоит из xhml-страниц с использованием Javascript и MathML для отображения формул. MathML (Mathematical Markup Language, язык математической разметки) – это язык разметки, основанный на XML, используемый для представления математических символов и формул в среде Интернет [3]. В качестве основы для серверной части используется язык программирования Python. Это высокоуровневый язык программирования общего назначения с акцентом на производительность разработчика и читаемость кода. Для выполнения математических вычислений и преобразований используется библиотека для Python с открытым исходным кодом SymPy.

Выводы

В рамках данной работы был произведён анализ потенциальных конкурентов, выявлены их сильные и слабые стороны. В результате этого анализа был сделан упор на следующие ключевые моменты при разработке данного web-сервиса:

1) использование современных web-технологий;
2) коллективная работа;
3) организация сайта в виде виртуального сообщества;
4) предоставление пользователям вычислительных мощностей сервера.

Web-сервис является доступным широкому кругу пользователей через Интернет. Целью его разработки является повышение эффективности работы над математическими моделями за счёт применения web-технологий.

Литература

1. А.В. Иващенко, А.Ю. Орлов, С.И. Вольман, И.А. Минаков Виртуальные сообщества в сети Интернет. Организация и Управление // Самара: СНЦ РАН, 2008 – 99 с., ил.
2. Облачные вычисления, http://ru.wikipedia.org/wiki/Облачные_вычисления
3. MathML, http://ru.wikipedia.org/wiki/MathML

Модель для создания системы поддержки принятия решений в области планирования и управления инновационными программами

В.Б. Киреев[1], Т.В. Кондранин[1], А.В. Николаев[1], Д.И. Озол[1], Д.Б. Пайсон[2], А.В. Киселёв[1], Ю.И. Полищук[1], С.С. Семёнов[1]

[1] *ГОУВПО Московский физико-технический институт (государственный университет) (МФТИ), Москва*

[2] *ФГУП «Центральный научно-исследовательский институт машиностроения», Королёв*

Введение

Целью данной работы является рассмотрение подходов и методологии создания систем поддержки принятия решений (ППР) для планирования и управления долгосрочными инновационными программами и проектами и описание одного из её вариантов с использованием рубежно-целевого подхода, применимого для планирования крупномасштабных программ освоения космоса.

В каждый момент времени стратегическое планирование инновационной деятельности по периодам будет осуществляться на основе баланса между необходимостью осуществления генеральной миссии национального стратегического планирования по обеспечению устойчивого развития и национальной безопасности и необходимостью решения прикладных задач, продиктованных временем, что в каждый момент определения плановых показателей потребует проведения оптимизации, критерии которой могут и будут меняться во времени. Для последовательного решения этой задачи у лиц, принимающих решения, должен быть гибкий настраиваемый инструмент, позволяющий легко изменять текущие критерии принятия решений и варианты развития инновационной деятельности. Рассмотрению и развитию методик и инструментов ППР в условиях изменяющейся обстановки и неопределённости применительно к инновационной деятельности и посвящена настоящая работа.

Требования к созданию информационных моделей, алгоритмов и программ расчета на ЭВМ оптимальных вариантов реализации инновационных программ

В результате рассмотрения подходов к рассмотрению стратегического прогнозирования, планирования, распределения ресурсов и формирования (синтеза) альтернатив в рамках современной теория выбора решений, основанных на различных парадигмах современной математической теории ППР: критериально – экстремизационный выбор, парно – доминантный выбор, псевдокритериально – экстремизационный выбор, псевдографовый выбор, коллективный выбор на основе процедур голосования, сделан вывод о необходимости создания прикладного инструментария, выходящего за рамки использования стандартных экспертных подходов, способного обеспечить долгосрочное планирование и управление сложными поисковыми НИР и НИОКР при отсутствии надёжной информации, когда изначально нет однозначных траекторий для программ инновационного развития.

Инновационное развитие в целом, так же как и программы освоения космоса не имеют единой чётко структурированной цели, не имеют границ, и, следовательно, для них возможно лишь планирование достижения ряда последовательных рубежей-целей. В долгосрочном периоде могут изменяться приоритеты и конкретные промежуточные цели.

Всё это должны учитывать информационные модели, алгоритмы и программы для ЭВМ, направленные на поддержку принятия решений по определению оптимальных вариантов реализации инновационных программ и проектов

Информационная модель выбора оптимальных миссий и формирование программ инновационного развития

В рамках предлагаемой в данной работе модели формализация описания возможных путей построения и реализации планов развития инновационной деятельности и обеспечивающей её инфраструктуры основана на использовании понятий «миссия» и «сущность», которые объединяются в потенциальный план развития инновационной деятельности. Одним из компонентов потенциального плана является «миссия». Под миссией субъекты, участвующие в инновационной деятельности, обычно понимают конкретный проект. Потенциальный план развития инновационной деятельности включает полный набор возможных (предлагаемых к реализации или рассматриваемых как потенциально возможные) инновационных миссий и набор необходимых для этого элементов инфраструктуры – сущностей.

Центральным компонентом (элементом) потенциального плана развития является «сущность». Сущностью может быть технология, объект инфраструктуры, однородная совокупность объектов, законодательный акт, идеологическая или культурологическая установка, индивидуум с уникальными способностями, другой объект или субъект (совокупность объектов или субъектов), наличие которого (совокупности которых) необходимо для реализации инновационной миссии или реализации другой сущности, включённой в иерархически организованную сеть, описывающую причинно-следственные связи возможностей достижения тех или иных сущностей более высокого иерархического уровня.

На рис.1 представлен фрагмент потенциального плана развития инновационной деятельности, включающий конкретную сущность, логические операторы «И» и «ИЛИ», и вектора связей данной сущности с сущностями других иерархических уровней: более низкого (вектора разрешающих блоков сущностей, объединяемых между собой операторов «ИЛИ», а внутри каждого разрешающего блока оператором «И») и более высокого (разрешающие вектора, формирующие разрешающий блок, для сущности следующего иерархического уровня организации).

Формализованный потенциальный план развития инновационной (космической) деятельности представляет собой однонаправленный граф, описывающий переходы (в рамках всех возможных рассматриваемых миссий) внутри иерархически организованной системы сущностей с учётом возможных альтернатив реализации миссий. Разрешающие вектора от данной сущности направлены к тем сущностям (в рамках выполнения некой миссиям), достижение которых возможно при наличии данной сущности (наряду с наличием ряда других сущностей) и формируют разрешающий блок (возможно один из альтернативных) для сущности более высокого уровня иерархической организации. Логический оператор «ИЛИ» объединяет возможные альтернативы достижения (реализации) данной сущности (объединяет разрешающие блоки различного состава). Логический оператор «И» объединяет вектора, идущие от альтернативного набора сущностей, наличие которых обеспечивает возможность реализации рассматриваемой на рис.1. сущности. Для целей анализа вариантов развития сущность может иметь два состояния «воспрещена» и «разрешена».

Сущность находится в состоянии «разрешена», когда при реализации алгоритма (программы анализа) на выходе «Р» схемы разрешающего блока (рис. 1) есть логическая 1. В противном случае сущность находится в состоянии «воспрещена». Разрешающие блоки НП1, НП2, …, НПN — достаточные наборы предшествующих сущностей 1, 2, N. При наличии реализации всех сущностей, входящих хотя бы в один из этих разрешающих наборов, возможна реализации рассматриваемой сущности. Таким образом, реализация всех сущностей из одного или более таких наборов,

формирующих разрешающий блок, разрешает реализацию рассматриваемой сущности. Выход(ы) сущности будем называть разрешающими векторами (РВ).

На каждом разрешающем векторе (РВ) сущность в соответствии с алгоритмом (программой анализа) выдает логическую 1 в состоянии «разрешена» и 0 в состоянии «воспрещена». Разрешающий вектор (РВ) входит в (замыкается на) один или более набор предшествующих (НП) сущностей для одной или более других сущностей более высокого иерархического уровня. Воспрещены циклические зависимости в наборе сущностей. Ситуация, когда несколько разрешающие вектора (РВ) одной и той же сущности замыкается на различные наборы предшествующих (НП) сущностей другой сущности более высокого иерархического уровня разрешена.

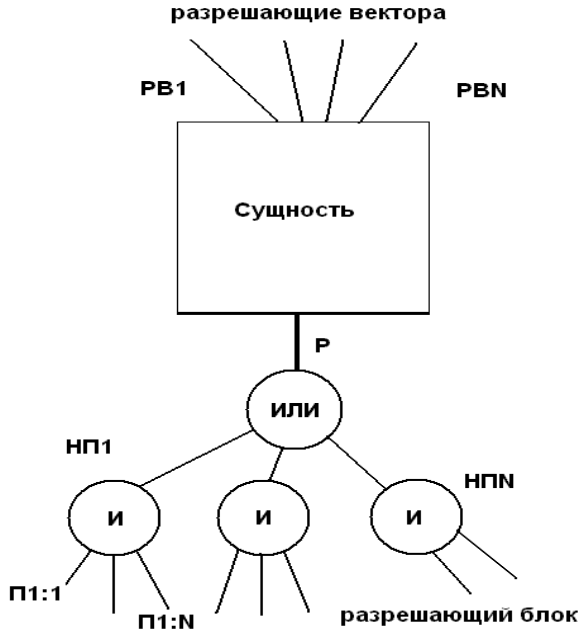


Рис.1. Фрагмент потенциального плана развития инновационной (космической) деятельности

Таким образом, любой конкретный план развития инновационной деятельности представляет собой однонаправленный граф, объединяющий сущности (необходимые для реализации соответствующих миссий), разрешающие связи (вектора), и логические операторы, отвечающие причинно-следственным связям осуществления миссии и достижения в ходе её реализации соответствующих сущностей.

Полный план развития инновационной деятельности, представляющий собой всю возможную (включённую в данное рассмотрение с целью выбора оптимальной стратегии развития) совокупность сущностей, разрешающих связей между ними и логических операторов назовем полной сетью вариантов развития или полным альтернативным графом (ПАГ).

Из полной сети (полного альтернативного графа – ПАГ) по тем или иным критериям выбираются маршруты развития или маршрутные графы (МГ) или миссии. Таким образом, в рамках нашей формализации миссия это совокупность конкретной последовательности шагов (этапов реализации конкретного набора сущностей) необходимых для реализации конкретного проекта.

Целью назовем сущность, наличие которой в маршрутном графе обязательно. При этом, такая сущность может находиться в любой части маршрутного графа и входить в наборы предшествующих сущностей (НПС) или сущностей наиболее высокого иерархического уровня для данного полного альтернативного графа. Набором целей назовем набор из сущностей, отмеченных как обязательные. Для каждого из маршрутных графов, исходя из набора параметров, описывающих сущности, включенные в данный МГ могут быть построены по выбираемому алгоритму наборы

характеристик МГ, опираясь на которые по алгоритмам теории принятия решений могут быть выбраны оптимальные маршрутные графы при заданном полном альтернативном графе и заданном наборе целей.

Набор выбранных миссий (конкретных проектов) и есть стратегия инновационной деятельности. В работе рассмотрен способ формализации выбора и уточнения стратегии на каждом из этапов развития инновационной, в частности космической деятельности.

Всякое дополнительное введение в рассмотрение новых целей, новых альтернатив их достижения, а значит, как правило, и новых сущностей приводит к модификации полного альтернативного графа, на котором в соответствии с выбранным алгоритмом определения эффективности маршрутных графов и набором ограничений, которые в совокупностями с параметрами описания сущностей (см. ниже) также могут уточняться на каждом из этапов, находится совокупность оптимальных миссий, т.е. оптимальная стратегия развития.

Параметры в условиях планирования инновационной деятельности характеризуются большими неопределённостями и, как правило, существенно уточняются по ходу реализации программы работ, следовательно, необходим отказ от однозначного плана работ и формирование специальным образом описанного обобщенного плана, охватывающего все возможные планы, совместимые с рассматриваемыми условиями плановой задачи, причем обобщенный план в рамках рассмотренного подхода должен автоматически уточняться по мере поступления новой информации, в том числе и при изменении системы целей. В нашем случае таким обобщённым планом является полный альтернативный граф – ПАГ.

Важным этапом построения модели является определение целей (рубежей), т.е. сущностей обязательных для достижения (посещения) на ПАГ. Для определения целей существует несколько подходов. В работе рассмотрены различные варианты определения рубежей. При создании тестовой программы для поддержки принятия решений в данном работе мы исходили из рассмотрения следующих типов рубежей:

- базовые рубежи (определённые как все сущности);
- априорные рубежи, (эти цели, т.е. сущности, обязательные для достижения, определяются априори, исходя из некоторых внешних условий и соображений, например, политического характера);
- промежуточные рубежи или рубежи принятия альтернативных решений (сущности, на которых происходит выбор использования соответствующих разрешающих векторов, а вместе с ними и выбор дальнейшей траектории маршрутного графа);
- верификационные рубежи или рубежи мониторинга, (директивные рубежи, определяемые для контроля и/или демонстрации достижений при реализации проекта).

В работе также рассматриваются возможные подходы к параметрическому описанию сущностей. Важно, что сущности могут иметь различную природу, а следовательно, могут и должны характеризоваться различными по своей природе параметрами. Поскольку сущности, включённые в полный альтернативный граф (ПАГ), в совокупности и создают инновационную инфраструктуру, то их параметрическое описание возможно в соответствии с принятой классификацией элементов инновационной инфраструктуры.

Для каждого из разделов классификации возможно выделение сущностей различного типа, характеризующихся различным набором параметров. На основании сформулированного и выбранного для проведения конкретного анализа набора требований выбираются наборы параметров, с помощью которых описывают свойства сущностей, включённых в ПАГ, и рассчитываются характеристики МГ.

В рамках предлагаемой нами информационной модели, представляется логичным ввести следующие обязательные критерии оценки инновационных проектов: оценка цены, оценка времени, затрачиваемого на проект, оценка вероятности успеха.

Но как уже отмечалось выше, наряду с основными критериями могут так же понадобиться различные технические политические и экономические характеристики. Возможно использование и других групп критериев.

Для решения проблемы многокритериального выбора оптимальных миссий предлагается использовать методы аналогичные методу анализа иерархий (МАИ) а именно, как и в МАИ строится матрица попарных сравнений между значениями критериев на основе экспертных оценок в единицах экспертной шкалы. Затем строится вектор приоритетов и эксперт ставит оценки каждому проекту (миссии) по нужному критерию. Затем в соответствии с методом МАИ может быть сформирована интегральная оценка для выбора оптимальных путей развития инновационной программы.

В соответствии с развитым подходом при использовании всех или части классификационных признаков для инфраструктуры инновационной деятельности формируется для выбранных целей полный альтернативный граф (ПАГ) сущностей с параметрическим описанием всех сущностей включённых в ПАГ. Таким образом, возможный набор конкретных планов развития инновационной (космической) деятельности представляется в виде однонаправленного графа, объединяющего нужные для достижения набора целей сущности (с необходимым набором параметров для реализации соответствующих миссий), разрешающие связи (вектора), и логические операторы, отвечающие причинно-следственным связям осуществления миссий и достижения в ходе их реализации соответствующих целей.

Сразу обратим внимание на то, что состав сущностей, включённых в полный альтернативный граф, наборы выбранных для их описания параметров и диапазон допустимых значений для каждого из этих параметров носят относительно субъективный характер и определяются с учётом не только объективных (определяемых на основе технологических требований и физических законов) значений соответствующих параметров, но также на основе субъективных мнений лиц принимающих решения, являющихся ответственными за проведение того или иного анализа (иногда с учётом также достаточно субъективных мнений экспертов-специалистов).

Заметим также, что параметрическое описание может быть на отдельных этапах проводимого анализа далеко не полным, то есть в рассмотрение может быть включена только часть сущностей и часть возможных параметров, а часть значений параметров может быть определена с невысокой точностью (например, в рамках некоторого, иногда достаточно широкого диапазона).

В предлагаемой нами информационной модели степень детализации описания сущности или (в соответствии с введённой нами терминологией) базового рубежа может быть различной. Так как выбор рассматриваемых сущностей (базовых рубежей) изначально относителен, то базовый рубеж может быть декомпозирован на составляющие его сущности (при том, как правило, альтернативными способами), т.е. на базовые рубежи (сущности) более низкого иерархического уровня. С другой стороны совокупность, некоторых рубежей может быть объединена в единую сущность, то есть в базовый рубеж более высокого иерархического уровня организации. Последний процесс называется кластеризацией.

Аналогичным образом степень детализации инновационных миссий также может быть различна, т.е. проект (миссия) может описываться с разной степенью подробности, включая или исключая различные варианты его реализации. В свою

очередь различные варианты реализации проекта фактически и/или формально могут при необходимости быть рассмотрены как отдельные миссии, конкурирующие за право реализации. С другой стороны при более «грубом» рассмотрении ряд вариантов проектов (которые можно рассматривать как различные самостоятельные миссии), направленных на достижение одной и той же сущности-цели (с точностью до детализации) может быть объединён в единую миссию достижения указанной цели.

В этой связи в работе рассмотрены алгоритмы и критерии кластеризации базовых и априорных рубежей инновационной инфраструктуры (сущностей её составляющих) и инновационных проектов (миссий) в параметрическом пространстве инфраструктуры для формирования необходимых для достижения рубежей развития.

Описан также пример программной реализации указанного подхода и предложенных алгоритмов, включая алгоритмы визуализации сложных альтернативных графов, для планирования развития космической деятельности.

О некоторых методах прогнозирования нестационарных временных рядов

А.А. Курбакова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Прогнозирование – определение тенденций и перспектив развития тех или иных процессов на основе анализа данных об их прошлом и нынешнем состояниях.

При решении научно-технических задач зачастую приходится сталкиваться с ситуациями, когда исследуемый процесс представлен нестационарным по какой-либо характеристике временным рядом. Обработка экспериментальных данных подобного характера усложняется.

Общая достоверность прогноза зависит не только от полноты статистической информации, но и от правильности выделения основных тенденций развития процесса.

Таким образом, можно выделить пять задач, решаемых специалистом при прогнозировании:

1. сбор данных,
2. редукция или уплотнение данных,
3. построение модели и её оценка,
4. прогноз,
5. оценка полученного прогноза.

Данные в зависимости от их модели можно разделить на

- стационарные,
- трендовые,
- сезонные,
- циклические.

В данной работе будем рассматривать сезонные модели.

Методы прогнозирования состоят в изучении и анализе полученных данных с целью нахождения моделей, которые могут быть эффективно продолжены в будущее.

Для прогнозирования временных рядов рассморим две модели: простое экспоненциальное сглаживание и экспоненциальное сглаживание с учетом тренда и сезонной компоненты (метод Винтерса).

Новый прогноз с использованием метода экспоненциального сглаживания рассчитывается по формуле

$$\hat{Y}_{t+1} = \alpha Y_t - (1-\alpha)\hat{Y}_t,$$

где

α - постоянная сглаживания ($0 < \alpha < 1$);

Y_t - наблюдение величины за текущий период *t*;

$\hat{Y}_t$ – предыдущий сглаженный прогноз величины на период *t*.

Постоянная сглаживания служит взвешивающим фактором. Если в прогнозе необходимо существенно учитывать величину ошибки последнего прогнозирования, то α необходимо задать ближе к 1. При малых значениях α прогнозируемая величина наиболее близка к предыдущему прогнозу. Одним из критериев выбора α является ее подбор, исходя из критерия минимизации среднеквадратической ошибки.

В качестве начального значения положим, что первое спрогнозированое значение равно первому значению наблюдаемой величины.

Новый прогноз на p периодов вперед с использованием метода Винтерса рассчитывается по формуле

$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p}$,

где

$L_t = \alpha \frac{Y_t}{S_{t-s}} - (1-\alpha)(L_{t-1} + T_{t-1})$ - экспоненциально сглаженные ряды,

$T_t = \beta(L_{t-1} + T_{t-1}) - (1-\beta)T_{t-1}$ - оценка тренда,

$S_t = \gamma \frac{Y_t}{L_t} - (1-\gamma)S_{t-s}$ - оценка сезонности,

α – постоянная сглаживания для этого уровня ($0 < \alpha < 1$);

β – постоянная сглаживания для оценки тренда ($0 < \beta < 1$);

γ – постоянная сглаживания для оценки сезонности ($0 < \gamma < 1$);

s – длительность периода сезонного колебания.

Для выбора постоянных сглаживания α, β и γ воспользуемся критерием минимизации среднеквадратической погрешности.

Существует несколько методов определения начальных значений. Воспользуемся следующим. Предположим, есть данные наблюдений за *k* сезонов по *M* периодов. Тогда индексы сезонности рассчитываются следующим образом:

$I_i = \frac{\sum_{i=1}^{M} \frac{Y_{ji}}{A_j}}{k}, 1 \le i \le M$ - индекс сезонности для *i*-го периода,

где

$A_j = \frac{\sum_{j=1}^{M} Y_{ji}}{M} 1 \le j \le k$ - среднее значение сезона,

Y_{ji} – наблюдение, соответствующее *i*-му периоду *j*-го сезона.

Для получения качественных результатов необходимо правильно рассчитать начальную тенденцию. Располагая данными только за один сезон, довольно сложно отличить тенденцию от сезонных колебаний. Таким образом, с точки зрения определения начального коэффициента тенденции, метод Винтерса требует наличия минимум *2M* наблюдений. Располагая данными за два сезона, можно выявить тенденцию, проводя сравнение соответствующих периодов сезонов.

Таким образом, оценка тенденции рассчитывается по следующей формуле:

$$b = \frac{1}{M}\left(\frac{Y_{M+1} - Y_1}{M} + \frac{Y_{M+2} - Y_2}{M} + \ldots + \frac{Y_{M+M} - Y_M}{M}\right).$$

Чем больше полных сезонов имеется в наличии, тем более точно можно задать начальные значения.

Для оценки ошибок прогнозов воспользуемся следующими методами:

- среднее абсолютное отклонение $MAD = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$,
- среднеквадратическая ошибка $MSE = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2$,

- средняя абсолютная ошибка в процентах $MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{Y_t}\cdot 100\%$,
- средняя процентная ошибка $MPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left(Y_t - \hat{Y}_t\right)}{Y_t}\cdot 100\%$.

Рассмотрим результат обработки реальных данных, полученных с сайта холдинга «ФИНАМ». Загрузим в автоматизированную систему данные котировок акций АО «ГАЗПРОМ» за два последних года, с 9 сентября 2008 года по 9 сентября 2010 года. Данные представляют собой нестационарный временной ряд без пропусков наблюдений.

С помощью подсистемы прогнозирования рассчитаем прогноз, используя описанные выше методы простого экспоненциального сглаживания и Винтерса. Результаты работы подсистемы представлены на рисунках ниже (см. рис. 1, 2).

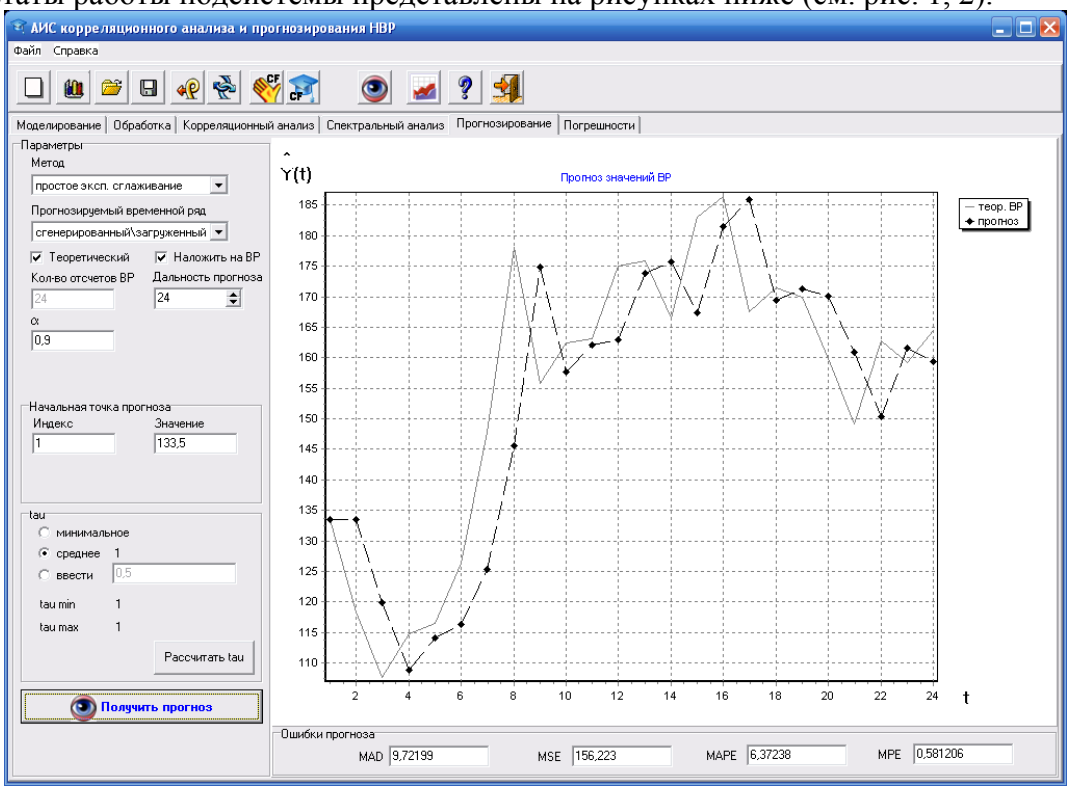


Рис.1.

Проведем анализ ошибок прогнозов.

Значение среднего абсолютного отклонения свидетельствует о том, что каждый прогноз рассматриваемых методов в среднем отличается от реального значения на 9,7 и 4,7 единиц соответственно.

Среднеквадратическая ошибка и средняя абсолютная ошибка в процентах необходимы для подбора метода прогнозирования и в данном случае свидетельствуют о том, что, используя метод Винтерса, можно рассчитать прогноз более точно.

Близость значения средней процентной ошибки к нулю говорит о том, что рассматриваемые методы не являются систематически переоценивающими или недооценивающими обрабатываемые данные.

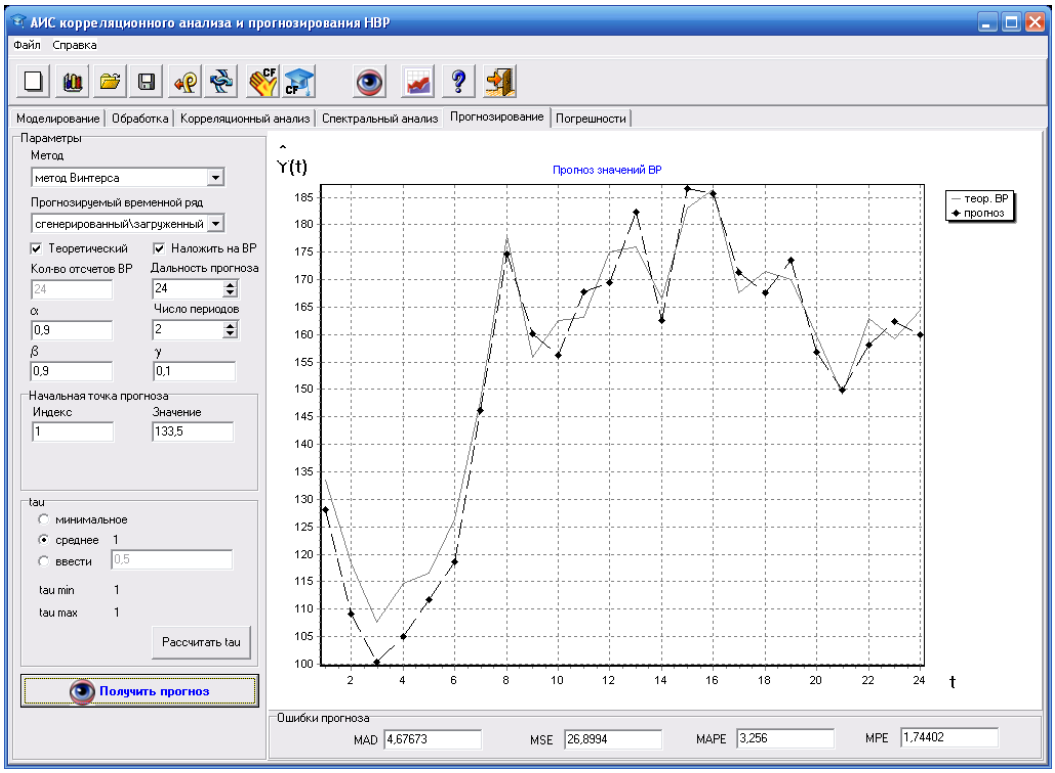


Рис.2.

Литература

1. *Прохоров, С.А.* Математическое описание и моделирование случайных процессов – Самарский государственный аэрокосмический университет, 2001 – 329 с.;
2. *Ханк Д.Э., Уичерн Д.У., Райтс А.Дж.* Бизнес-прогнозирование, 7-е издание: Пер. с англ. – М.Издательский дом «Вильямс», 2003 – 656 с.;
3. Холдинг «ФИНАМ» [Электронный ресурс] – http://www.finam.ru/analysis/export/default.asp.

Применение сети Вольтерри для решения задачи идентификации случайных последовательностей

И.В. Лёзина
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Одной из наиболее распространенных задач, решение которой требуется на практике, является задача идентификации – отнесение объекта из какой-либо предметной области к некоторому классу. Системы идентификации можно условно разделить на две большие группы – системы с явным заданием классификационных правил и системы с неявным заданием классификационных правил. Основное их отличие друг от друга состоит в том, что в системах с явным заданием классификационных правил существует база знаний правил. Заданные в явном виде правила, указывают, какие из поступивших сведений позволяют сделать тот или иной вывод. Сделанные выводы могут быть использованы другими правилами для получения выводов второго уровня; затем правилами для получения выводов третьего уровня и так далее.

Системы с неявным заданием классификационных правил не содержат заданные в явном виде классификационные правила, но включают в себя набор фактов о предметной области. Подобные системы способны самостоятельно сделать вывод о том, какие из представленных сведений позволяют сделать те или иные выводы.

Среди систем с неявным заданием классификационных правил наибольшее распространение получили нейронные сети. Наиболее распространена послойная реализация нейронной сети. В таком виде нейронная сеть – это последовательность слоев нейронов; выход каждого нейрона одного слоя поступает на взвешенный вход каждого нейрона последующего входа.

Основными достоинствами нейронных сетей являются параллельная обработка информации одновременно всеми нейронами, что обеспечивает значительное ускорение процесса обработки информации, нечувствительность к ошибкам, возникающих в отдельных контактах (благодаря большому количеству межнейронных соединений), способность к обобщению, за счет которого нейронная сеть вырабатывает ожидаемую реакцию применительно к данным, не обрабатывавшимся в процессе обучения. Нейронная сеть используется тогда, когда неизвестен точный вид связей между входами и выходами, то есть не работают классические методы. Эти достоинства нейронных сетей обеспечивают предпочтение нейронных моделей перед классическими методами.

Нейронная сеть Вольтерри относится к сетям со специализированной структурой. Это динамическая сеть для нелинейной обработки последовательности сигналов, задержанных относительно друг друга. Возбуждением для сети в момент t служит вектор $x = \left[x_t, x_{t-1}, ..., x_{t-L}\right]^T$, где L - количество единичных задержек, а $(L+1)$ означает длину вектора. В соответствии с определением ряда Вольтерри выходной сигнал y генерируется по формуле:

$$y(t) = \sum_{i_1=1}^{L} w_{i_1} x(t-i_1) + \sum_{i_1=1}^{L}\sum_{i_2=1}^{L} w_{i_1 i_2} x(t-i_1)x(t-i_2) + ... + \sum_{i_1=1}^{L}...\sum_{i_K=1}^{L} w_{i_1 i_2 ... i_K} x(t-i_1)x(t-i_2)...x(t-i_K) \quad (1)$$

где x обозначает входной сигнал, а веса $w_{i_1}, w_{i_2}, ..., w_{i_1 i_2 ... i_K}$, называемые ядрами Вольтерри, соответствуют реакциям высших порядков. Порядок полинома Вольтерри K также называется степенью ряда Вольтерри.

Пусть целевая функция для одного обучающего вектора выражается формулой:

$$E = \frac{1}{2}\big(y(t) - d(t)\big)^2 \tag{2}$$

В этом случае можно минимизировать значение целевой функции градиентными методами, которые сводятся к решению системы дифференциальных уравнений вида:

$$\begin{aligned}
&\frac{dw_{i_1}}{dt} = -\mu(y(t) - d(t))x(t - i_1),\\
&\frac{dw_{i_1 i_2}}{dt} = -\mu(y(t) - d(t))x(t - i_1)x(t - i_2),\\
&\frac{dw_{i_1 i_2 i_3}}{dt} = -\mu(y(t) - d(t))x(t - i_1)x(t - i_2)x(t - x_3),\\
&\frac{dw_{i_1 i_2 ... i_{K-1} i_K}}{dt} = -\mu(y(t) - d(t))x(t - i_1)x(t - i_2)...x(t - i_k).
\end{aligned} \tag{3}$$

Для упрощения структуры сети представленное разложение Вольтерри можно записать в следующей форме:

$$y_t = \sum_{i_1=0}^{L} x_{t-1}\left[w_{i_1} + \sum_{i_2=0}^{L} x_{t-i_2}\left[w_{i_1 i_2} + \sum_{i_3=0}^{L} x_{t-i_3}(w_{i_1 i_2 i_3} + ...) \right] \right] \tag{4}$$

где используются обозначения $y_t = y(t), x_{t-i_1} = x(t - i_1)$ и т.д. Каждое слагаемое в квадратных скобках представляет собой линейный фильтр первого порядка, в котором соответствующие веса представляют импульсную реакцию другого линейного фильтра следующего уровня. Количество уровней, на которых создаются фильтры, равно порядку K.

Система представляет собой структуру типичной многослойной однонаравленной динамической нейронной сети. Это сеть с полиномиальной нелинейностью. Подбор весов производится последовательно слой за слоем, причём эти процессы независимы друг от друга, и, увеличение числа весов в слое и числа самих слоёв в незначительной степени сказывается на обусловленности задачи. Это даёт возможность существенно увеличить длину L и порядок K системы при её практической реализации. Обучение нейронной сети Вольтерри лучше всего приводить с использованием технологии сопряжённых графов [1].

В соответствии с обозначениями, принятыми на рисунках, возбуждением сопряженного графа служит разностный сигнал $(y_t - d_t)$, где d_t обозначает ожидаемое, а y_t – фактическое значение в выходном узле системы в момент t. Формулы для определения конкретных компонентов вектора градиента имеют вид [2]:

$$
\begin{aligned}
&\frac{\partial E}{\partial w_{i_1}} = x_{t-i_1}(y_t - d_t),\\
&\frac{\partial E}{\partial w_{i_1 i_2}} = x_{t-i_2}\,\hat{y}_{i_1},\\
&\frac{\partial E}{\partial w_{i_1 i_2 i_3}} = x_{t-i_3}\,\hat{y}_{i_1 i_2},\\
&\frac{\partial E}{\partial w_{i_1 i_2 \ldots i_{K-1} i_K}} = x_{t-i_k}\,\hat{y}_{i_1 i_2 \ldots i_{K-1}}\\
&\hat{y}_{i_1} = (x_{t-i_1}(y_t - d_t))\\
&\hat{y}_{i_1 i_2} = (x_{t-i_2}\,\hat{y}_{i_1})\\
&\hat{y}_{i_1 i_2 i_3} = (x_{t-i_3}\,\hat{y}_{i_2}))
\end{aligned}
\qquad (5)
$$

В приведенных формулах сигналы, обозначенные символом ^, соответствуют сопряженному, а остальные – исходному графу системы. После определения конкретных компонентов градиента обучение сети с применением оптимизационного метода наискорейшего спуска может быть сведено к решению дифференциальных уравнений:

$$
\begin{aligned}
&\frac{dw_{i_1}}{dt} = -\mu\frac{\partial E}{\partial w_{i1}},\\
&\frac{dw_{i_1 i_2}}{dt} = -\mu\frac{\partial E}{\partial w_{i_1 i_2}},\\
&\frac{dw_{i_1 i_2 i_3}}{dt} = -\mu\frac{\partial E}{\partial w_{i_1 i_2 i_3}},\\
&\frac{dw_{i_1 i_2 \ldots i_{K-1} i_K}}{dt} = -\mu\frac{\partial E}{\partial w_{i_1 i_2 \ldots i_{K-1} i_K}}.
\end{aligned}
\qquad (6)
$$

где μ – коэффициент обучения.

Важным достоинством метода сопряженных графов считается простота учета равных значений весов в различных ветвях сети. Симметрия ядер Вольтерри приводит к равенству весов $w_{i_1 i_2 \ldots i_K}$ для всех перестановок индексов $i_1, i_2, \ldots, i_K$. Для двухиндексных весов это означает, что $w_{i_1 i_2} = w_{i_2 i_1}$. Подобные соотношения можно написать и для трёхиндексных и четырёхиндексных и так далее весов. Анализ обозначений весов сети, изображённой на рис. 1, позволяет найти веса, которые имеют одни и те же значения.

С учетом симметрии ядер Вольтерри выражения, описывающие компоненты градиента относительно весов $w_{i_1 i_2 \ldots i_K}$, могут быть определенным образом модифицированы [1]:

$$
\begin{aligned}
&\frac{\partial E}{\partial w_{i_1 i_2}} = x_{t-i_2}\,\hat{y}_{i_1} + x_{t-i_1}\,\hat{y}_{i_2},\\
&\frac{\partial E}{\partial w_{i_1 i_2 i_3}} = x_{t-i_3}\,\hat{y}_{i_1 i_2} + x_{t-i_2}\,\hat{y}_{i_1 i_3} + x_{t-i_3}\,\hat{y}_{i_2 i_1} + x_{t-i_1}\,\hat{y}_{i_2 i_3} + x_{t-i_2}\,\hat{y}_{i_3 i_1} + x_{t-i_1}\,\hat{y}_{i_3 i_2}
\end{aligned}
\qquad (7)
$$

В реализованном программном комплексе сеть Вольтерри идентифицирует случайные последовательности по законам распределения, структура сети задается

размерностью входного, выходного векторов и количеством слоев. Входным вектором являются частоты попадания анализируемой случайной величины в j-ый дифференциальный коридор p_j, количество коридоров M задаётся пользователем. Количество слоев K также задается пользователем. Размерность выходного вектора – число классов, по которым производится идентификация (число законов распределения с различными параметрами), параметрами алгоритма обучения являются коэффициент обучения μ и количество обучающих выборок N.

При работе пользователь задает число отсчетов для генерации случайной последовательности (объем выборки V). Затем по полученной случайной последовательности строится гистограмма, определяются частоты попадания анализируемой случайной величины в j-ый дифференциальный коридор, и производится идентификация.

Естественно полагать, что если тестировать сеть на том законе распределения (с конкретным параметром), на котором сеть не была обучена, результат идентификации будет ошибочным.

Алгоритм обучения для сети Вольтерри предназначен для подбора весовых коэффициентов нейросети Вольтерри по обучающим выборкам. В качестве входных данных алгоритм использует инициализированную структуру сети, объем обучающих выборок, коэффициент обучения. Обучение производится по формулам 5–7.

Обучим нейронную сеть, выбрав некоторые законы распределения с различными параметрами при различных значениях M, K, μ, N и V. Затем проведем по 3 исследования, генерируя по 100 раз каждый закон распределения с конкретным параметром. Данные, полученные в ходе исследований, сведены в табл. 1 – 4.

Таблица 1. Зависимость качества идентификации от количества слоев при M=20, μ=0,25, N=100 и V= 1000

Название закона распределения	% правильного определения				
	K = 2	K = 3	K = 4	K = 5	K = 6
Равномерный	97	100	99	96	88
Симпсона	96	99	95	90	85
Арксинуса	98	99	97	92	90

Как видно из табл. 1, качество идентификации зависит от количества слоев. Подборка количества слоёв является самой сложной задачей при проектировании нейронной сети и для каждого случая решение является уникальным. В нашем случае остановимся на K = 3.

Таблица 2. Зависимость качества идентификации от количества обучающих выборок при M=20, μ=0,25, K=3 и V= 1000

Название закона распределения	% правильного определения				
	N = 130	N = 160	N = 190	N = 220	N = 250
Равномерный	94	98	100	100	100
Симпсона	97	98	99	100	100
Арксинуса	95	97	98	99	100

Чем больше количество обучающих выборок, тем лучше результат идентификации чего и следовало ожидать. Но при увеличении количества выборок также возрастает и сложность расчетов, поэтому их необоснованное увеличение нецелесообразно. В данном случае N = 220 вполне достаточно.

Таблица 3. Зависимость качества идентификации от коэффициента обучения при *M*=20, *K*=3, *N*=220 и *V*= 1000

Название закона распределения	% правильного определения				
	μ = 0,1	μ = 0,25	μ = 0,35	μ = 0,75	μ = 0,9
Равномерный	94	100	100	99	95
Симпсона	97	99	99	98	93
Арксинуса	95	100	98	95	94

Как видно из результатов исследований, качество идентификации зависит от значения коэффициента обучения μ. При увеличение коэффициента происходит «расшатывание системы». Этого и следовало ожидать, ведь по правилу корректировки весов коэффициент обучения влияет на скорость схождения алгоритма, так как количество обучающих выборок ограниченно алгоритм заканчивает свое действие, не достигнув оптимального решения. Однако выбор слишком маленького коэффициента замедлит скорость обучения сети, поэтому необходимо подбирать оптимальное решение, остановимся на μ = 0,25.

Как видно из табл. 4, результаты идентификации зависят от размерности входного вектора. Это можно объяснить тем, что при увеличении числа коридоров для построения гистограммы происходит растяжение графика плотности распределения, что вносит дополнительную погрешность при идентификации. В нашем случае остановимся на *M*=20.

Таблица 4. Зависимость качества идентификации от размерности входного вектора при *K*=3, μ =0,25, *N*=220 и *V*= 1000

Название закона распределения	% правильного определения			
	M = 10	M = 20	M = 40	M = 80
Равномерный	99	100	98	97
Симпсона	100	100	100	99
Арксинуса	99	100	99	97

Литература

1. Осовский, С. Нейронные сети для обработки информации [Текст]/С. Осовский. – М.: Финансы и статистика, 2002. – 344 с.

2. Солдатова, О.П. Основы нейроинформатики [Текст] / Солдатова О.П. // СГАУ, 2006 г. – 131 с.

Система научной визуализации химических процессов

А.В. Смирнов, Л.С. Зеленко, В.В. Варфоломеева, А.В. Терентьев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Как показывает практика, при проведении научных исследований часто встречаются задачи анализа различных данных. В решении таких задач большую роль играет *метод научной визуализации*, который находит широкое применение, как в теоретических, так и экспериментальных исследованиях. *Целью* использования визуальных компьютерных моделей является перевод набора формальных определений в наглядные геометрические свойства, за которыми можно наблюдать, динамически меняя параметры. Последнее дает возможность непосредственно увидеть свойства и взаимосвязи, что приводит к формулировке гипотез, далеко не очевидных без компьютерного эксперимента и геометрической визуализации [1]. Таким образом, задача визуализации является *актуальной* в современной науке.

Бурный прогресс вычислительной техники и развитие программного обеспечения привели к научной революции в области компьютерного моделирования в химии. Теперь можно теоретически исследовать неизвестные молекулы, интермедиаты, переходные состояния в химических реакциях и другие, не менее экзотические, объекты.

Применение Z-матриц для представления геометрической структуры молекулы

Для проведения различного вида расчётов молекулу удобно представлять в графическом виде, задавая геометрические параметры молекулы, в первую очередь межатомные расстояния r, валентные углы α и торсионные (двухгранные) углы τ (рис. 1).

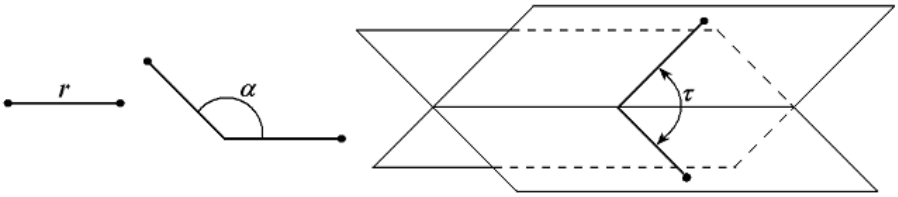


Рис. 1. Основные структурные параметры молекул

Одним из способов представления координат атомов молекулярной системы является *Z-матрица,* которая задает внутренние координаты молекулы. Это представление определяет каждый атом системы через атомный номер n, длину связи r, валентный угол α и двугранный угол τ. Под связью в данном случае подразумевается не химическая связь, а просто вектор, направленный от одного атома к другому, хотя они могут и совпадать. Тем не менее, принято записывать Z-матрицу через длины и углы химических связей, так как такая запись позволяет описать не только относительное расположение атомов друг относительно друга, но и связи этих атомов. Z-матрица называется так потому, что второй атом всегда располагается вдоль оси аппликат (оси Z).

Z-матрицы используются для описания молекулярных систем и принимаются многими программами моделирования молекулярной динамики и вычислительной химии. Кроме того, многие программы визуализации молекулярной структуры предоставляют графические редакторы Z-матриц [2].

Современные технологии визуализации

Современные графические адапторы позволяют использовать сложные алгоритмы визуализации, позволяющие добиваться реалистичного и эффектного изображения. В системе применяются следующие методы и подходы визуализации:

– попиксельное освещение по Блинну-Фонгу;
– антиалиасинг — устранение «зубчатых» границ полигонов;
– ray-tracing — техника вычисления пути луча света в сцене, учитывающая многочисленные отражения луча;
– мягкие тени;
– эффект свечения.

Попиксельное освещение по методу Блинна-Фонга

Модель отражения Фонга является довольно популярной и широко используемой в компьютерной графике [3]. Каждая точка поверхности имеет свои координаты и в ней определена нормаль к поверхности. Её освещенность складывается из трех компонент: фоновое освещение (ambient), рассеянный свет (diffuse) и бликовая составляющая (specular). Свойства источника определяют мощность излучения для каждой из этих компонент, а свойства материала поверхности определяют её способность воспринимать каждый вид освещения:

фоновая составляющая на рис. 2, рассеянная составляющая на рис. 3, зеркальная составляющая на рис. 4, суммарная составляющая на рис. 5.



Рис. 2. Фоновая составляющая

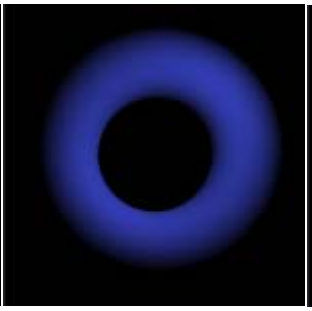


Рис. 3. Рассеянная составляющая

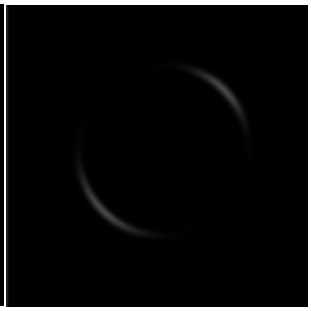


Рис. 4. Зеркальная составляющая



Рис. 5. Суммарное освещение

Фоновое освещение

Это постоянная в каждой точке величина надбавки к освещению. Вычисляется фоновая составляющая освещения как:

$$I_a = k_a i_a , \qquad (1)$$

где i_a - фоновая составляющая освещенности в точке, k_a – свойство материала воспринимать фоновое освещение, I_a – мощность фонового освещения.

Рассеянный свет

При попадании на поверхность свет рассеивается равномерно во все стороны. При расчете такого освещения учитывается только ориентация поверхности (рис. 6) и направление на источник света. Рассеянная составляющая рассчитывается по закону косинусов - закон Ламберта

$$I_d = k_d \cos\left(\vec{L}, \vec{N}\right) i_d = k_d \left(\vec{L} \cdot \vec{N}\right) i_d \qquad (2),$$

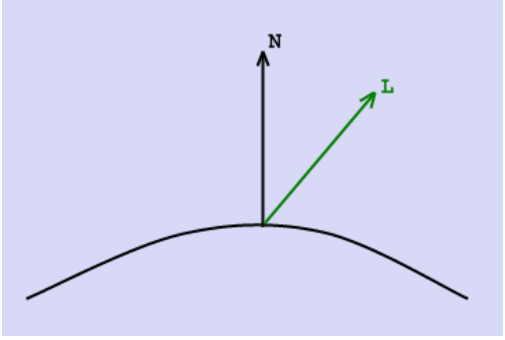


Рис. 6. Нормаль и отраженный луч

где I_d – рассеянная составляющая освещенности в точке, k_d – свойство материала воспринимать рассеянное освещение, i_d – мощность рассеянного освещения, $\vec{L}$ – направление из точки на источник, $\vec{N}$ - вектор нормали в точке.

Зеркальный свет

При попадании на поверхность подчиняется следующему закону: «Падающий и отраженный лучи лежат в одной плоскости с нормалью к отражающей поверхности в точке падения, и эта нормаль делит угол между лучами на две равные части» (рис. 7). Таким образом, отраженная составляющая освещенности в точке зависит от того, насколько близки направления на наблюдателя и отраженного луча. Это можно выразить следующей формулой:

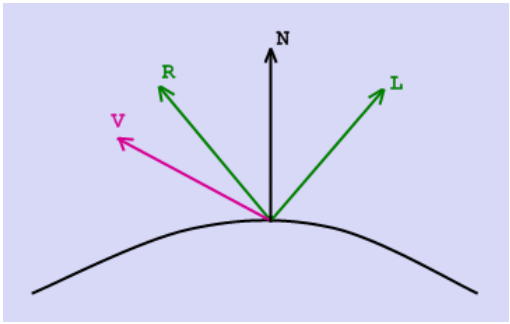


Рис. 7. Зеркальная составляющая

$$I_s = k_s \cos^{\alpha}\left(\vec{R},\vec{V}\right) i_s = k_s \left(\vec{R}\cdot\vec{V}\right)^{\alpha} i_s \qquad (3)$$

где I_s – зеркальная составляющая освещенности в точке, k_s – коэффициент зеркального отражения, i_d – мощность зеркального освещения, $\vec{R}$ – направление отраженного луча, $\vec{V}$ - направление на наблюдателя, α - коэффициент блеска, свойство материала.

Именно зеркальное отражение представляет наибольший интерес, но в то же время его расчет требует больших вычислительных затрат.

Антиалиасинг

Сглаживание - технология, использующаяся в обработке изображений с целью сделать границы кривых линий более гладкими, убирая «зубцы», возникающие на краях объектов. Основной принцип сглаживания - использование возможностей устройства вывода для показа оттенков цвета, которым нарисована кривая. В этом случае пикселы, соседние с граничным пикселом изображения, принимают промежуточное значение между цветом изображения и цветом фона, создавая градиент и размывая границу [4].

В системе применяется два варианта сглаживания:

– общее сглаживание отрисовкой излишне крупного несглаженного изображения с последующим уменьшением разрешения.

– специализированные алгоритмы сглаживания, работающие на изображениях определённого типа (например, алгоритм Ву для отрисовки отрезков).

Ray-tracing

Трассировка лучей - один из методов геометрической оптики, применяемый для исследования оптических систем путём отслеживания взаимодействия отдельных лучей с поверхностями [5]. В узком смысле – это технология построения изображения трёхмерных моделей в компьютерных программах, при которых отслеживается обратная траектория распространения луча (от экрана к источнику).

Возможности системы визуализации химических процессов

Разработанная система предназначена для научной визуализации химических процессов и представляет собой по существу редактор молекул, имеющий графический интерфейс и позволяющий создать новые молекулярные соединения и редактировать существующие. Кроме того система предоставляет пользователю возможности

визуального восприятия 3-х мерных изображений молекул и молекулярных структур, дает средства для анимации протекания химических процессов.

Для решения перечисленных задач в системе предусмотрены следующие возможности:

– добавление новых молекул в проект, задание их структуры с помощью встроенного редактора Z-матрицы и редактора связей в молекуле, подгрузка готовых конформаций молекул из химических файлов, созданных в других системах (gzmat, xyz, pdb, games и др.);

– редактирование связей в молекуле путем последовательного выделения двух атомов и задания типа связи между ними (одинарные, двойные, полуторные, тройные, водородные);

– задание позиции просмотра сцены на каждом кадре анимации, для удобства использования эта возможность реализована с помощью отдельного режима навигации по сцене, когда пользователь сможет управлять направлением взгляда камеры с помощью мыши, а положением в пространстве с помощью клавиатуры;

– редактирование положения и поворота каждой из молекул в сцене задается пользователем на каждом этапе анимации: пользователь указывает на атом молекулы, относительно которого будут происходить дальнейшие преобразования, затем выбирает требуемое действие – смещение или поворот.

– сохранение результирующего проекта анимации в виде видео-файла.

Выводы

Современные программные системы, построенные на высокопроизводительной вычислительной технике, становятся более надежными, удобными в работе и позволяют использовать новейшие информационные технологии для решения задач, связанных с обработкой больших объемов информации, их визуализации и управления. Разработанная система всецело использует возможности современной вычислительной техники и позволяет получить наглядное представление об исследуемых химических процессах.

Литература

1. Смирнов, А.В. Разработка средств компьютерной визуализации геометрической структуры молекулы для моделирования явлений физической адсорбции [Текст] / А.В. Смирнов // Труды XLVIII Международной научной студенческой конференции «Студент и научно-технический прогресс» / Новосибирский гос. ун-т . – Новосибирск, 2010. – С. 253.

2. Молекулярное моделирование. Z-матрица [Электронный ресурс] / ru.wikipedia.org/wiki.

3. Кулагин, Д. Закон Ламберта. Закон отражения Фонга. Закон отражения Блинна-Фонга [Электронный ресурс] / http://compgraphics.info/3D/lighting/phong_reflection_model.php // раздел Компьютерная графика. – 2007.

4. Антиалиасинг [Электронный ресурс] / http://ru.wikipedia.org/wiki/Антиалиасинг // раздел Трехмерная графика.

5. Трассировка лучей [Электронный ресурс] / http://ru.wikipedia.org/wiki/Трассировка_лучей // раздел Трехмерная графика.

Разработка пакетов прикладных программ и использующих их приложений на основе онтологического подхода

С.В. Смирнов
Институт проблем управления сложными системами РАН, Самара

Введение

Известно, что ценность любого «цивилизованного» пакета прикладных программ (ППП) состоит, прежде всего, в предоставлении пользователю развитой концепции для описания ситуаций в некоторой предметной области (ПрО), а уже затем – в собственно палитре, «пакете» способов преобразования ситуаций в нужном направлении, интерпретируемом как решение задач. Фактически этот взгляд абсолютизировали еще пионеры искусственного интеллекта в знаменитом «Общем решателе задач» (General Problem Solver, GPS). А именитый ученый-программист А.П. Ершов неоднократно формулировал мысль о ППП как методологии решения прикладных задач. Следует также отметить, что идею первичности представления знаний о ПрО (точнее, о нескольких ПрО) при создании программных систем последовательно реализуют CASE-средства.

Подобный анализ приводит к видению ППП как системы, базирующейся на знаниях, основная функция которой – формирование для специалиста предметно-ориентированной среды моделирования (ПОСМ) с некоторым работающим в ней арсеналом методов решения задач или ведения научных исследований. В этом смысле для разработчиков ППП вполне мотивирован интерес к онтологиям, которые сегодня признаются одной из конструктивных форм структурирования знаний о ПрО.

В докладе исследуется различные аспекты разработки ППП и использующих их приложений на основе онтологий.

Онтология как объектная модель предметной области

Имеются различные варианты толкования онтологии; обычно ей отводится место посредника между тезаурусом и той или иной формальной моделью ПрО, но практический опыт дает все основания считать, что онтология и есть модель ПрО. Пессимизм в отношении самой возможности построения онтологий выражает известный философский тезис об онтологической относительности. На наш взгляд, одним из достаточно строгих решений этой проблемы может служить прагматическая интерпретация онтологии как объектно-ориентированной модели данных ППП и использующего его приложения.

Имеется широкий консенсус в том, что при формулировке знаний, составляющих содержание научных теорий и инженерных дисциплин, фундаментальную роль играют два когнитивных суждения о мире: наличие (возможность различения) дискретных объектов и существование отношений между ними. Множество объектов, рассматриваемых в контексте всякой задачи как приложения некоторой теории, образует ПрО задачи. На этой основе (т.е. на множестве объектов ПрО) задаются отношения: унарные, которые интерпретируются как имманентные агрегируемые свойства (или атрибуты, функции) объектов, и произвольной арности, или собственно отношения, которые описывают различные ассоциации объектов. Мощность моделирования, присущая бинарным отношениям, позволяет ограничиться ими при описании ПрО, и тогда возникает возможность говорить о наличии в ПрО связей между отдельными парами объектов. Картина ПрО в виде сети связанных объектов в теории представления знаний известна как семантическая сеть, наглядным изображением которой служит раскрашенный орграф. Вершины сети, представляющие объекты ПрО, распадаются на две категории:

экземпляры и классы, – которые находятся в бинарном отношении экземплификации. Подсеть, состоящая из вершин-классов и соединяющих эти вершины дуг, описывает понятийную структуру ПрО, т.е. задает онтологию ПрО. Подсеть вершин-экземпляров с эпистемологической точки зрения представляет собой денотативную объектную модель ПрО, в которой типы вершин и дуг определяются онтологией ПрО.

При решении задач онтология представляет теорию, а денотативная модель конкретизирует эту теорию применительно к актуальной ситуации в моделируемой ПрО. Множественность таких ситуаций делает выгодным:

- иметь раздельные спецификации для онтологической и денотативной компонент объектной модели ПрО;
- реализовывать отношение экземплификации «внемодельными», технологическими методами представления знаний в компьютере;
- использовать при решении задач понятие контекста моделирования как актуальной пары ⟨онтология ПрО, денотативная модель ПрО⟩.

В качестве термина, объединяющего представления об онтологиях и денотативных моделях ПрО, пригодно словосочетание «онтологические модели», которые однородны в том смысле, что они являются денотативными объектными моделями своих особенных предметных областей. Для всякой денотативной модели это ПрО, в которой реализуются объемы понятий, описываемых некоторой определенной онтологией. Для онтологий – это ПрО, в которой располагаются, прежде всего, объемы понятий «класс объектов» и «атрибут», а онтология, где эти понятия описаны, т.е. «онтология онтологий», или метаонтология, должна быть семантически замкнута, вполне описывая саму себя. В этом и состоит «технологическая реакция» на онтологическую относительность описания реального мира.

Парадигма создания и эксплуатации ППП: автоматизированное конструирование предметно-ориентированных сред моделирования

Понятие ПОСМ является естественным развитием концепции предметно-ориентированных ППП и систем моделирования, обобщающим эволюцию известных схем автоматизации программирования. Состав и назначение системы эпистемологических единиц ПОСМ, а также порядок их порождения и использования определяется единством и логической определенностью осуществляемого любой отраслью науки познавательного процесса. Другими словами, решение задачи с помощью ЭВМ как процесс включает в себя ряд ключевых этапов, определяющих совместно с продуктами их выполнения общую схему обработки информации при создании и использовании прикладных ППП.

Онтология как концептуальная (К-) модель ПрО является продуктом этапа концептуального (онтологического) анализа, основным феноменом которой является возможность формирования и сосуществование нескольких различных К-моделей для данной ПрО. При этом основу К-модели составляет понятийная структура ПрО - таксономические иерархии концептов ПрО, где свойства и способности каждого концепта описываются с помощью механизма «атрибут-значение» и могут быть установлены специальные (т.е. отличные от таксономической) связи между концептами. Именно этот этап резонно сопоставить с разработкой ППП а, следовательно, ППП отождествить с онтологией (здесь возможны исключения). С прагматических позиций важно то, что одновременно с декларативным описанием актуальной ПрО онтология непосредственно задает:

- совокупность абстрактных типов данных, необходимых при разработке алгоритмов (а затем программ) для решения задач;
- схему объектно-ориентированной базы данных для размещения моделей объектов из соответствующей ПрО.

Остальные этапы обработки информации в общей схеме имеют характер синтеза:

- результатом денотативного (Д-) моделирования является отражение, воспроизведение, описание конкретного объекта или системы объектов ПрО, выполненное в рамках избранной онтологии (т.е. на «языке» этой К-модели, или «языке» ПрО). Получаемая здесь модель суть экзогенная объектная модель ПрО;
- параллельно Д-моделированию могут конструироваться трансформационные (Т-) модели, реализующие знания о допустимых «вычислениях», связанных с актуальными Д-моделями. В привычном смысле именно этому этапу соответствует разработка приложения, использующего ППП, т.е. спецификация сценария решения задачи, описания воздействий на экзогенную объектную модель, в результате которых она приобретает некоторые удовлетворяющие пользователя свойства. Операционным базисом подобных трансформаций, аппаратом решения служит исчисление, основу которого составляют функциональные части К-моделей, а множественность Т-моделей закономерно отражает как множество различных задач в каждой ПрО, так и множество способов их решения;
- на этапе собственно вычислений, или вычислительного эксперимента (прогон программного приложения, использующего ППП, В-моделирование), реализуются (интерпретируются физической или виртуальной машиной) знания о возможных трансформациях экзогенной денотативной модели ПрО в ее эндогенную Д-модель, которая с общих позиций и есть результат работы приложения.

Перечисленные модели (т.е. специфические формы ППП, использующих его приложений и получаемых результатов) составляют систему эпистомологических единиц ПОСМ. Однородность (на уровне представления в базах данных) всех перечисленных моделей обеспечивают две априори заданные К-модели ПОСМ:

- метаонтологии для конструирования онтологий ПрО;
- онтологии приложений, описывающей их на уровне, абстрагированном от конкретного содержания решаемых задач.

Интеграция знаний при разработке ППП на основе онтологического подхода

Созданные теоретические основы и инструментарий онтологического анализа предоставляют эффективные подходы к решению проблемы интеграции разнородных знаний при разработке ППП и использующих их приложений.

На этапе онтологического анализа ПрО наиболее значимыми являются следующие приемы такой интеграции:

- возможность конструирования понятийных структур ППП на основе экспериментальных данных о ПрО расширенными методами формального анализа понятий;
- формирование новых понятийных структур путем специализации или композиции имеющихся.

Другие способы интеграции знаний могут быть использовании при разработке использующих ППП приложений (т.е. в трансформационных моделях) и реализованы при выполнении вычислений:

- одновременная работа в ходе решения задачи с несколькими объектными моделями ПрО, представленной некоторой одной онтологией;
- совмещение нескольких взглядов на актуальную ПрО, которые выражены в приложении объектными моделями, построенными согласно различным онтологиям этой ПрО;

- совместное использование объектных моделей из разных ПрО, включая модели, конкретизирующие метод решения задачи (речь идет о том, что в методическом плане целесообразно различать предметно- и методо-ориентированные ППП);
- возможность построения приложений, автоматизирующих программирование ППП и использующих их приложений, на основе соответствующих метамоделей;
- организация взаимосвязанных вычислительных экспериментов для реализации альтернативных и эволюционных исследований ПрО.

Все эти методы интеграции обеспечиваются двумя «контурами» управления моделями различных типов на этапе вычислений:

- первый, задачный (т.к. его обеспечивают методы онтологии задач), контур путем переключения контекстов моделирования - надлежащего определения, «текущей» пары ⟨К-модель, Д-модель⟩;
- второй, проектный, контур не является обязательным и связан с макроуправлением вычислениями. Для его поддержки необходимо специальное приложение – супервизор вычислений, который организует под управлением пользователя взаимосвязанные серии прогонов программ, структурируя сведения о выполненных вычислениях в форме растущего ациклического графа. Такая структура, по сути, определяет отдельный проект по обработке информации, и указанный супервизор, контролируя «правильный» рост этой структуры, обеспечивает транзакционные свойства приложений ППП.

Реализация предложенной концепции

В ИПУСС РАН в течение ряда лет развивается проект по созданию на массовой вычислительной платформе программного инструментария (*gB*-система), реализующего разработку ППП и использующих их приложений на основе онтологического подхода.

Создан ряд экспериментальных и промышленных программных систем (включая и инструментальные средства, и специализированные прикладные интеллектуальные системы) в машиностроении, экологии, организационного управления.

Выводы

При создании и эксплуатации ППП разработчик и пользователь оперируют онтологиями как основной разновидностью эпистемических и соответствующих им программно-технических единиц. Разработчику предлагается использовать расширенную трактовку приложений, использующих ППП, как предметно-ориентированных сред моделирования. Конструирование подобных сред допускает создание средств автоматизации программирования и управления вычислениями, эффективно мобилизующих когнитивные и интеллектуальные ресурсы человека.

В целом онтологический подход дает достаточно ясный взгляд на состав, назначение и структуру программно-технических компонентов и методов управления ими в ходе создании и эксплуатации ППП. При этом охватывается большинство содержательных проблем разработки прикладного программного обеспечения: организации системы знаний об актуальной ПрО и о способах решения задач в этой ПрО, планирования решения задач и управления вычислениями, методологии отчуждения ППП от разработчика. Решения всех этих проблем оказываются в высокой степени унифицированными, а используемые программно-технические составляющие ППП и использующих их приложений в высокой степени однородными.

Графо-аналитический метод моделирования сигма-дельта преобразователей информации замкнутой структуры

Б.В. Чувыкин, А.В. Селезнев, Р.Г. Тер-Аракелян
ГОУ ВПО «Пензенский государственный университет», Пенза

Сигма-дельта преобразователи информации замкнутой структуры широко используются в информационно-измерительных системах благодаря наличию уникальных свойств: высокой точности и линейности функции преобразования (24-разрядный сигма-дельта АЦП), технологичности (возможность реализации по технологии «система на кристалле»), большому разнообразию структурно-алгоритмических решений [1]. Однако, при их проектировании возникает проблема аналитического описания динамических процессов в замкнутой структуре, состоящей из аналоговой части (интегрирующие звенья), цифровой части (цифровые фильтры) и однобитного аналого-цифрового и цифро-аналогового преобразователя. Ниже, на рис. 1 представлен пример структуры сигма-дельта АЦП с дельта-модулятом 3-го порядка.

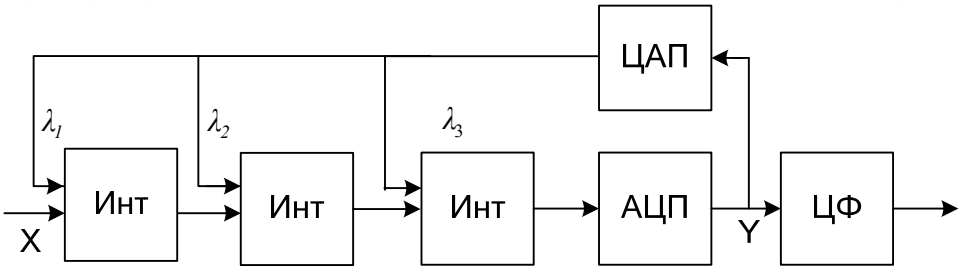


Рис. 1

Сложность описания связана с нелинейным преобразованием в однобитном АЦП-ЦАП. На практике данная проблема решается методом компьютерного моделирования с использованием известных алгоритмов оптимизации динамических процессов в многомерном пространстве дискретного времени. Формально задача сводится к решению систем нелинейных уравнений высокого порядка. На этом этапе необходимо аналитически записать уравнения состояний инерционных звеньев (интеграторов) преобразователей информации (ПИ) замкнутой структуры. Предлагается для решения этой задачи использовать графо-аналитический метод [2].

Ниже приводятся основные элементы и процедуры моделирования, необходимые для построения графа и перехода к системе нелинейных уравнений, описывающих динамические процессы в сигма-дельта ПИ замкнутой структуры.

Если предположить, что в каждом цикле преобразования входная величина X_0 не изменяется, то в дискретной модели состояние интеграторов может быть задано в дискретные моменты времени, совпадающие с шагом дискретизации импульсного элемента (АЦП) и описывается разностным уравнением:

$$U_{k[n]} = U_{k[n-1]} + \varepsilon X_{[n-1]} \qquad (1)$$

где $U_{k[n]}$ – напряжение на выходе k-го интегратора в момент времени $t_n = nh$, где h – шаг дискретизации, n – номер шага; $\varepsilon = \dfrac{h}{\tau}$ – относительная постоянная времени интегратора; $X_{[n-1]}$ – напряжение на входе интегратора на предыдущем шаге дискретизации.

Переходя от итерационного уравнения к операторной форме, получим:

$$U_k = U_k z^{-1} + \varepsilon X z^{-1} \qquad (2)$$

Такой подход позволяет каждый интегратор заменить сумматором. Используя теорию графов, сумматор можно представить в виде следующей структуры (рис. 2).

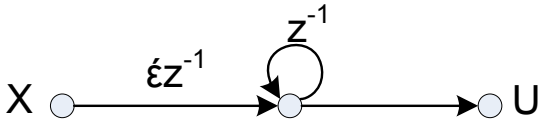


Рис. 2.

В качестве примера на рис. 3 приведена математическая модель сигма-дельта модулятора первого порядка в виде графа.

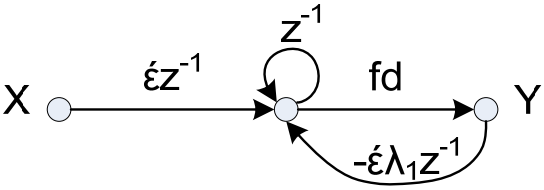


Рис. 3.

Операция **fd** – дискретизация и формирование импульсов обратной связи, которая описывается следующим выражением:

$$y(t) = \Big|_{t \in (h_{(n-1)};\, h_{(n)}))} = U_{1[n]} \tag{3}$$

Для сигма-дельта модулятора с однобитной обратной связью выражение (3) примет следующий вид:

$$y(t) = \Big|_{t \in (h_{(n-1)};\, h_{(n)}))} = U_0 sign(U_{1[n]}) \tag{4}$$

Такой подход позволяет достаточно просто перейти от функциональной схемы ПИ с нелинейной обратной связью к соответствующему графу и системе разностных уравнений, описывающих динамические процессы в дискретном пространстве времени.

Особенностью этих структур является то, что реакция на входной сигнал «распространяется» по всей цепочке инерционных звеньев в прямой цепи преобразования, это относится и к сигналам в цепи обратной связи.

Поясним на конкретном примере. Цепочка из *k*–интеграторов, имеющих одинаковую постоянную времени τ, реагирует на входное ступенчатое воздействие в виде монотонно нарастающей функции, которая описывается параболой *k*–го порядка:

$$U_k(t) = \frac{t^k X_0}{k!\tau^k} \tag{5}$$

Следовательно, в конце цикла преобразования $t = h$ и каждый из интеграторов находится в состоянии:

$$U_{k[1]} = \frac{\varepsilon^k X_0}{k!} \tag{6}$$

Используя принцип суперпозиции, для следующего такта преобразования можно воспользоваться уравнением (6) если допустить, что каждый из интеграторов имеет ненулевое начальное состояние, равное конечному состоянию предыдущего такта. Рассматривая X_0 как входной сигнал, приложенный к первому интегратору или сигнал обратной связи, приложенный к соответствующему интегратору и учитывая масштабные коэффициенты обратной связи λ_1, λ_2, …, λ_k можно получить систему разностных уравнений в общем виде:

$$\begin{cases} U_{1[n]} = U_{1[n-1]} + \varepsilon X_{[n-1]} - \varepsilon\lambda_1 U_{OC[n-1]} \\ U_{2[n]} = U_{2[n-1]} + \varepsilon U_{1[n-1]} - \varepsilon\lambda_2 U_{OC[n-1]} - \dfrac{\varepsilon^2}{2!}\lambda_1 U_{OC[n-1]} + \dfrac{\varepsilon^2}{2!} X_{[n-1]} \\ \vdots \\ U_{k[n]} = U_{k[n-1]} + \varepsilon U_{k-1[n-1]} + \dfrac{\varepsilon^2}{2!} U_{k-2[n-1]} + \ldots + \dfrac{\varepsilon^{k-1}}{(k-1)!} U_{1[n-1]} - \varepsilon\lambda_k U_{OC[n-1]} \\ - \dfrac{\varepsilon^{k-1}}{(k-1)!}\lambda_{k-1} U_{OC[n-1]} - \ldots - \lambda_1 \dfrac{\varepsilon^k}{k!} U_{OC[n-1]} + \dfrac{\varepsilon^k}{k!} X_{[n-1]} \end{cases} \quad (7)$$

Поскольку ПИ является нелинейной системой, функциональная связь между сигналом обратной связи и сигналом с выхода *k*–го интегратора определяется в зависимости от вида нелинейности, следовательно система из *k*–уравнений (7) должна быть дополнена уравнением:

$$U_{\mathrm{OC}[n-1]} = U_0 \cdot sign\left(U_{k[n-1]}\right) \quad (8)$$

Рассмотрим применение данного метода при анализе сигма-дельта модулятора 3–го порядка. Его математическая модель в виде графа показана на рисунке 4:

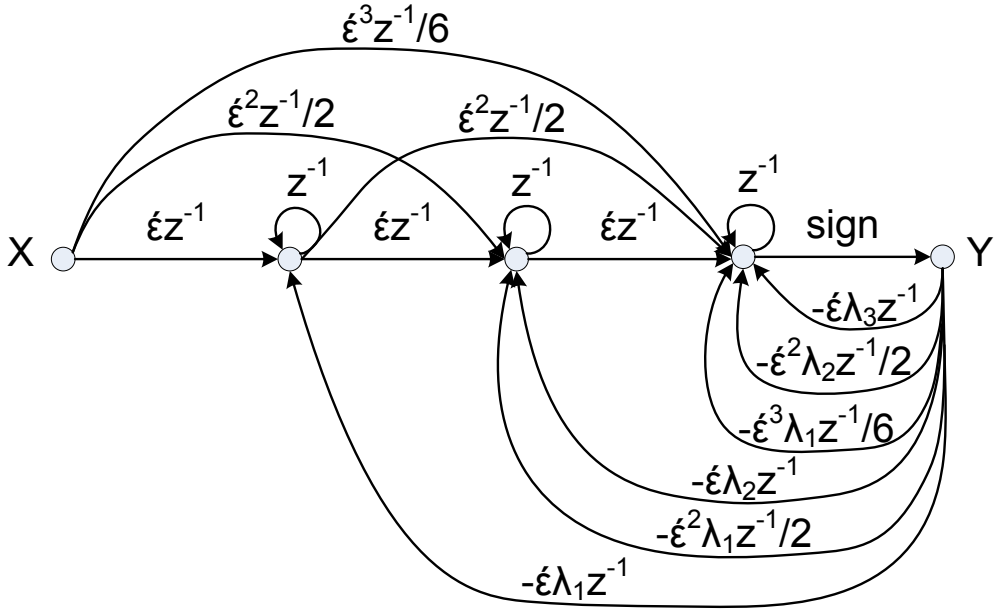


Рис. 4.

Математическая модель сигма-дельта модулятора 3–го порядка в виде системы разностных уравнений будет иметь следующий вид (9).

$$\begin{cases} U_{1[n]} = U_{1[n-1]} + \varepsilon X_{[n-1]} - \varepsilon\lambda_1 U_{OC[n-1]} \\ U_{2[n]} = U_{2[n-1]} + \varepsilon U_{1[n-1]} - \varepsilon\lambda_2 U_{OC[n-1]} - \dfrac{\varepsilon^2}{2!}\lambda_1 U_{OC[n-1]} + \dfrac{\varepsilon^2}{2!} X_{[n-1]} \\ U_{3[n]} = U_{3[n-1]} + \varepsilon U_{2[n-1]} + \dfrac{\varepsilon^2}{2!} U_{1[n-1]} - \varepsilon\lambda_3 U_{OC[n-1]} - \dfrac{\varepsilon^2}{2}\lambda_2 U_{OC[n-1]} - \dfrac{\varepsilon^3}{6}\lambda_1 U_{OC[n-1]} + \dfrac{\varepsilon^3}{6} X_{[n-1]} \\ U_{OC[n]} = sign(U_{3[n]}) \end{cases} \quad (9)$$

Данный метод может быть использован и в более сложных случаях нелинейной функции обратной связи, однако, нужно учитывать, что если информативный параметр сигнала обратной связи не амплитуда, а время (ШИМ – сигнал), частота, фаза, то реакция интеграторов на сигнал обратной связи будет описываться нелинейной функцией, одним из параметров которой будет информативная составляющая сигнала (время, фаза). Например, если информативным параметром сигнала обратной связи является длительность импульса при неизменной амплитуде, то реакция интеграторов на такой сигнал будет нелинейной функцией (начиная со второго интегратора). В этом случае каждый коэффициент обратной связи $\lambda_1, \lambda_2, \ldots, \lambda_k$ будет функционально связан с амплитудным значением выходного сигнала. Это количественно усложняет граф и систему уравнений, однако, качественно не меняет подход, позволяя моделировать любые нелинейности, а также любые прямые и обратные связи в цепи преобразования [3].

Литература

1. Data Conversion Handbook / Walt Kester, Editor. – Newnes, 2004. – 976 p.
2. Мэзон С., Циммерман Г. Электронные цепи, сигналы и системы / Под ред. П.А Ионкина. - М., Изд-во иностр. лит-ры. - 1963. – 620 с.
3. Ашанин В.Н. ΣΔ-аналого-цифровые преобразователи: основы теории и проектирование / Ашанин В.Н., Чувыкин Б.В., Шахов Э.К. // Монография – Пенза: Информационно-издательский центр ПенГУ, 2009. – 188 с.

Организация информационной поддержки процесса принятия решений в условиях неопределенности и риска

А.А. Шкромадо
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Современные предприятия представляют собой сложные науко- и ресурсоемкие производства, в связи с чем, все процессы функционирования: от проектирования изделия до его продажи тесно взаимосвязаны и требуют системного подхода для организации информационного обеспечения и четкого централизованного управления на его основе. Основные решения, принимаемые на уровне ответственных лиц предприятия, и особенно его руководителя, невозможно реализовать без развитой многофункциональной системы информационной поддержки. Качество информационного обеспечения управления – один из важных факторов, определяющих действенность принимаемых управленческих решений [13]. Отсутствие качественной системы информационной поддержки принимаемых решений приводит к вероятностному характеру принимаемых управленческих действий и, как следствие, к невысокой эффективности управления[1].

Информационную поддержку принимаемых решений обеспечивают советующие системы, экспертные системы (ЭС) и непосредственно системы поддержки принятия решений (СППР). Первые предполагают последовательное интерактивное взаимодействие с целью выявления параметров проблемы и поэтапной выдачи рекомендаций по устранению проблемы [12]. Вторые представляют собой базу знаний экспертов в определенной предметной области и особенно ценны для повторяющихся задач [12]. Экспертные системы являются своего рода эмуляторами опыта и интуиции человека, однако, результат принятия того или иного решения можно предсказать только при идеальном совпадении всех факторов (повторении задачи), что на практике практически не достижимо.

Компьютерная СППР [10] - это интерактивная автоматизированная система, использующая модели выработки решений, обеспечивающая пользователям эффективный доступ к распределенной базе данных и предоставляющая им разнообразные возможности по отображению информации. В таком понимании система поддержки принятия решений представляет собой совокупность следующих подсистем: комплекса распределенных технических средств; комплекса математических моделей; анализа состояний и выработки решений; базы данных; систем управления моделями, языков моделирования, обработки и отображения информации. В состав СППР входят три главных компонента: база данных, база моделей и программная подсистема [1]. Благодаря возможности имитационного моделирования возможных исходов, СППР оказываются ценным и эффективным инструментом руководителей различного уровня (ЛПР).

Цели, функции и уровни применения СППР

Основной целью применения системы поддержки принятия решений является повышение качества оперативных, тактических и стратегических решений. Все большее развитие данные системы получают в связи с актуальностью проблемы квалифицированности принятых решений в процессе управления.

Несмотря на важность и более чем сорокалетнюю историю СППР (в конце 60-х годов впервые появляется новый тип информациолнных систем — модель-ориентированные СППР (Model-oriented Decision Support Systems — DSS) или системы

управленческих решений (Management Decision Systems — MDS)) до сих пор не существует даже точного определения данных систем или общепризнанной классификации [10], в тоже время, интерес к разработке и совершенствованию не только не угасает, но постоянно растет. Разрабатываются как индивидуальные, так и групповые системы, особенно ценные для расширения взаимодействия и поддержки принятия коллегиальных решений (например, совета директоров). [11]

Наиболее важными направлениями совершенствования собственно СППР являются разработка эффективных математических методов, моделей, алгоритмов, схем и программ для реализации целей поддержки принятия решений в условиях многокритериальности выбора, а также слабой структурированности исходной информации.

Отдельно следует сказать о весьма важном направлении принятия решений для руководителей – принятии решений в условиях неопределенности и риска. Учитывая, что перспективы развития предприятий отрасли сильно зависят от научно-технического прогресса и общей политической ситуации в России и мире, результаты принимаемых руководителями решений можно «предсказать» основываясь исключительно на развитой системе математического аппарата и алгоритмов выбора, а также механизмах имитационного моделировании возможных исходов.

При этом необходимо отдельно рассматривать три уровня принимаемых решений: оперативный, тактический и стратегический [1, 6, 13]. Соответственно, для каждого уровня применяемые СППР методы и алгоритмы различны. Оперативный уровень системы предназначен для сравнительно несложного анализа среднего объема структурированной информации, отобранного по определенному критерию за известный временной промежуток. Динамика принимаемых решений на данном уровне, однако, весьма высока. Характерен для низовых звеньев организации. Результат, как правило, отображается в виде отчета. На тактическом уровне число принимаемых решений меньше, однако, значительно возрастают объемы обрабатываемой информации и сложность. Стратегический же уровень предназначен для поддержки принятия решений на долгосрочную перспективу с учетом влияния различных факторов и затрагивающих принципиально важные для предприятия вопросы (высший уровень руководства).

Организация процесса принятия решений в условиях неопределенности и риска

Неопределенность – это ситуация, когда последствия принимаемых решений неизвестны и их можно лишь приблизительно оценить или по [4] – явления, не поддающиеся анализу и измерению со сколь угодно большой точностью. Учитывая, что каждое решение, во-первых, предполагает использование тех или иных ресурсов на свое принятие и, во-вторых, имеют конечную ценность результативности, следует говорить о критерии минимакса, т.е. поиске решения в условиях неопределенности, максимизирующего возможную прибыль (положительный результат) и минимизирующего возможный убыток (отрицательный результат).

На сам процесс принятия решений оказывают активные группы (actors), чье мнение учитывается ЛПР [11]. Для снижения степени неопределенности и риска активные группы зачастую набираются из экспертов, т.е. высококвалифицированных специалистов, привлекающих для решения проблемы весь опыт и интуицию [13]. Кроме того, при принятии сложных стратегических решений работу групп координирует консультант по принятию решений. Таким образом, количество (и качество) полученных альтернативных вариантов будет сильно зависеть от подбора экспертов комиссией и организации работы между различными экспертными группами. Следует упомянуть, что альтернативные варианты могут быть как независимыми, так и

зависимыми, т.е. решения по одной из них оказывают влияние на качество других [10, 12, 13]. Выбор той или иной альтернативы осуществляется на основе удовлетворения критериям. Современные методы принятия решений ориентированы на учет всех отличительных особенностей качеств альтернатив, что существенно приближает формальные схемы к реальному миру. Поэтому в настоящее время многокритериальное описание альтернатив становится все более принятым.

Таким образом, поддержка принятия решения в условиях неопределенности и риска включает [9, 11, 12]: помощь ЛПР при анализе объективной составляющей, т.е. в понимании и оценке сложившейся ситуации и ограничений накладываемой внешней средой; выявление предпочтений ЛПР, т.е. в выявление и ранжирование приоритетов, учет неопределенности в оценках ЛПР и формировании его предпочтений; генерацию возможных решений, т.е. формировании списка альтернатив, оценку возможных альтернатив, исходя из предпочтений ЛПР и ограничений, накладываемых внешней средой; анализ последствий принимаемых решений, определение риска; выбор лучшего, с точки зрения ЛПР, варианта.

Сущность, модели и методы принятия решений в условиях неопределенности и риска

Обычно выделяют три типа задач, решение которых возможно средствами СППР: хорошо структурированные или количественно сформулированные проблемы, слабо структурированные или смешанные (содержат количественные и качественные критерии), а также неструктурированные или качественно выраженные проблемы [10]. Для решения каждой из этих задач свойственно применение своего математического и молельного аппарата, а также специфичных алгоритмов и методов. Сложность неструктурированности исходных данных порождает проблему неопределенности как исходной информации, так и результатов принимаемых на основе этой информации решений [4,5].

Долгое время неопределенности просто отбрасывались или заменялись на кажущуюся определенную (детерминированную) информацию на основе экспертных мнений, однако на самом деле, как показано выше, такая подмена не всегда корректна.

В 1965 г. Лотфи Заде предложил теорию нечетких или размытых множеств, получивших также название нечеткой логики (fuzzy logic). Теория нечетких множеств дала схему решения проблем, в которых субъективное суждение или оценка играют существенную роль при оценке факта неясности и неопределенности. Она прошла путь от разработки формальных средств представления плохо определяемых понятий, используемых человеком, и аппарата для их обработки до моделирования приближенных рассуждений, к которым человек (руководитель, ЛПР) прибегает в повседневной и профессиональной деятельности. При этом следует различать методы, модели и математический аппарат теории вероятности и теории нечетких множеств. Наряду с теорией вероятности как раздел прикладной математики также применяется теория игр и исследование операций [3].

Для обращения с неточно известными величинами обычно применяется аппарат теории вероятностей. Однако случайность связана с неопределенностью, касающейся принадлежности некоторого объекта к обычному множеству [5], в связи с чем случайность и неопределенность – это не одно и то же. Понятию вероятностной меры в теории вероятностей соответствует более простое понятие функции принадлежности в теории нечетких множеств. По этой причине обычно удобнее оперировать с неопределенностью методами теории нечетких множеств без привлечения аппарата теории вероятностей.

Подход на основе теории нечетких множеств является, по сути дела, альтернативой общепринятым количественным методам анализа систем. Он имеет три

основные отличительные черты: вместо или в дополнение к числовым переменным используются нечеткие величины и, так называемые, "лингвистические" переменные, простые отношения между переменными описываются с помощью нечетких высказываний, сложные отношения описываются нечеткими алгоритмами.

Задачи выбора наиболее выгодного варианта поведения из нескольких имеющихся вариантов в условиях неопределённости рассматриваются в теории игр и могут быть описаны матричными играми особого типа, в которых игрок взаимодействует не со вторым игроком, а с окружающей средой. Матричная игра в теории игр описывается с помощью платежной матрицы со строками (x_i), представляющими стратегии ЛПР и столбцами (y_i) – состояниями окружающей среды, представляющими собой неопределенный фактор с неизвестным законом распределения вероятностей. Решение задач в условиях неопределенности зависит от выбранного критерия оптимальности $Z = Z(x_i, y_i)$. Выбор решения основывается на принципе получения максимальной гарантированной величины показателя эффективности, отраженного значениями z_{ij}. При этом каждой стратегии ЛПР (x_i) приписывается некоторый результат Z_i (оценочная целевая функция), характеризующий все последствия этого решения. Из массива результатов принятия решений ЛПР выбирает элемент Z, который наилучшим образом отражает мотивацию его поведения.

Существует несколько критериев оптимизации решения [4, 8, 9, 15]. Критерий максимального математического ожидания выигрыша применяется, когда известны вероятности состояний окружающей среды. Критерий следует применять, когда максимизация среднего выигрыша более приоритетна, чем минимизация среднего проигрыша, т.е. оптимальной считается стратегия, при которой значение математического ожидания выигрыша максимально $Z = \max(Z_i)$.

Критерий недостаточного основания Лапласа используется при наличии неполной информации о вероятностях состояний окружающей среды или ее отсутствии (реальные вероятности состояний окружающей среды должны быть близки по значениям). Вероятности состояний окружающей среды принимаются равными, и по каждой стратегии ЛПР в платёжной матрице определяется среднее значение выигрыша. Данный критерий также ориентирован на максимальный средний выигрыш, как и предыдущий критерий. Максиминный критерий Вальда (критерий осторожного наблюдателя) применяется для стратегии, полностью исключающей риск, т.е. ЛПР не получит результат хуже, чем тот, на который ориентируется. Оптимальной по данному критерию считается та стратегия ЛПР, при выборе которой минимальное значение выигрыша максимально $Z = \max(Z_i)$. Подходит для ситуаций с неизвестными состояниями окружающей среды и однократной реализацией решения.

Критерий пессимизма-оптимизма Гурвица демонстрирует промежуточную ситуацию между крайним пессимизмом и оптимизмом путем введения промежуточного статистического коэффициента пессимизма в диапазоне [0,1]. Оптимальной по данному критерию считается стратегия, в которой значение Z_i максимально: $Z = \max(Z_i)$. Критерий допускает некоторый риск и реализуется при малом количестве решений в условиях неизвестной или недостоверной информации об окружающей среде.

Ключевым моментом принятия решения в условиях неопределенности является оценка эффективности принятых решений [8]. Показатель эффективности *Z* зависит от трех групп факторов $Z=Z(\alpha, X, \xi)$, где α – заданные заранее известные факторы (условия

выполнения задачи); X – зависящие от нас элементы решения, образующие в своей совокупности множества решений; ξ – неизвестные факторы. Так как величина X зависит от неизвестных факторов ξ она уже не может быть вычислена и остается неопределенной. Поэтому в этом случае задачу можно сформулировать так: при заданных условиях α, с учетом неизвестных факторов ξ, найти такое решение X, которое, по возможности, обеспечивает максимальное значение показателя эффективности Z, $Z(\alpha, X, \xi) \rightarrow Z_{max}$ [1].

Модели принятия решений в условиях неопределенности можно разделить на две большие группы [3, 14, 15]: модели линейного упорядочивания и модели группового упорядочивания. Из первой группы наиболее часто применяемыми являются модель функции доминируемости, модель Бержа-Брука-Буркова и модель векторной оценки альтернатив. Первый тип моделей порождает метод принятия решений на базе нечеткой логики, второй – метод анализа иерархий (МАИ), а модель векторной оценки альтернатив приводит к применению целой группы качественных методов принтия решений: метода ЗАПРОС (ЗАмкнутые ПРОцедуры у Опорных Ситуаций), метода ПАРК (парная компенсация) и метода ОРКЛАСС (Ординальная КЛАССификация). Выбор модели упорядочивания с теми свойствами, которые особенно желательны в данном конкретном случае, представляется весьма полезным в системах поддержки принятия решений [15]. Модели типа: модель функции доминируемости, Брэдли-Терри, Бержа-Брука-Буркова, стохастическая модель Ушакова, модель равномерного сглаживания предлагают гораздо более убедительные доводы в пользу соответствующих оптимальных упорядочиваний. Модель Брэдли-Терри пригодна для простых структур и целочисленных турнирных матриц, которые не учитывают неточность, неопределенность в оценках экспертов. Стохастическая модель Ушакова скорее ориентирована на вероятностный класс неопределенностей, в отличие от нее модель Бержа-Брука-Буркова и модель функции доминируемости позволяют учитывать неопределенность явлений, не поддающихся измерению со сколь угодно большой точностью и с учетом нечеткости соответственно. Для линейного упорядочивания в условиях неопределенности предпочтительнее пользоваться моделями функции доминируемости и Бержа-Брука-Буркова.

Выводы

Эффективная поддержка принятия решений является необходимым компонентом, определяющим действенность управления на предприятии. Практически всегда решения приходится принимать в условиях неопределенности, что порождает риск. Неопределенность вызвана как недостатком (недостоверностью) исходной информации, так и сложностями в процессе принятия решений, вызванными многокритериальностью альтернатив, а также субъективностью подбора экспертов и выдаваемых ими результатов. Из систем компьютерной поддержки принятия решений в условиях неопределенности чаще всего применяются СППР. Это связано с широкими возможностями моделирования, а также использованием развитого математического аппарата теории вероятности и теории нечетких множеств (нечеткой логики).

Литература

1. Блюмин С.Л., Шуйкова И.А. Модели и методы принятия решений в условиях неопределенности. - Липецк: ЛЭГИ, 2001. – 138 с.

2. Кофман А., Хил Алуха Х. Введение теории нечетких множеств в управлении предприятиями: Пер. с исп. - Минск: Выш. шк., 1992. - 224 с.

3. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. - М.: Наука, 1981. - 194 с.

4. Трахтенгерц Э.А. Компьютерная поддержка принятия решений: Научно-практическое издание. Сер. Информатизация России на пороге XXI века. - М.: СИНТЕГ, 1998. - 376 с.

5. Эддоус М, Стэнфилд Р. Методы принятия решений/ Пер. с англ. под ред. член-корр. РАН И.И.Елисеевой. – М.: Аудит, ЮНИТИ, 1997 – 590 с.

Подсекция 2
«Автоматизированные информационные системы»

Автоматизированная информационная система обслуживания посетителей и статистического анализа библиотечных процессов

В.В. Васильев
Самарский государственный архитектурно-строительный университет, Самара

В ходе происходящих преобразований в России библиотеки как хранилища знаний и информации, являющиеся важнейшими факторами развития экономики, неуклонно повышают свою значимость. Интернет содержит огромное количество информации, но до сих пор самую качественную информацию по многим областям знания можно найти только в книгах. К сожалению, большинство книг не доступны в Интернете, а те, что доступны, являются незаконными. Прочитать необходимые книги возможно только посещая библиотеки.

Дела в библиотеках обстоят не самым лучшим образом. Уровень информатизации многих библиотек остался на уровне 90х годов. К счастью так не везде. В некоторых библиотеках в читальных залах установлены компьютеры для поиска различных изданий, реализована технология электронного читательского билета, сформированы базы данных о фонде библиотеки. Но как это все сказывается на читателе? Никак. Искать издания через библиотечную систему поиска так же удобно, как искать эти издания с помощью старого карточного каталога. Поиск по электронным каталогам множества современных библиотек (у которых есть Интернет-версии поиска) практически не отличается и является чересчур неудобным для пользователей. Статистика в большинстве библиотек ведется вручную, позже обработкой накопленных данных занимается как правило один человек, которому как и всем людям свойственно ошибаться.

Исходя из всего этого, была сформулирована цель работы - разработка информационно-аналитической интернет-системы, позволяющей повысить эффективность работы пользователя (читателя и сотрудника) с существующей базой данных библиотеки, автоматизация сбора статистических показателей, а также создание своей базы данных, многократно расширяющей информативность и возможности системы.

В результате такая система была создана. С 1 по 30 мая 2010 года система проходила тестирование в ГУК «СОУНБ». В течение этого времени сотрудники библиотеки знакомились с автоматизированной системой и учились ей пользоваться. Руководством библиотеки был проведен анализ эффективности работы автоматизированной системы мониторинга статистики. По результатам анализа было установлено, что тестовое внедрение статистической части системы в ГУК «СОУНБ» проведено успешно. Для дальнейшего полного внедрения система должна быть полностью настроена и должны быть обучены все сотрудники ГУК «СОУНБ».

Разработанная система предоставляет читателю и работникам библиотеки эффективный интерфейс, разработанный с учетом создания ошибок в подобных системах, с помощью которого возможно использовать Интернет для заказа литературы, просмотра электронного библиотечного каталога, ведения и анализа

статистических показателей. Как в любой современной информационной интернет-системе, у каждого пользователя существует профиль, который содержит всю информацию о пользователе, права доступа, а также историю заказов литературы. В системе присутствует мощный аналитический модуль сбора различной статистической информации. Для управления всем этим в системе присутствует мощный, эффективный и гибкий административный модуль, позволяющий управлять как отдельными частями, так и всей системой в целом. Все это позволяет сотрудникам библиотеки всегда быть в курсе их работы и ее эффективности.

При полном внедрении система позволит:

- избавиться от ошибок в статистических отчетах;
- сократить время на ведение статистики в связи с автоматизацией сбора и подсчета статистических данных;
- проводить анализ востребованности фондов ГУК «СОУНБ»;
- обезопасить статистические данные от внешнего воздействия;
- предоставить читателям возможность заказа книг в библиотеке через сеть Интернет;
- предоставить читателям советы в виде наиболее интересной для них литературы.

Автоматизированная система мониторинга состояния здоровья и уровня физического развития студентов ВУЗов

В.В. Горожанов, В.М. Богданов, А.В.Соловов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Проблема мониторинга уровня физического состояния студентов и, как следствие, проблема эффективно физического воспитания по-прежнему являются одними из актуальных задач для университета. Разработка современной автоматизированной информационно-аналитической системы мониторинга уровня физического здоровья студентов является исключительно сложной задачей, требует привлечения больших материальных и интеллектуальных ресурсов, применения самых современных информационных технологий. Существуют методы, позволяющие оценить некоторые параметры физического состояния, но большинство из них не автоматизировано. А среди автоматизированных систем нет систем с аппаратом статистической обработки результатов тестов. Несмотря на определенные трудности, связанные с организационными, материально-техническими, научно-методическими аспектами разработки и внедрения современных информационных технологий в область физической культуры и спорта, они вызывают определенный интерес у ряда специалистов, так как здесь, как и в других областях, назрела необходимость перехода о традиционных средств к использованию современных информационных и коммуникационных технологий, позволяющих значительно эффективнее осуществлять сбор, обработку и передачу информации, вести самостоятельную работу и самообразование, качественно изменить содержание, методы и организационные формы обучения студентов, проведения с ними физкультурно-оздоровительной работы. В этой связи научный поиск в создании и использовании в учебно-тренировочном процессе современных информационных технологий приобретает особую актуальность.

Одним из главных факторов повышения эффективности спортивной деятельности, является знание каждым студентом собственного уровня физического состояния и вид картины физического состояния студентов в целом, для правильного составления программы физического воспитания. Эффективность в данном случае повышается за счет использования современных информационных технологий для качественной и всесторонней оценки уровня здоровья.

Информационная система «АИС мониторинга состояния здоровья и уровня физического развития студентов вузов» разрабатывается в виде сайта, чтобы каждый мог, пройдя тест, узнать о состоянии своего здоровья. Система должна обеспечивать автоматизацию мониторинга уровня физического состояния студентов: проводить комплексное тестирование уровня физического состояния студентов; выдавать качественную и количественную оценку показателей здоровья; осуществлять анонимный сбор полученных данных в базе данных; на основе собранных данных проводить статистическую обработку результатов.

Для реализации нашей системы была выбрана система управления содержимым сайта CMS Drupal, в качестве СУБД - MySQL, язык разработки – PHP.

Предполагается архивирование данных из базы данных каждые пять лет во избежание разрастания базы данных до очень больших размеров, что может сильно сказаться на производительности системы в целом.

В настоящее время система «АИС мониторинга» находится на завершающей стадии разработки в Самарском областном центре новых информационных технологий при СГАУ им. академика С.П.Королева (ЦНИТ СГАУ).

Программное обеспечение процесса сертификации систем менеджмента

А.В. Графкин, В.В. Графкин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Программный комплекс «АРМ – Сертификация СМ» (автоматизированное рабочее место – сертификация систем менеджмента) разработан с целью автоматизации процесса планирования, регистрации результатов, автоматизации учета, формирования отчетов для анализа результативности и предоставления электронных отчетов в соответствующие контролирующие организации при выполнении работ по сертификации систем менеджмента (СМ). Программный комплекс предназначен для специалистов осуществляющих работы по сертификации СМ и позволяет автоматизировать весь процесс сертификации СМ начиная от заявки клиента на проведение сертификации вплоть до формирования бланков сертификатов и автоматического составления отчетов, а также их отсылки контролирующим организациям.

«АРМ – Сертификация СМ» позволяет автоматизировать процесс подтверждения соответствия по следующим системам менеджмента:

- ISO 9001 Система менеджмента качества;
- ISO 14001 Система экологического менеджмента;
- OHSAS 18001 Система менеджмента профессиональной безопасности и здоровья;
- ISO 22000 Система менеджмента рисков при производстве пищевой продукции;
- SA 8000 Система социальной ответственности.

Сведения о системах менеджмента заносятся в специальный «Справочник стандартов», информация из которого используется при составлении заявки и дальнейшей работе с ней. Ведение данного справочника решает проблему актуализации сведений о сертифицируемых системах менеджмента: устаревшая информация может быть удалена, а сведения о новой СМ – добавлены, при этом отсутствует необходимость модернизации программного комплекса.

В программном комплексе помимо «Справочника стандартов» также реализованы: «Справочник стран», содержащий названия стран и их коды; «Справочник областей», содержащий наименования областей, их коды и при этом имеется связь с соответствующей страной; «Справочник банков», содержащий наименования банков и их реквизиты; «Справочник органов по подтверждению соответствия», данные из которого используются, например, при заполнении заявки в случае ресертификации системы менеджмента; «Справочник экспертов-аудиторов»; «Справочник отраслей» и прочие. Внедрение данных справочников позволяет существенно сократить время, затрачиваемое специалистом на операции, необходимость в которых возникает в ходе подтверждения соответствия СМ.

Работа в «АРМ – Сертификация СМ» по обработке информации, получаемой от клиента, начинается с составления заявки в «Журнале регистрации заявок и сертификатов». Если были выявлены какие-либо неточности при заполнении или же произошли перемены, связанные с изменением пожеланий клиента, имеется возможность редактирования заявок в последующем либо, вообще, полное удаление заявки. Далее для каждой заявки формируется график аудитов с внесением полной информации, включающей: «Область аудита»; «Руководителя группы аудиторов»;

«Стандарты», по которым проводится проверка соответствия; информация о договоре с клиентом и все прочие сведения, требуемые на данном этапе. Также как и информация о заявке, информация о графике аудитов может быть скорректирована при возникновении подобной необходимости.

По результатам работы аудиторов составляются соответствующие акты и могут быть сделаны первые заключения по результатам аудита. Далее приводится полный отчет, на основе которого решается вопрос о выдаче, переоформлении, отказе либо приостановке сертификата, и, в итоге, в случае благополучного прохождения аудита, а также выполнения условий и требований договора, составляется бланк сертификата.

Необходимо отметить, что структура отчетов, генерируемых в качестве документов Microsoft Word, может быть изменена простым переносом соответствующих полей заполнения информации на требуемые новые позиции. То есть данные видоизменения бланков документов возможно осуществить любым пользователем программного комплекса «АРМ – Сертификация СМ». Благодаря данным особенностям, позволяющим не прибегать к модернизации программного комплекса (что может потребоваться при незначительных изменениях в оформлении каких-либо документов), данная система является весьма гибким средством автоматизации процесса подтверждения соответствия системам менеджмента.

Автоматизированная система оценки эффективности венозного оттока

Д.А. Загуменнов[1], Л.С. Зеленко[1], П.Ф. Кравцов[2]
[1]*Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*
[2]*Самарский государственный медицинский университет, Самара*

Введение

Варикозная болезнь – наиболее распространенное заболевание периферических сосудов, характеризующееся нарушением оттока крови из венозного бассейна, что влечет за собой каскад патологических изменений на молекулярном, клеточном и тканевом уровне. Согласно обобщенным данным эпидемиологических исследований, этим заболеванием в разных странах страдают 15–30% трудоспособного населения. В России различные формы варикозной болезни наблюдаются у 35 млн. человек, причем у 15% из них имеются трофические изменения кожи, открытые или рецидивирующие трофические язвы. Подобная распространенность заболеваний делает работу специалистов-флебологов особенно важной и повышает ее трудоемкость. В связи с этим остро встает вопрос о повышении эффективности и рационализации деятельности врачей, в том числе с помощью современных достижений вычислительной техники и информационных технологий. Разработка специализированной медицинской информационной системы, которая бы позволяла автоматизировать спектр задач, связанный с комплексным анализом объективных и субъективных критериев обследования пациентов с нарушениями венозного оттока, становится очень *актуальной*.

Описание критериев обследования пациентов с нарушениями венозного оттока

Обследование венозной системы пациента - процесс долгий и дорогостоящий, правильная постановка диагноза предполагает проведение исследований по нескольким десяткам различных параметров. Их можно сгруппировать по трем уровням:

1) *данные макрогемодинамических показателей* (ультразвуковое исследование показателей венозного оттока и артериального притока);

2) *данные о работе мышечно-венозной помпы* (параметры цикла шага, миография мышц, формирующих мышечно-венозную помпу);

3) *показатели функционального состояния глубоких вен* (данные функциональной флеботонодебитометрии).

Каждый параметр характеризуется набором числовых показателей (среднее значение или норма, допустимое отклонение от нормы), по которым можно сделать вывод о стадии заболевания. Специалисты-флебологи выделяют три характерные стадии венозной болезни:

1) компенсация (можно ограничиться консервативным лечением);

2) субкомпенсация (возможна хирургическая коррекция по косметическим показаниям);

3) декомпенсация (требуется хирургическое вмешательство).

Обследование пациентов начинается с неинвазивных общедоступных методов, включенных в первый уровень параметров. В зависимости от полученных результатов, части пациентов для уточнения тактики лечения требуется более детальное исследование функционального состояния мышечно-венозной помпы нижних конечностей (второй уровень параметров). В случае значительного отклонения показателей этого уровня, для уточнения объема требуемой оперативной коррекции венозного русла, пациентам проводятся инвазивные исследования третьего уровня.

Описание возможностей системы

Автоматизированная система оценки эффективности венозного оттока по существу представляет собой автоматизированное рабочее место врача-флеболога. В системе предусмотрено две роли пользователей: *врач* и *администратор*. При запуске системы необходимо пройти процедуру аутентификации, после ввода личных данных программа настраивает интерфейс пользователя в соответствии с его правами. Разделение ролей обеспечивает защиту информации: так как каждый пациент приписан к определенному лечащему врачу, врачи - пользователи системы - получают возможность работать только с данными своих пациентов и не могут изменять другую хранящуюся в базе данных информацию.

Один из пользователей обладает статусом администратора: он может управлять учетными записями, добавляя, изменяя или удаляя пользователей-врачей, а также распределять между ними пациентов, кроме того, он может редактировать параметры обследования и состав рекомендаций.

Навигация по системе осуществляется через меню главного окна программы (рис. 1). В нем же находится список пациентов, записанных за текущим пользователем, а также форма поиска, которая позволяет отбирать больных по различным параметрам, таким как пол, возраст, диагноз, место проживания и др.

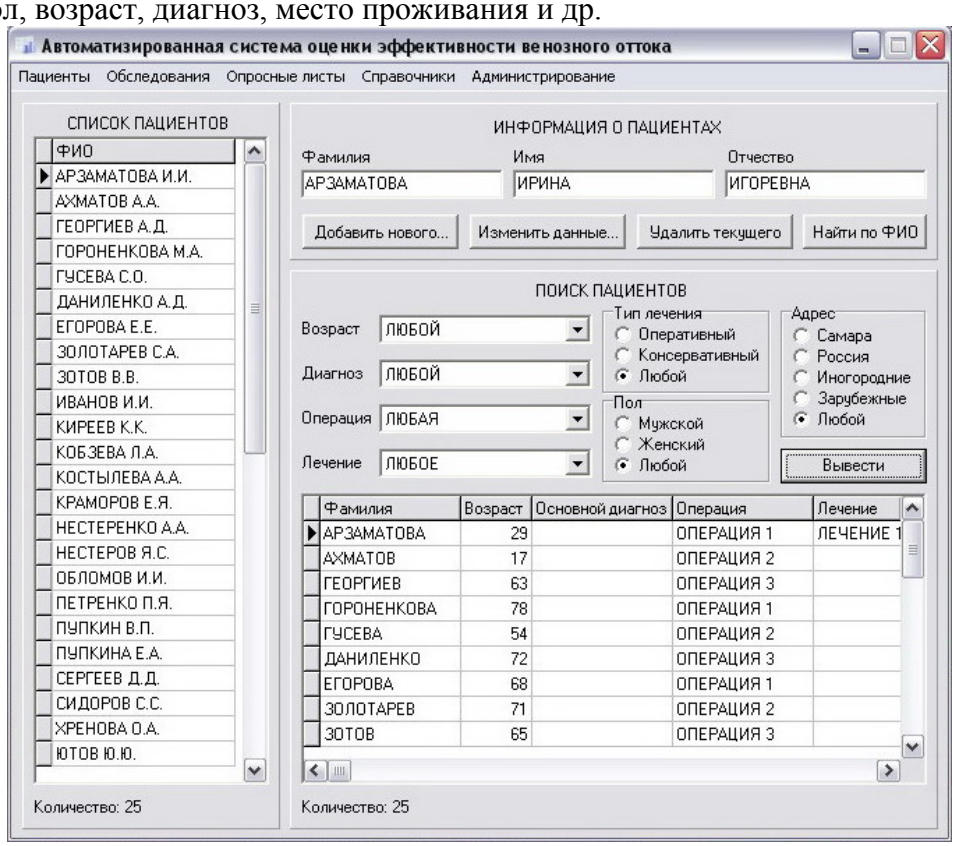


Рис. 1. Главная экранная форма системы

Система обладает широкими возможностями по настройке *состава обследования*, позволяя добавлять новые, либо изменять и удалять текущие параметры в любом из трех уровней. Число уровней фиксировано, но их наименования поддаются изменению. Экранная форма, отвечающая за составление обследований и расчет их результатов, представлена на рис. 2.

После проведения текущего обследования в систему вводятся данные о полученных параметрах, в зависимости от значений нормы и допустимых отклонений от нее система определяет, к какой стадии заболевания относится каждый параметр.

Кроме того, по всем данным проводятся общие расчеты и представляются в процентном отношении для всех стадий заболевания (компенсация, субкомпенсация и декомпенсация) и для всех уровней. Система также выдает текстовую рекомендацию по лечению (к примеру, «требуется коррекция клапанного аппарата венозной системы» или «рекомендуется консервативное лечение»). Текст рекомендаций доступен для редактирования.

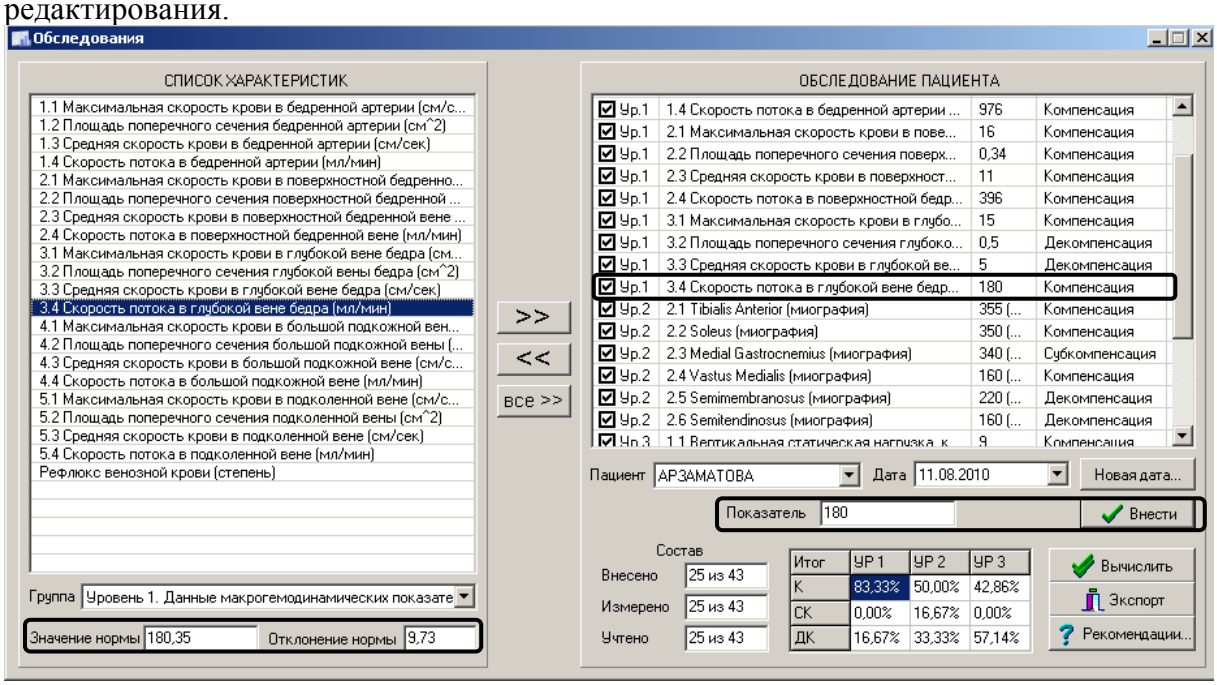


Рис. 2. Экранная форма для данных проведенных обследований

Помимо профильных данных, в системе ведется учет общей информации о пациентах, в базе данных хранятся все личные карточки. Форма личной карточки приведена на рис. 3. Для повышения эффективности и удобства работы пользователей в системе ведутся территориальные (страны, области, города, районы, улицы) и медицинские (диагнозы, операции, консервативные схемы лечения) справочники, а также справочники возрастных групп и типовых диагнозов. При необходимости личную карточку пациента можно экспортировать в текстовый процессор MS Word.

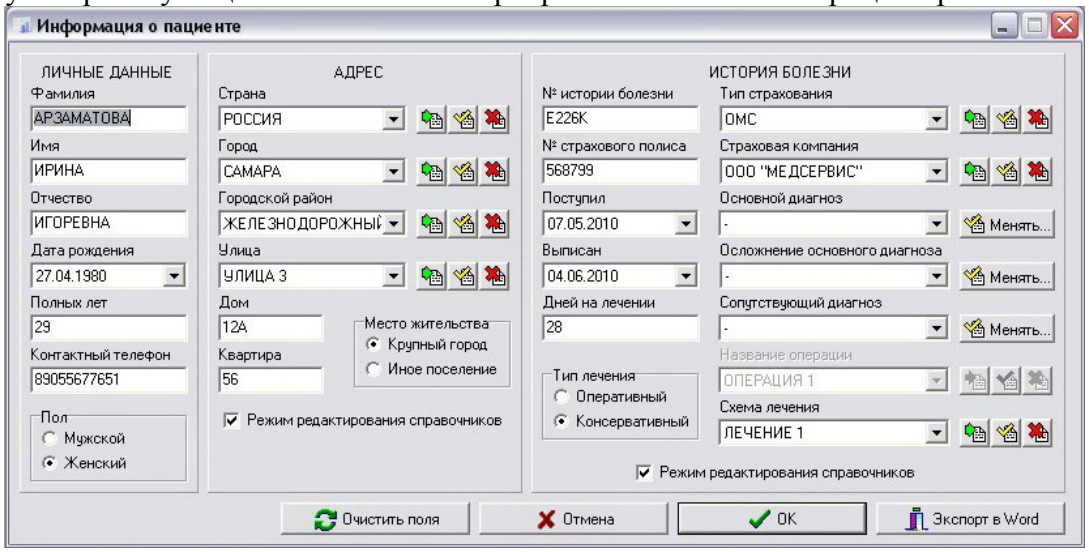


Рис. 3. Форма личной карточки пациента

Кроме обработки объективных данных (результатов инструментальных методов обследования: ультрасонодопплерографии, функциональной флеботонодебитометрии, функциональной миографии, подометрии), в системе предусмотрена возможность работы с субъективными данными, которые формируются на основе анализа самостоятельной работы пациентов с предлагаемым опросником по оценке качества их жизни. Формат опросника фиксирован – в качестве него используется международный стандарт CIVIQ, содержащий 20 вопросов. Вопросы разбиты на четыре группы: психологические, болевые, физические и социальные проявления. Ответ на каждый из вопросов выражается в виде числа от 1 до 5. Результат каждого опроса вычисляется системой, он включает общую сумму баллов, количество учтенных вопросов, а также промежуточные суммы по группам. Результаты всех опросов, хранящихся в базе данных, могут быть сведены в единую таблицу и отсортированы по фамилиям пациентов, датам проведения либо по набранным баллам. Кроме того, любой опросный лист может быть экспортирован в MS Word аналогично личной карточке. Экранная форма для работы с опросными листами представлена на рис. 4.

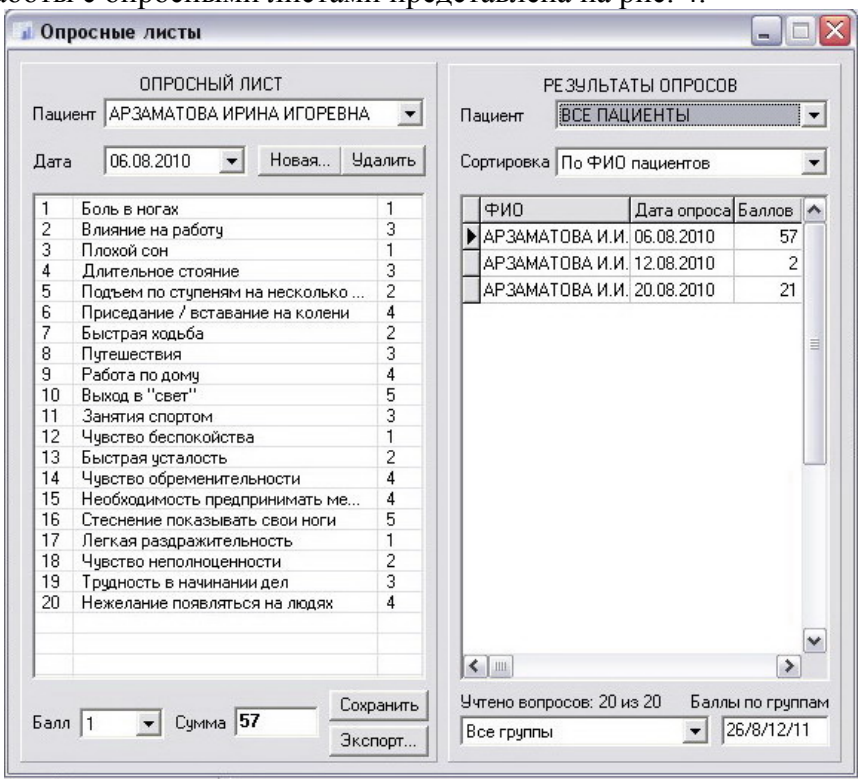


Рис. 4. Экранная форма для работы с опросными листами

Вся требующая сохранения информация в системе сведена в единую базу данных, работающую под управлением СУБД MySQL Server версии 5.1. Логическая модель данных представлена на рис. 5.

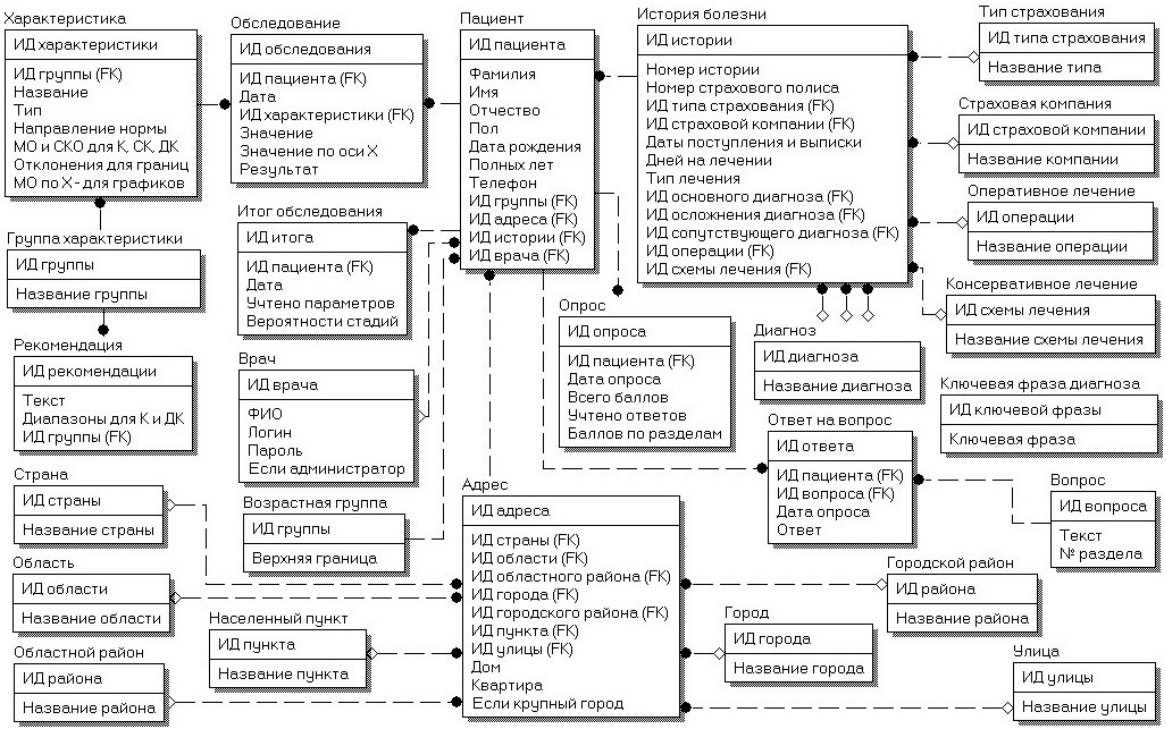


Рис. 5. Логическая модель данных

Выводы

Реализация системы позволила автоматизировать деятельность врачей-флебологов, значительно сократив объем бумажной работы и арифметических расчетов. Учет личных карточек дал возможность вести единую базу пациентов, а определение стадии заболевания и выдача рекомендаций – быстро анализировать результаты проведенных обследований и принимать верное решение о тактике лечения.

Система внедрена в работу сосудистого отделения кафедры госпитальной хирургии клиник Самарского государственного медицинского университета и прошла проверку в реальных условиях.

Автоматизированная система управления ремонтами оборудования

И.В. Занин
ООО «СМС - Информационные технологии», Самара

Введение

Одной из задач, решаемой для обеспечения бесперебойного функционирования Единой Энергетической Системой (ЕЭС) России, является управление ремонтами оборудования. Использование автоматизированных систем при этом является ключевым фактором, без которого качественное и своевременное решение данной задачи невозможно.

С точки зрения сроков принятия решения выделяют две подзадачи: оперативное управление и планирование. Ниже речь пойдет об особенностях каждой из подзадач (задач) и применяемых при их решении программных комплексах.

Характеристика объекта автоматизации

При решении задачи оперативного управления объектом автоматизации является процесс формирования, обмена, рассмотрения, согласования и утверждения диспетчерских заявок на ремонт оборудования. Заявкой при этом называют электронный документ, оформленный и переданный на соответствующий уровень диспетчерского управления, в котором оформляется ответственное намерение организации изменить оперативное состояние оборудования или режим его работы.

В задаче планирования объектом автоматизации является процесс работы с графиками ремонтов оборудования. График ремонта состоит из плановых ремонтных заявок, каждая из которых описывает планируемый ремонт единицы оборудования с указанием срока начала и конца ремонта, вида ремонта и других характеристик.

Субъектами обеих задач являются собственники оборудования (сетевые организации и их филиалы, владельцы объектов электросетевого хозяйства, генерирующие компании и их филиалы, электростанции), региональные диспетчерские еправления (РДУ), объединенные диспетчерские управления (ОДУ), центральное диспетчерское управление (ЦДУ). Уровни диспетчерского управления связаны в единую технологическую сеть и имеют строгую иерархию. Верхним уровнем иерархии является ЦДУ.

Объектами задач является энергетическое оборудование генерирующих и сетевых предприятий.

Рассмотрим отличительные особенности процессов планирования и оперативного управления ремонтами оборудования. Различия между процессами можно увидеть, обратившись к таблице (см. табл. 1).

Таблица 1. Различия процессов оперативного управления и планирования ремонтов

Характеристика	Оперативное управление	Планирование
Передаваемая информация	Оперативная заявка	График ремонта
Принятие решений	Индивидуальное	Групповое
Основание для принятия решения	Содержимое заявки	Согласованность сроков ремонта
Горизонт контроля	Неделя	Месяц \год
Поток информации	200 заявок \день	13 графиков \год

Исходя из написанного выше, мы получаем следующий вывод – процессы отличаются друг от друга, но при этом базируются на одном и том же оборудовании и направлены на решение задач в одних и тех же организациях. Соответственно каждая задача решается самостоятельным программным комплексом, имеющим общую часть в виде оборудования.

Опишем основные функции каждого программного комплекса.

Оперативное управление

Задача оперативного управления решается с помощью программного комплекса «Заявки». Основными функциями данного программного комплекса являются:

Подготовка заявок. При подготовке задаются основные реквизиты заявки (вид ремонта, категория и т.д.), время ремонта, текстовое содержание работ и, создается маршрут прохождения заявки, включающий службы предприятия, которые должны рассмотреть данную заявку, а также перечень предприятий, в ведении которых находится данная единица оборудования.

Маршрутизация заявок. Автоматическое прохождение заявки по маршруту: по службам и пользователям внутри предприятия, а также передача заявок на другие предприятия - уровни диспетчерского управления. Передача заявок между предприятиями происходит по каналам прямого доступа и/или с использованием электронной почты.

Оповещение о вновь поступивших заявках и сообщениях. В случае наступления в системе события, требующего внимания пользователя, выдается соответствующее сообщение. Присутствует цветовая и звуковая сигнализация.

Отображение списка заявок в удобном для пользователя виде. Реализована возможность настройки компонент пользовательского интерфейса. Все настройки сохраняются в профиле пользователя и доступны с любого рабочего места. Поддерживается сортировка списка заявок по одному или нескольким столбцам, быстрый поиск по значениям полей. Реализована группировка заявок по оборудованию, предоставлены механизмы управления группами.

Возможность индивидуальной настройки экранных и отчетных форм.

Планирование

В функции программного комплекса «Планы ремонтов», предназначенного для решения задачи планирования, входит:

Создание графика. Подготовка графика ремонтов заключается в указании периода планирования, выбора типа графика и наполнения его плановыми ремонтными заявками. Наполнение графика возможно по одной заявке или массированно в виде группы заявок.

Прием входящих графиков и формирование сводного. Решение о сроках проводимых ремонтов принимается на основании сформированного сводного графика, который компилируется из входящих графиков предприятий нижнего уровня иерархии.

Отображение. Сформированный сводный график для удобства рассмотрения и принятия решения отображается в виде диаграммы Ганта.

Совместимость \несовместимость. При принятии решения о сроках ремонтов важным является соблюдение условий согласованности их сроков. Для этого предусмотрены таблицы совместимости и несовместимости, в которых, соответственно настраиваются взаимосвязи единиц оборудования с точки зрения ремонтов. Режимы отслеживания конфликтов совместимости \несовместимости отображаются при рассмотрении графика и позволяют сделать вывод о необходимости смещения сроков.

Маршрутизация графика. Каждый созданный график обладает маршрутом прохождения, который состоит из внутренней и внешней части. Элементами внутренних частей являются пользователи и службы предприятия, внешних – другие предприятия. Существенным является тот факт, что внутренняя часть является общей для всего рассматриваемого графика, внешняя зависит от каждой конкретной ремонтируемой единицы оборудования.

Общая часть

К функциям части, вынесенной за пределы обоих программных комплексов, относятся:

Ведение нормативно-справочной информации (НСИ), справочников оборудования. Подсистема ведения НСИ и справочников оборудования обеспечивает создание, редактирование, печать, пересылку между предприятиями общих технологических справочников (вид заявки, категория заявки, вид ремонта, условия производства работ, программа переключений и др.). Для оборудования поддерживается управление-ведение и наследуемые дополнительные характеристики.

Механизм идентификации и контроль прав доступа пользователей к функциям системы. Права пользователей по доступу к функциям системы централизованно настраиваются администратором комплекса. Проверка прав пользователей осуществляется как на стороне АРМа пользователя, так и на стороне сервера.

Журнализация действий пользователя и системных событий.

Отслеживание истории работы с данными.

Формирование отчетов.

Выводы

Представлено описание задач оперативного управления и планирования ремонтов как составляющих задачи управления ремонтами.

Показаны различия задач, выделена общая часть в виде оборудования. Описаны функции программных комплексов, решающих указанные задачи.

Применение информационных технологий при выборе схемы расположения опор в конструктивно-силовой системе ГТД

В.А. Зрелов, А.С. Миронов, М.Е. Проданов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В процессе формирования технического задания (ТЗ) на предэскизном этапе проектирования газотурбинного двигателя (ГТД) проектировщик решает множество задач, одна из которых – выбор схемы расположения опор в конструктивно-силовой системе будущего двигателя. При выборе прототипа для проектирования будущего двигателя необходимо проводить анализ конструктивных решений, как на схемном уровне, так и на уровне конструктивных решений конкретных элементов двигателей. При этом проектировщику приходится хранить и обрабатывать большой объем бумажной и электронной документации. Из-за многообразия реализованных схем, начинающему инженеру, не обладающему большим опытом проектирования, выбрать прототип особенно трудно. Существует вероятность, что в процессе выбора он может пропустить важное решение. Ошибка инженера на предэскизном и эскизном этапах проектирования может стоить колоссальных затрат на ее исправление на последующих этапах жизненного цикла двигателя. Это может оказать негативное воздействие на его конкурентоспособность.

Исходя из выше сказанного, можно сделать вывод, что для помощи инженеру-проектировщику необходимо создавать комплекс программ автоматизации проектных действий.

Комплекс программ, поддерживающий функции проектировщика, возможно реализовать с помощью программного модуля, работающего в среде систем управления данными об изделии (PDM-системы). Данный модуль имеет интерфейс взаимодействия с PDM системой Smarteam. Работа программного модуля основана на SQL запросах в базу данных Smarteam. Для практической реализации комплекса выбираем PDM-систему Smarteam. В нашем случае используется функция PDM, как хранилища информации – база данных о двигателях (БД), а разрабатываемый программный модуль обеспечивает процесс поиска рационального решения.

Алгоритм поиска рационального схемного решения включает в себя следующий механизм:

Формирование списка конструктивных решений

Сформировать список конструктивных решений позволяет использование теории графов. Формализация задачи требует задействовать ряд множеств описаний конструктивных схем: множества схем роторов компрессоров и турбин с различным расположением опор и множество схем силового замыкания турбокомпрессора ГТД. Используя данные множества, строится ориентированный граф.

Представление процесса поиска возможных решений в виде ориентированного графа дает возможность последовательно объединять в различных сочетаниях вышеописанные множества.

Далее, используя элементы теории графов, создается список возможных конструктивных решений.

Формирование таблицы применяемости

Таблицу применяемости множества конструктивных решений составляется на основе ретроспективного анализа реализованных схем ГТД. Она является информационным полем множества реализованных схем ГТД. Полученные в графе пути решения, сравниваются с существующими в таблице применяемости и удаляются

пути, не нашедшие практической реализации на современном этапе развития технологий.

При формировании данной таблицы необходимо учитывать ограничения:

– данные из ТЗ (тяга двигателя, безвозвратные потери масла, виброскорость колебаний по опорам и т.д.);

– данные по результатам термогазодинамического расчета (расход воздуха, габаритные размеры и т.д.).

Формирование таблицы применяемости осуществляется средствами PDM-системы, которая позволяет рационально хранить данные и создавать алгоритмы их обработки.

Выбор наиболее рационального конструктивного решения в качестве прототипа

На данном этапе выбираются оценочные параметры для анализа конструктивных решений. Состав и количество оценочных параметров определяется методикой проектирования конкретной конструкторской организации.

Далее необходимо определить коэффициенты приоритета выбранных оценочных параметров. Коэффициенты рассчитываются с помощью метода парного сравнения. Метод парного сравнения применяется обычно, чтобы выявить «предпочтения» экспертов. Существует несколько способов попарного сравнения и упорядочения признаков. В работе программного модуля используется метод попарного сравнения параметров по важности, в результате которого определяются коэффициенты приоритета оценочных параметров. Это обусловлено тем, что экспертам необходимо оценить большое количество свойств исследуемого объекта и установить, что либо один из показателей важнее другого, либо они равновесны.

После расчета коэффициентов приоритета, происходит сравнение нормированных значений оценочных параметров в соответствии с техническим заданием по каждому выбранному конструктивному решению. При этом значения параметров извлекаются из базы данных под управлением PDM-системы. Расчет проводится по следующей формуле

$$\delta y_к = \rho_i \frac{y_{iк} - y_i^{ТЗ}}{y_i^{ТЗ}},$$

где ρ_i – коэффициент приоритета *i*-го параметра.

За прототип выбирается конструктивное решение, которое по совокупности нормированных значений оценочных параметров имеет наименьшее суммарное значение.

Расчет линейных размеров расположения подшипников в проектируемом двигателе

После выбора конструктивно-схемного решения в качестве прототипа необходимо рассчитать рациональное расположение подшипников в опорах. Для этого введем понятие относительных линейных коэффициентов, значения которых будут являться критериями выбора места расположения подшипников:

$\bar{L}_К = \frac{L_К}{L_{ТК}}$ - относительная длина компрессора;

$\bar{L}_Т = \frac{L_Т}{L_{ТК}}$ - относительная длина турбины;

$\bar{L}_{ОПК} = \frac{L_{ОПК}}{L_{ТК}}$ - относительное расстояние от подшипника до входа в компрессор;

$\bar{L}_{ОТК} = \frac{L_{ОТК}}{L_{ТК}}$ - относительное расстояние между подшипниками ТК.

На этом этапе инженеру отводиться роль наблюдателя. Расчет производится программой в автоматическом режиме. Программный модуль берет из хранилища данных информацию об относительных размерах расположения опор в двигателе прототипе. Далее она умножает относительные размеры расположения опор двигателя прототипа на длину турбокомпрессора проектируемого двигателя. В результате определяются линейные расстояния расположения опор в проектируемом двигателе.

Таким образом, применение информационных технологий управления данными об изделии и разработанного прикладного комплекса программ, позволяет на этапе создания ТЗ получить структуру конструктивно-силовой схемы турбокомпрессора с уточненным количеством опор и их расположением, а также подобрать прототип для дальнейшего проектирования газотурбинного двигателя.

Создание многофункциональных средств измерения физических величин для промышленных систем управления

С.В. Ионов, В.В. Козлов, М.Г. Мясникова, А.А. Трофимов
Пензенский государственный университет, Пенза

Развитие многофункционального приборостроения для промышленных систем управления возможно по двум направлениям, которые могут дополнять друг друга. Это разработка ряда унифицированных преобразователей многофункциональных датчиков и создание многофункциональной виртуальной системы на основе сбора информации с типовых измерительных преобразователей с последующей обработкой измерительной информации с помощью средств персонального или промышленного компьютера. Первое направление предполагает оснащение системы управления рядом датчиков, каждый из которых способен измерять две и более физические величины (например, скорость и ускорение, линейное и угловое перемещение и т.д.). Второе направление предполагает создание многофункционального прибора, который позволяет реализовать алгоритм любой сложности, обеспечить его универсальность и удобство использования в АСУ путем встраивания в ее программное обеспечение. Такие приборы используют современные методы цифровой обработки и позволяют определять различные параметры и характеристики сигналов сложной формы.

Оба эти направления развиваются на кафедре «Информационно-измерительная техника» Пензенского государственного университета.

Эффективность развития и эксплуатации современных автоматизированных систем контроля и управления различными технологическими процессами во многом определяется технико-экономическими характеристиками первичных средств сбора и обработки информации, среди которых значительный удельный вес имеют первичные преобразователи (датчики) давления, перемещений, расхода.

Информация с датчиков о контролируемых процессах поступает в автоматизированные системы управления, и её достоверность определяется метрологическими и надёжностными характеристиками датчиков в экстремальных условиях эксплуатации. При этом к датчикам предъявляется широкий комплекс метрологических и эксплуатационных требований, зачастую противоречащих, а иногда и исключающих друг друга.

По результатам анализа мировых тенденций совершенствования датчиков и на их основе систем телеизмерений, контроля, диагностики и управления можно отметить следующее:

– возможности современных промышленных систем управления в значительной степени зависят от эффективности использования датчиков. В будущем эта зависимость увеличится в связи с возрастанием потребности в получении больших объемов высококачественной информации и развитием технологии датчиков в направлении повышения их функциональных возможностей;

– важнейшим направлением развития датчиков является использование технологии микроэлектромеханических систем – MEMS, являющейся логическим продолжением развития микроэлектроники. Она сочетает в себе преимущества по надежности, миниатюризации, многокомпонентности и высокой степени интеграции. Вследствие гибкости, присущей интеграции электрических и механических компонентов, потенциальные возможности применения MEMS-технологий, в т.ч. в датчикостроении, представляются безграничными;

– к важнейшим общим технологиям датчиков относятся технологии полупроводников, сверхпроводников, цифровые, компьютерные технологии, а также алгоритмы и программы обработки информации;

– наиболее очевидной общей тенденцией в развитии технологии датчиков является все более возрастающая и доминирующая роль твердотельной технологии. Совершенствование датчиков будет следовать по тому же пути, что и твердотельная технология: миниатюризация, повышение быстродействия, снижение энергопотребления на функцию, повышение плотности и сложности монтажа приборов и реализация концепции интегральной сборки интегральных чувствительных элементов и модулей датчиков и, естественно, снижение их стоимости;

– характерной тенденцией в современной технологии систем датчиков является одновременное использование большого числа источников информации - в том числе с помощью пространственно рассредоточенного множества датчиков одного и того же типа, либо системы одиночных датчиков, обеспечивающих восприятие и преобразование информации о множестве физических величин;

– развитие технологии цифровых устройств первичного преобразователя является единственным наиболее важным фактором, который окажет влияние на технологии датчиков в обозримом будущем. Преимущества цифровой, а не аналоговой, формы сигналов или информации является определяющим. Отдельно остановимся на необходимости ускоренного развития датчиков с цифровым выходом.

В настоящее время разработка и производство аппаратуры представляют собой сложнейшие наукоемкие задачи и невозможны без создания новых теорий проектирования, использующих методы математического моделирования, без решения вопросов стандартизации и унификации изделий, без разработки специальных технологий, без автоматизации производства и процессов создания проектной документации, развития производственной, метрологической и испытательной баз.

Второе направление развития многофункционального приборостроения связано с созданием виртуальных средств измерения. Работы по данному направлению проводятся на кафедре «Информационно-измерительная техника» на протяжении последних лет. Виртуальное средство измерения представляет собой программно-аппаратный комплекс, создается на базе компьютера и ключевую роль в его функционировании наряду с аппаратурой играет программное обеспечение, которое и определяет его функциональные возможности. Анализ современной технической литературы показывает, что в виртуальных средствах измерения фактически воспроизводятся те же методы, которые использовались и аналоговых приборах, а современные методы цифровой обработки сигналов не нашли широкого распространения в измерительной технике. Поэтому необходимо рассмотреть методы и алгоритмы, позволяющие решить задачу определения параметров и характеристик электрических сигналов. А для возможности применения наиболее перспективных методов в задачах измерения также необходимо оценить метрологические характеристики каждого из них в отдельности и выявить возможные направления их усовершенствования.

Известно, что прямых математических методов, позволяющих связать результаты измерения с физическими параметрами составляющих сигналов сложной формы, не существует. Эта задача решается за счет применения параметрических методов, в которых на основе перехода от аналогового процесса к дискретному, описываемому в общем случае авторегрессионным уравнением скользящего среднего, осуществляется определение вектора параметров колебательных составляющих. Авторами разработана методика измерения, позволяющая указать погрешности, обеспечиваемые данным методом, а также установить значения параметров

регистрации и оцифровки, при которых можно гарантировать обеспечение заданной точности – метрологическая поддержка прибора.

На кафедре был разработан программно-аппаратный комплекс, в котором использовался один из параметрических методов – метод Прони, который традиционно применяется лишь как метод спектрального оценивания. Известно, что в методе используется наиболее естественная для описания свободных и вынужденных колебаний модель – сумма колебательных составляющих разной частоты с соответствующими амплитудами, фазами и затуханиями (инерционная и постоянная составляющие – частный случай). Следовательно, использование метода Прони в виртуальном приборе позволяет оценить параметры сигналов и частотных характеристик. А исследование методических погрешностей с последующей разработкой методики измерений позволит говорить об измерении параметров с требуемой точностью.

Однако параметрические методы являются достаточно трудоемкими, а выбор порядка авторегрессионной модели субъективен. Поэтому дальнейшее совершенствование методов обработки измерительной информации авторы видят в поиске новых перспективных в плане расширения функциональных возможностей, повышения точности и быстродействия методов цифровой обработки и построении алгоритмов измерения параметров и характеристик на основе их рационального сочетания.

Так, в настоящее время получил распространение метод декомпозиции сигнала на эмпирические моды (EMD), применение которого позволяет разложить сложный процесс на колебательные и (или) инерционные составляющие. Такое разложение позволит: 1) сразу определить порядок; 2) синтезировать сигнал, отбрасывая неинформативные составляющие; 3) существенно снизить трудоемкость последующего анализа, так как можно свести одну сложную задачу к простым задачам оценивания параметров составляющих первого и второго порядков (вместо применения предопределенного порядка к исходной выборке).

Сочетание двух направлений, развиваемых на кафедре, позволит создавать многофункциональные средства измерения различного назначения для промышленных систем управления. А привлечение к выполнению научно-исследовательской работы студентов, аспирантов и молодых специалистов позволит сформировать молодежный коллектив научных кадров, способный развивать данное направление измерительной техники.

Компьютерное моделирование цифроуправляемых генераторов опорных сигналов

А.М. Косолапов, В.П. Рузанов
Самарский Государственный Университет Путей Сообщения, Самара

В системах управления мобильными объектами, используются генераторы опорных синусоидальных сигналов. Моделирование устройств и автоматизированных комплексов для управления мобильными объектами, является актуальным, так как позволяет обеспечить определение оптимальных параметров.

Для воспроизведения цифроуправляемого сигнала используется двоично-ортогональное разложений в ряд Фурье – Уолша.

Функция $\sin \pi x$ может быть представлена рядом Тейлора в окрестности точки $x_k = k2^{-n}$, где $k = 0,1,2...,2^n - 1$; n – число старших разрядов входного управляющего кода

$$y = \sin \pi x = \sin \pi x_k + \pi (x - x_k) \cos \pi k - \frac{\pi^2}{2}(x - x_k)^2 \sin \pi x_k + ... \quad (1)$$

Функция $y_k = \sin \pi x_k$ задаётся с помощью ряда Уолша в точках x_k в виде:

$$y_k = \sum_{i=0}^{2^{n-2}-1} \alpha_{2i} wal_c(2i, x_k) + \sum_{i=0}^{2^{n-2}} \alpha_{2i-1} wal_c(2i-1, x_k) \quad (2)$$

где α_i – коэффициент разложения по функциям Уолша; $wal_c(i, x_k)$ – чётная i - ая функция Уолша.

Отличительной характеристикой представления гармонических сигналов рядами Уолша является ступенчатая аппроксимация воспроизводимого гармонического сигнала функции $\sin \pi x$.

В этом случае погрешность аппроксимации сигнала зависит от числа разрядов и определяется по формуле:

$$\Delta_a = \frac{\mathrm{A}}{2^{n+1}}, \quad (3)$$

где A – амплитуда. По заданной погрешности находятся координаты узлов интерполяции, и производится кусочно-линейная аппроксимация функции:

$$\sin \pi \hat{x} = y_k + (y_{k+1} - y_k) j 2^{-m}, \quad (4)$$

где $\hat{x}$ является дискретным аргументом, принимающим 2^{n+m} значений, при условии, что $0 \le \hat{x} < 1$; $j = 0,1,2,3...,2^m - 1$; m - число младших разрядов входного кода. В этом случае погрешность кусочно-линейной аппроксимации выражается

$$\Delta_{кла} = \frac{\Delta}{2^m}, \quad (5)$$

В результате при воспроизведении функции $\sin \pi x$, амплитуды первой и высших гармоник зависят от числа разрядов.

На рис. 1 представлена осциллограмма функции $\sin \pi x$, построенная в среде Matlab, с учетом соотношений (2) и (4).

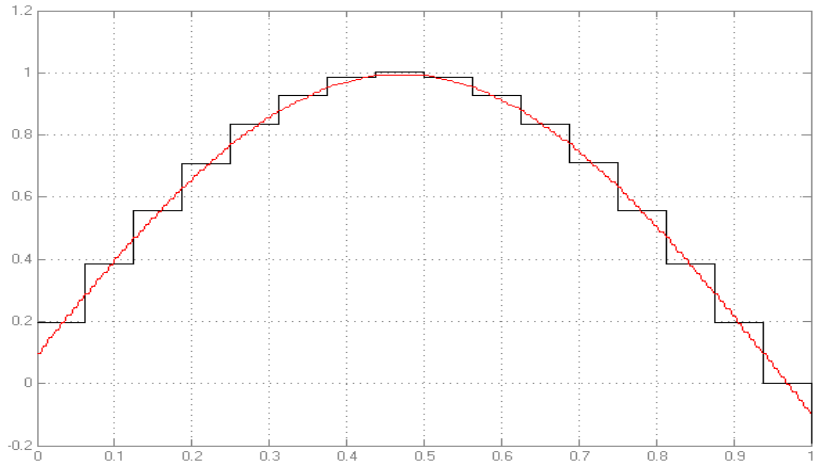


Рис. 1. Кусочно-линейная аппроксимация ступенчатого сигнала $\sin \pi x$

На рис. 2 представлена зависимость допустимого значения коэффициента гармоник $K_г$ воспроизводимого сигнала $\sin \pi x$, от параметров *m* и *n*, с учетом (2), (4) и следующего соотношения:

$$K_г = \frac{\sqrt{\sum_{i=2}^{n} Y_i^2}}{Y_1} * 100\%, \qquad (6)$$

где $K_г$ – коэффициент гармоник; Y_1 – действующее значение соответствующей гармоники.

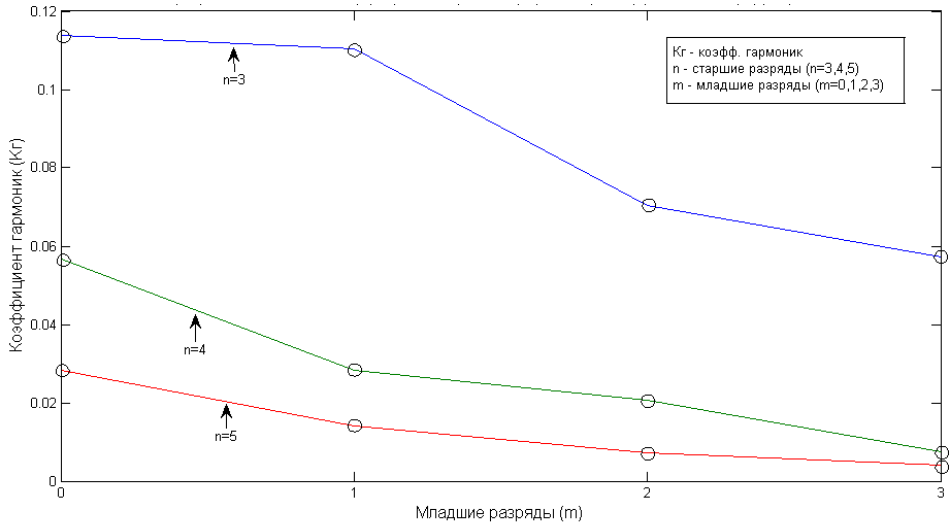


Рис. 2. Зависимость коэффициента гармоник от параметров *m* и *n*

При $K_г$ в диапазоне 0.1-0.3%, для воспроизведения гармонического сигнала целесообразно применять метод на основе ступенчатой аппроксимации при n = 3 и m = 0. При $K_г \le 0.05\%$, подход на основе ступенчатой аппроксимации ведет к существенному усложнению оборудования. Более предпочтительным является применение метода на основе кусочно-линейной аппроксимации при n = 4 или n = 5 и соответствующих значений *m* как видно из рис. 2.

Однако с повышением точности аппроксимации ступенчатого сигнала требуется пропорциональное увеличение частоты дискретизации, определяемой соотношением

$$f_Д = \pi * \frac{f}{\gamma_{a0}}, \qquad (7)$$

где f – частота воспроизводимого гармонического синусоидального сигнала $\sin \pi x$; γ_{a0} – приведенная погрешность аппроксимации, при этом

$$\gamma_{a0} = \frac{C_к}{2A}, \qquad (10)$$

где $C_к$ – ступень квантования.

Для генерирования ступенчато аппроксимированных сигналов с частотой $2*10^3$Гц и приведенной погрешности 10^{-3} необходимо, чтобы используемые счетчики и преобразователи кода в напряжение могли бы работать, учитывая (7), на частоте порядка $6{,}4*10^6$Гц. В этом случае возможности повышения частоты и точности ограничены.

При использовании фильтра нижних частот, достигается существенное расширение частотного диапазона за счет того, что при меньшем числе ступеней квантования, возможно, использовать более сильное сглаживание. При этом нечувствительность к амплитудным и фазовым погрешностям, возникающим при сглаживании, обеспечивается методом обратной связи, путем изменения амплитудной и фазовой характеристики сглаживающего фильтра.

Представим ступенчато аппроксимированный сигнал функции $\sin \pi x$ соотношением

$$y_k = A * \sin(\omega t + \varphi_н) + \sum_{i=1}^{\infty} (a_i \sin iM\omega t + b_i \cos iM\omega t), \qquad (9)$$

где $M = \pi^{-1} 2^{-n+1}$ – отношение частоты следования ступеней аппроксимирующего сигнала к частоте ω; a_i, b_i – амплитуды соответствующих гармоник; ω – частота первой гармоники; $\varphi_н$ – начальный фазовый сдвиг воспроизводимого сигнала.

После сглаживания ступенчато аппроксимированного сигнала составляющие с частотами $iM\omega$ уменьшаются. Изменение фазовой характеристики сглаживающего фильтра обеспечивает равенство фаз сигналов, т.е. обеспечивается компенсация фазовой погрешности, возникающей при сглаживании.

Воспроизводимый гармонический сигнал, определяемый соотношением (9) определяется только параметрами первой гармоники ступенчато аппроксимированного синусоидального сигнала, а именно частотой, фазой и амплитудой.

Пусть передаточная функция фильтра соответствует последовательному соединению l одинаковых апериодических звеньев с постоянной времени τ, тогда учитывая, что погрешность γ_c сглаживания

$$\gamma_c = 1 - K_c^{-1}(\omega), \qquad (10)$$

можно записать соотношение в виде

$$\varphi_c(\omega) = l * arctg\omega\tau = l\sqrt{(1-\gamma_c)^{-\frac{2}{l}} - 1}, \qquad (11)$$

где φ_c – фазовый сдвиг.

Амплитуда первой гармоники a_1 погрешности квантования до сглаживания определяется соотношением

$$a_1 \le \frac{C_к}{\pi} = A * 2^{-n+1}, \qquad (12)$$

Отношение амплитуд гармоник после сглаживания можно определить в виде выражения

$$Z_{\frac{a_i}{A}} = \frac{a_1}{A * K_c(\omega)} = 2^{-n+1} * K_c(\omega 2^n), \qquad (13)$$

где $K_c(\omega)$ – коэффициент передачи усилителя сглаживающего фильтра.

При одинаковых значениях $Z_{\frac{a_i}{A}}$ положительный эффект от сглаживания в отношении частоты дискретизации к частоте воспроизводимого сигнала, с учетом выражения (13), равен $K_c(\omega 2^n)$.Например, при γ_c= 3% выбирая величину *l*=4 из соотношения (11) и *n*=5, можно обеспечить $Z_{\frac{a_i}{A}} < 10^{-3}$. Следовательно, для воспроизведения гармонического сигнала с частотой $2*10^3$Гц, необходимо значение частоты дискретизации в 100 раз меньше, чем при использовании ступенчатой аппроксимации.

Результаты компьютерного моделирования позволяют оптимизировать число разрядов и обоснованно выбрать метод сглаживания для каждого конкретного случая.

Компьютерное моделирование датчиков тока мобильных объектов

А.М. Косолапов, Д.Н. Франтасов
Самарский Государственный Университет Путей Сообщения, Самара

При эксплуатации электрооборудования на мобильных объектах важным параметром являются весогабаритные характеристики устройства. Как известно, чем выше частота, тем меньше габариты электрооборудования переменного тока при одинаковой мощности, поэтому в бортовой аппаратуре в основном используется частоты от 400 Гц и выше. При перемагничивании ферромагнитных сердечников такого оборудования неизбежны потери энергии, и значит возникают потери информации при преобразовании сигналов. Определение этих погрешностей аналитическими методами весьма сложно, учитывая гистерезисный характер кривой намагничивания сердечников, особенно в динамическом режиме. Поэтому, компьютерное моделирование является наиболее эффективным способом исследования характеристик таких систем.

Наибольшее применение получили датчики тока на основе трансформаторов. При использовании таких типовых датчиков на мобильных объектах возникает проблема, так как соотношение точности и весогабаритных характеристик зачастую неприемлемо. Для решения проблемы предлагается введение корректирующего канала для датчика тока.

Обобщенная структурная схема трансформатора тока с коррекцией погрешности, пригодного для использования в широком диапазоне входных токов и частот, приведена на рис. 1.

На рис. 1 обозначено: I_1 – входной ток, $W_1 - W_6$ – обмотки трансформаторов, ТТ – типовой измерительный трансформатор тока, ТК – корректирующий трансформатор, $R_н$ – нагрузка, БК – блок корректирующего сигнала.

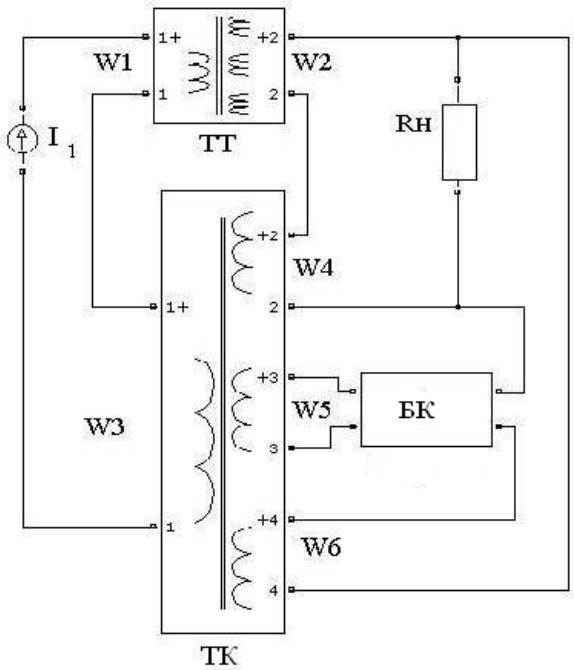


Рис. 1. Обобщенная структурная схема трансформатора тока с коррекцией погрешности (ТТК)

Процесс оптимизации не только статических, но и динамических параметров элементов при физическом моделировании довольно трудоёмкий. Программное обеспечение Matlab позволяют в значительной степени решить эти проблемы [1, 2].

Для исследования был выбран трансформатор тока класса точности 0.5 На рис. 2 показаны графики зависимостей токовой погрешности fi для измерительного токового трансформатора (ТТ) и для трансформатора тока с коррекцией погрешности (ТТК) от частоты входного сигнала f при номинальной величине первичного тока I_1=100 А и сопротивление нагрузки $R_н$=0.2 Ом.

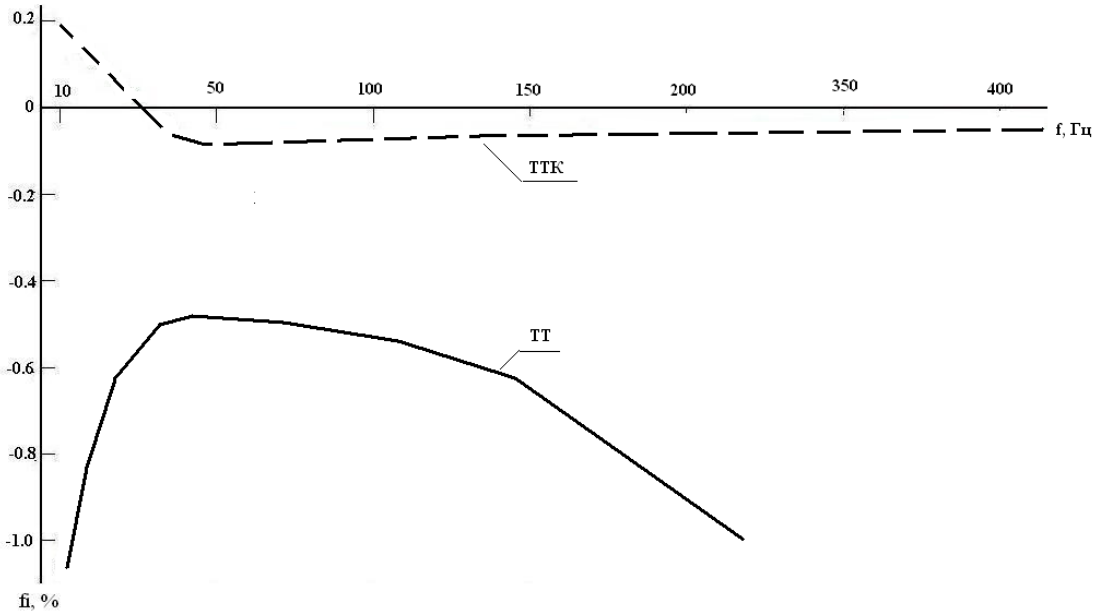


Рис. 2. График зависимости показателей токовой погрешности от частоты входного тока

Из графика видно, что ТТК имеет на порядок лучшие характеристики по сравнению с ТТ в достаточно широком диапазоне частот, при этом погрешность преобразования ПТК не превышает 0.05 %. А высокая точность устройства на частоте 400 Гц. позволит применять в мобильных объектах.

Ещё одним способом уменьшения весогабаритных характеристик может стать применение цифрового блока коррекции вместо аналогового, если корректирующий ток создавать с помощью электронных ключей, управляемых микроконтроллером. Такое решение также позволит значительно уменьшить рассеивание мощности усилителем.

На рис. 3 приведены зависимости номинальной токовой погрешности от величины первичного тока I_1 в диапазоне от 0 до 300 А, при использовании аналогового (ПТК) и цифрового (ПТЦК) блоков коррекции погрешности на частоте 400 Гц.

Использования цифрового блока коррекции погрешности вместо аналогового хоть и даёт незначительное увеличение токовой погрешности его результаты всё равно остаются на порядок лучшими по сравнению с некорректируемым трансформатором.

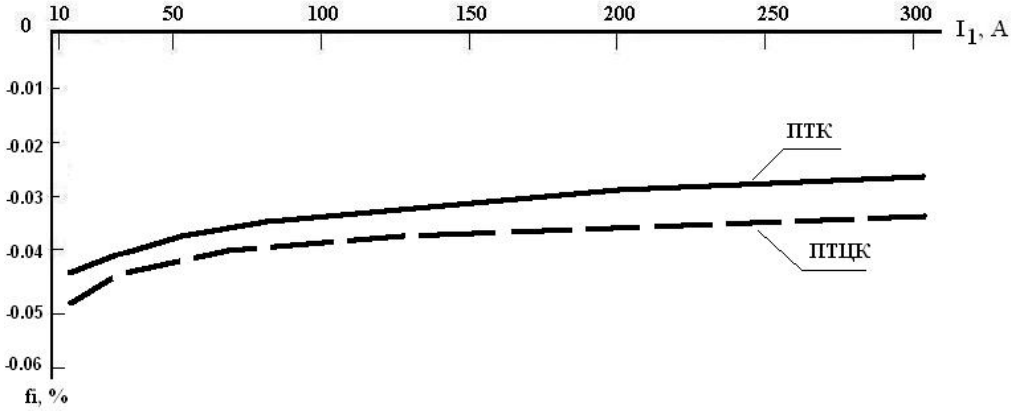


Рис. 3. Зависимость номинальной токовой погрешности от величины входного тока при использовании различных блоков коррекции погрешности

Из вышеизложенного можно сделать вывод, что существует возможность создания высокоточных приборов, входящих в состав систем измерения и управления, пригодных для эксплуатации в ограниченном пространстве мобильной техники.

Литература

1. Черных И.В. Simulink:среда создания инженерных приложений. - М.: Диалог-МИФИ, 2004. - 496 с.

2. Черных И.В. Моделирование электротехнических устройств в Matlab, SimPowerSystems и Simulink - М.: ДМК Пресс; СПб.: Питер, 2008. – 288 с.

Анализ погрешности измерения геометрических размеров лопаток газотурбинных двигателей

И.А. Лёзин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Программный комплекс аппроксимации двумерных плотностей вероятности ортогональными функциями и нейронными сетями позволяет решать широкий спектр задач, которые опираются на анализ двумерных плотностей вероятности. В том числе и задачи, связанные с обработкой комбинаций нескольких двумерных характеристик [1], распределение которых восстанавливается в аналитическом виде с помощью системы аппроксимативного анализа.

Аналитическое выражение двумерной плотности вероятности используется при решении множества задач: получение выражения условных плотностей по одной из осей при задании конкретного значения по другой оси; использование выражения для нескольких параметров при нахождении совместной плотности распределения нескольких величин и т.д.

Данный программный комплекс используется при обработке результатов натурного эксперимента по оценке погрешности базирования соединения типа «ласточкин хвост» лопаток компрессора ГТД при установке в наладке ПОМКЛ-БЛИК с использованием координатно-измерительной машины (см. рис. 1).

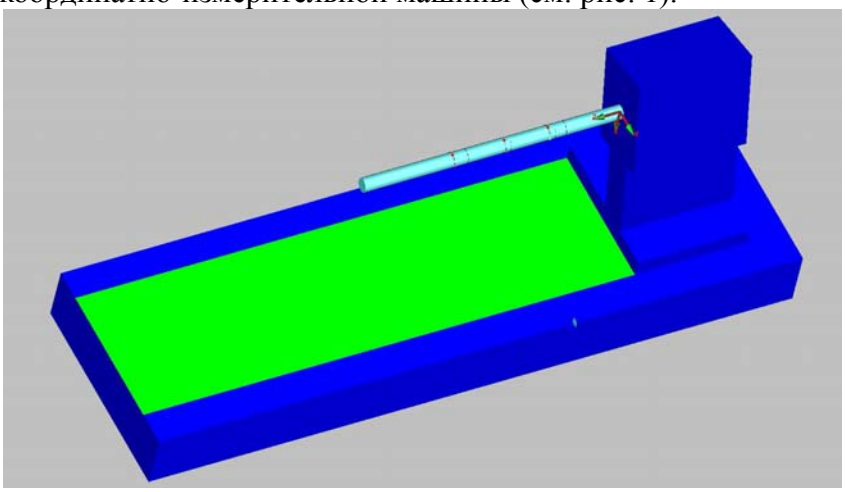


Рис. 1. Объемная модель наладки и контрольного эталона

КИМ DEA Global Performance 07.10.07 используется на кафедре производства двигателей летательных аппаратов СГАУ для измерения параметров различных деталей авиационных двигателей, в частности лопаток компрессора. С ее помощью для каждой лопатки проводится серия измерений, которая дает набор результатов, являющихся координатами измерения определенных точек для нескольких сечений лопатки (см. рис. 2).

С целью оценки качества измерений была поставлена задача определения характеристик набора полученных значений. С помощью программного комплекса аппроксимативного анализа строится аналитическая модель для плотности вероятности, характеризующей распределение отклонений измеренных координат точек лопаток от их эталонных значений.

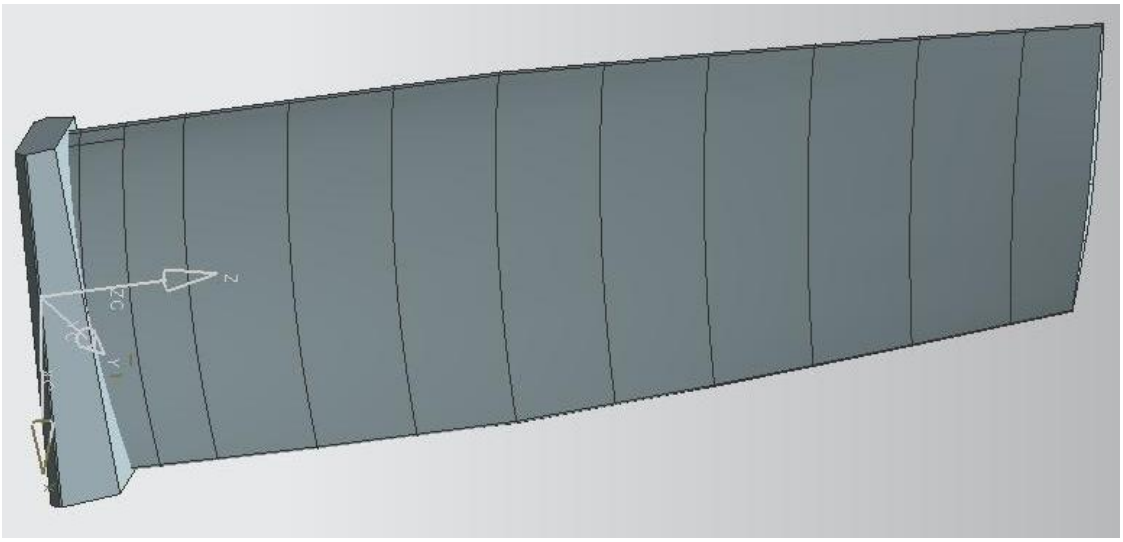


Рис. 2. Лопатка ГТД с поперечными сечениями по оси Z

Измерения с помощью КИМ проводятся с базированием лопатки в приспособлении, аналогичном используемому на приборе ПОМКЛ-БЛИК, что означает наследование части погрешностей базирования из прибора ПОМКЛ, который используется на ОАО «Моторостроитель».

Серии измерений, проведенных с помощью КИМ, позволяют нам оперировать двумя наборами данных – набором отклонений измеренных машиной координат точек лопаток и набором отклонений измеренных координат точек контрольного эталона.

Каждая серия измерений представляет собой многократный повтор снятия координат в заранее определенных точках. Отклонения полученных результатов от требуемых и есть тот набор данных, которым мы оперируем при оценивании погрешности измерений.

Сначала разберем погрешность, полученную при измерении эталона:

$$\Delta_{эт.} = \Delta_{нал.} + \Delta_{КИМ} . \quad (1)$$

Данная погрешность является суммой двух составляющих – погрешности базирования наладки ПОМКЛ в системе координат КИМ и погрешностью самой измерительной машины, которая допускает ошибки при снятии координат точек деталей.

Погрешность же измерения точек на спинке и корытце лопаток ГТД является более сложной, поскольку в отличие от эталона имеет больше составляющих:

$$\Delta_{лоп.} = \Delta_{нал.} + \Delta_{КИМ} + \Delta_{изг.} . \quad (2)$$

К уже имеющимся составляющим погрешности добавляется погрешность изготовления поверхности самой лопатки. Учитывая одинаковую природу первых двух составляющих погрешности (2) с погрешностью (1), ее можно записать в виде:

$$\Delta_{лоп.} = \Delta_{эт.} + \Delta_{изг.} . \quad (3)$$

Погрешность представляет собой набор большого числа измерений, результат которого случаен, то есть уместно говорить о том, что мы имеем дело со случайными величинами. Если обозначить погрешность (1) как случайную величину A, а погрешность (2) как случайную величину B, то через них можно оценить погрешность изготовления лопаток ГТД:

$$Z = B - A . \quad (4)$$

В силу формы эталона набор погрешностей (1) представляет собой последовательность точек на плоскости с координатами x и y. Аналогичным образом

представляется набор погрешностей измерений точек лопаток, разделенный по сечениям лопаток.

Оперируя данными, полученными в ходе экспериментов, с помощью системы аппроксимации двумерных плотностей вероятности находим выражения плотностей $f_A(x,y)$ и $f_B(x,y)$ для двумерных случайных величин A и B.

Учитывая, что для случайной величины и монотонной функции, для которых справедливо соотношение $y=\psi(x)$, плотность вероятности рассчитывается как

$$g(y)=f(\psi(x))|\psi'(x)|. \tag{5}$$

Таким образом

$$f_{-A}(x,y)=f_A(-x,-y). \tag{6}$$

Теперь величина погрешности Z является суммой погрешностей $-A$ и B:

$$(X_Z,Y_Z)=(X_{-A}+X_B,Y_{-A}+Y_B). \tag{7}$$

Функция распределения величины Z определяется как:

$$F_Z(x,y)=P(X_Z<x,Y_Z<y)=P(X_{-A}+X_B<x,Y_{-A}+Y_B<y). \tag{8}$$

Если при аппроксимации плотностей вероятности величин погрешностей $-A$ и B используются функции Эрмита [2], искомая плотность вероятности погрешности Z в таком случае определяется через известные выражения плотностей $f_{-A}(x,y)$ и $f_B(x,y)$:

$$\begin{aligned} f_Z(x,y)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{-A}(x_A,y_A)f_B(x-x_A,y-y_A)dx_Ady_A= \\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_A(-x_A,-y_A)f_B(x-x_A,y-y_A)dx_Ady_A \end{aligned} \tag{9}$$

или

$$\begin{aligned} f_Z(x,y)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_B(x_B,y_B)f_{-A}(x-x_B,y-y_B)dx_Bdy_B= \\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_B(x_B,y_B)f_A(x_B-x,y_B-y)dx_Bdy_B. \end{aligned} \tag{10}$$

Таким образом, мы получили оценку плотности распределения случайной величины Z, которая представляет собой распределение отклонений точек на поверхности лопатки ГТД от их требуемых значений.

В общем виде аналитическое выражение плотностей $f_{-A}(x,y)$ и $f_B(x,y)$ записывается следующим образом:

$$f(x,y)=\sum_{n=0}^{N}\sum_{m=0}^{M}\gamma_{nm}x^n y^m\varphi_x\left(\frac{x-\beta_x}{\alpha_x}\right)\varphi_y\left(\frac{y-\beta_y}{\alpha_y}\right). \tag{11}$$

Получаем следующее выражение для расчета плотности величины Z:

$$\begin{aligned} f_Z(x,y)=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{n=0}^{N_A}\sum_{m=0}^{M_A}\gamma_{nm,A}(-x_A)^n(-y_A)^m\varphi_{x,A}\left(\frac{x_A-\beta_{x,A}}{\alpha_{x,A}}\right)\varphi_{y,A}\left(\frac{y_A-\beta_{y,A}}{\alpha_{y,A}}\right)\cdot \\ &\sum_{k=0}^{N_B}\sum_{l=0}^{M_B}\gamma_{kl,B}(x-x_A)^k(y-y_A)^l\varphi_{x,B}\left(\frac{x-x_A-\beta_{x,B}}{\alpha_{x,B}}\right)\varphi_{y,B}\left(\frac{y-y_A-\beta_{y,B}}{\alpha_{y,B}}\right)dx_Ady_A \end{aligned} \tag{12}$$

Выбранной формуле определения плотности суммы двух величин соответствуют выражения плотности, восстановленные с использованием базисных функций Эрмита. После раскрытия скобок и подстановки сомножителей выражение (12) примет вид:

$$f_Z(x,y)=\sum_{n=0}^{N_A}\sum_{m=0}^{M_A}\sum_{k=0}^{N_B}\sum_{l=0}^{M_B}\sum_{i=0}^{k}\sum_{j=0}^{l}(-1)^{n+m+k-i+l-j}C_k^iC_l^j\gamma_{nm,A}\gamma_{kl,B}x^iy^j\int_{-\infty-\infty}^{\infty}\int x_A^{n+k-i}y_A^{m+l}$$
$$e^{-\left(x_A-\beta_{x,A}/\alpha_{x,A}\right)^2/2-\left(y_A-\beta_{y,A}/\alpha_{y,A}\right)^2/2-\left(x-x_A-\beta_{x,B}/\alpha_{x,B}\right)^2/2-\left(y-y_A-\beta_{y,B}/\alpha_{y,B}\right)^2/2}dx_Ady_A. \quad (13)$$

Для получения конечного выражения необходимо рассчитать интеграл под знаком суммы, который распадается на произведение интегралов:

$$I=\int_{-\infty}^{\infty}x_A^{n+k-i}e^{-\left(x_A-\beta_{x,A}/\alpha_{x,A}\right)^2/2-\left(x-x_A-\beta_{x,B}/\alpha_{x,B}\right)^2/2}dx_A\int_{-\infty}^{\infty}y_A^{m+l-j}e^{-\left(y_A-\beta_{y,A}/\alpha_{y,A}\right)^2/2-\left(y-y_A-\beta_{y,B}/\alpha_{y,B}\right)^2} \quad (14)$$

.

Раскрыв скобки и приведя сложение коэффициентов, для каждого из двух интегралов мы приходим к виду:

$$I_m(x,a_x,b_x,c_x)=\int_{-\infty}^{\infty}x^me^{-a_x(x-b_x)^2-c_x}dx, \quad (15)$$

где

$$a_x=\frac{1}{2\alpha_{x,A}^2}+\frac{1}{2\alpha_{x,B}^2},$$
$$b_x=\left(\frac{\beta_{x,A}}{\alpha_{x,A}^2}+\frac{x-\beta_{x,B}}{\alpha_{x,B}^2}\right)\frac{1}{2a_x}=\frac{\alpha_{x,B}^2\beta_{x,A}+\alpha_{x,A}^2\left(x-\beta_{x,B}\right)}{\alpha_{x,A}^2+\alpha_{x,B}^2},$$
$$c_x=\frac{\beta_{x,A}^2}{2\alpha_{x,A}^2}+\frac{\left(x-\beta_{x,B}\right)^2}{2\alpha_{x,B}^2}-b_x^2a_x=\frac{\beta_{x,A}^2}{2\alpha_{x,A}^2}+\frac{\left(x-\beta_{x,B}\right)^2}{2\alpha_{x,B}^2}-\frac{\left(\alpha_{x,B}^2\beta_{x,A}+\alpha_{x,A}^2\left(x-\right.\right.}{2\left(\alpha_{x,A}^2+\alpha_{x,B}^2\right)\alpha_{x,A}^2\alpha} \quad (16)$$
$$=\frac{\left(x-\beta_{x,A}-\beta_{x,B}\right)^2}{2\left(\alpha_{x,A}^2+\alpha_{x,B}^2\right)}.$$

После замены переменных в (15):

$$I_m(x,a_x,b_x,c_x)=e^{-c_x}\int_{-\infty}^{\infty}(u+b_x)^me^{-a_xu^2}du=e^{-c_x}\sum_{p=0}^{m}C_m^pb_x^{m-p}\int_{-\infty}^{\infty}u^pe^{-a_xu^2}du=$$
$$=e^{-c_x}\sum_{p=0}^{m}C_m^pb_x^{m-p}I_p(u,a_x). \quad (17)$$

Определенный интеграл $I_p(u,a_x)$:

$$I_0(u,a_x)=\int_{-\infty}^{\infty}e^{-a_xu^2}du=\sqrt{\frac{\pi}{a_x}},$$
$$I_{2k-1}(u,a_x)=\int_{-\infty}^{\infty}u^{2k-1}e^{-a_xu^2}du=0, \quad (18)$$
$$I_{2k}(u,a_x)=\int_{-\infty}^{\infty}u^{2k}e^{-a_xu^2}du=\frac{2k-1}{2a_x}\int_{-\infty}^{\infty}u^{2k-2}e^{-a_xu^2}du=\frac{2k-1}{2a_x}I_{2k-2}(u,a_x)=\frac{(2k-1)!!}{(2a_x)^k}$$

После расчета интегралов (15), раскрытия скобок и приведения слагаемых выражение (13) приводится к виду (11).

Используя данное выражение, можно определить различные характеристики: точку с максимальной вероятностью, математическое ожидание, границы области, в которой измеренная характеристика окажется с заданной вероятностью, вероятность выхода за пределы зоны допустимой погрешности и т.д. Например, вероятность того, что погрешность изготовления лопатки попадает в некую заданную область D, определяется через рассчитанную оценку:

$$p(Z \in D) = \iint_D f_Z(x, y)dxdy . \qquad (19)$$

Для каждого сечения лопатки по результатам серии измерений вычисляется плотность распределения погрешности изготовления и рассчитывается так называемое «пятно рассеивания» измеренных координат по осям x и y с центром в точке математического ожидания. Для пятна рассеивания, которое имеет форму эллипса, рассчитываются уравнения линий, на которых лежат его меньшая и большая полуоси, а также величины этих полуосей.

Таким образом, с вероятностной точки зрения оценивается качество изготовления лопатки ГТД через анализ погрешностей для каждого из ее сечений.

Литература

1. Муха, В.С. Анализ многомерных данных [Текст] / Монография // Мн.: УП «Технопринт», 2004. – 368 с. – 250 экз. – ISBN 985-464-676-9.

2. Суетин, П.К. Ортогональные многочлены по двум переменным [Текст] / М.: Наука. Гл. ред. физ.-мат. лит-ры, 1988. – 384 с. – 3000 экз. – ISBN 5-02-13757-X.

Автоматизированная система обработки акустических сигналов при диагностике сварных соединений

А.Ю. Модин
Самарский государственный технический университет, Самара

Ультразвуковой контроль нашел широкое применение в практике дефектоскопии изделий. Согласно отчету института доктора Ферстера (INSTITUTE DR. FORSTER) [1], в восточно-европейских странах объемы неразрушающего контроля, разделенные по методам, распределяются следующим образом – рис. 1.

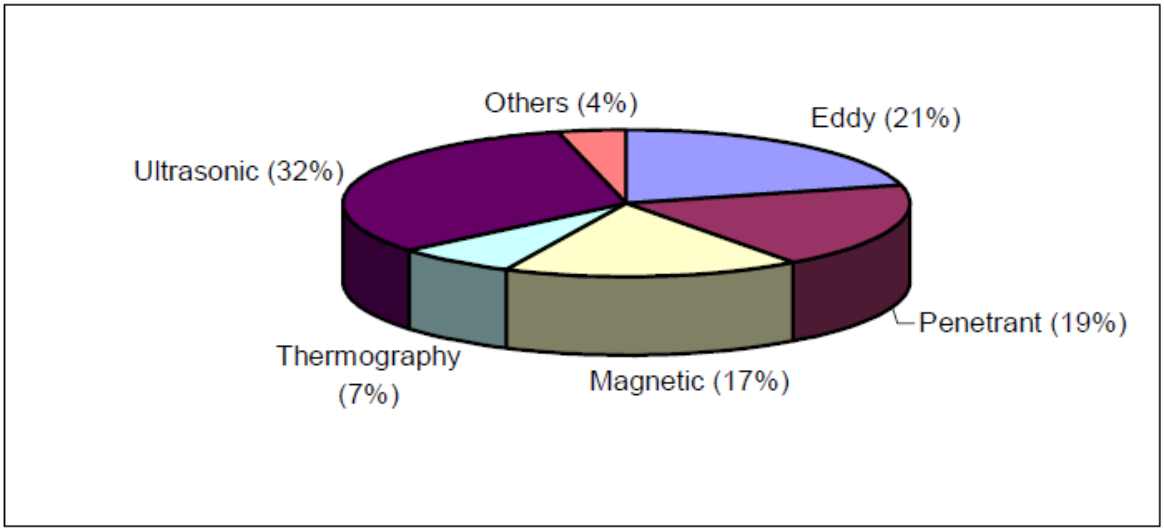


Рис. 1. Распределение объемов неразрушающего контроля

Ультразвуковой контроль здесь занимает главенствующее положение. На его долю приходится 32% объема контроля всех изделий. В отчете так же отмечается, что, несмотря на высокий уровень автоматизации труда в развитых странах, доля ручного ультразвукового контроля остается наибольшей. Это связано с тем, что большие объемы работ проводят на объектах, находящихся в эксплуатации: атомные и тепловые электростанции, трубопроводы различного назначения, металлоконструкции, транспортные средства и т.д. Характерной чертой указанных объектов является большое разнообразие конструкций, и следовательно, методик ультразвукового контроля, поэтому работы по контролю плохо поддаются автоматизации.

Самой распространенной формой представления отраженных сигналов при ручном ультразвуковом контроле является развертка типа А-скан. Она используется дефектоскопистами для оценки отраженного сигнала, а именно для определения размеров и координат дефекта. Размеры оценивают по амплитуде, координаты по времени. Т.е. чем больше амплитуда, тем больше дефект и чем больше время, тем дальше дефект. Система обработки акустических сигналов в данном конкретном случае берет картинку с экрана дефектоскопа с помощью видеоинформационных средств, например веб-камеры, и передает ее на ЭВМ, которая в свою очередь по заранее определенному алгоритму извлекает эти два параметра из развертки А-скана. Основной алгоритм построен на принципе последовательного приближения с использованием чашечных весов и гирь, который восходит к 1500-м годам [2] (см. рис. 2).

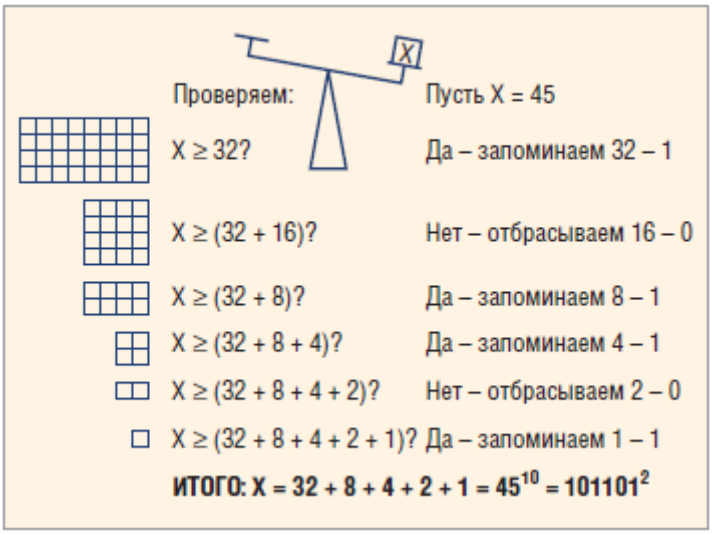


Рис.2. Алгоритм последовательного приближения с использованием чашечных весов и гирь

Он связан с решением полезной математической головоломки – определение неизвестного веса путем минимальной последовательности операций взвешивания. По условию этой задачи необходимо найти наименьшее число гирь для взвешивания предметов весом от 1 до 40 фунтов на чашечных весах. Одно из решений нашел математик Тарталья в 1556г. Он предложил использовать двоичную серию гирь весом 1, 2, 4, 8, 16 и 32 фунта. Предложенный алгоритм тот же, что и в современных АЦП последовательного приближения, который был выбран за алгоритм определения амплитуды эхо-сигнала. Но прежде чем определять амплитуду отраженного сигнала с помощью данного алгоритма картинка видеосигнала проходит последовательную обработку, блок-схема которой изображена на рис. 3.

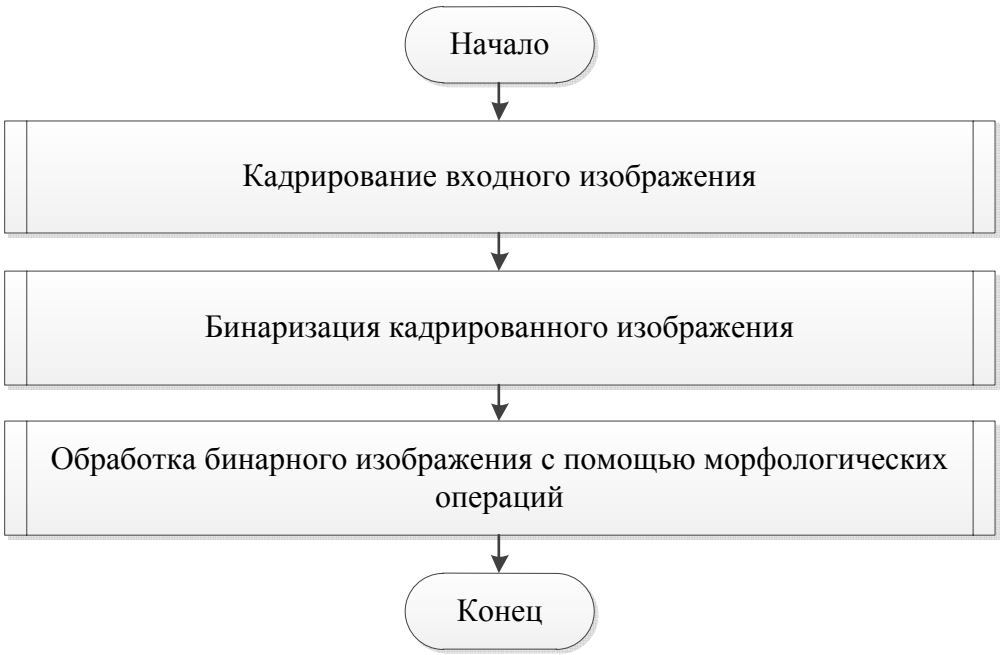
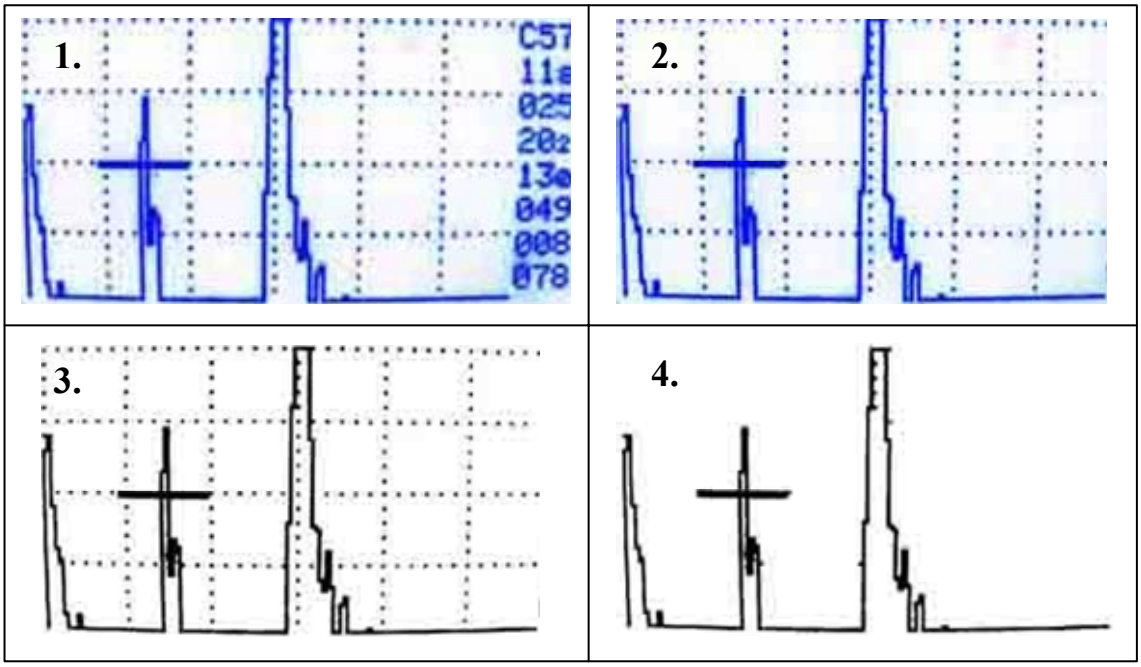


Рис. 3. Блок-схема обработки входного видеосигнала



1 – изображение входного видеосигнала;
2 – кадрированное изображение видеосигнала;
3 – бинарное изображение видеосигнала;
4 – изображение видеосигнала после обработки морфологическими функциями;

Рис. 4. Этапы обработки входного видеосигнала

Вычисленные координаты дефекта – это координаты относительно ультразвукового преобразователя (УЗП). Зная вычисленные координаты дефекта, но не зная где находился УЗП в это время, мы не знаем ничего о том, где находится дефект. Т.е. для того чтобы определить положение дефекта нужно знать вычисленные координаты дефекта и координаты УЗП относительно какой-нибудь постоянной точки отсчета. Для этой цели разрабатывается устройство, за основу которого взята оптическая мышь, позволяющее определять смещение УЗП от начальной точки. Нулевая точка отсчета или начальная точка это место помеченное маркером на объекте контроля, на котором был установлен УЗП и с которого был начат ультразвуковой контроль.

Условно систему обработки акустических сигналов можно разделить на две части программную и аппаратную. Программная часть на первом этапе была представлена в виде алгоритма и на сегодняшний день реализована в математическом пакете MATLAB [3]. С помощью разработанного программного обеспечения можно строить различные виды ультразвуковых изображений, таких как W-скан, B-скан, C-скан, квазитрехмерное и полутоновое изображения дефектных зон (см. рис. 5).

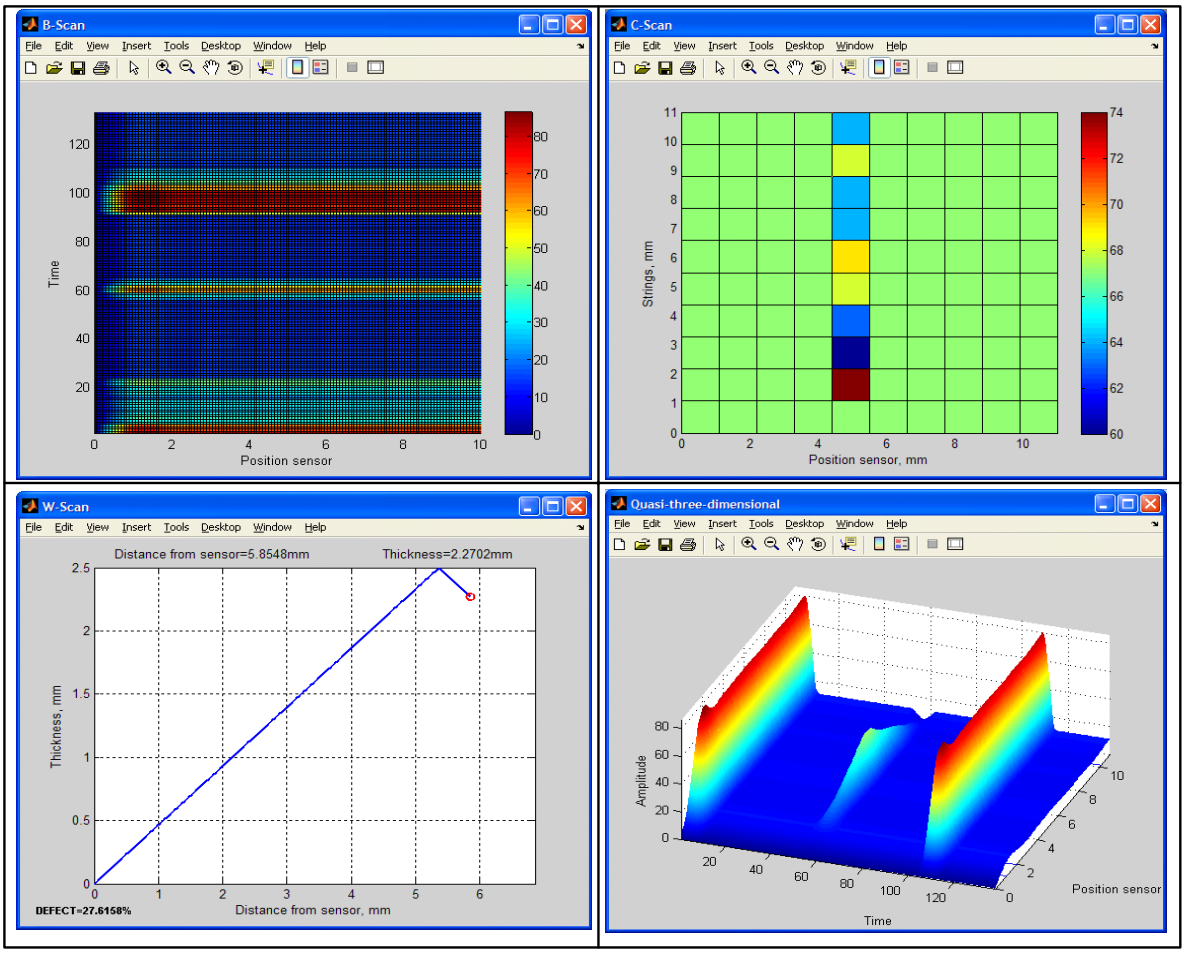


Рис. 5. Различные виды ультразвуковых изображений

К аппаратной части относится видео аппаратура и планарный оптический датчик перемещения. В качестве видео аппаратуры была выбрана веб-камера. В первую очередь это обусловлено ее дешевизной и простотой установки. Что касается характеристик, то их с запасом хватает на то, чтобы решить поставленную задачу. В результате эксперимента растяжения полиэтиленовой пленки проведенного в работе [4] было показано, что оптическая мышь может служить датчиком двумерного перемещения. Так как размеры УЗП предназначенных для контроля труб малого диаметра в несколько раз меньше оптической мыши, то она была поделена на две части: первичный преобразователь, который по размерам сопоставим с УЗП и физический интерфейс для соединения первичного преобразователя с ЭВМ. Для измерения смещения от нулевой точки отсчета первичный преобразователь оптической мыши жестко соединен с УЗП. Запуск и останов всех устройств синхронизирован. Наиболее удобной оказалась аппаратная реализация с помощью кнопки расположенной на нижней части первичного преобразователя оптической мыши. Таким образом, при контакте с объектом контроля происходит запуск записи видеоряда А-сканов и измерение смещения УЗП. Останов же произойдет, как только акустический контакт с объектом контроля будет нарушен.



Рис. 6. Размеры УЗП предназначенного для контроля труб диаметром 16мм

При использовании системы обработки акустических сигналов результатом проведенного УЗК будет запись всех зафиксированных эхо-сигналов попавших в стробируемую область экрана дефектоскопа со всеми необходимыми параметрами для их последующей идентификации, а именно: амплитуда, время и смещение относительно нулевой точки. Используя полученные данные можно строить различные виды изображений ультразвуковых сигналов, таких как W-развертка, полутоновые изображения дефектных зон, квазитрехмерные изображения, вплоть до трехмерной модели объекта с расположенными в нем дефектами.

В результате к любому дефектоскопу с возможностью представлять на своем экране отраженные сигналы в виде развертки А-скан может быть применена данная система обработки акустических сигналов стоимость аппаратной части, которой не будет превышать 100$.

Литература

1. Report on the actual situation of INSTITUTE DR. FORSTER. Information for customer and friends of INSTITUTE DR. FORSTER. №12/Dec, 1993.

2. Уолт Кестер Какая архитектура АЦП подходит для вашей задачи // Современная электроника. №3 – 2008 – С. 14-21.

3. Консультационный центр MATLAB компании Softline [Электронный ресурс]. – Режим доступа: http://matlab.exponenta.ru, свободный. – Загл. с экрана.

4. Ng, T.W. The optical mouse as a two-dimensional displacement sensor // Science Direct. A 107. Sensor and actuators. – 2003 – C. 21-25.

Функциональная модель береговой системы наблюдения

А.И. Моисеев
ФНПЦ ОАО «НПО «Марс», Ульяновск

Введение

Этапы анализа и проектирования системы являются определяющими вне зависимости от модели жизненного цикла. Так моделирование процесса функционирования системы при ее разработке позволяет проанализировать поведение объекта как в целом, так и на уровне его составных узлов. Изучение функциональной модели системы позволяет определить «узкие» места рассматриваемого объекта, отрицательно влияющие на эффективность функционирования БСН; сформировать и обосновать нормативной модели БСН; разработать и реализовать мероприятий по переходу от существующей («as is») к нормативной («to be») моделям; разработать проект автоматизированной системы управления БСН.

Процесс функционального моделирования БСН целесообразно разделить на два этапа:

1) Выбор и обоснование методологии функционального моделирования.
2) Моделирование БСН.

Выбор методологии моделирования

В настоящее время используется большое количество подходов к созданию функциональных моделей. Важнейшими из подходов являются структурный (IDEF), объектно-ориентированный (UML), отдельно выделяемая методология ARIS.

Для выбора методологии необходимо определить критерии их сравнения. Учитывая цель моделирования, а именно возможность оценки эффективности БСН, определяющими факторами при сравнении будут следующие:

1) Возможность глубокого анализа функционала системы с возможностью детализации от общего к частному.

2) Логическая целостность и полнота описания изучаемого процесса, а именно объединение в одной методологии организационного, управленческого, информационного аспектов функционирования системы.

3) Сравнительная наглядность представления информации.

Учитывая критерии сравнения и особенности подходов к функциональному моделированию, наиболее подходящей методологией для анализа функционирования БСН является IDEF. Данный выбор обусловлен прежде всего возможностями иерархического представления функционала системы и комплексным взглядом на процесс наблюдения, сочетаемые с отсутствием избыточных элементов, что в свою очередь обеспечивает легкость восприятия модели.

Разработка функциональной модели БСН

Береговая система наблюдения (БСН) – это организационно-техническая система, осуществляющая добывание, сбор, обработку и передачу на вышестоящую систему информации по наблюдаемым объектам. Функционирование БСН характеризуется протеканием двух основных процессов: информационного и командного.

Смысл первого из них заключается в выполнении основной задачи системы – получении информации о наблюдаемых объектах (рис. 1). Под информацией в данном случае понимаются элементы движения целей (координаты, курс, скорость), их физическая (лодка, корабль, летательный аппарат и т.д.) и тактическая принадлежность (гражданский, сторожевой, штурмовой, обеспечивающий и т.д.). Получение данных по цели с помощью средств наблюдения – не единственная составляющая

информационного процесса. Большое значение здесь имеет доставка информации в центр обработки информации и ее последующая своевременная выдача в вышестоящую систему.

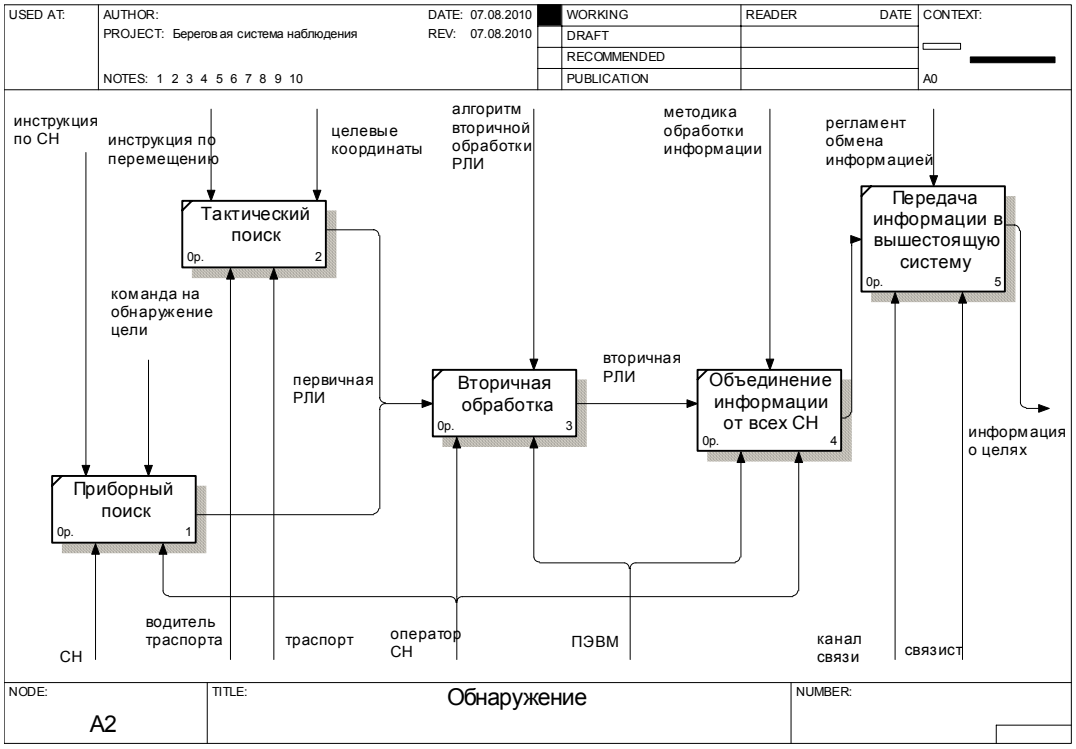


Рис. 1. Информационный процесс

Рассматривая информационный процесс детальнее, выделим два параллельных подпроцесса приборного и тактического поиска, протекающих в первую очередь. За обследование пространства наблюдения с помощью использования режимов и методов работы средств наблюдения отвечает подпроцесс приборного поиска. Подпроцесс тактического поиска отличается использованием переноса зоны действия средств наблюдения в пространстве путем соответствующего маневрирования их носителей. Результатом подпроцессов поиска является генерация первичной радиолокационной информации о наблюдаемом объекте. Затем данная информация проходит вторичную обработку с целью отождествления информации, вычисления курса и скорости объекта и борьбы со случайными всплесками и временными пропаданиями целей. Следующий подпроцесс – объединение информации от всех СН – имеет на входе вторично обработанную информацию. На его выходе генерируются уточненные характеристики и трассы движения целей, а также точности разных источников информации. Завершающим информационным подпроцессом является передача данных по сопровождаемым целям в вышестоящую систему. На данном этапе производится подготовительная обработка информации в зависимости от задачи, поставленной вышестоящей системой. Кроме того, на основе учета регламента связи формируется пакет данных и выдается в канал связи.

Используя методологию IDEF моделирования, проанализируем ресурсы и управляющие воздействия на информационный процесс. В процессе наблюдения для обеспечения информационного процесса необходимы следующие ресурсы: непосредственно самими СН, операторы СН, носитель источника информации, канал связи и связист. При этом требуются управляющие воздействия в части алгоритмов и инструкций по перемещению СН, обработке и передаче информации. Инициирующим сигналом являются команды на организацию приборного и тактического поиска.

Особенностью командного процесса (рис. 2) является анализ и изменение параметров БСН. Так в ходе функционирования БСН изменяются позиции расположения, форма диаграммы направленности и режимы работы СН. Командный процесс характеризуется следующей последовательностью действий: получение команды от вышестоящей системы на смену параметров БСН, оценка текущего значения и вычисление оптимального значения параметра, составление плана и смена состояния БСН за счет изменения значения управляемого параметра. В результате протекания процесса управления генерируются управляющие воздействия подсистемы управления на процесс обнаружения целей.

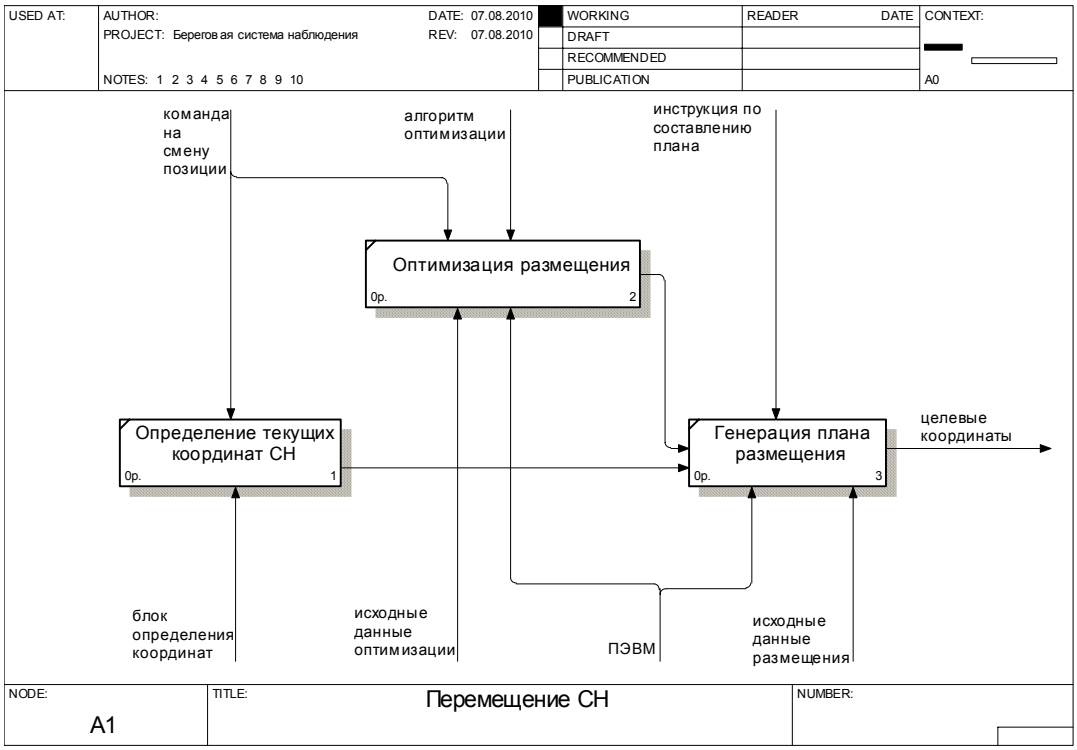


Рис. 2. Командный процесс

Используя методологию IDEF моделирования, проанализируем ресурсы и управляющие воздействия на командный процесс. Данный процесс принимает от вышестоящей системы команды на выполнение задач. Кроме того, важным управляющим воздействием являются алгоритм по составлению плана и инструкцию по составлению плана. В части ресурсообеспеченности ключевую роль играют средства и исходные данные оптимизации.

Выводы

Полученная функциональная модель БСН позволяет получить целостный взгляд на процесс наблюдения. При этом такой подход позволяет ориентироваться не только на создание прикладных информационных систем, а открывает возможность решения системных проблем, присущих самой предметной области.

Анализ алгоритмов восстановления частотной синхронизации с использованием преамбулы

И.М. Нестеров
Московский физико-технический институт (государственный университет), город Долгопрудный

Введение

Спутниковый канал связи представляет собой канал с белым гауссовым шумом. Для детектирования на демодуляторе необходимо восстановление синхронизации: тактовой, частотной и фазовой. Существует три класса алгоритмов восстановления частотной синхронизации: с добавлением данных, с управлением по решению, без добавления данных. Рассмотрен класс алгоритмов с добавлением данных на примере трех алгоритмов Кэя, Фитца, Луиса и Реджианнини. Алгоритмы данного класса выполняют оценку частотного сдвига по известной приемнику преамбуле [1].

Формулы для оценки частотного сдвига

На передатчике формируется сигнал вида $s(t)=\sum_i c_i g(t-iT-\tau)$, где c_i - символы данных, $g(t)$ - характеристика фильтра на передатчике. Для MPSK модуляции символы данных имеют вид $\left\{c_k=e^{j\alpha_k}:\alpha_k=0,2\pi/M,...,2\pi(M-1)/M\right\}$ [2]. Принятый сигнал имеет вид $r(t)=e^{j[2\pi\nu t+\theta]}s(t)+w(t)$, где ν - частотный сдвиг, $w(t)$ - белый гауссов шум.

В приемнике осуществляется согласованная фильтрация. В результате $z(k)=e^{j[2\pi\nu(kT+\tau)+\theta]}+w'(k)$. Схема исключения символов данных на приемнике изображена на рис.1.

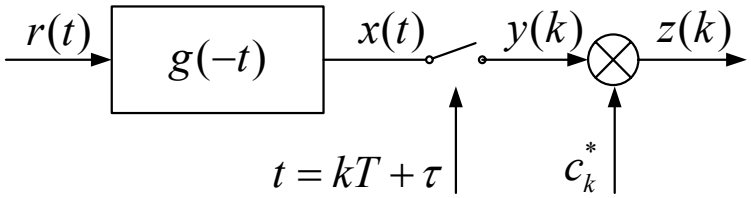


Рис.1. Схема исключения символов данных на приемнике

В зависимости от типа усреднения, алгоритмы имеют разные характеристики и вычислительные сложности. Ниже приведены формулы для оценки частотного сдвига для трех алгоритмов Кэя, Фитца, Луиса и Реджианнини [3].

$$\hat{\nu}=\frac{1}{2\pi T}\sum_{k=1}^{L_0-1}\gamma(k)\arg\left\{z(k)z^*(k-1)\right\}$$ – алгоритм Кэя,

где $\gamma(k)=\frac{6k(L_0-k)}{L_0(L_0^2-1)}$ $k=1,2,...,L_0-1$, L_0 – длина преамбулы,

$$\hat{\nu}=\frac{1}{\pi N(N+1)T}\sum_{m=1}^{N}\arg\left\{\frac{1}{L_0-m}\sum_{k=m}^{L_0-1}z(k)z^*(k-m)\right\}$$ – алгоритм Фитца,

где $N<\frac{1}{2|\nu_{\max}|T}$ – параметр.

$$\hat{\nu}=\frac{1}{\pi(N+1)T}\arg\left\{\sum_{m=1}^{N}\frac{1}{L_0-m}\sum_{k=m}^{L_0-1}z(k)z^*(k-m)\right\}$$ – алгоритм Луиса и Реджианнини,

где $N<\frac{1}{|\nu_{\max}|T}$ – параметр.

Результаты моделирования

Качественной характеристикой работы алгоритмов служит зависимость дисперсии нормированной частоты (нормирована на символьную скорость) от отношения сигнал/шум. Величина дисперсии нормированной частоты для рассматриваемой модели не должна превышать 10^{-6}. В результате моделирования спутникового канала с QPSK модуляцией в среде MATLAB были получены характеристики алгоритмов Кэя, Фитца, Луиса и Реджианнини (рис.2).

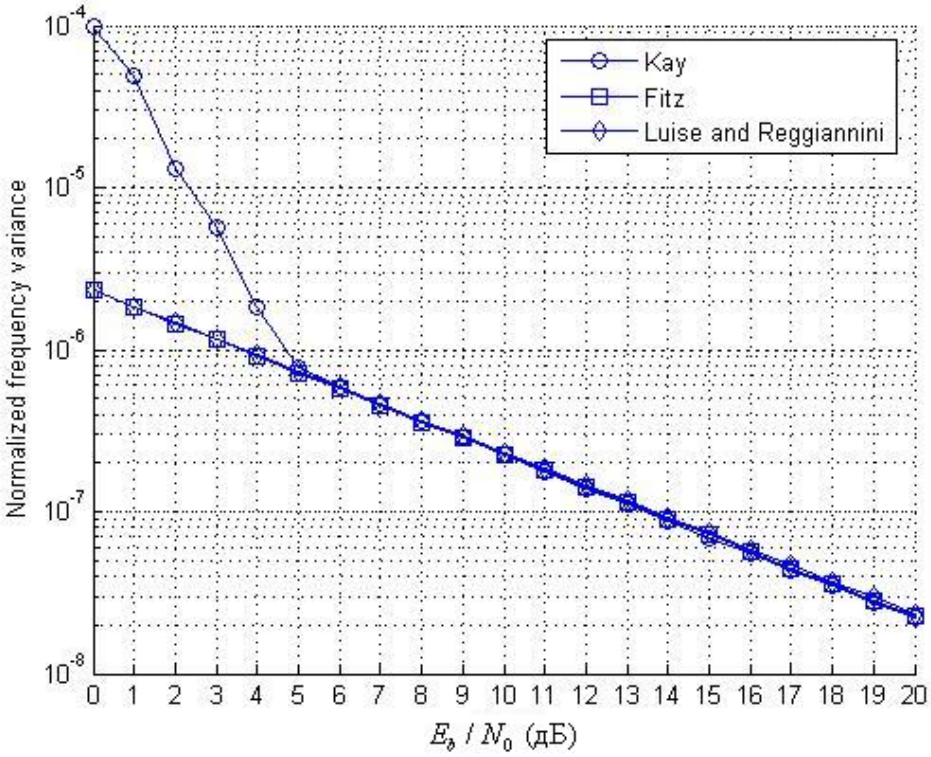


Рис.2. Графики зависимости дисперсии нормированной частоты от отношения сигнал/шум для алгоритмов Кэя, Фитца, Луиса и Реджианнини (длина преамбулы 32 символа, частотный сдвиг 0.01).

Сложность реализации алгоритма зависит от его вычислительной сложности (количество комплексных умножений). Формулы для нахождения вычислительных сложностей алгоритмов представлены в таблице 1.

Таблица 1

Алгоритм	Вычислительная сложность
Кэя	$O(L_0\text{-}1)$
Фитца	$O(N(2L_0\text{-}N\text{-}1)/2)$
Луиса и Реджианнини	$O(N(2L_0\text{-}N\text{-}1)/2)$

Результаты моделирования для длины преамбулы 32 символа представлены в таблице 2.

Таблица 2

Алгоритм	Дисперсия нормированной частоты	E_b/N_0, дБ	Максимальный частотный сдвиг	Длина оценочной последовательности	Вычислительная сложность
Кэя	10^{-6}	4,7	0.15	32	31
Фитца	10^{-6}	3,8	0.025	32	496
Луиса и Реджиан-нини	10^{-6}	3,8	0.05	32	496

Выводы

Для спутникового радиоканала, работающего на частоте 8 ГГц, со скоростью передачи данных 50 Мсимв/сек, нормированный частотный сдвиг составляет порядка 0.004 (нормирован на символьную скорость). Для данной системы подходят все вышеперечисленные алгоритмы, но при этом алгоритм Кэя имеет меньшую вычислительную сложность. Алгоритмы можно применить для восстановления частотной синхронизации в системах со сверточным кодированием, и в некоторых случаях для систем с турбо-кодированием и LDPC-кодированием.

Литература

1. *Meyr H., Moeneclaey M., Fechtel S. A.,* Digital Communication Receivers, // New York: Wiley, 1998;

2. *Скляр Б.,* Теоретические основы и практическое применение. // М.: Издательский дом “Вильямс”, 2003. – 1104 с.

3. *Umberto Mengali and Aldo N. D’Andrea,* Synchronization techniques for digital receivers, // Plenum Press, NY, 1997.

Автоматизированная система обработки результатов исследований по влиянию предпосевного воздействия на жизнеспособность растений

М.В. Пантюхина, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Влияние научно-технического прогресса на все сферы деятельности человека очевидно, инновации появляются не только в технике, но и даже в такой устоявшейся и медленно меняющейся области как агрономия. Повышение урожайности зерновых культур - важная государственная задача. В связи с этим ученые-агрономы проводят разнообразные эксперименты и изучают влияние предпосевной обработки семенного материала на урожайность зерновых. Подготовка высококачественных семян невозможна без их обеззараживания и защиты проростков от внешних патогенных факторов, а также без стимулирования роста и развития растений. Обработка семян ведется электромагнитными волнами короткой частоты, импульсным магнитным полем, химическими препаратами в отдельности или в различных сочетаниях [1]. После обработки семена высеваются на поля, созревают, а после сбора урожая обследуются каждый стебель и каждое зернышко, на основании полученных сведений проводятся разнообразные расчеты. Для того, чтобы результаты обработки считались достоверными и по ним можно было сделать вывод о влиянии предпосевной обработки на качество семян, эксперимент необходимо проводить в течение нескольких лет.

Проведением таких экспериментов занимаются на кафедре защиты растений Самарской государственной сельскохозяйственной академии и на полях Поволжского НИИСС им. Константинова. На сегодняшний день ученые академии имеют хорошие результаты, накоплен большой объем данных о растениях, и эксперимент набирает обороты. В связи с этим появилась необходимость перевода всех экспериментальных данных с бумажных на электронные носители, а задача автоматизации обработки результатов исследований по влиянию предпосевного воздействия на жизнеспособность растений стала *актуальной*.

Информация об эксперименте

Развитие, или интенсивность развития болезней, отражающих среднюю интенсивность поражения, определяется по формуле, предложенной А.Е. Чумаковым, Т.И. Захаровой [2]:

$$U = \frac{\sum_{i=1}^{k}(n_i \times в_i) \times 100}{NK}, \quad (1)$$

где U – степень развития болезни, в %;

n_i – число растений, относящихся к i-ой группе пораженности;

b_i - балл поражения, соответствующий i-ой группе пораженности;

k – количество групп пораженности;

N – общее количество учтенных растений (здоровых и больных);

К – высший балл шкалы учета.

Пораженность растений задается баллами b_i по определенной шкале:

0 баллов – здоровое растение;

1 балл – слабое побурение основания стебля или подземного междоузлия;

2 балла – сильное побурение основного и подземного междоузлия;

3 балла – сильное побурение и белостебельность;

4 балла – погибшие или не выколосившиеся растения.

Распространенность или частота встречаемости болезней вычисляется по формуле:

$$P = \frac{n \cdot 100}{N}, \quad (2)$$

где Р – распространенность болезни, в %;

N – общее число обследованных растений в пробе;

n – количество больных растений в пробе.

Густота всходов подсчитывается в фазу полных всходов. По данным числа взошедших растений на 1 м^2 определяется полевая всхожесть, как процент взошедших растений к общему числу всех высеянных семян:

$$П = \frac{Д \times Б}{К \times 100}, \quad (3)$$

где Д – число взошедших растений на 1 м^2;

Б – лабораторная всхожесть в %;

К – норма высева лабораторно-всхожих семян;

П – полевая всхожесть в %.

Для определения биологического урожая в фазу восковой спелости отбираются снопы для анализа и установления структуры урожая, к основным элементам которой относят: число растений на единице площади, их общую и продуктивную кустистость, длину колоса, число и вес зерен в колосе, массу 1000 зерен урожая. Эти данные позволяют определить, какой из названных элементов структуры обеспечил полученный уровень урожая, сложился ли он в результате большого числа растений, или их хорошей продуктивной кустистости, или за счет элементов колоса (длина, число зерен), или большой массы 1000 зерен.

Из полученных данных структурного анализа выводятся средние показатели, и определяется биологический урожай при фактической влажности по формуле:

$$У = \frac{Р \times К \times З \times А}{10000}, \quad (4)$$

где У – урожай;

Р – число растений на 1 м^2 при уборке;

К – продуктивная кустистость;

З – количество зерен в колосе;

А – масса 1000 зерен, г.

Описание возможностей системы

Основным требованием ученых-аграриев, не имеющих большого опыта работы с компьютером, было о том, чтобы программа была проста в установке и легко переносима с компьютера на компьютер, а взаимодействие с системой было как можно более простым и понятным пользователям. Все эти требования были учтены при разработке системы.

В системе предусмотрено две роли пользователей: оператор и администратор. При запуске системы необходимо пройти процедуру аутентификации, после ввода личных данных программа настраивает интерфейс пользователя в соответствии с его правами. Разделение ролей обеспечивает защиту важной информации об уже проведенных экспериментах, к ней имеют доступ только руководители эксперимента: в *режиме администратора* пользователь имеет возможность редактировать справочники базы данных об эксперименте (см. рис. 1), проводить расчеты интенсивности развития болезней, распространенности болезней, полевой всхожести, биологического урожая. Кроме того, администратор имеет возможность редактировать

список пользователей, добавлять новых пользователей, изменять их пароли (см. рис. 2). В обязанности *оператора* входит лишь ввод новой информации в базу данных.

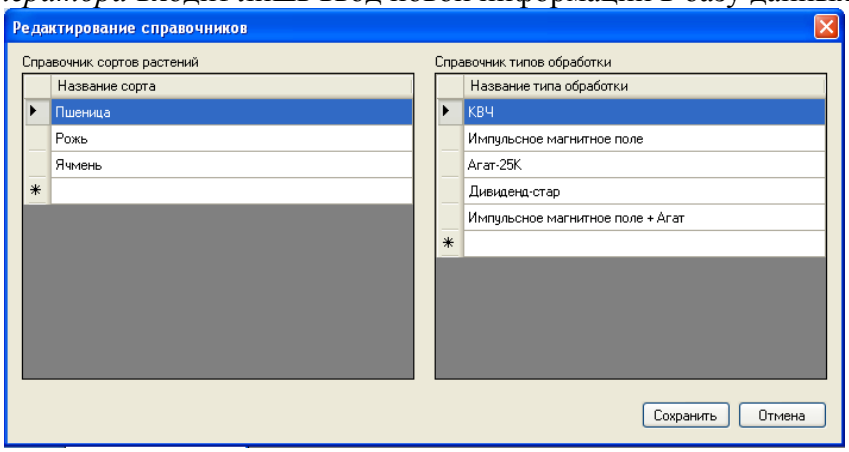


Рис. 1. Экранная форма редактирования справочников

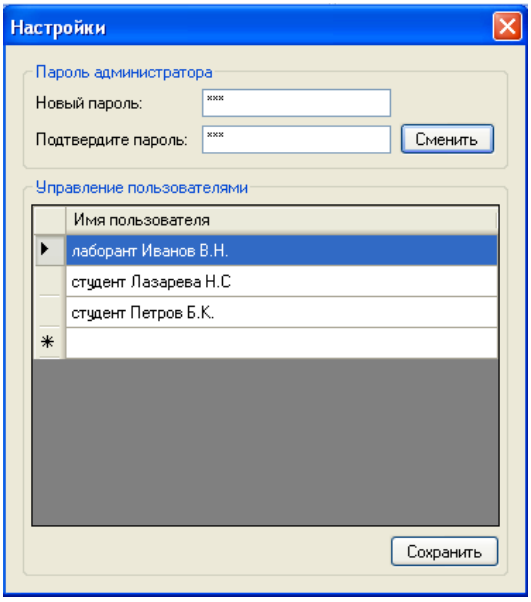


Рис. 2. Экранная форма изменения настроек системы

После ввода данных (см. рис. 3, 4) система производит расчеты по формулам (1) – (4). На рис. 5 изображена экранная форма, в которой отображаются результаты расчетов.

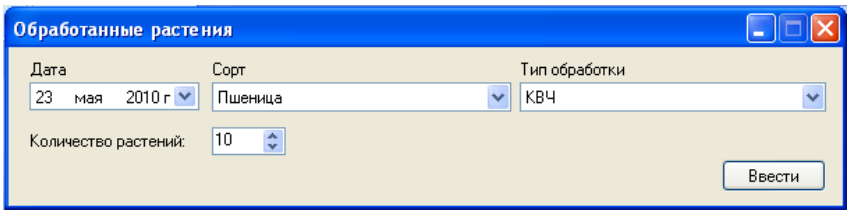


Рис. 3. Экранная форма ввода информации о выборке растений

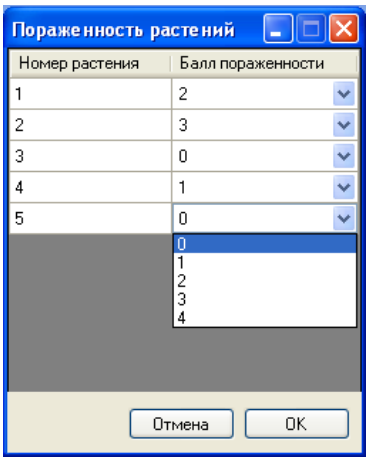


Рис. 4. Экранная форма ввода информации о пораженности растений

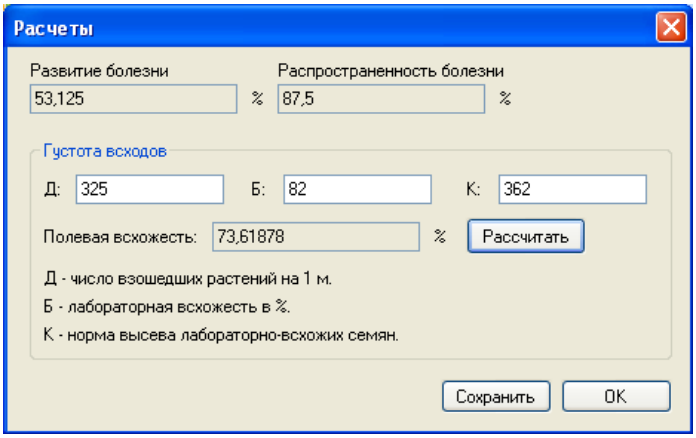


Рис. 5. Экранная форма отображения результатов расчетов

Для работы с базой данных используется библиотека SQLite, главным достоинством которой при решении поставленной задачи является хранение базы данных в одном файле на диске, без необходимости установки системы управления базами данных, что упрощает процесс развертывания программы и позволяет производить его пользователям, не имеющим прав администратора в системе.

Выводы

Разработанная система удовлетворяет всем поставленным требованиям, с ее помощью можно проводить мониторинг пораженности растений на разных фазах вегетации в различных районах земледелия и оценивать устойчивость растений к неблагоприятным условиям среды. Система проходит опытную эксплуатацию на кафедре защиты растений Самарской государственной сельскохозяйственной академии, предполагается дальнейшее развитие системы.

В качестве перспектив развития системы можно отметить следующее: предполагается автоматизировать расчет других показателей результативности эксперимента (водоудерживающая способность листьев, общая оводненность листьев, различные статические показатели) и проведение анализа результатов эксперимента, в систему будут включены различные отчеты о ходе проведения эксперимента.

Литература

1 Нижарадзе, Т.С. Сравнительная оценка влияния физических, химических и биологических методов предпосевной обработки семян на устойчивость к болезням, развитие и продуктивность зерновых культур в лесостепи Среднего Поволжья [Текст]: дисс. канд. биол. наук: 06.01.11: защищена 23.11.04: утв. 06.03.05/Нижарадзе Татьяна Сергеевна. – Кинель, 2004.- 241 с. – РГБ ОД, 61:05-3/371.

2 Чумаков, А.Е. Вредоносность болезней сельскохозяйственных культур [Текст]/А.Е. Чумаков, Т.И.Захарова. – М.: Агропромиздат, 1990. – 127с.

Разработка систем измерений и анализа сигналов на базе свободного программного обеспечения

А.И. Попов, А.В. Родионов
Поморский государственный университет имени М.В. Ломоносова, Архангельск
ЗАО «Инженерный центр «Энергосервис», Архангельск

Введение

Работа посвящена вопросам типового проектирования информационных систем (ИС), автоматизирующих процессы регистрации сигналов, их хранения, поиска, выполнения некоторых процедур цифровой обработки (ЦОС), отображения результатов анализа сигналов в виде чисел и графиков, экспорта данных. Такие ИС востребованы в самых разных областях науки и техники, в частности, в медицинской диагностике [4, 11]. Зная о споре метрологов и разработчиков ИС [1, 3] о том, следует или нет рассматривать эти ИС как средства измерений (СИ), введем понятие системы измерений и анализа сигналов (СИАС). СИАС определим как ИС, содержащую среди прочих компонентов измерительную часть (подсистему измерений) и способную применять к сохраняемым цифровым сигналам вычислительные процедуры анализа. Последнее с точки зрения метрологии относится к косвенным измерениям [5].

Типовое проектирование на основе свободного программного обеспечения

Распространен подход, при котором СИАС создаются «с нуля». Небольшую СИАС вполне способен разработать один человек. Однако при этом усилия разработчика направлены на реализацию требуемой функциональности, а не на удобство для конечного пользователя или совместимость с другими системами.

При другом широко распространенном подходе система строится из отдельных программ; некоторые из них пишутся специально для интеграции остальных, часто зарытых и весьма дорогостоящих, таких как MATLAB. В этом случае теряется ощущение целостности продукта, возникают некоторые трудности, например, при обработке сигналов в режиме реального времени с графическим отображением результатов обработки. Такой продукт является дорогим, что часто не оправдано.

Со стороны специалистов в области проектирования ИС возрастает интерес к возможности построения ИС из готовых типовых блоков (типовых проектных решений), имеющихся на рынке. Соответствующий подход имеет название типового проектирования ИС [2] и позволяет соблюдать принципы многократного использования компонентов ИС, унификации типовых операционных процессов, ликвидации дублирования функциональности в программном обеспечении (ПО).

Цель, преследуемая авторами, заключается в объединении мировых открытых разработок в сфере ЦОС и создании инструмента, позволяющего в короткие сроки конструировать СИАС из типовых блоков.

Коллекция строительных блоков для СИАС может быть получена на основе свободного ПО. В мире активно разрабатываются открытые инструменты ЦОС, среди которых можно выделить: функциональный язык программирования процедур ЦОС Faust [7]; инструмент для редактирования блочных диаграмм и симуляции Scicos, входящий в состав SciLab [8]; редактор звуковых файлов Audacity [6]; многочисленные библиотеки процедур ЦОС [10, 14] и т.д.

Речь идет не просто о замене проприетарных компонентов системы на свободные аналоги (например MATLAB на SciLab), а о разработке технологии построения ИС определенного класса, подразумевающей, как минимум, построение модели системы, разработку программных и аппаратно-программных интерфейсов, разработку формата хранения данных. Наличие стандартной модели системы упрощает

формирование требований и верификацию, наличие стандартных интерфейсов и форматов - интеграцию, сопровождение. Инструменты, поставленные в соответствие фрагментам модели, упрощают и ускоряют реализацию ИС. Другими словами, обеспечиваются улучшения на всех этапах жизненного цикла ИС.

Основные задачи

Состав и взаимосвязи типовых блоков СИАС позволяет определить технология структурного анализа и проектирования (SADT) [2]. Авторами разработана SADT-модель типовой СИАС, выделены подсистемы, информационные и управляющие связи между ними.

Важной задачей является унификация аппаратно-программного интерфейса для взаимодействия с измерительной аппаратурой, к чему стремятся и разработчики ПО, и производители аппаратуры. Так, российское предприятие L-Card [9] предлагает единую библиотеку подпрограмм для организации взаимодействия со всеми выпускаемыми им модулями АЦП. На базе открытой библиотеки Qt [12] и предоставляемого ею механизма создания подключаемых модулей авторами разработан интерфейсный класс, позволяющий производить замену модулей аналого-цифрового преобразования без перекомпиляции программ ИС.

Не менее важная задача - организация хранения данных динамических измерений. Требуется обеспечить хранение, как минимум, следующих данных: собственно цифровой сигнал, метаданные (данные о масштабах и конфигурации формата), прочие сопутствующие данные. Важно учитывать возможность изменения параметров оцифровки в рамках одного сеанса измерений. Обязательной является совместимость формата хранения со стандартными форматами, применяемыми в конкретных предметных областях, например, форматами описания геофизических данных SEG-Y [15] и PSN [13]. Основной объем в СИАС составляют бинарные данные, поэтому построение классической базы данных кажется неоправданным. Мета- и сопутствующие данные предлагается хранить в XML-файлах, содержащих ссылки на бинарные файлы. Такая схема применяется в открытом звуковом редакторе Audacity [6].

Выводы

Работа в рамках проекта по созданию открытой технологии типового проектирования СИАС на базе свободного ПО началась в 2009 году на кафедре информационных технологий Поморского государственного университета им. М.В. Ломоносова (ПГУ). Проведен анализ стандартов в области измерений и свободного ПО ЦОС; разработана модель типовой СИАС; выработаны принципы организации взаимодействия ПО с измерительной аппаратурой, а также хранения сигналов и результатов их обработки. Проект находится в начальной стадии, многие положения носят рекомендательный характер. Основные результаты обсуждаются на занятиях научного семинара математического факультета ПГУ «Прикладные информационные технологии».

Литература

1. ГОСТ Р 8.596-2002 Государственная система обеспечения единства измерений. Метрологическое обеспечение измерительных систем. Основные положения.

2. Грекул В.И., Денищенко Г.Н., Коровкина Н.Л. Проектирование информационных систем. М.: ИТУИТ.РУ, 2008. 304 с.

3. Гуртовцев А. Территория конфликта: метрологические требования к АСКУЭЭ. "Информационные технологиии и измерение. Приложение к журналу". № 68. 2009. С.94 — 96.

4. Свиньин С.Ф., Хаймина Л.Э., Латухина Е.А., Попов А.И. Медицинская информационно-измерительная система функциональной диагностики в гастрологии. "Известия вузов - Приборостроение". 2009. №12. С.11 — 16.

5. Эрастов В.Е. Метрология, стандартизация и сертификация. М.: ФОРУМ. 2008. 208 с.

6. Audacity: свободный звуковой редактор. http://audacity.sourceforge.net/

7. Faust: signal processing language. http://sourceforge.net/projects/faudiostream/
8. Home - Scilab WebSite. http://www.scilab.org/
9. L-Card - оборудование для автоматизации измерений и сбора данных, АСУТП, АЦП, ЦАП. http://www.lcard.ru/.
10. LOGIS - Custom software development and outsourcing. http://www.logis-pro.com/index.php?lang=ru&cat_id=3&page_id=24.
11. Neurobotics.ru - Главная. http://neurobotics.ru/.
12. Qt - A cross-platform application and UI framework. http://qt.nokia.com.
13. Redwood city public seismic network. http://psn.quake.net/.
14. Signal processing using C++. Project itpp. http://sourceforge.net/apps/wordpress/itpp/.
15. The society of exploration geophysics. http://www.seg.org/.

Адаптивный квадратно-корневой ковариационный алгоритм фильтрации для навигационных комплексов

И.В. Семушин, Ю.В. Цыганова
Ульяновский государственный университет, Ульяновск

Введение

Обработка измерительной информации в навигационных комплексах опирается на развитый аппарат теории оптимальной фильтрации и идентификации моделей систем. Зарубежные материалы и отечественная литература дают широкое представление об этих методах. За последние десятилетия созданы численно устойчивые реализации фильтра Калмана, включающие класс квадратно-корневых алгоритмов, которые позволяют обрабатывать данные с двойной точностью. Это свойство позволяет считать данный класс алгоритмов предпочтительным для практики [1]. Вместе с тем в отечественной литературе эффективные квадратно-корневые алгоритмы фильтрации освещены слабо. Еще менее подробно представлены так называемые совмещенные алгоритмы (Kailath [1] и др.). Важное место занимают адаптивные методы; их применение — вынужденное из-за априорной неопределенности, и им посвящено большое число работ.

Методы адаптации систем могут быть отнесены к двум крупным категориям [2]: пассивные или активные. Согласно пассивному принципу, для адаптации фильтра сначала оценивают неизвестные характеристики внешней (сигнально-помеховой) обстановки и затем используют эти оценки в алгоритмах как «истинные». Принцип пассивен в том смысле, что оценивание происходит в разомкнутой цепи — без слежения за фактическим уровнем качества системы. Второй принцип активен в том смысле, что адаптивный фильтр контролирует степень удаления своего критерия качества от оптимума. Эта концепция либо пассивного, либо активного принципа адаптации отличается от терминов «пассивный» / «активный» в других работах, например, в известной монографии Дж. Саридиса и недавней книге В.И. Денисова и др. (серия «Монографии НГТУ», 2009).

Активный принцип адаптации, как он понимается здесь, предложен и описан в ряде работ (например, [3], [4] или [5]). Он детально характеризуется как метод вспомогательного функционала качества и в этой работе, но с новым содержанием.

Традиционно, для построения вспомогательного функционала качества применяли адаптивный фильтр в форме стандартного алгоритма Калмана и градиент наблюдаемого функционала вычисляли по отношению к параметрам этого фильтра. Есть два обстоятельства, побуждающие искать лучшее решение. Первое – стандартный алгоритм может быть численно неустойчив; выход – применять упомянутые выше устойчивые реализации фильтра. Второе – градиент по настраиваемым параметрам адаптивного фильтра приводит к моделям чувствительности с сильно разреженными матрицами, т.е. дифференцированные по параметрам фильтра алгоритмы адаптации получаются неэффективными и неустойчивыми для вычислений; выход – применять дифференцирование не по параметрам фильтра, а по параметрам неопределенности в модели поступающей информации. Целью настоящей работы является улучшение качества динамического процесса активной адаптации фильтра на основе отмеченных двух подходов.

1. Постановка задачи

Рассмотрим данные в виде стохастической линейной дискретной системы

$$x_t = \Phi(\Theta)x_{t-1} + \Gamma(\Theta)w_{t-1}, \qquad (1)$$

$$z_t = H(\Theta)x_t + v_t \qquad (2)$$

с вектором состояния $x_t \in \Re^n$ и вектором измерения $z_t \in \Re^m$, где $\{w_0,w_1,\ldots\}$ и $\{v_1,v_2,\ldots\}$ суть q-мерная и m-мерная независимые последовательности независимых и одинаково распределенных случайных векторов с нулевыми математическими ожиданиями и ковариациями $Q(\Theta)$ и $R(\Theta)$ соответственно. Эти последовательности не зависят от случайного начального состояния системы $x_0 \in \mathrm{N}(\bar{x}_0(\Theta), P_0(\Theta))$. Предположим, что описывающие систему (1), (2) матрицы зависят от неизвестного параметра $\Theta \in R^p$, причем элементы указанных матриц являются дифференцируемыми по θ_i функциями, где θ_i – i-й элемент вектора Θ, i=1,…,p. Таким образом, параметрами адаптивного фильтра можно считать оценки θ_i. Тогда возникает задача их отыскания, но в активном принципе это должно происходить как процесс отыскания оптимума вспомогательного (одновременно, и исходного) критерия качества функционирования фильтра.

Для решения задачи по методу вспомогательного функционала качества введем ряд предположений: пусть матрица $R>0$, матрица $P_0 \geq 0$ и $\Gamma Q \Gamma^T \geq 0$, система (1), (2) стабилизируема и полностью наблюдаема, причем все частные индексы наблюдаемости [4] p_k (k=1,…,m) известны и не зависят от параметра Θ.

Так как состояние недоступно наблюдению, адаптивный фильтр строим в форме фильтра Калмана. Для повышения точности алгоритма вместо стандартного фильтра Калмана будем рассматривать квадратно-корневой ковариационный фильтр [8], который зависит от неизвестного параметра Θ. В случае его совпадения с оптимальным («истинным») значением Θ_*, полученная оценка вектора состояния «автоматически» будет оптимальной в среднеквадратическом смысле, т. е. будет доставлять минимум также и исходному функционалу качества (ИФК)

$$J_{ИФК} = E\{\|e_t\|^2\}. \qquad (3)$$

Область определения ИФК включает в себя все одношаговые оценки предсказания вектора состояния x_t, полученные по доступным измерениям $Z_1^t = [z_1 \mid z_2 \mid \ldots \mid z_t]^T$, а e_t – ошибка предсказания. Препятствием для активной адаптации фильтра является то, что функционал (3) – теоретический, но не практический: ошибка e_t принципиально не известна, т.е. этот ИФК для численных методов оптимизации не доступен. Препятствие преодолимо, если удастся сформировать некоторый вспомогательный функционал качества (ВФК)

$$J_{ВФК} = E\{\|\varepsilon_t\|^2\}, \qquad (4)$$

который зависит только от доступных величин ε_t и достигает минимума одновременно с ИФК на множестве значений параметра Θ.

Ниже будем использовать следующие обозначения: квадратный корень матрицы A, полученный в разложении Холесского [10], обозначим как S_A ($A=S_AS_A^T$), при этом S_A – нижняя треугольная матрица; квадратные корни ковариационных матриц $\widetilde{P}$ и $\hat{P}$, участвующие в уравнениях фильтра Калмана, обозначим $\widetilde{S}$ и $\hat{S}$; и пусть $\overline{L}$ и $\overline{U}$ – строго нижняя и строго верхняя треугольные матрицы, D — диагональная матрица.

2. Построение адаптивного квадратно-корневого ковариационного алгоритма

Для построения ВФК перейдем от представления системы в виде (1), (2) к стандартной наблюдаемой модели, т. е. выполним соответствующее преобразование базиса в пространстве состояний. Далее, используя результаты работ [4, 5], построим вспомогательный функционал (4), а вспомогательный процесс определим следующим образом:

$$\varepsilon_t = \varphi(Z_{t+1-s}^t) - \tilde{x}_{t+1-s}, \qquad (5)$$

где преобразование φ определено в [4], $Z_{t+1-s}^t = [z_{t+1-s} \mid z_{t+2-s} \mid ... \mid z_t]^T$ – составной вектор измерений (s=max$\{p_k \mid k$=1,…,$m\}$), $\tilde{x}_{t+1-s}$ – оценка предсказания, полученная от адаптивного фильтра.

Теорема 1. [5] Пусть задан исходный функционал качества в виде (3). Тогда существует вспомогательный функционал качества (4), который зависит только от наблюдаемого процесса (5), и при этом выполняется условие $J_{ВФК} = J_{ИФК} + const$, в котором величина *const* не зависит от параметров фильтра и, следовательно, ВФК достигает минимума одновременно с ИФК на множестве значений параметра Θ.

Таким образом, теорема 1 дает возможность применять вместо недоступного ИФК реализуемый ВФК. Для его минимизации можно воспользоваться известными численными методами [7], которые требуют вычисления градиента ВФК или градиента его оценки. В качестве оценки ВФК рассмотрим величину $\hat{J}_{ВФК} = \frac{1}{2N} \sum_{k=t-N+1}^{t} \varepsilon_k^T \varepsilon_k$. Тогда

$$\frac{\partial \hat{J}_{ВФК}}{\partial \theta_i} = \frac{1}{N} \sum_{k=t-N+1}^{t} \varepsilon_k^T \frac{\partial \varepsilon_k}{\partial \theta_i}, \; i = 1,\ldots, p.$$

Вычисление частных производных вспомогательного процесса требует записи уравнений дифференцированного фильтра, или так называемых уравнений чувствительности. Различные методы вычисления уравнений чувствительности на основе стандартного фильтра Калмана подробно обсуждаются в работе [9]. Однако поскольку хорошо известно, что стандартная форма фильтра Калмана является неустойчивой по отношению к ошибкам машинного округления, то следует предположить, что тем же недостатком будет обладать и дифференцированный фильтр.

Поэтому применим другой подход и разработаем алгоритм для вычисления величин $(\varepsilon_k)'_{\theta_i}$ непосредственно в терминах квадратно-корневого ковариационного фильтра (ККФ). В работе [8] проведен численный анализ нескольких модификаций фильтра Калмана, среди которых в качестве наиболее устойчивого к ошибкам машинного округления рекомендован ККФ. Таким образом, получим следующий результат:

Теорема 2. Пусть элементы матриц Ф, Г, Q, H, R и P_0 в системе (1), (2) являются дифференцируемыми функциями по неизвестному параметру $\Theta \in R^p$. Тогда идентификация неизвестного параметра Θ и оценивание вектора состояния системы (1), (2) по критерию (4) могут быть выполнены с помощью адаптивного квадратно-корневого ковариационного фильтра (АККФ):

0. *Инициализация*. Пусть $\hat{x}_0 = \bar{x}_0(\Theta)$, $\Phi = \Phi(\Theta)$, $\Gamma = \Gamma(\Theta)$, R=$R(\Theta)$>0, $R = S_R S_R^T$, Q=$Q(\Theta)$≥0, $Q = S_Q S_Q^T$, P_0=$P_0(\Theta)$>0, $P_0 = \hat{S}_0 \hat{S}_0^T$. Вычислить $\partial \hat{S}_0 / \partial \theta_i$, $\partial \hat{x}_0 / \partial \theta_i$ и положить $\hat{J}_{ВФК} = 0$, $\partial \hat{J}_{ВФК} / \partial \theta_i$ (i=1,…,p).

При заданных начальных условиях рекуррентно вычислять (t=1,…,N):

1. *Экстраполяция*.

1.1. Вычислить оценку вектора состояния: $\tilde{x}_t = \Phi \hat{x}_{t-1}$.

1.2. Для каждого θ_i, i=1,…,p вычислить: $\frac{\partial \tilde{x}_t}{\partial \theta_i} = \frac{\partial \Phi}{\partial \theta_i} \hat{x}_{t-1} + \Phi \frac{\partial \hat{x}_{t-1}}{\partial \theta_i}$.

1.3. Вычислить квадратный корень ковариационной матрицы $\widetilde{S}_t$:

$$\widetilde{T}_t \begin{bmatrix} \hat{S}_{t-1}^T \Phi^T \\ S_Q^T \Gamma^T \end{bmatrix} = \begin{bmatrix} \widetilde{S}_t^T \\ 0 \end{bmatrix}, \tag{6}$$

где $\widetilde{T}_t$ – матрица ортогонального преобразования к верхнему треугольному виду правой части выражения (6).

1.4. Для каждого θ_i, i=1,…,p применить ортогональное преобразование $\widetilde{T}_t$ из (6):

$$\widetilde{T}_t \begin{bmatrix} \dfrac{\partial\left(\Phi \hat{S}_{t-1}\right)^T}{\partial\theta_i} \\ \dfrac{\partial\left(\Gamma S_Q\right)^T}{\partial\theta_i} \end{bmatrix} = \begin{bmatrix} A_i \\ C_i \end{bmatrix}, \tag{7}$$

где через A_i и C_i обозначены соответствующие матричные блоки размеров $n\times n$ и $s\times n$ в правой части выражения (7).

1.5. Для каждого θ_i, i=1,…,p вычислить:

$$\frac{\partial \widetilde{S}_t^T}{\partial\theta_i} = (\overline{L}_i^T + D_i + \overline{U}_i)\widetilde{S}_t^T, \qquad \text{где } \overline{L}_i + D_i + \overline{U}_i = A_i \widetilde{S}_t^{-T}.$$

2. *Вычисление ВФК и его градиента*.

2.1. Вычислить вспомогательный процесс: $\varepsilon_t = \varphi(Z_{t+1-s}^t) - \widetilde{x}_{t+1-s}$.

2.2. Вычислить ВФК: $\hat{J}_{ВФК} := \hat{J}_{ВФК} + \dfrac{1}{2t}\varepsilon_t^T \varepsilon_t$.

2.2. Для каждого θ_i, i=1,…,p вычислить:

$$\frac{\partial \varepsilon_t}{\partial\theta_i} = -\frac{\partial \widetilde{x}_{t+1-s}}{\partial\theta_i}, \qquad \frac{\partial \hat{J}_{ВФК}}{\partial\theta_i} = \frac{\partial \hat{J}_{ВФК}}{\partial\theta_i} + \frac{1}{t}\left(\varepsilon_t^T \frac{\partial\varepsilon_t}{\partial\theta_i}\right).$$

3. *Фильтрация*.

3.1. Вычислить квадратный корень ковариационной матрицы $\hat{S}_t$:

$$\hat{T}_t \begin{bmatrix} S_R^T & 0 \\ \widetilde{S}_t^T H^T & \widetilde{S}_t^T \end{bmatrix} = \begin{bmatrix} S_E^T & \Psi_t^T \\ 0 & \hat{S}_t^T \end{bmatrix}, \tag{8}$$

где $\hat{T}_t$ – матрица ортогонального преобразования к верхнему треугольному виду правой части выражения (8).

3.2. Вычислить оценку вектора состояния: $\hat{x}_t = \widetilde{x}_t + \Psi_t \overline{e}_t$, где $\overline{e}_t = S_E^{-1}(z_t - H\widetilde{x}_t)$.

3.3. Для каждого θ_i, i=1,…,p применить ортогональное преобразование $\hat{T}_t$ из (8):

$$\hat{T}_t \begin{bmatrix} \dfrac{\partial S_R^T}{\partial\theta_i} & 0 \\ \dfrac{\partial(H\widetilde{S}_t)^T}{\partial\theta_i} & \dfrac{\partial\widetilde{S}_t^T}{\partial\theta_i} \end{bmatrix} = \begin{bmatrix} X_i & Y_i \\ V_i & W_i \end{bmatrix}, \tag{9}$$

где через X_i, Y_i и V_i, W_i обозначены соответствующие матричные блоки размеров $m\times m$, $m\times n$ и $n\times m$, $n\times n$ в правой части выражения (9).

3.4. Для каждого θ_i, i=1,…,p вычислить матрицу G:

$$G = \begin{bmatrix} X_i & Y_i \\ V_i & W_i \end{bmatrix} \begin{bmatrix} S_E^T & \Psi_t^T \\ 0 & \hat{S}_t^T \end{bmatrix}^{-1} = \bar{L}_i + D_i + \bar{U}_i . \qquad (10)$$

3.5. Далее для i=1,…,p вычислить:

$$\frac{\partial \hat{S}_t^T}{\partial \theta_i} = \left(\bar{L}_i^{*T} + D_i^* + \bar{U}_i^* \right) \hat{S}_t^T , \qquad \frac{\partial S_E^T}{\partial \theta_i} = \left(\bar{L}_i^{**T} + D_i^{**} + \bar{U}_i^{**} \right) S_E^T ,$$

$$\frac{\partial \Psi_t^T}{\partial \theta_i} = Y_i^T + (\bar{L}_i^{**T} - \bar{L}_i^{**}) \Psi_t^T + (V_i S_E^{-T})^T \hat{S}_i^T ,$$

где $\bar{L}_i^*$, D_i^* и $\bar{U}_i^*$ – составляющие части подматрицы $[G]_{col(m+1):(m+n)}^{row(m+1):(m+n)}$ (верхний и нижний индексы указывают область расположения матричного блока), $\bar{L}_i^{**}$, D_i^{**} и $\bar{U}_i^{**}$ – составляющие части подматрицы $[G]_{col\,1:m}^{row\,1:m}$.

3.6. Для каждого θ_i, i=1,…,p вычислить:

$$\frac{\partial \hat{x}_t}{\partial \theta_i} = \frac{\partial \tilde{x}_t}{\partial \theta_i} + \frac{\partial \Psi_t}{\partial \theta_i} \bar{e}_t + \Psi_t \frac{\partial \bar{e}_t}{\partial \theta_i} , \qquad \text{где } \frac{\partial \bar{e}_t}{\partial \theta_i} = -S_E^{-1} \left(\frac{\partial S_E}{\partial \theta_i} \bar{e}_t + H \frac{\partial \tilde{x}_t}{\partial \theta_i} \right) .$$

4. *Оценка неизвестного параметра* $\Theta \in R^p$: $\hat{\Theta}_{min} = \underset{\Theta \in D(\Theta)}{arg\,min}\, \hat{J}_{ВФК}(\Theta)$.

Доказательство основано на свойствах ортогональных преобразований и ортогональных матриц, на правилах перемножения блочных матриц, на правилах дифференцирования матриц, а также на методах работы [6].

Выводы

Разработан численно устойчивый адаптивный алгоритм, который позволяет одновременно решать задачу идентификации параметров линейной системы и задачу адаптивного оценивания ее состояния по измерительной информации. Вычислительные эксперименты подтвердили работоспособность алгоритма и повышенное качество процесса активной адаптации.

Литература

1. Kailath T. Array algorithms for structured matrices. // Presented at the conference of the International Linear Algebra Society. Winnipeg, USA, 1998.
2. Семушин И. В. Активные методы адаптации и контроля нарушений в дискретных системах управления. – Докторская диссертация, ЛИАП, 1987.
3. Понырко С. А., Семушин И. В. Построение обучающихся винеровских фильтров при ограниченном объеме априорной информации. // Техническая кибернетика, Изв. АН СССР, N 5, 1971. – с. 215-220.
4. Семушин И. В. Адаптивное управление стохастическим линейным объектом в условиях неопределенности. // Нелинейные динамические системы: качественный анализ и управление: Сборник трудов. М.: ИСА РАН, 1994. Вып. 2. – с. 104-110.
5. Semoushin I. V., Tsyganova J. V. Auxiliary Performance Functional Approach to Adaptive and Learning Filtering and Control. // Conference Proceedings of European Control Conference ECC'99. Karlsruhe. Germany. 31 August-3 September 1999.
6. Bierman G. J., Belzer M. R., Vandercraft J. S., Porter D. W. Maximum Likelihood Estimation Using Square Root Information Filters // IEEE Trans. On Automat. Contr. 1990. V. 35. N 12. P. 1293-1299.
7. Mosca E. Optimal, Predictive and Adaptive Control. – Prentice Hall, Inc., 1995.
8. Verhaegen M., Van Dooren P. Numerical aspects of different Kalman filter implementations. // IEEE Trans. On Automat. Contr. 1986. V. AC-31. No. 10. P. 907-917.
9. Gupta N. K., Mehra R. K. Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations. // IEEE Trans. On Automat. Contr. 1974. V. AC-19. No. 6. P. 774-783.
10. Bierman G. J. Factorization Methods For Discrete Sequential Estimation. New York: Academic Press, 1977.

Исследование и разработка алгоритмов распознавания номерных знаков для системы контроля доступом

А.В. Симкин
МГТУ им. Баумана, Москва

Введение

Идея создания комплексной системы автоматизации (далее Система) контрольно-пропускного пункта (далее КПП) основана на выводах, полученных из оценки современного состояния и анализа ситуации сложившейся на российском рынке автоматизации КПП.

Существующие на рынке решения обладают хорошей технологической базой, однако им не хватает универсальности в функциональных возможностях для «кастомизации» своих решений под нужды Заказчика. В свою очередь это не дает полной возможности (без доработок) построить автоматизированную систему учета и контроля доступа на основании системы распознавания номерных знаков для автостоянки или КПП.

Проведение научно-исследовательской работы (далее НИР) в области разработки универсальной автоматизированной системы, которая будет обладать всеми преимуществами различных видов решений, является на сегодняшний день актуальной задачей.

В современных условиях функционирование любых участников экономической деятельности, в качестве которых выступают предприятия, организации и прочие учреждения, невозможно без наличия эффективной автоматизированной системы управления. Разработка подобной системы повышающей эффективность учета автотранспортных средств на КПП и является основной целью ее создания. Эффективность, в данном случае, выражается в упрощении регистрации и контроля автотранспортных средств, пересекающих контрольно-пропускной пункт за счет автоматизации процесса учета. Данный процесс обеспечивает своевременность и актуальность получения информации о текущем состоянии объекта автоматизации не только владельцам, но и всем участникам процесса.

Алгоритмы и методы решения задачи

Основная задача проведенных исследований заключалась в изучении, применении, а также разработке новых алгоритмов обработки информации, удовлетворяющие требованиям Системы.

Процесс распознавания изображений является сложной многоэтапной задачей. Иерархичность и последовательность решения задачи обусловлена тем, что различные этапы обработки очень тесно связаны и качество решения одной из них влияет на выбор метода решения остальных. Например, выбор метода распознавания зависит от свойств изображения (характера фона, помех) и связан с выбором методов предобработки, сегментации, фильтрации [1].

С учетом проведенных исследований была предложена наиболее оптимальная иерархическая структура распознавания для данной задачи (см. рис. 1). Которая заключается в том, что после предварительных этапов обработки изображений, таких как получение изображения, сегментация и фильтрация, рекомендуется применять несколько методов распознавания.

На первом этапе распознавания рекомендуется применять наименее трудоемкие признаковые алгоритмы для решения задачи о неэквивалентности входных изображений и эталонов. Входное множество изображений или эталонов при этом существенно сокращается (в зависимости от типа задачи). Если входные изображения

содержат геометрические и аморфные искажения, на следующем этапе необходимо применять нормализацию. После классифицировать входные изображения одним из конструктивных способов, например корреляционным. При такой структуре распознавания время решения задач сокращается в сотни раз.

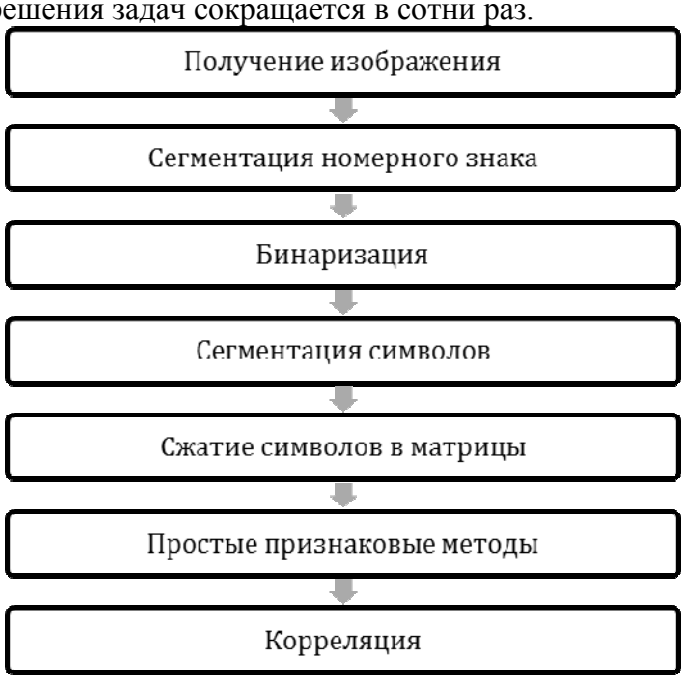


Рис. 1. Структура обработки изображения в поставленной Системе

Более детально в данной работе исследовались и разрабатывались алгоритмы распознавания автомобильных номерных знаков. Как показывает практика, сегментация символов является одной из сложнейших задач в алгоритмах распознавания.

Исследование исходных данных показало, что в данной задаче наиболее оптимальны два подхода к сегментации отдельных символов на номерном знаке. Первый заключатся в выделении определенных областей исходного изображения за счет определения координат предполагаемого положения символов, основываясь на геометрических признаках. Второй, более универсальный метод, относящийся к методам наращивания областей, а именно центроидное связывание [2].

В результате проведенных экспериментальных исследований появились модифицированные версия алгоритма центроидного связывания, и признаковый алгоритм сегментации.

Основная идея метода центроидного связывания состоит в том, что информация о связываемых областях учитывается на этапе выбора начальных точек. На каждом последующем шаге рассматриваются точки соседние с размеченными пикселями, и если разность их яркостей меньше порога, то они присоединяются к области. В силу того, что сегментация изображения происходит уже после бинаризации, сравнивать соседние области по яркости не имеет смысла, достаточно ограничиться логической операцией сравнения, что существенно ускорит процесс. Вторая необходимая модификация, состоит в том, что мы не можем заранее выбрать стартовые (начальные) точки. В данной задаче необходимо просматривать всю область и инициализировать каждый новый контур, когда в текущем пикселе находится черный цвет.

Признаковый метод сегментации основан на геометрических характеристиках исходного изображения. Согласно ГОСТ Р 50577-93 можно выделить геометрические признаки сегментации. После проведения анализа различных входных данных, были замечены определенные признаки, по которым возможно сегментировать символы

на изображении. Основная идея алгоритма основывается на последовательном анализе распределения средних значений яркости вертикальных групп пикселей и выделении по полученным признакам областей с символами.

Последним этапом в поставленной задаче является распознавание [1]. Распознавание лежит в основе процессов интерпретации и понимания. Существует множество методов распознавания, однако с учетом проведенных исследований было рекомендовано пользоваться корреляционными методами [3].

Результаты исследований и экспериментов

Результатом проведенной НИР в области обработки изображений с последующей их интерпретацией стала разработанная подсистема контроля доступа на основании системы распознавания номерных знаков.

Для поставленной задачи была разработана структура системы, модулей и схемы их взаимодействия. Предложена среда разработки и реализации автоматизированной системы. Сделан выбор технических и трудовых средств и ресурсов, обслуживающих автоматизированную систему.

Были разработаны две процедуры распознавания номерных знаков и проведены их экспериментальные исследования зависимости качества распознавания от шума, изменения яркости и геометрических искажений. Разработанный корреляционный алгоритм сегментации показал лучшие результаты в ходе проведенных экспериментов, чем корреляционно-признаковый алгоритм. Особенностью алгоритма основанного на сегментации является лежащей в основе модифицированный алгоритм центроидного связывания пикселей. Данный алгоритм обладает хорошей универсальностью и может обработать «номерными метками» изображение любого размера. Таким образом, были сделаны выводы по полученным результатам и предложена наиболее оптимальная процедура распознавания (см. рис. 2).

В проведенных экспериментах была исследована зависимость качества распознавания от влияния различных параметров и настроек. Следовательно, основываясь на результатах экспериментов можно предложить наиболее эффективные настройки и методы распознавания.

Результатом проведенных исследований стал вывод о том, что система обладает хорошими качеством распознавания (до 10% ошибок) и отличной скоростью работы (до 100 мс), что позволяет системе работать в режиме реального времени.

Выводы

Таким образом, в результате проведенных исследований были детально исследованы алгоритмы корреляционного распознавания, изменения размера изображения, методы бинаризации и сегментации. Была предложена структура решения задачи и в последующем модифицирована. Проведены экспериментальные исследования и тестирование части реализованной задачи.

На основании экспериментов были сделаны выводы о том, что система обладает хорошими качеством распознавания и отличной скоростью работы, что позволяет использовать эту систему в режиме реального времени.

Одним из важнейших плюсов предложенной архитектуры системы является то, что в отличие от нее подобные автоматизированные учетные системы редко интегрируются с существующими бизнес-процессами организации и часто не затрагивают их деятельность, т.е. являются охранными. Поэтому предложенная идея создания тиражного решения является приоритетной задачей направленной на дальнейшее развитие и разработку.

На сегодняшний день в России подобные системы редко используются в реальном времени на КПП и паркингах. Поэтому в дальнейшем будет сделана

попытка углубиться в исследования, провести НИОКР и продолжить разработку Системы с целью внедрения новых разработанных технологий.



Рис. 2. Структуры алгоритмов обработки изображений

Литература

1. Путятин Е.П., Аверин С.И. Обработка изображений в робототехнике. – М: Машиностроение, 1990. – 320 с.

2. И.С. Грузман, В.С. Киричук и др. Цифровая обработка изображений в информационных системах: Учеб. пособие. – Новосибирск: Изд-во НГТУ, 2002. – 352 с.

3. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – Техносфера, 2005. – 1072 с.

Подсистема расчета молекулярной площадки адсорбата

М.С. Стребков, Л.С. Зеленко, В.В. Варфоломеева, А.В. Терентьев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Исследования процесса физической адсорбции на поверхности раздела сред газ – твердое тело является *актуальной задачей* и представляет интерес с практической точки зрения. Адсорбция занимает ведущее место среди способов защиты биосферы от вредных промышленных выбросов, ее широко используют для глубокой очистки и осушки технологических потоков, улучшения качества сырья, воздуха, получения кислорода и т.д. В науке и технике приобрёл большое значение хроматографический метод анализа, основанный на различной способности компонентов анализируемой смеси к адсорбции. Одним из способов изучения процесса адсорбции является моделирование, в связи с этим ведется разработка эффективных инструментов и методов моделирования процесса адсорбции, способных учитывать влияние других процессов, происходящих на поверхности раздела сред, например, поверхностную диффузию.

Широко применяемые в настоящее время методы, основанные на решении систем дифференциальных уравнений в частных производных, оказываются недостаточно эффективным при моделировании процессов адсорбции. Указанные трудности стимулируют поиск принципиально новых методов анализа подобных процессов и систем, которые позволят существенно расширить спектр рассматриваемых физико-химических процессов.

Основные понятия процесса адсорбции

Адсорбция - поглощение какого-либо вещества из газообразной среды или раствора поверхностным слоем жидкости или твёрдого тела. Вещество, на поверхности которого происходит адсорбция, называется адсорбентом, а поглощаемое из объёмной фазы - адсорбатом. Адсорбированные молекулы рано или поздно покидают поверхность - десорбируются. Время, в течение которого молекула находится на поверхности, называется временем адсорбции. Времена адсорбции могут колебаться в очень широких пределах. Скоростью адсорбции (соответственно скоростью десорбции) называется количество молекул, адсорбирующихся (или десорбирующихся) за единицу времени, оба значения величин относят к единице поверхности или массы адсорбента. Если скорости адсорбции и десорбции равны друг другу, то говорят, что установилось адсорбционное равновесие. Адсорбированные молекулы колеблются у поверхности адсорбента, то приближаясь, то удаляясь от неё. Чем выше температура, тем интенсивнее колебательное движение, и больше вероятность того, что в процессе таких колебаний связь молекулы с поверхностью будет разорвана и молекула десорбируется. Благодаря этому с ростом температуры уменьшается время адсорбции и равновесное количество адсорбированных молекул. С ростом концентрации или давления адсорбата в объёме увеличивается частота попаданий молекул адсорбата на поверхность адсорбента; пропорционально ей возрастает скорость адсорбции и увеличивается равновесное количество адсорбированных молекул. Кривые зависимости равновесной адсорбции от концентрации или давления адсорбата при постоянной температуре называются изотермами адсорбции. Почти всегда процесс адсорбции сопровождается выделением тепла, называемой теплотой адсорбции.

Описание математической модели процесса адсорбции

Для моделирования процесса физической адсорбции был выбран метод вероятностного клеточного автомата (ВКА), который заключается в следующем: поверхность адсорбента представляется в виде решетки, каждая ячейка которой может находиться в двух состояниях - занята адсорбированной молекулой либо свободна. Таким образом, состояние поверхности адсорбента определяется совокупностью состояний каждой ячейки. В свою очередь совокупность состояний поверхности в каждый момент времени определяет кинетику процесса. Шаг дискретизации по времени τ называется временным шагом ВКА, за его значение принимается период колебания молекулы адсорбата на поверхности адсорбента $\approx 10^{-12}$ с. Для каждой ячейки переход между состояниями носит вероятностный характер и определяется формулами

$$P_{адс} = \tau d^2 p \sqrt{\frac{N_A}{2\pi k T_Г M}},\ P_{дес} = e^{-\frac{\Delta E_{дес}}{kT_П}},$$

где τ - временной шаг ВКА, d - размер ячейки ВКА, p - давление адсорбируемого газа, N_A - число Авогадро, k - постоянная Больцмана, $T_Г$ - температура газа, M - молярная масса газа, $\Delta E_{дес}$ - энергия активации десорбции, $T_П$ - температура поверхности, $P_{адс}$ и $P_{дес}$ - соответственно вероятности адсорбции и десорбции.

Как следует из формулы вероятности адсорбции, размер ячейки ВКА является существенным параметром моделирования. В случае однородной поверхности размер ячейки ВКА определяется площадью поверхности адсорбента, занятой одной молекулой адсорбата, и называется молекулярной площадкой. Для определения этого параметра на основании данных о геометрическом строении молекулы была спроектирована и реализована подсистема расчета молекулярной площадки, которая вошла в состав системы моделирования процесса физической адсорбции на поверхности сорбента.

Методы вычисления молекулярной площадки

В подсистеме предусмотрена возможность вычисления молекулярной площадки несколькими методами:

- метод описанной сферы,
- метод описанного куба,
- метод проекции.

Метод описанной сферы основан на построении описанной сферы минимального радиуса, в объеме которой молекула адсорбата помещается целиком. Далее вычисляется площадь проекции сферы на плоскость адсорбента, и ее значение принимается за площадь ячейки ВКА. Поиск минимальной описанной сферы производится модифицированным случайным инкрементным алгоритмом Е. Велзла, обобщенным на трехмерный случай [1].

Алгоритм Е. Велзла для двумерного случая основан на утверждении, что наименьший описанный круг должен быть касательным к одному, двум или трем кругам. Нахождение круга, касательного к одному или двум кругам очевидно, для нахождения круга, касательного к трем кругам решается система уравнений:

$$(x-a_1)^2 + (y-b_1)^2 = (R-r_1)^2, \quad (1)$$

$$(x-a_2)^2 + (y-b_2)^2 = (R-r_2)^2, \quad (2)$$

$$(x-a_3)^2 + (y-b_3)^2 = (R-r_3)^2. \quad (3)$$

Путем несложных математических преобразований можно выразить x и y как функцию от R (формулы 4 и 5).

Упрощенная схема алгоритма miniCircle(C,B) для вычисления описанной сферы минимального радиуса представлена на рис. 3, где C – массив кругов, которые надо описать, а B – массив касательных кругов.

$$x = \left(\frac{r_1^2 - a_1^2 - b_1^2 - (r_2^2 - a_2^2 - b_2^2)}{2(b_2 - b_1)} - \frac{r_1^2 - a_1^2 - b_1^2 - (r_3^2 - a_3^2 - b_3^2)}{2(b_3 - b_1)} \right) \Bigg/ \left(\frac{a_2 - a_1}{b_2 - b_1} - \frac{a_3 - a_1}{b_3 - b_1} \right) -$$

$$- R \left(\frac{r_1 - r_2}{b_2 - b_1} - \frac{r_1 - r_3}{b_3 - b_1} \right) \Bigg/ \left(\frac{a_2 - a_1}{b_2 - b_1} - \frac{a_3 - a_1}{b_3 - b_1} \right), \quad (4)$$

$$y = \left(\frac{r_1^2 - a_1^2 - b_1^2 - (r_2^2 - a_2^2 - b_2^2)}{2(a_2 - a_1)} - \frac{r_1^2 - a_1^2 - b_1^2 - (r_3^2 - a_3^2 - b_3^2)}{2(a_3 - a_1)} \right) \Bigg/ \left(\frac{b_2 - b_1}{a_2 - a_1} - \frac{b_3 - b_1}{a_3 - a_1} \right) -$$

$$- R \left(\frac{r_1 - r_2}{a_2 - a_1} - \frac{r_1 - r_3}{a_3 - a_1} \right) \Bigg/ \left(\frac{b_2 - b_1}{a_2 - a_1} - \frac{b_3 - b_1}{a_3 - a_1} \right). \quad (5)$$

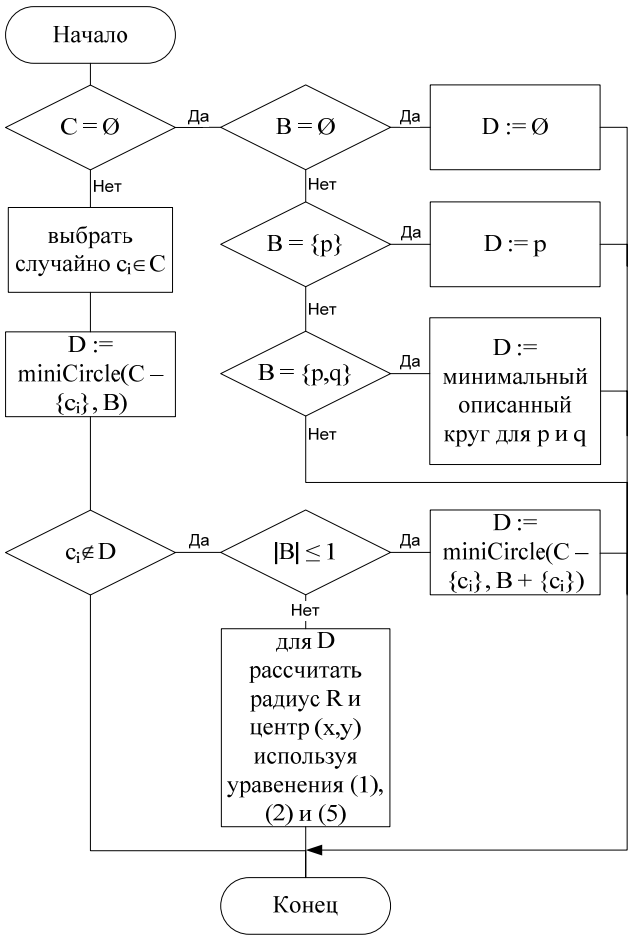


Рис. 3. Схема алгоритма miniCircle(C,B) для нахождения описанного круга минимального радиуса

Метод описанного куба аналогичен методу описанной сферы: сначала находится описанная сфера минимального радиуса, затем находится описанный куб для данной сферы, за размер молекулярной площадки принимается площадь стороны этого куба.

Метод проекции основан на нахождении площади проекции трехмерной модели молекулы на плоскость адсорбента.

Описание подсистемы расчета молекулярной площадки адсорбента

Основная форма подсистемы - вычисление молекулярной площадки - представлена на рис. 4. Пользователь должен ввести координаты *x, y, z* и радиусы *r* атомов молекулы в таблицу. Атомы и их ван-дер-ваальсовы радиусы выбираются из соответствующих справочников. В подсистеме предусмотрена возможность загрузки структуры молекулы из предварительно подготовленного файла Microsoft Excel (опция «Загрузить»). Если вся необходимая информация введена корректно, то после действия «Рассчитать» производится расчет значения молекулярной площадки, который выводится в поле «Размер ячейки», в противном случае выводится соответствующее предупреждение.

В подсистеме имеется возможность редактирования справочников сортов атомов и ван-дер-ваальсовых радиусов, экранная форма которых представлена на рисунке 5. Справочник ван-дер-ваальсовых радиусов может содержать несколько записей для одного химического элемента, полученных, например, из разных источников. При этом пользователь может менять порядок следования данных из разных источников, например, для сортировки их по достоверности.

Для хранения таблиц справочников используется библиотека SQLite. При использовании данной библиотеки нет необходимости в установке системы управления базами данных на компьютер пользователя. Все данные хранятся в одном файле на диске. При этом данная библиотека предоставляет подсистеме все преимущества использования реляционных баз данных и языка SQL.

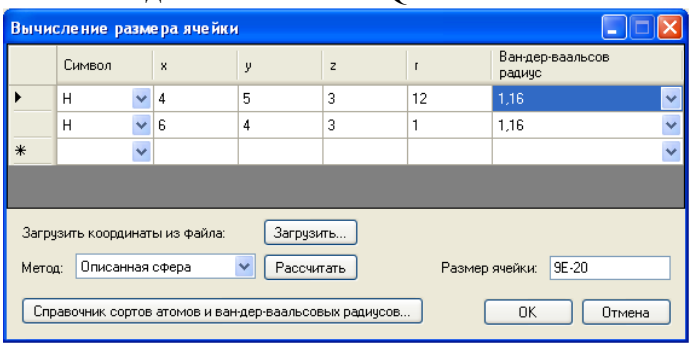


Рис. 4. Экранная форма вычисление размера молекулярной площадки

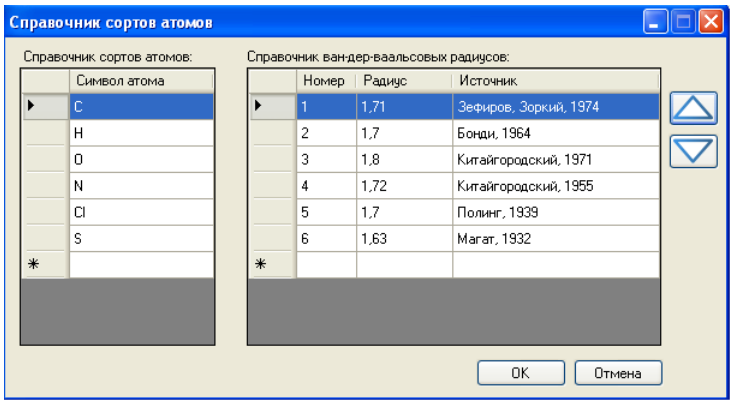


Рис. 5. Окно просмотра и редактирования справочников

Выводы

Проблема определения молекулярных площадок молекул возникла очень давно и имеет большое прикладное значение. Но возможность получения «точных» значений площадок молекул можно считать иллюзорной [2]. Экспериментальные методы определения площади, приходящейся на молекулу (метод Брунауэра-Эммета-Теллера и определение из плотности жидкости при плоском расположении молекул на поверхности), хоть и имеют хорошее согласие [3], но используют для определения

микроскопической величины макроскопические параметры и, по сути, являются примером обратной задачи [4]. Это приводит к необходимости использования эталонов и выбора «собственной шкалы» для каждого типа адсорбента [2].

Рассмотрение различных способов определения молекулярной площадки на микроскопическом уровне позволит углубиться в понимание процесса адсорбции. Разработанная подсистема является инструментом для выбора наиболее оправданной с физической точки зрения модели формирования молекулярной площадки, которая будет давать хорошее согласие с экспериментом. Понимание принципов выбора молекулярной площадки позволит уменьшить число измеряемых параметров при сохранении предсказательной способности метода моделирования процесса адсорбции, что приведет к увеличению «ценности» теории [5].

Кроме того, для всестороннего изучения многих процессов на границе газ – твердое тело (полимолекулярная физическая адсорбция, возникновение новых адсорбционных центров при хемосорбции и др.) необходимо рассматривать процесс адсорбции на молекулярном, микроскопическом уровне. В перспективе, развитие подсистемы расчета молекулярной площадки в комплексе с расширением возможностей систем расчета методом вероятностного клеточного автомата и расчета термодинамических характеристик адсорбции [6], позволит понять более тонкие механизмы адсорбционных процессов на поверхности твердых тел.

Литература

1. Xu, Sheng. Solution Methodologies for the Smallest Enclosing Circle Problem [Текст]/ Sheng Xu, Robert M. Freund , Jie Sun // Computational Optimization and Applications, v.25 n.1-3, p.283-292
2. Лопаткин, А.А. Теоретические основы физической адсорбции. [Текст]/ А.А. Лопаткин– М.: Изд-во Моск. ун-та, 1983 - 344 с.
3. Авгуль, Н.Н. Адсорбция газов и паров на однородных поверхностях. [Текст]/ Н.Н. Авгуль, А.В. Киселев, Д.П. Пошкус – М.: Химия, 1975 – 384 с.
4. Грибов, Л.А. Квантовая химия. [Текст]/Л.А Грибов, С.П. Муштакова – М.: Градарики, 1999 – 390 с.
5. Китайгородский, А.И. // Успехи физ. наук. 1979. № 3. С. 391-419.
6. Зеленко Л.С., Варфоломеева В.В., Кузнецов В.Ю. и др. // Вестн. СГАУ. 2008. № 1. С. 208-218.

Применение преобразование Гильберта-Хуанга в обработке медицинских изображений

А.Ю. Тычков
Пензенский государственный университет, Пенза

Введение

Преобразование Гильберта-Хуанга (Hilbert-Huang Transform - HHT) это метод обработки и анализа нелинейных и нестационарных данных. Разновидностью нелинейных и нестационарных данных являются одно и двумерные сигналы (изображения), особенности обработки которых, рассматриваются в представленной работе.

Преобразование HHT предложено Хуангом Н.Е. в 1995 в США (NASA) для изучения поверхностных волн тайфунов [1]. HHT основано на методе декомпозиции на эмпирические моды (empirical mode decomposition - EMD) и преобразование Гильберта (Hilbert Transform - HT). В последующие годы, по мере расширения возможностей преобразования HHT, нашло широкое применение в различных областях науки: прикладных и фундаментальных. Большой вклад практического использования HHT отмечено в следующих научно-технических направлениях: речевой анализ и идентификация источников; анализ звуковых волн, электромагнитных волн, теплопроводность, и механические колебания; анализ биомедицинских данных; анализ спектральных сигналов и т.д. Наибольший интерес представляет направление по обработке и анализу биомедицинских данных. Одним из ключевых направлений которого, является обработка медицинских изображений.

Постановка задачи

В настоящее время, проблема обработки медицинских изображений сложилась в отдельное направление, разветвлённое на множество частных задач, связанных с различными аспектами их обработки: корректировка цветовой гаммы, удаление помех, сегментация, выделение отдельных областей и контуров на изображении [2].

Применение преобразования Гильберта-Хуанга в обработке медицинских изображений является новой вехой в истории развития медицинских информационных систем. Согласно статистическому анализу и обзору работ, применение преобразования HHT для обработки медицинских изображений не обнаружено. Возможности HHT позволяют с высокой эффективностью выполнять широкий спектр задач по обработке и анализу медицинских изображений.

Решение

HHT представляет собой частотно-временное преобразование данных, которое не требует априорного функционального базиса. Базис, используемый для разложения изображения на составные части (набор двумерных эмпирических мод), конструируется непосредственно из самого входного сигнала. Это позволяет учесть локальные особенности входных данных, параметры сигнала, присутствие помех различного вида и интенсивности, выделять отдельные области и контуры деталей изображения, и т.д. Кроме адаптивности, HHT обладает и другими важными для практических приложений свойствами: ортогональностью и полнотой [3, 4.].

Преобразование HHT в обработке медицинских изображений включает в себя последовательное выполнение двух действий: декомпозицию сигнала на эмпирические моды (EMD) и преобразование Гильберта (HT). Эмпирические моды – это монокомпонентные составляющие сигнала, модулированные по амплитуде и частоте, т.е. их амплитуда и частота меняются во времени. Базовый алгоритм декомпозиции сигнала на эмпирические моды [1] предложенный Хуангом Н.Е. в 1998 не позволяет

обрабатывать двумерные сигналы, к которым относятся изображения, в том числе и медицинские. Возможности базового алгоритма ограничены обработкой одномерных сигналов. Поэтому в 2002 году коллективом авторов предложено использовать технологию 2D EMD или bidimensional empirical mode decomposition (BEMD) [5]. Данный подход позволяет раскладывать изображение на эмпирические моды, однако как показала практика, для обработки медицинских изображений высокого качества данный алгоритм является не эффективным. Время, которое требуется для выполнения действий декомпозиции алгоритмом BEMD к примеру, флюорографического снимка размером 1200x900 пикселей, составляет порядка более часа, что может отрицательно сказаться на ***оперативной*** постановки диагностического заключения. В 2004 году предложен алгоритм ensemble empirical mode decomposition (EEMD) на базе BEMD [6]. Алгоритм EEMD позволил снизить время проведения декомпозиции входного сигнала на эмпирические моды до 40 минут в зависимости от качества входного изображения. Однако уже в 2009 году был предложен алгоритм multi-dimensional ensemble empirical mode decomposition (MEEMD) [7], который позволяет производить декомпозицию двумерного сигнала высокого качества менее чем за 20 минут, что значительно повысило его конкурентоспособность по сравнению с аналогами. Как показала практика, применение алгоритма MEEMD для обработки медицинских изображений ранее не обнаружено, что позволило на его основе разработать алгоритм обработки медицинских изображений (см. рисунок 1), на примере флюорографического снимка (ФОС).

Зарегистрированный флюорографический снимок приведен на рисунке 2а, в нашем случае он представлен в виде матрицы (см. рисунок 2б) значений градации яркости x_{ij}, по строкам и столбцам (i – номер отсчета по строке, i_{max}= 350, j – номер строки, j_{max}= 350).

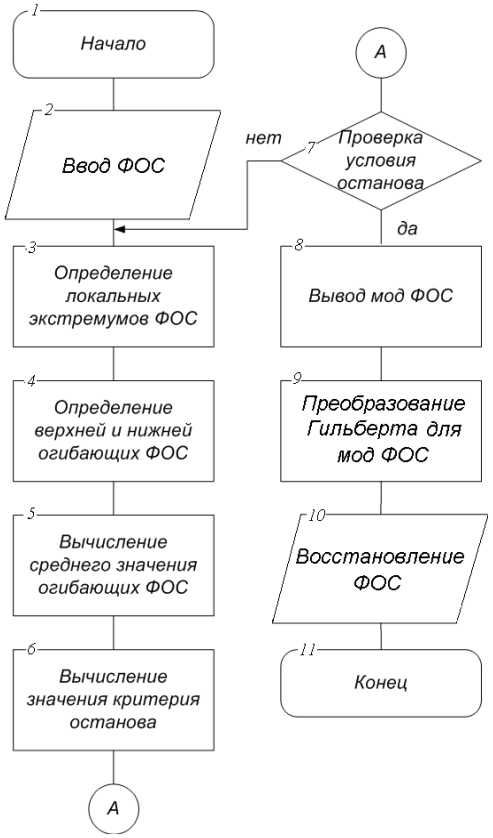
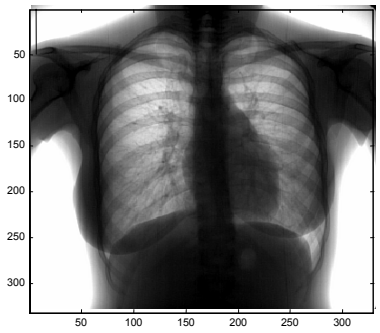


Рис. 1. Алгоритм обработки медицинских изображений, на примере флюорографического снимка

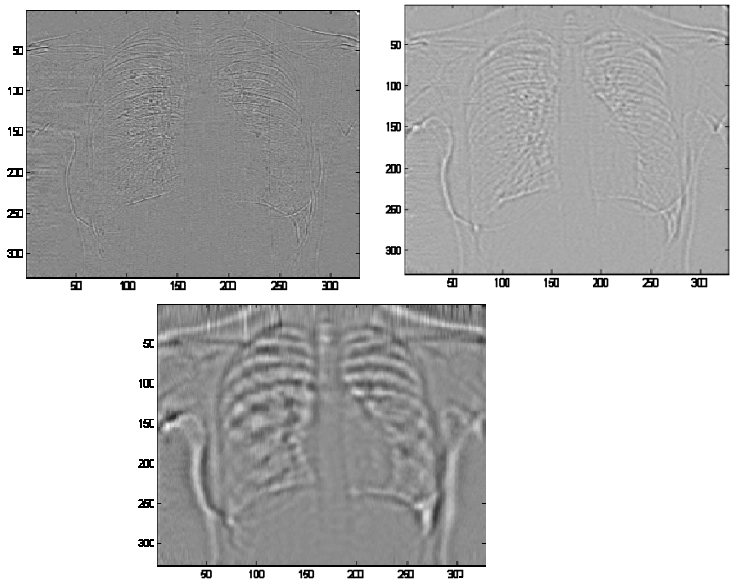


$$\begin{bmatrix} x11 & x21 & \dots & xi1 \\ x12 & x22 & \dots & xi2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x1j & x2j & \dots & xij \end{bmatrix}$$

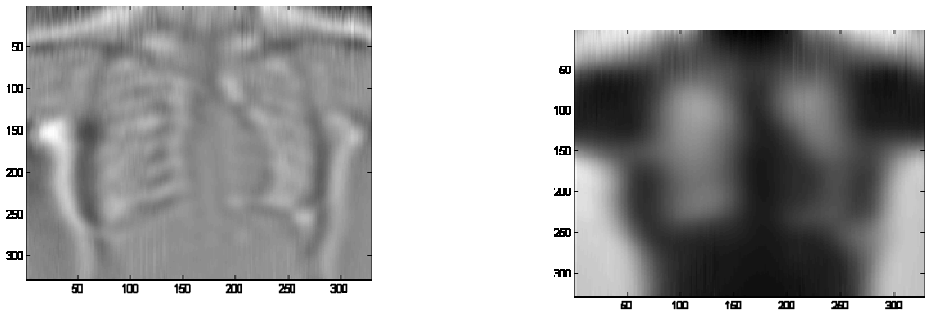
а б

Рис. 2. Зарегистрированный флюорографический снимок (а); матрица значений градации яркости снимка (б)

Декомпозиция флюорографического снимка на эмпирические моды производиться согласно алгоритму MEEMD. Последовательность действий можно проследить на рисунке 1: действия 3-8 в точности повторяют работу базового алгоритма декомпозиции на эмпирические моды, с учетом обработки входного сигнала в двух плоскостях. Результатом действий 3-8 алгоритма обработки медицинских изображений является формирование эмпирических мод флюорографического снимка (см. рис. 3).



а 500 Гц б 250 Гц в 125 Гц



г 62,5 Гц д 31,25 Гц

Рис. 3.

Декомпозиция флюорографического снимка производится путем разложения его на аппроксимирующую и детализирующую функцию. Представленные на рисунке 3 эмпирические моды являются детализирующими, они получены путем применения

алгоритма MEEMD последовательно для каждого изображения, начиная с исходного флюорографического снимка, зарегистрированного с частотой дискретизации 1000 Гц. Из флюорографического снимка (см. рисунок 1а) получена первая мода с частотой дискретизации 500 Гц (см. рисунок 3а), а из нее получена следующая с частотой дискретизации 250 Гц (см. рисунок 3б) и так далее. Частота дискретизации каждого сигнала, прошедшего через алгоритм декомпозиции, делится пополам. Каждая мода, представленная на рисунке 3, получена в результате разложения предшествующей ей моды. Мода, представленная на рисунке 3д, получена с частотой дискретизации 31,25 Гц и дальнейшему разложению не подлежит, потому что не выполняются условия процесса декомпозиции [3, 4].

Следующим действием алгоритма обработки медицинских изображений является преобразование Гильберта для ранее полученных мод входного изображения, в нашем случае это моды флюорографического снимка.

Значения мгновенных частот эмпирических мод флюорографического снимка, полученных в результате выполнения действий декомпозиции, представляют собой спектр Гильберта, который позволяет [8]:

– определить по его виду наличие и характер амплитудной и частотной модуляций;

– идентифицировать временные и частотные диапазоны, где концентрируется энергия исходного сигнала;

– построить трехмерную поверхность энергетической плотности в системе координат энергия-частота-время.

Построение трехмерной поверхности сигнала в системе координат энергия-частота-время осуществляется посредством преобразования Гильберта эмпирических мод сигнала следующим образом [9, 10]:

$$M_k(t_i) = \frac{1}{\pi}\int \frac{m_k(\tau)}{t_i - \tau} d\tau \,, \tag{1}$$

где $m_k(t_i)$ – эмпирическая мода, подвергнутая преобразованию Гильберта;

$M_k(t_i)$ – сопряженный по Гильберту сигнал, соответствующий моде $m_k(t_i)$;

k – номер моды.

τ – независимая переменная.

Затем для каждой моды сигнала определяется аналитический (комплексный) сигнал:

$$Z = m_k(t_i) + qM_k(t_i), \tag{2}$$

где $q = \sqrt{-1}$ – мнимая единица.

Далее определяется мгновенное значение амплитуды каждой моды сигнала:

$$a_k(t) = \sqrt{m_k(t_i)^2 + M_k(t_i)^2} \tag{3}$$

и мгновенное значение частоты каждой моды сигнала

$$\omega_k(t) = 1/t_i \,. \tag{4}$$

После выполнения преобразования Гильберта в каждой эмпирической моде изображения в виде поверхности в системе координат амплитуда-частота-время может быть выражена следующим образом [10]:

$$f_j(t_i) = \sum_k a_k(t) \times \exp\left(q\int \omega_k(t)dt\right), \tag{5}$$

Чтобы представить энергетическую плотность частотных составляющих сигнала и построить поверхность в системе координат энергия-частота-время мгновенной амплитуде $a_k(t)$ в выражении (3) придают квадратичную форму

$$E(t_i) = \sum_k a_k{}^2(t) \times \exp\left(q \int \omega_k(t) dt\right). \quad (6)$$

Результатом выполнения преобразования Гильберта мод флюорографического снимка является поверхность в системе координат энергия-частота-время. Трехмерная поверхность позволяет эффективно выделять локальные области с низкой энергией, производить фильтрацию, подавление помех и искажений в сигнале, и т.д.

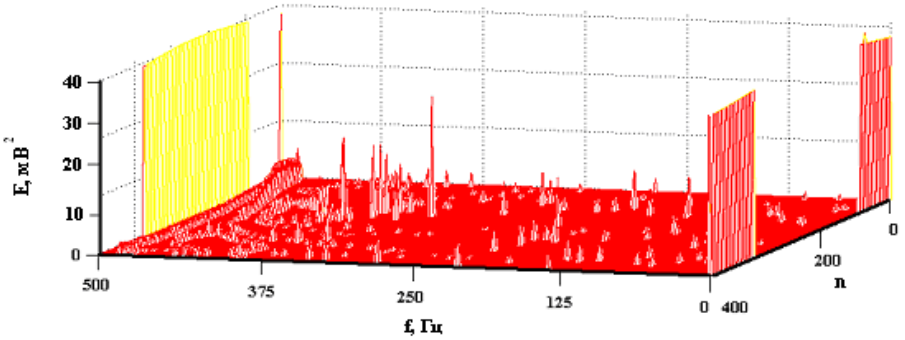


Рис. 4. Поверхность мод ФОС построенных посредством преобразования Гильберта

Выводы

Согласно алгоритму обработки медицинских изображений, произведена декомпозиция ФОС на эмпирические моды. Анализ частотных составляющих сигнала открывает возможность выполнения процедур подавления помех и искажений на медицинских изображениях, которые могут настолько изменить входной сигнал, что он становится непригодным даже для визуальной интерпретации. Выполнение процедур подавления помех и искажений может быть достигнуто за счет удаления высокочастотных мод, в зависимости от свойств помех, пороговой обработки отдельных мод или их локальному анализу с последующей фильтрацией.

Выполнение преобразования Гильберта позволяет уже в трехмерном пространстве анализировать входную информацию, как сам исходный флюорографический снимок, так и его частные составляющие (эмпирические моды). Трехмерная поверхность позволяет эффективно выделять локальные области с низкой энергией по трем координатам энергия-время-частота. Кроме того, анализ трехмерной поверхности открывает возможность постановки диагностического заключения на ранней стадии развития заболевания. Это может быть достигнуто за счет анализа энергетических точек на плоскости, выделение точек, соответствующих определенному заболеванию и их статистической обработки, с целью выявления динамики течения заболевания у пациента. Выполнение последнего действия, по предположению автора, может быть достигнуто за счет наличия верифицированной базы данных по тем или иным патологическим заболеваниям, с учетом особенностей их локализации в трехмерном пространстве в системе координат энергия-время-частота.

Направлением дальнейших исследований следует считать моделирование и исследование процессов обработки и анализа медицинских изображений, разработку и исследование новых методов и алгоритмов подавления помех на медицинских изображениях, применение которых возможно для всего многообразия исследуемых сигналов, а так же алгоритмов выделение отдельных областей и деталей изображения.

Литература

1. Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. London, Ser. A, 454, 903-995.

2. Путятин, Е.П. Нормализация и распознавание изображений / http://sumschool.sumdu.edu.ua/is-02/rus/lectures/pytyatin/pytyatin.htm.

3. Huang N.E., Shen S.S. The Hilbert-Huang transform and its applications – World Scientific Publication, 2005, pp. 323.
4. Huang, N.E., Attoh-Okine Nii O. The Hilbert-Huang transform in engineering – Taylor and Francis, 2005, pp. 95.
5. Linderhed, A., 2D empirical mode decompositions in the spirit of image compression, Wavelet Independ. Compon. Analy. Appl. IX, SPIE Proc, 4738, 1-8, 2002.
6. Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," Tech. Rep. 193, Centre for Ocean-Land-Atmosphere Studies, Calverton, Md, USA, 2005.
7. Z. Wu, N. E. Huang and X. Chen, The multi-dimensional ensemble empirical mode decomposition method, Advances in Adaptive Data Analysis 1 (2009), 339–372
8. Huang, N.E., Shen, Z., and Long, S.R., A new view of water waves—the Hilbert spectrum, *Ann. Rev. Fluid Mech.,* 31, 417-457, 1999
9. Cohen A., Zigel Y. Compression of Multichannel Through Multichannel Long Term Prediction, IEEE BME magazine, Vol. 17, 1998, pp. 109-115.
10. Poularicas, A.D. The transform and application handbook – Library of congress cataloging in publication data, CIP 99-16719, 2000, pp. 1335.

Роль систем Service Desk в IT–менеджменте предприятия

Т.О. Шевина, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

По независимым оценкам, до 80% ресурсов инфраструктуры информационных технологий уходит на обеспечение сервисов и их поддержку. Именно поэтому как в мире, так и в России все более востребованными становятся системы класса Service Desk, которые позволяют обеспечить наиболее эффективную работу сервисов предприятий широкого класса.

Service desk - информационная система, обеспечивающая техническую поддержку пользователей при их работе с аппаратным и/или программным обеспечением. Это важная составляющая ITIL (Information Technology Infrastructure Library – стандарт в области организации и управления информационными технологиями, обобщающий в себе лучший международный опыт). Системы класса Service Desk позволяют выявить проблемные участки IT-инфраструктуры и оценить эффективность работы предприятия [1].

Появление Service Desk является следствием сервисного подхода: если IT-отдел предоставляет пользователям какие-либо сервисы, то нужна «точка контакта» в случае любых затруднений у пользователей при использовании этих сервисов, поэтому Service Desk можно рассматривать как интерфейс между пользователями и провайдерами IT-услуг.

Круг задач, решаемых с помощью Service Desk

Цель создания Service Desk - обеспечение IT-подразделений удобным инструментом для организации технической поддержки сотрудников организации, управления событиями, инцидентами, запросами на обслуживание и проблемами, такие системы позволяют оптимизировать внутренние процессы обслуживания персонала компании и повысить качество предоставляемых сервисов (рис. 1).

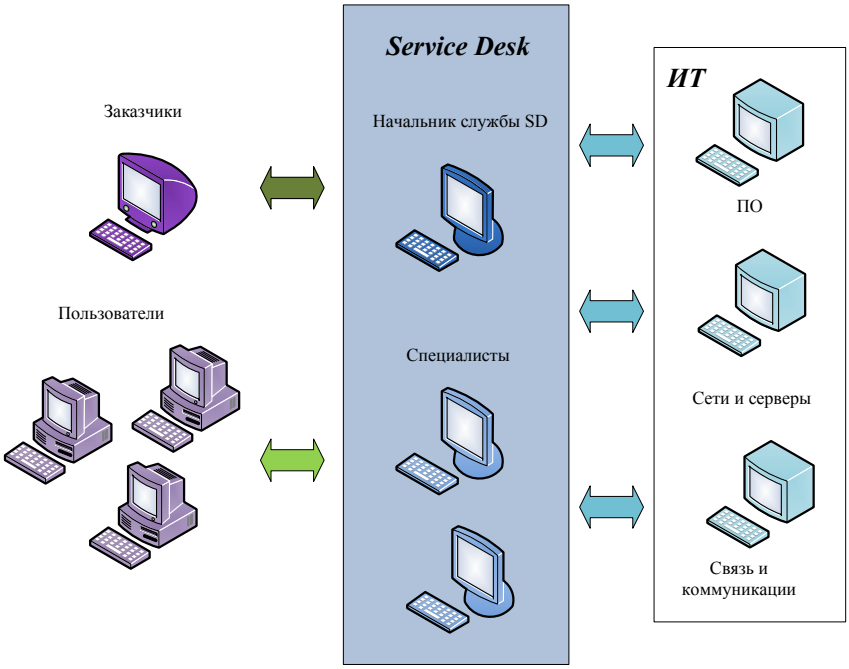


Рис. 1. Схема взаимодействия Service Desk с пользователями услуг и ИТ

Обработка любого обращения пользователя к системе происходит по следующему сценарию (рис. 2): специалисты Call-центра принимают звонки от пользователей, осуществляют первоначальное взаимодействие с клиентами. Затем формируется заявка, сотрудник группы диспетчеров регистрирует ее и классифицирует по одному из типов (инцидент, заявка на обслуживание, проблема и т.д.). Сотрудник первой линии поддержки проводит оценку обращения, идентифицирует проблему и устраняет ее. В случае невозможности решения вопроса на заданном уровне он совершает эскалацию, т.е. передает обращение на более высокую линию поддержки. Информирование клиентов о состоянии и процессе обработки их обращения происходит в течение всего процесса обработки. После решения проблемы заявка закрывается, определяется необходимость обучения клиента, а затем руководителю департамента информационных технологий (ДИТ) представляется окончательный отчет о результатах обработки обращения.

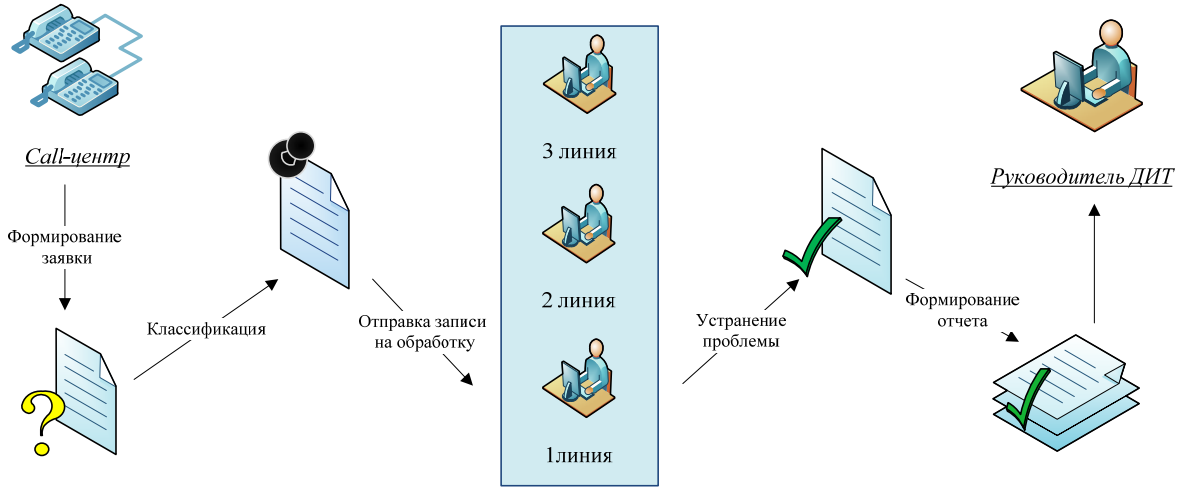


Рис. 2. Сценарий обработки обращения пользователя

Таким образом, системы Service Desk обеспечивают:

– *единую «точку обращения» к службе поддержки* - удобный и понятный для пользователей механизм направляет запросы в службу поддержки, минуя менее эффективные способы разрешения проблем (попытки решить проблему самостоятельно или при помощи коллег, обращение к первому попавшемуся сотруднику, даже если тот не обязан заниматься поддержкой и т.п.);

– *стандартный способ регистрации* и выдачи заданий специалистам;

– *контроль последовательности исполнения работ*, потраченного времени и ресурсов и формирование соответствующей отчётности, это позволяет формализовать отношения между пользователями информационных систем предприятия и службой IT, а также выявлять закономерности в потоке поступающих запросов, выделяя таким образом «узкие места» в инфраструктуре предприятия;

– *назначение приоритетов запросам* в зависимости от типа запроса, конкретного пользователя, загруженности системы или других обстоятельств;

– *эскалацию запросов и инцидентов*, оповещение соответствующих администраторов;

– *формирование базы знаний* по прошлым запросам, позволяющей специалистам быстро разрешать проблемы, схожие с уже возникавшими [2].

Особенности структурной организации и функционирования

Системы класса Service Desk строятся по модульному принципу и состоят из следующих стандартных компонентов (рис. 3):

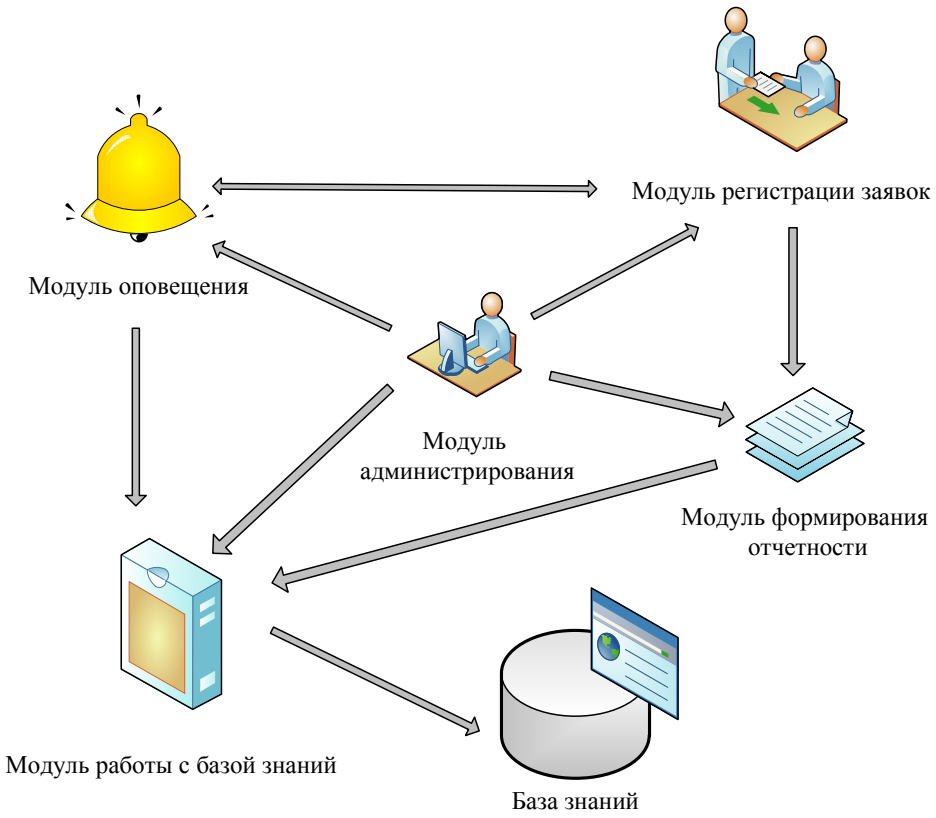


Рис. 3. Стандартные компоненты Service Desk

- модуль регистрации заявок об инцидентах;
- модуль отслеживания статуса заявки и оповещения;
- модуль работы с базой знаний;
- модуль настройки и администрирования;
- модуль формирования отчетности.

Модуль регистрации заявок об инцидентах хранит записи об инцидентах, обеспечивает выполнение операций с ними, добавление вложенных файлов (скан-копий документов, документы в текстовых или графических форматах) и ведение обсуждений в записях об инцидентах.

Модуль отслеживания статуса заявки и оповещения предназначен для обеспечения автоматической отправки уведомлений пользователям системы при изменении статусов записей.

База знаний служит хранилищем информации по всем известным на текущий момент решениям запросов на обслуживание, инцидентов и проблем, связанных с ошибками ПО, отказами оборудования и т.д. (рис. 4).

Модуль настройки и администрирования обеспечивает возможность задания параметров конфигурации системы.

Модуль формирования отчетности позволяет руководителям IT-подразделений контролировать эффективность обслуживания с помощью отчетов.

Системы Service Desk могут также интегрироваться с другими информационными системами, например, со средствами учёта компьютерного оборудования или инструментами управления сетью.

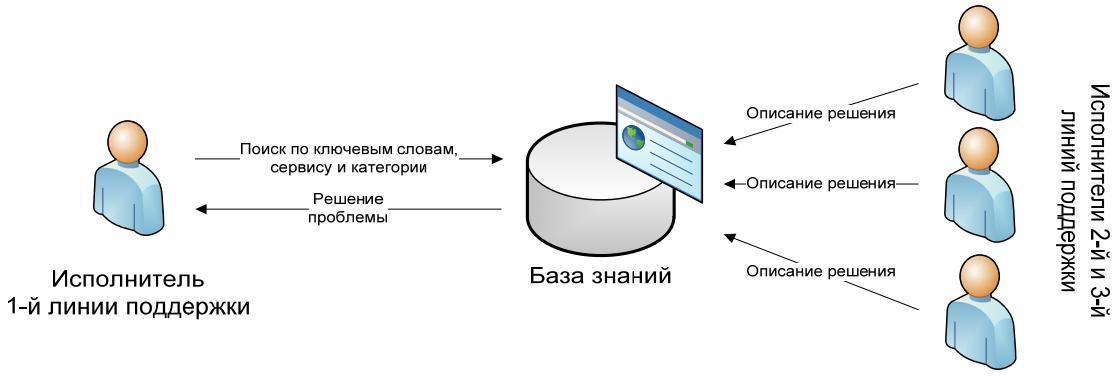


Рис. 4. Организация обслуживания базы знаний

Заключение

Для получения эффекта от внедрения системы Service Desk предприятие должно предварительно организовать качественную работу традиционных служб (сервисов). Если эффективность работы предприятия низка, отсутствует общеорганизационная культура оказания услуг и предоставления сервисов, то никакая автоматизированная система не исправит ситуацию. Сам по себе продукт Service Desk не решает ни одной из перечисленных проблем. Для того чтобы компания ощутила реальную пользу от его внедрения, необходимо правильно организовать работу сотрудников и четко разделить круг задач для различных групп персонала.

Литература

1. ITIL V3. Glossary Russian Translation v0.92. [Электронный ресурс]// www.itsmforum.ru/news/all_news/2009_05_12_Glossary_Russian

2. Ян Ван Бон, Георгес Кеммерлинг, Дик Пондман, «Введение в ИТ сервис-менеджмент». Издательство «Открытые системы», Москва, 2005 г. — 225 стр.

Методы оперативного рационального перестрахования особо серьёзного риска на базе эволюционных алгоритмов как новое научное направление. Риски авиационные и космические

М.В. Шептунов
Москва

Значительного внимания заслуживают вопросы управления риском – в технических и социально-экономических системах, вопросы оперативного рационального перестрахования особо серьёзного риска между членами страховых либо перестраховочных пулов. Требующими перестрахования, как правило, являются авиационные и космические риски вследствие их особой значимости. Существенные результаты в данной сфере возможны при использовании достижений различных областей знания.

Страхование воздушных судов как вид авиационного страхования осуществляется наряду со страхованием различных видов ответственности эксплуатантов воздушных судов и других летательных аппаратов перед третьими лицами на земле и в полёте, перед пассажирами на случай смерти или телесных повреждений, за багаж и ручную кладь, за груз общегражданской ответственности.

При осуществлении космической деятельности страхование имущественных интересов широко распространено в мировой практике, начиная с 1965 г. При этом страхованием охвачены как запуски отдельных космических аппаратов, так и космические программы, включающие разработку и запуск серии спутников, проведение различных экспериментов в космосе, жизнь и здоровье космонавтов и наземного персонала, потери доходов, ответственность перед третьей стороной и проч. Такое страхование часто называют страхованием космической деятельности или космическим страхованием.

Относительно авиационного страхования следует отметить, что международная практика относит к нему также многие другие виды страхования, включая страхование авиазапчастей и двигателей, пассажиров от несчастных случаев, членов экипажей, авиадиспетчеров и других специалистов от риска утраты способности к своей профессиональной деятельности, страхование ответственности аэропортов и производителей летательных аппаратов.

С самого начала выявилась специфика авиационного страхования – его катастрофическая природа. В отличие от страхования морских и речных судов, где случаев частных аварий намного больше, чем полной гибели судов, в авиационном страховании чаще всего наблюдается обратная ситуация.

В страховании космической деятельности наибольший риск связан с возможностью полной гибели дорогостоящей ракеты космического назначения [1, 3]. Возможно также страхование наземных комплексов и страхование ответственности перед третьими лицами за ущерб, нанесённый им при эксплуатации космической техники. Последнее может иметь место в случаях: падения элементов ракеты космического назначения вне зоны отчуждения вследствие как случайных факторов, так и возникновения аварийной ситуации; возникновения экологической катастрофы в результате аварии ракеты космического назначения с токсичными компонентами топлива.

Перестрахование, представляющее собой операцию между двумя страховыми

компаниями по приёму-передаче части (или всей суммы) риска по страховому договору в обмен на выплату страховой премии, облегчает положение страховщиков при страховом событии – за счёт дробления крупного риска между

членами страхового пула. Сам процесс передачи рисков по-другому называют цессией; перестрахователя (страховую компанию), передающую риски в перестрахование – цедентом; перестраховщика (страховое либо перестраховочное общество, пул), принимающее определённый риск в перестраховании – цессионером (либо цессионарием). Соответственно, под ретроцессией подразумевают передачу рисков, принятых в перестрахование, в дальнейшее перестрахование, а под ретроцедентом – страховую либо перестраховочную компанию, передающую принятые в перестрахование риски в ретроцессию.

Страхователями воздушных судов могут выступать операторы воздушных такси, аэроклубы, авиазаводы, различные организации, фирмы, занимающиеся продажей и перепродажей авиатехники, частные владельцы воздушных судов и др.

При передаче крупного риска на перестрахование при большом количестве членов пула могут иметь место значительные вычислительные трудности, связанные с оперативным выбором страховщиком приемлемых вариантов перестрахования. Эта задача может рассматриваться как модифицированная задача коммивояжера, имеющая отличающийся экономический смысл и эффективно решаемая на основе специализированных эволюционных алгоритмов.

Безусловно, важна оперативность принятия рациональных решений в сфере управления особо серьёзным риском путём его перестрахования.

В настоящее время во многих областях человеческого знания интенсивно развивается тематика, связанная с исследованиями эволюционных алгоритмов и возможностями их применения. Не является исключением и область перестрахования рисков, в том числе в авиации и космонавтике.

Принципы действия эволюционных алгоритмов аналогичны процессам, происходящим в природе. Таковыми являются генетические, муравьиные алгоритмы и некоторые другие, названия которых являются производными соответствующих природных процессов либо систем.

Сущность генетических алгоритмов в реализации целенаправленного процесса размножения-гибели, при котором размножению соответствует появление новых объектов, а гибели – удаление объектов в соответствии с определёнными критериями отбора. Эти алгоритмы сочетают в себе механизмы выживания сильнейших особей и некоторый упорядоченный перебор, в котором присутствует элемент случайности; осуществляются операции копирования строк, обмен их частями.

Следует выделить достоинства генетических алгоритмов: 1) они слабо чувствительны к выбору начальных условий; 2) скорость работы этих алгоритмов приемлема при решении задач большой размерности.

В свою очередь, основная идея муравьиного алгоритма имитирует движение муравья (одного либо нескольких, в зависимости от специфики решаемой задачи). Хотя муравьи и слепы, они способны перемещаться по сложной местности, отыскивать пищу на большом расстоянии от муравейника и успешно возвращаться в него.

Чем больше муравьёв используют один и тот же путь, тем выше концентрация специального выделяемого фермента на этом пути. Чем ближе внешняя точка к муравейнику, тем больше раз к ней перемещались муравьи. Чем выше концентрация ферментов на рассматриваемом пути, тем предпочтительнее он для муравьёв по сравнению со всеми прочими из множества доступных.

Таким образом, эволюционные генетические и муравьиные алгоритмы представляют существенный интерес при решении широкого класса теоретических и научно-практических задач.

Пусть, например, страховой пул включает в себя всего пять членов, причём a – страховщик, b, c, d, e – перестраховщики. В рассматриваемой задаче сумма

страховой ответственности каждого последующего перестраховщика, являющаяся

частью страховых условий, зависит от “рейтинга” предыдущих перестраховщиков

в глазах каждого последующего:

- при первом варианте перестрахования

$SSO_c = f_{bc}(SSO_b)$, $SSO_d = f_{cd}(SSO_c, SSO_b)$, $SSO_e = f_{de}(SSO_d, SSO_c, SSO_b)$,

- при втором варианте перестрахования $SSO_d = f_{bd}(SSO_b)$, $SSO_e = f_{de}(SSO_d, SSO_b)$,…,

- при третьем варианте $SSO_c = f_{bc}(SSO_b)$, $SSO_e = f_{ce}(SSO_c, SSO_b)$,…, и т. д.,

где SSO_c, SSO_b, SSO_d, SSO_e – суммы планируемой страховой ответственности перестраховщиков *c*, *b*, *d*, *e* соответственно, f_{bc}, f_{cd}, f_{de}, f_{bd}, f_{ce} – некоторые функции, отражающие “рейтинг” перестраховщиков по отношению друг к другу и им известные.

“Рейтинги” перестраховщиков, выражаемые упомянутыми функциями, могут носить объективный либо субъективный характер.

В свете сказанного при большом числе членов страхового пула весьма актуальна оперативность принятия рациональных решений о порядке перестрахования особо серьёзного риска прямым страховщиком – что достижимо за счёт применения эволюционного (генетического либо муравьиного) алгоритма.

Следует отметить, что общее количество вариантов путей обхода в задаче коммивояжера составляет (*n*-1)!, где *n* – число пунктов обхода. При этом подчеркнём, что $10! = 3628800$; $12! = 479001600$; $15! = 1307674368000$.

В случае тотального перебора вариантов при *n* порядка 10 и более пунктов данная задача неразрешима даже за один год (!) при использовании ПЭВМ.

При значительном количестве пунктов (городов) часто применяют метод ветвей и границ, способствующий сокращению количества вариантов перебора. Однако, данный метод не гарантирует, что в ходе решения задачи не будет произведен полный перебор вариантов. В связи с этим при значительном числе перестраховщиков весьма актуально методическое отыскание приемлемых последовательностей перестрахования путём вычисления посредством эволюционных алгоритмов предлагаемого критерия оптимальности – для сферы перестрахования.

При переходе от классической задачи коммивояжера к упомянутой модифицированной будет иметь место отличающийся критерий оптимальности, а именно отличающаяся целевая функция задачи. Суть её в минимизации разности между некоторой фиксированной суммой, на которую осуществляется страхование данного серьёзного риска всем страховым пулом – и суммарным размером ответственности всех перестраховщиков, не считая прямого страховщика:

$$x_k^{safe} = \left(S - \sum_{q=1}^{n-1} SSO_q \right) \to \min, \qquad (1)$$

где x_k^{safe} – искомая приемлемая сумма страховой ответственности страховщика,

q – номер дуги по порядку обхода графа, в котором все *n* членов страхового пула изображены вершинами графа, соединёнными, при наличии партнёрства, между собой,

SSO_q – сумма страховой ответственности *q*-го члена страхового пула при наступлении страхового случая, $q = \overline{1,(n-1)}$, $k = \overline{1,(n-1)}$.

На рис. 1 обозначены: a – страховщик, b, c, d, e – перестраховщики, x_k – искомая сумма страховой ответственности (первого) страховщика при различных вариантах последовательности перестрахования, $k = \overline{1,(n-1)}, \langle 100 \rangle$ – сумма S (здесь условная), на которую страхуется риск страховщиком a при последующем перестраховании; остальные числа в пределах этой суммы означают планируемые суммы страховой

ответственности, соответствующие каждому из перестраховщиков.

О вопросе назначения упомянутых лимитов (приемлемого диапазона)

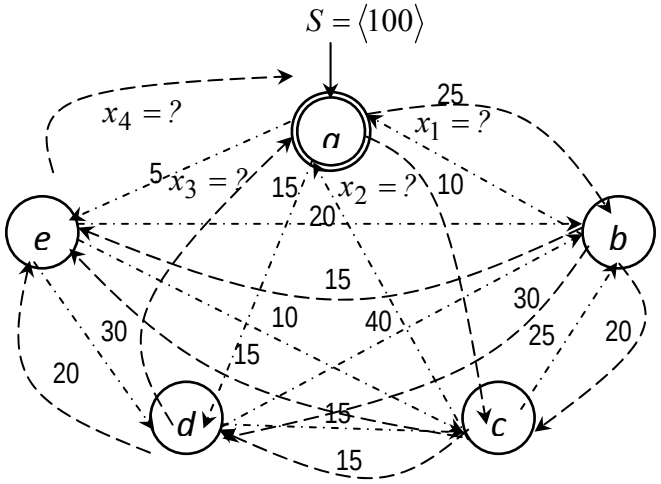


Рис. 1. Поясняющий числовой пример приведения задачи оптимизируемого перестрахования к модифицированной задаче коммивояжера

собственного удержания прямого страховщика. Оно зависит от тарифа (ставки страховой премии): предполагается, что при меньшей вероятности наступления убытка компания может увеличить размер собственного удержания.

Ещё один важный момент состоит в том, что цедент (страховщик), передавая часть риска перестраховщику (цессионеру – ретроцеденту), одновременно делится с ним премией.

Выплачиваются обязательно оригинальная, а также перестраховочная, комиссии: оригинальная является вычетом из премии в пользу цедента и в зависимости от распределяемых рисков достигает 20–40 % премии; перестраховочная выплачивается в пользу ретроцедента и составляет 10–15 % премии.

При участии страхового брокера – независимого посредника – имеет место также брокерская комиссия, представляющая собой вычет в его пользу.

Пример расчёта премии перестраховщика (цессионера) при участии страхового брокера для страховой суммы S=100 д. е. с. с., за вычетом комиссионных, представлен таблицей 1.

Таблица 1

Страховая сумма: 100 д. е. с. с. (денежных единиц страховой суммы)			
Цессия:	Ставка премии	8 %	8 д. е. с. с. (сумма премии)
	Оригинальная комиссия цеденту (перестрахователю)	25 %	2 д. е. с. с.
Ретроцессия:	Брокерская комиссия	5 %	0,4 д. е. с. с.
	Доля перестраховщика (цессионера – ретроцедента):	10 %	0,8 д. е. с. с.
	Вычет цеденту	10	0,2 д. е. с. с.

	(перестрахователю):	%	
	Вычет страховому брокеру:	10 %	0,04 д. е. с. с.
Итого перестраховщику:			0,56 д. е. с. с.: [0,8-0,2-0,04]

Отметим, что с помощью генетического алгоритма значение целевой функции вида (1) для поставленной задачи, в соответствии с рис. 1, возможно найти всего за две итерации: $x_k^{safe} = 5$ д.е.с.с. при перестановке *aedbc* (для случая приемлемого диапазона безопасного значения собственного удержания $5 \le x_k^{safe} \le 10$).

Перестрахование как таковое имеет международный характер. Не последнюю роль в определённых страховых случаях играет и фактор распределения рисков в пространстве, между перестраховочными и страховыми компаниями разных стран.

Отметим также, что согласно [8]: “Научное направление (или подход) – то направление в изучении явлений, в котором предполагается, что их понимание в основном достигается через систематическое применение научного метода”.

В то же время следует отметить различие между двумя близкими понятиями “научное направление” и “научная школа”. В науковедении под научным направлением (в узком смысле этого слова) обычно понимают чётко выраженный мировоззренческий, методологический, методический, даже технический подход к исследованию одного и того же объекта, который пытается развивать или один исследователь, или группа исследователей-единомышленников. Этот подход может быть воспринят, может быть не воспринят со стороны других членов дисциплинарного научного сообщества, как делающий существенный вклад в развитие взглядов на объект. Научная школа является более узким по содержанию понятием [2]. Это чётко оформленное научное направление, которое развивается поколениями единомышленников.

Учитывая, что перестрахование очень многогранно, а эволюционные (генетические, муравьиные) алгоритмы, которые, как установлено [5, 6], могут успешно применяться в перестраховании серьёзных рисков и весьма глубоки по своему научно-практическому содержанию, возможно из всего сказанного сделать вывод о факте основания нового научного направления. Оно обоснованно может быть названо как “Методы оперативного рационального перестрахования особо серьёзного риска на базе эволюционных алгоритмов”, находящееся на стыке экономики перестрахования, перспективных информационных технологий и естествознания.

Литература

1. Медведчиков Д. А. Организационно-экономические принципы страхования космических рисков. – М.: Анкил, 1998.
2. Смирнов В.В. Критерии сравнения научных направлений и школ социологии организаций // Рабочие тетради по компаративистике. Вып. 4: Сравнительные исследования в социальных и экономических науках. – С.-Пб., 2002.
3. Страхование от *А* до *Я* / Под ред. Л. И. Корчевской, К. Е. Турбиной. – М.: ИНФРА-М, 1996.
4. Холланд Дж. Генетические алгоритмы // В мире науки. 1992. №9-10.
5. Шептунов М. В. Задача коммивояжера и применение генетического алгоритма при перестраховании по критерию минимизации суммы ответственности прямого страховщика // Страховое дело. 2008. №11.
6. Шептунов М. В. Применение муравьиного алгоритма в модифицированной задаче коммивояжера при перестраховании по критерию минимизации суммы ответственности цедента // Страховое дело. 2009. №№11-12.
7. Bonavear F., Dorigo M. Swarm Intelligence: from Natural to Artificial Systems. – Oxford university Press, 1999.
8. http://mirslovarei.com.

Формат *.dxf как основа технологического проектирования

В.Н. Шерстобитова
Аэрокосмический институт Оренбургского государственного университета, Оренбург

Введение

В настоящее время в России на предприятиях машиностроения, авиации и космоса используют современные информационные технологии, связанные с внедрением отечественных автоматизированных систем проектирования и технологической подготовки. Новейшие образцы техники в авиации, космосе определяют и новые подходы к производству. Конструкторско-технологическая подготовка уже невозможна без программного обеспечения, которое обладает определенным уровнем интеллекта и включает математический аппарат, насыщенный возможностями управления современным оборудованием и технологиями.

Одной из самых сложных задач подготовки производства является проектирование экономичных технологий изготовления, в том числе и программирование оборудования для обработки деталей из современных авиационных материалов [1].

Основная проблема автоматизации проектирования в настоящее время связана не только и не столько с вопросами совершенствования средств вычислительной техники, сколько с тем обстоятельством, что в науке о конструировании новых технических средств не выявлены аналитические и логические зависимости, связывающие назначение технических средств с их структурой и характеристиками. Например, в технологической науке отсутствуют формализованные взаимосвязи между параметрами обрабатываемой детали, структурой и характеристиками технологического процесса.

Общая постановка задачи

Проведенный анализ различных САПР показал два пути передачи информации из конструкторской системы проектирования в технологическую: использование специализированных программ (20 %) или разработка чертежей в конструкторской системе, а затем ввод и обработка данных в технологических системах (80 %).

Используемые на большинстве предприятий автоматизированные системы имеют локальный характер, так как внедрялись разными подразделениями в разное время и для различных целей. В связи с этим возникает проблема ввода в разные системы одной и той же информации. Наиболее отчетливо видна проблема при совместном использовании автоматизированных систем конструкторского и технологического назначения.

На основе проведенного системного анализа подсистем определены наиболее распространенные форматы графических данных, их доля в общем рынке и формализованное описание автоматизированных систем. Установлено, что наиболее используемым форматом является *.dxf, поддерживаемый большинством систем и обладающий рядом преимуществ. На предприятиях имеются электронные архивы чертежей в форматах AutoCAD. В частности, данный формат поддерживается системами AutoCAD и Компас, наиболее распространенными на предприятиях г. Оренбурга.

Цель работы – повышение эффективности процесса передачи в САПР ТП информации, имеющейся на электронном чертеже детали.

Разработанная программа должна обеспечивать полное распознавание графического образа детали и текстовой информации (размеры, отклонения формы и расположения, шероховатость поверхностей, материал детали, технические требования).

Применяемые методы решения задач

В работе использованы методы системного анализа, аналитической геометрии, теории графов, множеств, распознавания образов, баз данных и аппарат полихроматических множеств. При разработке программного обеспечения применялись методы структурного и объектно-ориентированного программирования.

Анализ формата DXF-файла на практических примерах ввода чертежей различных валов позволил установить, что в данном формате нет общего представления о чертеже детали. Он позволяет определить лишь тип и координаты графических объектов, используемых при построении [2].

Для распознавания технологических и производственных данных на чертеже детали необходимо провести анализ модели. При этом особое внимание уделяется анализу геометрических связей и отношений элементов детали.

Идентификацию чертежа, хранящегося в формате DXF-файла, предлагается осуществлять считыванием графических примитивов и распознаванием их с помощью методов аналитической геометрии.

Для функционирования САПР ТП требуется входная информация, которая имеется на электронном чертеже детали. Разработано формализованное описание конструкторско-технологической информации, в котором определен необходимый и достаточный состав входных и выходных параметров для передачи между автоматизированными системами проектирования и технологической подготовки производства.

Осуществлена классификация информации, имеющейся на чертеже детали. Информация, необходимая для передачи из одной системы в другую, разделена на графическую и текстовую.

В разработанную информационную модель детали включены координаты примитивов, их взаимное расположение, конструктивные особенности детали, вид, тип, требования к поверхностям.

Предлагается преобразование информации для последующего использования в автоматизированные системы технологического назначения осуществлять следующим образом. Конструкторско-технологическая информация считывается с файла в формате *.dxf. Чертеж разбивается на примитивы, затем определяется их взаимное расположение с помощью методов аналитической геометрии. По типу линий и взаимному расположению примитивов делается вывод о назначении элемента детали. Конструктивные особенности детали записываются в виде множества, которое преобразуется в объекты. Их состав и описание представляется в виде таблицы базы данных. Для распознавания графических примитивов использованы векторный и структурный подходы теории распознавания образов, что позволило применять параметрическую настройку.

Выделим следующие этапы распознавания типов объектов [2]:

1) Считывание данных из файла, сохраненного в формате *.dxf.
2) Распознавание типов примитивов и их параметров.
3) Структурный поиск распознавания групп примитивов.
4) Определение взаимного расположения примитивов.
5) Определение конструктивных особенностей объектов.
6) Определение типа поверхностей.
7) Регистрация новых обозначений.
8) Идентификация параметров совокупности примитивов.
9) Регистрация новой совокупности примитивов.
10) Вывод отчета.

Деталь – это сложная техническая система, описание которой предполагает ее моделирование. Зрительная модель системы «Деталь», представленная в виде конструкторского чертежа удобна только для восприятия человека. Для распознавания конструкторско–технологических свойств детали используются символьные, информационные, математические, геометрические модели. Общая структурная модель детали описывается выражением [3]:

$$ТСД = \bigcup_{y=1}^{n} Э_y ,$$

где ТСД – техническая система «Деталь», $Э_y$ - символ элемента системы;

$y = \lfloor 1; n \rfloor$ - индекс элемента.

Для описания состава ТСД используются зрительные геометрические модели. Состав системы определяется набором элементов. Сложность системы предопределяется необходимостью использования при её анализе принципа постепенной многоуровневой декомпозиции, выделяя на каждом уровне совокупность элементов, характеризующихся общим признаком классификации.

Для деталей типа тел вращения принимается в качестве исходного элемента конфигурации (геометрического примитива, элемента первого уровня) – цилиндрическое тело. Объединение таких тел образует осесимметричное тело для любой детали этого класса. Другие элементы, являющиеся соосными с исходными (например, фаски, канавки, другие), относятся ко второму уровню элементов, так как они вписаны в тела первого уровня. Остальные элементы отличные от первой и второй группы относятся к элементам более высокого уровня (третьего так далее).

В результате распознавания получим состав элементов детали, описываемые упорядоченными множествами:

$$Э_{вр} = \left\langle {}_{I}Э_{10R},\ {}_{I}Э_{20R},\ {}_{I}Э_{2010R},\ {}_{I}Э_{2011R},\ {}_{I}Э_{30R},\ {}_{I}Э_{3010R} \right\rangle;$$

$$Э_{пл} = \left\langle {}_{I}Э_{10},\ {}_{I}Э_{20},\ {}_{I}Э_{30},\ {}_{II}Э_{2010},\ {}_{II}Э_{2020},\ {}_{II}Э_{2021},\ {}_{III}Э_{3010},\ {}_{III}Э_{3011} \right\rangle.$$

Затем в многомерный массив заносятся данные о распознанных элементах вращения и о плоскостных элементах в составе: параметр (мм); точность, квалитет; верхнее отклонение (мм); нижнее отклонение (мм); допуск (мм); шероховатость (мкм); физико-механические свойства; покрытие; наличие и число покрытий; индексы связанных элементов; признак связи.

Многомерная табличная модель информации о детали позволяет определить связанные элементы и ограничения между ними. В результате построения данной модели устраняется неопределенность взаимного расположения элементов детали внутри занимаемого ею пространства.

Для технической системы «Деталь» важно определить не только состав связей, но и сколько используется разнообразных структур. Число таких вариантов можно оценить выражением:

$$Vr = n^{n-2},$$

где Vr – число вариантов; n – число элементов в структуре взаимосвязанного множества.

Эта величина показывает специфику и сложность решаемой задачи. При выборе возможного варианта необходимо оценить технологичность конструкции и провести оценку структуры связей.

Для дальнейшего моделирования используется база знаний при представлении конструкторских и технологических свойств через полихроматические множества.

Результат работы при программной реализации предлагается сохранять в файл, описывающий тип детали и содержащий координаты распознанных объектов.

Сочетание поверхностей и их признаков необходимы для формирования траектории обработки при проектировании технологии. Этот обменный файл служит средством следующего этапа распознавания конструкторско-технологических свойств детали.

Выводы

В работе получено новое решение актуальной научной задачи – разработана методика распознавания конструкторско-технологической информации с электронного чертежа детали.

Дальнейшие исследования направлены на разработку методики проведения размерного анализа и построения размерной схемы технологического процесса по DXF-файлу.

Литература

1. ADEM в авиационной и аэрокосмической отраслях // САПР и графика. – 2010. – № 1. : Web - сервер журнала «САПР и графика». – Электрон. дан. – М. : Компьютерпресс, 2010. – Режим доступа : http://www.sapr.ru.

2. Шерстобитова, В.Н. Алгоритмы интеграции систем автоматизации конструкторского и технологического проектирования: автореф. дисс. … канд. техн. наук: защищена 04.12.2004: утв. 11.03.2005 / В. Н. Шерстобитова. – Оренбург: ГОУ ОГУ, 2004. – 16 с.

3. Шерстобитова, В.Н. Модель распознавания конструкции детали в условиях ИПИ – технологий: материалы всероссийской научно-практической конференции «Компьютерная интеграция производства и ИПИ-технологии» / В. Н. Шерстобитова. - Оренбург, ИПК ГОУ ОГУ, 2009. – С. 297 – 302.

Методы и средства обеспечения безопасности полетов по упреждающей технологии

Р.К. Мирзаев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Безопасность представляет собой состояние, при котором риск причинения вреда лицам или нанесения ущерба имуществу снижен до приемлемого уровня и поддерживается на этом либо более низком уровне посредством непрерывного процесса выявления источников опасности и контроля факторов риска [1].

В современном этапе развитие военной и гражданской авиации определяется использованием упреждающих технологии, технологией построения пятого поколения самолетов, созданием перспективных авиационных комплексов, в частности, комплексных систем безопасности полетов на основе высоких интеллектуальных информационных систем и технологий [6, 7, 8].

Упреждающие технологии в боевой авиации означает обеспечении превосходства над противником благодаря упреждению его в выработке решений и в действиях. Современная промышленность создала новую систему, чрезвычайно эффективную технологию, которая позволяют полностью нейтрализовать одну из самых совершенных в мире систем противовоздушной обороны и она – систему электронного противодействия (лазерная система) без труда подавить радары системы ПВО и т.д. Недавние конфликты между странами свидетельствует о применении упреждающей технологии – высокоточную лазерную систему.

В настоящее время система управления безопасности полетов (СУБП) предусматривает создание упреждающих технологии по предотвращению авиационных происшествий (АП) [1, 9, 11].

Система управления безопасностью полетов представляет собой системный подход к управлению безопасностью, включая организационные структуры, политику, сферу ответственности, и процедуры. Тремя основными аспектами системы являются:

− ориентированная на безопасность полетов системная деятельность в соответствии с предварительно составленным планом, применяемая в организации на постоянной основе;

− упреждающий подход к определению опасностей и управлению рисками и их исключения;

− характеризующаяся транспарентностью деятельность, связанная с безопасностью полетов.

Система SMS представляет собой упорядоченный подход к обеспечению безопасности полетов и включает необходимые организационные структуры, сферы ответственности, политику и процедуры. В соответствии с положениями Приложений 6, 11 и 14, государства требуют от отдельных эксплуатантов, служб технического обслуживания, поставщиков УВД и сертифицированных эксплуатантов аэродромов внедрения системы SMS на приемлемом уровне, включающей, как минимум, следующие элементы [1,2,9].:

a) определение опасностей, влияющих на обеспечение безопасности полетов;

b) принятие мер по исправлению недостатков в целях исключения рисков/опасностей;

c) обеспечение постоянного контроля и регулярной оценки достигнутого уровня обеспечения безопасности полетов.

Обеспечение БП предусматривает (по СУБП) [1].:

- мониторинг, измерение и анализ показателей БП.
- управление изменения в области БП
- непрерывное совершенствование систем БП.

На основании выше указанных проблем, нами были сформулированы в следующем виде цели исследования и постановки задачи:

1) разработка критериев и методов оценки степени опасности полетной ситуации на борту ЛА в режиме реального времени или с прогнозированием;

2) разработка методологии построения бортовых технических средств обеспечения безопасности полетов на основе упреждающей технологии;

Решение выше перечисленных проблем и оттуда вытекающих задач опираются прежде всего на теорию БП, используемых и вновь разрабатываемых методов, алгоритмов оценивания показателей БП. Исследования, проводимые в области обеспечения БП в последние годы, показывают необходимость существенного изменения и развития научно методических подходов, используемых в данной области. Теория БП отражая многоуровневую систему, включает концепцию, принцип, понятия, критерия, модели, метод и средства обеспечения БП, дают современные знания закономерности функционирования АС, определяющих БП[3].

Основным понятием, используемым в области БП, является понятие авиационного происшествия (АП). Основными методами являются методы теории вероятностей и математической статистики. Основные задачи проводимых исследований решаются на основе статических данных об АП и инцидентах и состоят в оценке вероятности их возникновения и выявления статистических закономерностей влияния на эти оценки различных факторов. [3,4]

В настоящее время требуется дальнейшее развитие теории безопасности и изменение традиционных взглядов на её роль, и место в процессах создания и функционирования сложных АС. По нашему мнению теория безопасности должна создаваться как наука о закономерностях возникновения опасности (угрозы) при функционировании сложных АС, средствах и методах её снижения. БП понимается как возможность отсутствия АП за время полета при осредненной частоте воздействия неблагоприятных факторов в заданных условиях эксплуатации данного типа ЛА. Осредненная частота, проявление неблагоприятных факторов определяется по совокупности реальных или моделируемых полетов с использованием статистических или вероятностных показателей.

В результате этого безопасность отдельного полета определяется исходя из осредненной безопасности в совокупности однородных полетов. Это приводит к стиранию различий между понятиями безопасность полета и безопасность полетов. Наиболее часто используется понятие уровня безопасности и показателя безопасности полетов. При таком подходе обеспечение безопасности понимается как профилактика АП, предполагающая проведение предупредительных мероприятий направленных на снижение вероятности возникновения АП при выполнении совокупности полетов. Вместе с тем, в каждом полете воздействие неблагоприятных факторов на летательный аппарат и экипаж является индивидуальным, и может приводить к различным последствиям в зависимости от состояния ЛА и экипажа, решаемой задачи и реальных условий полета. В этих случаях профилактические мероприятия оказываются малоэффективными.

Для повышения эффективности защиты ЛА и экипажа от воздействий конкретных неблагоприятных факторов в полете необходимо предать системе «экипаж – ЛА» способности вырабатывать защитные управляющие воздействия, адекватные сложившейся полетной ситуации и направленные на снижение её опасности. С этой целью современный ЛА оснащается бортовыми средствами обеспечения БП. [4].

В последнее время такие средства объединяется в комплексные системы ОБП (КСБП).

Назначением КСБП является выработка и реализация защитных управляемых воздействий, направленных на предотвращение АП в процессе полета. Для предотвращения АП в конкретном полете необходимо уметь оценивать опасность (угрозу) полета в каждый момент времени, выбирать наименее опасные способы выполнения полетного задания, а так же активно противодействовать возрастанию опасности в полете.

Поэтому для одного функционирования КСБП в ней должны быть реализованы основные этапы интеллектуального управления, представленные на рис. 1.

Основными этапами создания упреждающих научно-технических задач интеллектуального управления являются:

1. Восприятие информации о системе «экипаж – ЛА- среда» - факторы опасности.
2. Построение информационной модели полетной ситуации - модели ПС.
3. Оценка опасности (риска) ПС.
4. Выработка защитных управляющих воздействий.
5. Реализация защитных управляющих воздействий.

Для решения этого подхода (вышеперечисленных задач) могут быть применены следующие методы БП:

1. Аналитические
2. Вероятностные
3. Статистические
4. Экспертные
5. Нейросетевые.

Факторы и причины, связанные с ним полетные ситуации (ПС), их оценки , показатели и уровни БП выявляются следующими методами:

- **Аналитические** - расчетными методами, т.е. путем расчетов вероятности опасных ситуаций (ОС) возникающих при отказах в функциональных системах ЛА при оценивание уровня БП по фактору «воздушное судно» (ВС) выполняется разработчиком на этапе разработке, при оценке соответствия уровне задаваемый к нормированному на этапе производства и сертификация типа ВС. Приемлемые уровни БП должен поддерживаться на этапе эксплуатации[3].

- **статистическими** методами при апостериорных оценивании и прогнозировании уровни БП по результатом испытаний и эксплуатации. Известно , что АП редкое явление для оценивания уровня БП и факторов риска АП необходимо большой статистический материал. Этот метод худшее работает, когда ЛА новый и у него еще не набран нароботки и другие параметры, необходимые для количественного оценивания показателей БП[3,4].

- **экспертные** методы, используемые при априорном оценивании и прогнозировании уровня БП по заключению специалистов. При эксплуатации новой авиационной техники для ограниченного количество малосерийных типов ЛА , когда нет возможности использовать статистические методы. Наиболее приемлемыми и доступными для априорного оценивания уровня БП оказывается экспертные как основное методы качественного и количественного оценивания состояния сложных динамических систем при отсутствии описания закономерности динамики этих систем в виде аналитических зависимостей[5,8].

- **нейросетевые** -Методы искусственного интеллекта (и т.д.) применяется при разработке информационно-аналитических систем управления уровнем безопасности [8].

Кроме того при расчете различных статистических показателей БП предлагается **метод простого скользящего среднего**, поскольку именно этот метод используется при формирование требований к уровни БП государства в глобальном плане обеспечения БП ИКАО. Выбор того или иного метода ОБП осуществляется по критерию снижения уровни БП.

Современные технические средства и технологии – персональный компьютер, средств автоматизации и информационные технологии при мощных методах, алгоритмах и программных обеспечениях позволяет построить интеллектуальные системы различного назначения, в частности:

– бортовые интеллектуальные информационно – измерительные системы ;

– бортовые интеллектуальные информационно – управляющие системы;

– бортовые интеллектуальные информационные системы принятия решений и т.д.

XXI век может стать переходом к искусственному интеллекту в управлении. Однако существующие интегрированные комплексы и системы бортового цифрового оборудования нельзя считать комплексами, построенными на основе искусственного интеллекта, т.к. средства информационной поддержки принятия решения человеком

(СИППР) отсутствуют. Анализ подобных систем показывает, что в настоящее время пока не созданы бортовые информационно- управляющие системы (БИУС) с системой принятия решений (СИППР), построенные на базе искусственного интеллекта [4].

В настоящее время уже существует несколько поколений базовых интегрированных комплексов авионики (КСЦПНО 204/96 (I), ИКБО – 95 (II), ИКБО – 2005 (III), ИКБО – 2010 (IУ)).

Для перспективных комплексов III и последующих поколений вопрос обеспечения требуемым объемом памяти Q и быстродействия F не является ограничением с точки зрения принятия решения об архитектуре комплекса и распределения задач внутри него, не увеличивая при этом массу M и потребляемую мощность P.

Однако принятая сегодня идеология построения комплекса как системы резервированных обособленных функциональных систем, взаимодействующих между собой по жестко определенным протоколам , не позволяет в полной мере использовать вес комплекс вычислительных ресурсов , присущих современному комплексу, ограничивая показатели надежности возможностями микроэлектронной технологии. Необходимо разработать принципиально новые архитектуры интегрированных комплексов авионики ,основанные на многократных возросших ресурсах бортовой измерительно – вычислительной техники и новых технологиях проектирования и производства систем.

Все это должен быть реализовано на единой технологической базе с внедрением унифицированного модульного принципа во все цифровые системы и блоки.

Исследование по ЧФ, проведенные техническим комитетом ИКАО, причин летных происшествий показали, что при создании АТ практически не учитывались законы психической деятельности летчика, неожиданно попадающие в нештатную ситуацию, 82% тяжелых летных происшествий «по вине экипажа» является следствием перехода психики пилота в десинхронизированные режимы функционирования[10].

Таким образом, встает вопрос в необходимости с одной стороны глубокой аппаратно- программной интеграции оборудования и с другой, проведения эргономических и психофизиологических исследований с целью оптимизации объема и характера выдаваемой экипажу информации и сигнализации.

В мировой и отечественной практике, применительно к авионике, все шире используется понятие «сложная интегрированная система». Проектирование и сертификации сложных интегрированных систем авионики широко используется на западе и активно внедряется в отечественных разработках.

Большой опыт по сертификации сложных интегрированных систем имеет концерн AIRBUS. Процесс обеспечения безопасности/надежности реализованной на AIRBUS при создании самолета А-380, аналогичен порядку и процедурам, применяемым в отечественной практике создания и сертификации сложных интегральных бортовых систем АТ, в основу которых положен анализ функциональных отказов систем.

Однако при рассмотрении систем обеспечения БП А380 по [10] , процесс обеспечения безопасность/надежность в программе AIRBUS (по материалом J.P.Hechmann) состоит из анализа общих причин (CCA) (8 – пункт):

– анализ особых рисков, т.е. анализ влияния внешних факторов, которые могут оказывать влияние на несколько зон самолета: на двигатель, шоссе, и т.д. Некоторые из этих рисков служит одним из требований норм летный годности. Примером особых рисков могут быть взрыв или пожар двигателя, разрыв шины и т.д.

– зональный анализ - включает рассмотрение зон самолета с точки зрения размещения оборудования и снижения возможных рисков взаимного влияния бортовых систем;

– анализ общих режимов – включает доказательство независимости отказов (фактор ТС);

– анализ влияния человеческого фактора (анализ влияния ошибки летного экипажа и т.д.) [9].

Заключение

Таким образом, как показывает отечественный и мировой опыт, рассмотренные методы и средства обеспечения БП на основе бортовых ИИС реального времени на основе перехода от системного подхода к процессному позволяет успешно решить задачу управления уровнем БП ЛА. Использование искусственного интеллекта в базе интегрированных комплексов и систем бортового цифрового оборудования воздушных судов нового поколения может снять некоторые стандартные нагрузки экипажа при управлении в нормальных полетных условиях.

Литература

1. Руководство по управлению безопасностью полетов (РУБП) (Doc 9859-AN/460). Издание первое — 2006 год.— ИКАО, 2006.
2. Руководство по предотвращению авиационных происшествий. (Doc 9422-AN/923). Первое издание — 1984 год.— ИКАО, 1984.
3. Жулев В. И., Иванов В. С. Безопасность полетов летательных аппаратов: (Теория и анализ).— М.: Транспорт, 1986.— 224 с.
4. Федоров С. М. Михайлов О. И., Сухих Н. Н. Бортовые информационно-управляющие системы: Учеб. для вузов / Под ред. С. М. Федорова.— [illegible]-' Транспорт, 1994.— 262 с.
5. Гузий А. Г., Чуйко А. А. Методологический подход к экспертному прогнозированию уровня безопасности полетов.// Проблемы безопасности полетов./ Научный информационный сборник. Вып. 10, 2006.— М.: ВИНИТИ, 2006.
6. ГОСТ Р 51898-2002 Аспекты безопасности. Правила включения в стандарты.С,
7. Руководство по информационному обеспечению автоматизированной системы обеспечения безопасности полетов воздушных судов гражданской авиации Российской Федерации (АСОБП).— М.: Аэронавигационное консалтинговое агентство, 2002.
8. Гладкин С. М., Гузий А. Г., Онуфриенко В. В. Нейросетевой подход к разработке интеллектуальной информационно-аналитической системы управления уровнем безопасности полетов.// Проблемы безопасности полетов./ Научный информационный сборник. Вып. 9, 2007.— М.: ВИНИТИ, 2007.
9. Алексеев Э.П., Евгепов А.В., Перчаткин М.П. Интегрированный комплекс авионики ИКБО-95 / Авиакосмическое приборостроение, № 2, 2003.
10. Доклад Heckmann J.P. (Airbus) на заседаниях рабочих групп SAE и EUROCAE.
11. СУБП. Отдел производства полетов Международный аэропорт Бангалор (BIAL) – Индия, Сентябрь 2007 г.

Автоматизированная система обработки полетной информации: проблемы и перспективы

Р.К. Мирзаев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В современных условиях развитие нового поколения летательных аппаратов (ЛА) в военной и гражданской авиации предусматривает ускоренного внедрения как информационных технологий и телекоммуникаций, так и единую комплексную систему обеспечения безопасности (КСБП) на базе датчиковых аппаратур, пилотажно-навигационных устройств и приборов, бортовых ИИС и вычислительных комплексов, информационно-управляющих систем обеспечения безопасности полетов нового поколения, современных микропроцессоров, а также программных средств и в будущем использовать технологию с «открытой архитектурой».

Современные зарубежные военные самолеты (по источникам зарубежных военных обозрений), использующие геоинформационную технологию, оснащены электронным оборудованием с «открытой архитектурой», сверхмощными специальными компьютерными системами с программным обеспечением (на языке «АДА») и могут нести точечные удары, контролировать и управлять свое пространственное положение через спутниковый связь, передавать визуальные изображение местности в реальном масштабе времени и тем самым наземные системы полностью контролируют динамику полета, состояние ЛА, действие экипажа. Данные о параметрах полета также может быть записаны в «черный ящик» Вся информация о системе – техническое состояние, психофизиологическое и функциональное состояние пилота и динамика полета летательного аппарата – могут сохраниться в зарегистрированных сигналах вышеуказанных устройств и систем.

Несмотря на это, проблемы авиационной безопасности, то есть ее обеспечение, становятся первостепенной важности, и это касается не только систем и средств управления безопасности (системы предупреждения критических моментов), но и систем и средств обработки и анализа полетной информации, измеренных и полученных с помощью бортовых устройств регистрации средств объективного контроля полетов. (МСРП, ТЕСТЕР, БУР, из зарубежных аналогов – FDR , DFDR и т.д.), которые в зависимости от модели и типа самолета записывают от 6 до 88 параметров. При этом основными параметрами, определяющими безопасность полета самолета на режимах разбега и взлета, набора высоты и крейсерского полета, маневрирования и снижения, посадки и пробега, а также при бомбометании, стрельбе и управлении ракетным вооружением являются угол атаки, крен, тангаж, приборная воздушная скорость и число М, нормальная перегрузка и высота полета и может быть зарегистрирован в основном на ленточные накопители с магнитным принципом записи и на твердотельные накопители с электронным принципом записи (флэш-память) информации.

Необходимость перехода на новую технологии при построении КСБП, а именно на твердотельных накопителях полетной информации и другие технологии (ARING, новые технологии технического стандарта TSO администрации FAA) подталкивается постоянно ужесточающимися стандартами Международной организации гражданской авиации (ИКАО) и требованиями зарубежных заказчиков.

Статистика авиационных происшествий показывает, что в основном они происходить по вине летчика (75%), из-за неисправности и отказа оборудования(20%) и по другим причинам. Как показывает анализ мировой статистики летных происшествий

(ЛП), более 80% всех причин ЛП обусловлено человеческим фактором (ЧФ), а именно ошибками летных экипажей и лиц групп руководства полетами (ГРП) из служб управления воздушным движением (УВД). При этом меньшая доля ЛП вызвана ошибочной оценкой ситуации, неверными прогнозами и расчетами. Основная же часть ЛП связана с ошибками летчиков в технике пилотирования и отсутствием надлежащей и своевременной реакции на усложнение условий полета(см. табл. 1).

За 10 лет общие потери всей государственной авиации Российской Федерации составили 301 воздушное судно, при этом погибли достаточное количество человека. Относительный показатель (число авиационных происшествий на 100 тыс. часов налета), характеризующий уровень аварийности, в течение 30 лет находится на уровне 4 – 5 авиационных происшествий на 100 тыс. часов налета, в то время как в ведущих авиационных державах данный показатель в 2 и более раз ниже

В настоящее время ежегодный материальный ущерб, наносимый Российской Федерации авиационными происшествиями в государственной авиации, составляет более 2 млрд. рублей, а к 2015 году при сохранении существующей системы обеспечения безопасности полетов может достичь 30 млрд. рублей в год [2].

Таблица 1

	Отказы авиационной техники	Человеческий фактор	Неблагоприятные внешние условия
Аварии	41%	48%	11%
Катастрофы	6%	84%	10%

Такому положение дел привело:

- несовершенство законодательных актов и нормативные положения по ОБП;
- принятие управленческих решений в условиях отсутствия полной и достоверной информации о состоянии элементов авиационной системы и особенностях их взаимодействия в процессе организации, подготовки и выполнения полетов;
- конфликт интересов при проявлении опасных факторов;
- несоответствие выделяемых ресурсов (материально-технических, финансовых, административных, организационных, кадровых, информационных) масштабности и сложности задач обеспечения безопасности полетов;
- объективное снижение научно-производственного потенциала авиационной промышленности России;
- взгляды не отвечают требованиям времени на вопросы безопасности полетов.

Один из вариантов решения вышеуказанных проблем связана, в первую очередь, автоматизированной системой обработки полетной информации (АСОПИ), осуществляемой на базе бортовых измерительных систем полетной информации (БИИС) – САРПП, МСРП, БУР и т.д. и используемой в качестве контролирующей системы обеспечения БП. В настоящее время возникает на этой основе другая проблема – проблема ее использования непосредственно на борту – упреждающий подход для экипажа (до настоящего времени они используется в наземных условиях эксплантации).

В настоящее время на борту самолета успешно используются информационно-управляющие системы (ИУС) ОБП.

Недостатки существующих средств обеспечения безопасности полета

Запоздалое срабатывание систем ограничительных сигналов (СОС) и ограничителей предельных режимов (ОПР) при потере скорости на крутых траекториях.

Отсутствие прогноза динамики движения.

Отсутствие четких рекомендаций по принятию решения для вывода из сложных ситуаций

Выше перечисленные проблемы представляют серьезную научно-техническую задачу и ее решение носит первостепенной важности.

На основании выше указанных проблем, нами были сформулированы в следующем виде цели исследования и постановки задачи: разработка методологии построения бортовых технических средств обеспечения безопасности полетов на основе упреждающей технологии. В этом направлении нами удалось решить ряд важных моментов – разработаны классификация БП с учетом различных факторов опасности, методы, модели, структуры СУБП реального времени.

В настоящее время системы управления безопасностью полетов (СУБП) в гражданской авиации на основе Стандартов и Рекомендуемой практики (SARPS) Международной организации гражданской авиации (ICAO) активно внедряются ведущими авиационными державами как эффективный инструмент снижения аварийности. Понятие СУБП введено в 2006г. ИКАО, СУБП должен разрабатываться Эксплуатантом (касается всех участников Чикагской Конвенции) (в России еще не разработана) и должен внедряться с 2009 г. [1].

В современных условиях целенаправленная деятельность по выявлению, оценке и устранению опасных факторов или уклонению от них с целью предотвращения авиационных происшествий в государственной авиации может быть эффективна только в рамках единой системы управления безопасностью полетов.

Государственная Программа по ОБП (разработана в 2008г.) и Концепция ОБП (разработана по поручению Президента и Правительства Российской Федерации от 22 апреля 2009 г. N 554-р) учитывает вышеописанные проблемы и указывает на срочное решение этих проблем с 2010 -2015 г, где с целью повышения БП представлены целый ряд проблем, касающиеся разработке и **внедрения интегрированных бортовых интеллектуальных информационно-измерительных и управляющих систем** – системы постоянного многоуровневого мониторинга безопасности полетов.

XXI век может стать переходом к искусственному интеллекту в управлении. Однако существующие интегрированные комплексы и системы бортового цифрового оборудования нельзя считать комплексами, построенными на основе искусственного интеллекта, так как средства информационной поддержки принятия решения человеком (СИППР) отсутствуют. Анализ подобных систем показывает, что в настоящее время пока не созданы бортовые информационно-управляющие системы (БИУС) с системой СИППР, построенные на базе ИИ. Лишь в отдельных публикациях можно встретить упоминание о подобных системах.

Технология построения архитектуры систем и блоков - объединение приборного (АО), радиотехнического (РЭО) и пилотажно-навигационного оборудования (ПНО) в единое целое с цифровым двоичным исполнением по «открытой архитектуре», информационно связанных радиальными линиями связи (ARINC 429 и т.д.), с кратностью резервирования, определяемой категорией критичности системы и выполняемой ею функции, позволило выпускать промышленностью различных поколений АВИОНИКИ – интегрированных комплексов бортового оборудования (КСЦПНО-204/96(I), ИКБО-95 (II), ИКБО – 2005 (III), ИКБО – 2010 (IУ)) [3, 4]. Однако, поколений АВИОНИКИ не решает задачу ОБП.

В современных условиях развитие нового поколения летательных аппаратов (ЛА) предусматривает ускоренного внедрения как информационных технологий и телекоммуникаций, так и бортовых ИИС статистического анализа параметров полета и информационно-управляющих (ИУС) систем обеспечения безопасности полетов, их интеграцию – в единую систему обеспечения безопасности на базе современных

микропроцессоров, а также программных средств и в будущем использовать упреждающую технологию.

В качестве бортовых ИИС статистического контроля параметров полета настоящее время по выполняемым функциям могут быть **автоматизированная система обработки полетной информации (АСОПИ)** на базе бортовых измерительных систем полетной информации (БИИС) – САРПП, МСРП, БУР и т.д.

Рассмотрим **автоматизированную систему обработки полетной информации (АСОПИ)** на базе бортовых измерительных систем полетной информации (БИИС) – САРПП, МСРП, БУР и т.д.

В настоящее время для расследования авиационных происшествий (АП), для предполетного и после полетного контроля состояния ЛА применяются бортовые измерительные системы (БИИС) - системы типа САРПП (САРПП -12 ГМ, ДМ и т.д.), МСРП (МСРП-12-96, МСРП-64 и т.д.), БУР (БУР1-3,БУР-4-1и т.д.), Тестер-У3 (сер.2, сер.3, сер. Л), К3-62, МС-61, П-503 , позволяющие измерять и регистрировать большое количество аналоговые, цифровые и бинарные сигналы соответствующих параметров, характеризующих пространственное положение воздушных судов, действия экипажа и работу авиатехники, а также речевые сигналы – переговоры экипажей [3, 5].

Эти средства позволяют автоматически получить информацию непосредственно от объекта, измерять, преобразовать ее, передать, накопить и сохранить необходимую информацию об условиях полета, параметрах движения и состояния ЛА, техническом состоянии его силовых установок и оборудования, действиях экипажа по управлению ЛА. Бортовые устройства регистрации только регистрируют необходимые параметры, но не позволяют осуществить анализ их на борту ЛА.

Анализ параметров осуществляется после полета в процессе проведения экспресс – оперативной обработки.

Своевременный и полный анализ полетной информации бортовых ИИС (БУР) существенно способствует **повышению безопасности полетов** благодаря использованию этой информации в целях:

- **оценки режимов полета и правил летной эксплуатации ЛА;**
- **оценки работоспособности авиационной техники;**
- **оценки и определение причин авиационных происшествий и инцидентов.**

На современном этапе данные, получаемые с помощью этих средств, считаются полуфабрикатом из-за отсутствия четкого анализа этой информации с точки зрения измерения статистических характеристик (математическое ожидание, дисперсия, корреляционная функция), анализа погрешности измерений, описание закономерностей и модели процесса как с восстановлением, так и без восстановления сигнала.

Полетные данные, полученные и измеренные датчиками, преобразованные в цифровые сигналы с помощью АЦП и других устройств, записанные в твердотельные накопители - микропроцессорные запоминающие устройства, представляют собой дискретные случайные сигналы (процесс и поток, временной ряд) как с равномерным, так и неравномерным интервалом дискретизацией, которые имеют два массива значений: значение сигнала и момент времени.

Полетную информацию можно рассматривать как нестационарный случайный процесс – временной ряд, с целью ее анализа следует ее исследовать – измерить статистические характеристики, на основе ее разработать модели процесса, описывающей этот временной ряд, на основе статистического анализа временного ряда определить погрешность, законы их проявления, тем самым идентифицировать (распознать объект) и прогнозировать (состояние объекта). Кроме того, с целью обеспечения безопасности полетов, использование этих данных в оперативном порядке в процессе управления ЛА – своевременный прогноз о полете (о динамике движения

полета, режиме, этапе и определение критических моментов по установленным критериям и правилам и т. д.) и предупреждение об отказе (оценка состояние авиационной техники) по полетной информации – результаты обработки статистических данных делает процесс исследования актуальным и целенаправленным.

В настоящее время в военной и гражданской авиации применяемые методы, средства и алгоритмы, программы обработки («Monstr», “WinArm”, и т. д.) не позволяют анализировать в реальном масштабе времени и они осуществляются только в наземных условиях, т. е. обработка полетной информации осуществляется после приземления летательного аппарата. Исследования по дисплеям - по визуальному осмотру восстановленных непрерывных аналоговых сигналов критических моментов (чрезвычайные ситуации), отказ, выход за пределы эксплуатационных ограничений (ЭЛО) и этапы полетов (ЭП), тем более оценка технического состояния ЛА затруднена как из-за отсутствия методов, алгоритмов и средств оперативной обработки полетной информации, так и из-за отсутствия некоторых важных первичных преобразователей информации, например датчиков механических напряжений и деформации.

Эти проблемы могут быть решены при оборудовании самолета контролирующими по безопасности полетов средствами – автоматизированной информационной системой обеспечения безопасности (АИСОБ) на основе бортовых комплексов сбора и анализа полетной информации и наземной телеметрии, создаваемых с помощью современных информационных систем и технологии, использующими статистические методы, алгоритмы и средства обработки полетной информации.

Выводы

В результате внедрения информационной технологии в ЛА нового поколения встал вопрос **интеграции** т.е. объедение БУР и наземной обработки, т.е. на основе автоматизированной системы обработки полетной информации (АСОПИ) осуществляемой на базе бортовых измерительных систем полетной информации (БИИС) – САРПП, МСРП, БУР и т.д., тем самым создания бортовых информационно-измерительных систем специального назначения (кроме АВИОНИКА) – бортовую автоматизированную систему контроля состояния БП . Решение этой задачи – вопросы обеспечения безопасности полетов, построения комплексной автоматизированной информационной системы ОБП (мониторинг БП) могут быть на основе комплексного подхода – с учетом классификации факторов опасности (ЧФ, ТФ, ВФ), с учетом выше перечисленных систем (БИИС, ИУС). Это позволит обеспечить снижение рисков авиационных происшествий, чрезвычайных ситуаций и потерь человеческого, природного, экономического и оборонного потенциалов, создание условий для устойчивого развития государственной авиации, достижение показателей безопасности полетов, соответствующих уровню передовых авиационных держав.

Литература

1. Руководство по управлению безопасностью полетов (РУБП) (Doc 9859-AN/460). Издание первое — 2006.— ИКАО, 2006.

2. Федоров С. М., Михайлов О. И., Сухих Н. Н. Бортовые информационно-управляющие системы: Учеб. для вузов / Под ред. С. М. Федорова. — М.: Транспорт, 1994.— 262 с.

Подсекция 3
«Информационные технологии в образовании»

Интернет система мониторинга учебного процесса по информационным специальностям

И.Ф. Батаев, В.В. Козлов
Самарский государственный архитектурно-строительный университет, Самара

Современная система высшего образования предполагает высокий уровень информационной нагрузки на студента и с каждым годом эта нагрузка растет, что требует постоянного контроля за уровнем усвоения материала. В нынешнее время уже не обойтись итоговым экзаменом в конце семестра. Итоговый экзамен часто уже не дает возможности исправить положение, если материал был усвоен не достаточно, поэтому требуется вести непрерывный мониторинг качества знаний в течении семестра.

Кроме того, вузам предоставляется достаточная свобода для улучшения качества подготовки специалистов не только путем уточнения дисциплин, изучаемых в вузе, но и путем их оптимального расположения во времени. Рабочие учебные планы составляются ежегодно и вузам предоставляется возможность корректировать в определенных пределах объемы изучаемых дисциплин, содержание и структуру образования. Таким образом, просто необходимо иметь данные о успеваемости студентов. Имея данные по каждому предмету, факультет имеет возможность оптимизировать учебный процесс, правильно расположив дисциплины в сетке занятий.

Для своевременного контроля и принятия мер на факультете ФИСТ кафедры ПМиВТ Самарского архитектурно-строительного университета в 2007 году была введена система промежуточной аттестации по контрольным точкам, что позволило своевременно принимать меры по улучшению качества усвоения материала, и позволило прогнозировать итоговую оценку студента по дисциплинам учебного плана в конце семестра. Изначально результаты тестирования заносились в Excel-файл, где и происходила их обработка. По мере накопления данных начали возникать сложности. Большой объем файлов, выложенных на сайте факультета, не полная поддержка браузерами Excel-листов в формате web-страниц – вот только некоторые из возникших сложностей.

Поэтому нами разработана следующая версия системы на языке PHP и JavaScript. Результаты аттестации хранятся в базе MySql. Система промежуточной аттестации по контрольным точкам подразумевает прохождение студентом шести тестов по пройденному материалу по каждой дисциплине в течение всего семестра через приблизительно равные промежутки времени.

Результатом прохождения контрольной точки является оценка, выставляемая преподавателем по шкале от 0 до 2.

В данном случае оценка 0 подразумевает, что студент не прошел контрольную точку или получил не удовлетворительный результат, 1 подтверждает прохождение контрольной точки, а оценка 2 означает прохождение контрольной точки на отлично. Прохождение всех контрольных точек на отлично дает возможность студенту получить «автомат» по предмету.

Система аттестации и мониторинга на основе контрольных точек была опробована и оптимизирована за период с 2007 года, были уточнены количество контрольных точек, студенты и преподаватели привыкли к нововведению.

Для студентов были введены как меры поощрения, так и наказания. В начале семестра студентам «начисляется» тридцать баллов. Далее в процессе учебы к «счету» студента прибавляется процент контрольных точек сданных на «отлично» и вычитается процент пропусков. При уменьшении рейтинга до нуля студент отчисляется из университета.

Студенты привыкли к рейтинговой системе и среди них возникло дружеское соперничество за обладание высоким рейтингом, поэтому на факультете каждый семестр по результатам рейтинга называется лучший студент семестра.

Автоматизированная информационная система управления учебным процессом на факультетах СГАУ

М.С. Глянько, Д.С. Оплачко, Е.И. Чигарина
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Проблема обеспечения комплексной информационной поддержки развития управления учебным процессом по-прежнему является одной из актуальных задач информатизации факультета. Разработка современной автоматизированной информационно-аналитической системы управления учебным процессом на факультете является исключительно сложной задачей, требует привлечения больших материальных и интеллектуальных ресурсов, применения самых современных информационных технологий. Во многих ВУЗах существуют программные комплексы, в той или иной степени решающие задачи, относящиеся к этой сфере их деятельности, тем не менее, в настоящее время эта проблема далека от полного решения.

В последние годы в СГАУ наблюдается интенсивное развитие различных информационных ресурсов: телекоммуникационных сетей различного уровня, организационной и методологической инфраструктуры, содержательного наполнения – баз данных, программных средств поддержки работы подразделений и образовательного процесса ВУЗов. Появилась необходимость дальнейшего качественного улучшения информационного и телекоммуникационного обеспечения системы образования путем автоматизации документооборота и обеспечения информационной поддержки организационной деятельности ВУЗа.

Одним из главных факторов повышения эффективности деятельности любой организации, в том числе и ВУЗа, является использование автоматизированной информационной системы. Эффективность в данном случае повышается за счет использования современных информационных технологий для совместной работы, решения задач по отражению учебных планов, учебных графиков и расписания, своевременному отслеживанию выплат студентам, оперативному доступу к текущей успеваемости.

Для целей автоматизации делопроизводства и основного документооборота деканата, обеспечения оперативного получения информации всеми заинтересованными лицами путем доступа к базе данных, разрабатывается автоматизированная информационная система «АИС Деканат». «АИС Деканат» – единая информационная среда, предназначенная для хранения информации о студентах, кафедрах факультета ВУЗа, преподавателях, организации учебного и воспитательного процесса. Система позволяет принимать управленческие решения на основе полной, достоверной и оперативной информации о деятельности структурных подразделений факультета и студенческого контингента.

Принятие решения об отказе от использования систем сторонних производителей было обусловлено необходимостью доработок в соответствии со спецификой работы деканатов СГАУ, определенной формой документов и отчетов при управлении учебным процессом на факультете. Еще одним фактором в пользу разработки своей системы является то, что в случае использования системы от сторонних разработчиков возникает необходимость постоянной поддержки системы производителем, в противном случае система утрачивает свою актуальность уже через небольшой срок.

Информационная система «АИС Деканат» должна обеспечивать автоматизацию основных этапов работы деканата во время учебного и воспитательного процессов (см. рис. 1):

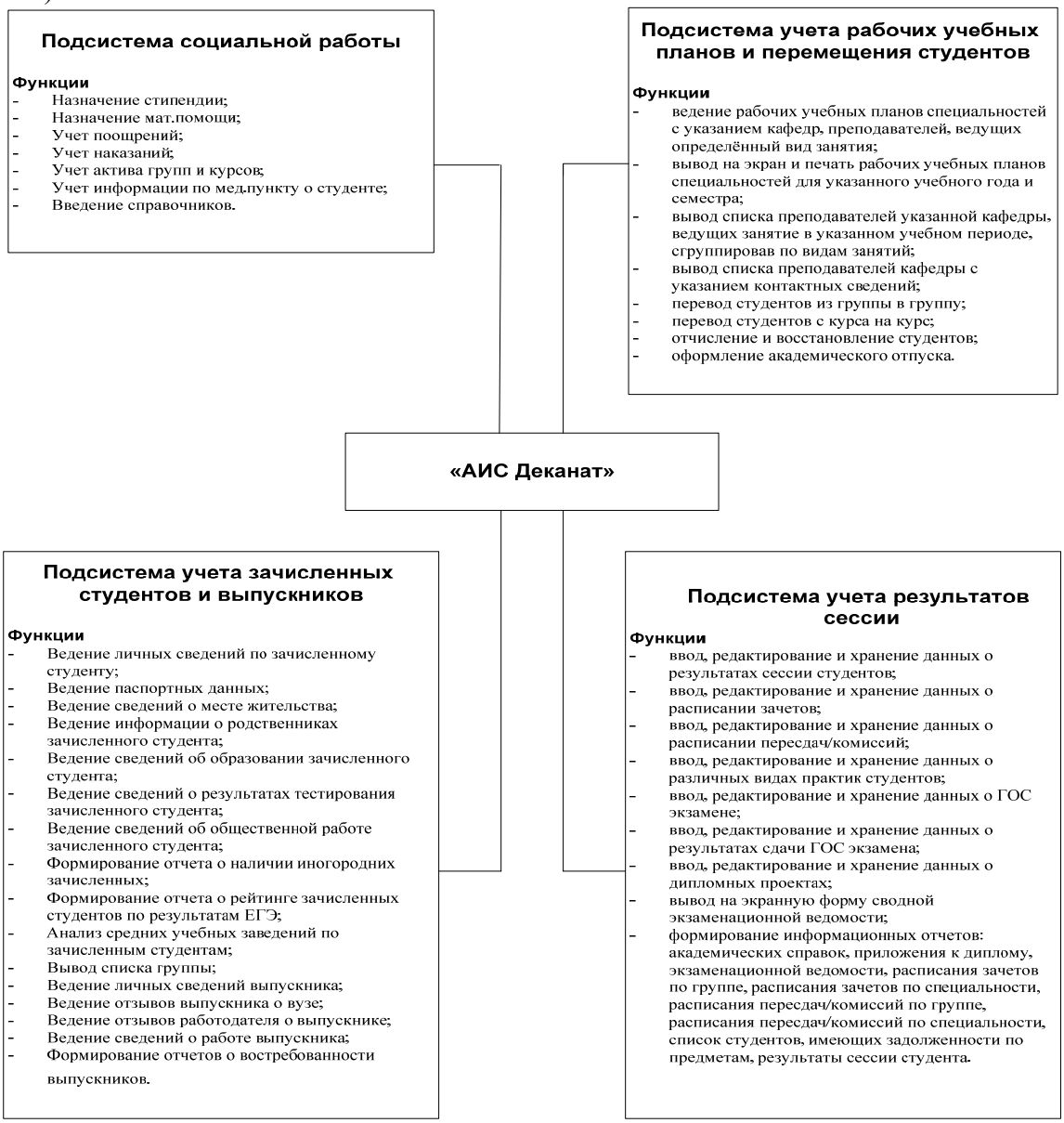


Рис. 1. Основные функции «АИС Деканат»

- зачисление студентов на первый курс;
- формирование групп по специальностям;
- работа, связанная с движением студентов (перевод из группы в группу, с курса на курс, отчисление, восстановление, оформление академического отпуска);
- ведение учебной карты студента;
- организация сессии (заполнение результатов сессии, составление расписания зачётов, расписания преподавателей);
- учёт рабочих учебных планов;
- работа с выпускниками (выпускная работа, государственный экзамен);
- социальная работа (назначение стипендии, матпомощи, поощрения, наказания для студентов);
- общественная работа и обеспечение связи с медпунктом;
- формирование отчётов и приказов.

Для реализации системы в качестве СУБД была выбрана MS SQL Server, средства разработки – Borland Developer Studio 2006, технология доступа к данным – ADO.

Схема базы данных включает около семидесяти сущностей. Подсхема базы данных, отображающая сведения о студентах, результатах сессии представлена на рис. 2.

Предполагается архивирование данных из базы данных каждые пять лет во избежание разрастания базы данных до очень больших размеров, что может сильно сказаться на производительности системы в целом.

Разрабатываемая система реализует клиент-серверную архитектуру, в которой на сервере хранится база данных, а клиентское приложение установлено на компьютере пользователя и выполняет все основные действия по обработке и представлению данных. Следовательно, в данном случае тип клиента – толстый, или Rich-клиент.

В системе предусмотрено разграничение прав доступа пользователей. Пользователи разделены на следующие группы:

- работники деканатов (декан, заместители декана, начальники курсов);
- оператор деканата.

Для каждой группы пользователей разработан собственный интерфейс и определено множество допустимых функций системы, решены вопросы авторизации и аутентификации. В системе реализованы ограничения целостности данных. При заполнении рабочих учебных планов реализованы два способа ведения – с применением визарда и через сводную форму. Кроме наличия меню в системе представлена древовидная структура всех функций, реализованных в системе. При возникновении ошибочных ситуаций система выводит информативно понятные сообщения. Вся терминология интерфейса понятна всем группам пользователей системы.

В настоящее время система «АИС Деканат» находится на стадии тестирования в деканате факультета информатики СГАУ им. академика С.П.Королева.

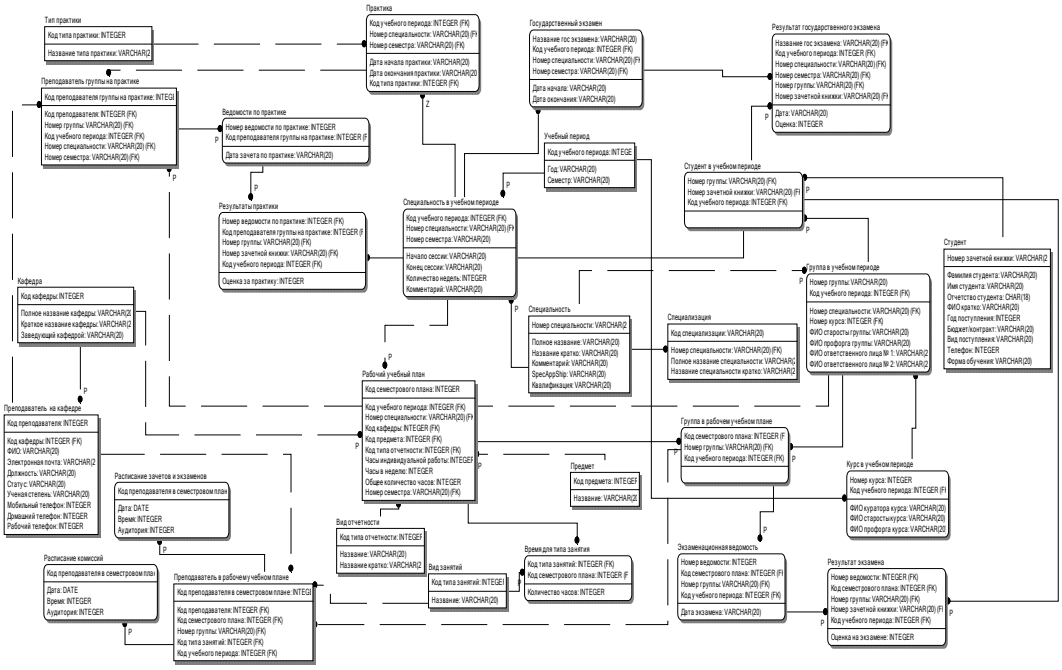


Рис. 2. Подсхема базы данных подсистемы учета рабочих учебных планов «АИС Деканат»

Опыт обучения информационным технологиям в условиях зарубежной сетевой академии

А.В. Горшков, У.М. Тулемисов
ФГУП «Системпром», Москва

Введение

Обучение специалистов в разных форматах (образование, повышение квалификации, профессиональная подготовка) позволяет решать важные социальные задачи как для тех, кто работает в сетевых компаниях, так и для тех, кто пользуется их услугами. В наши дни международная компания, производящая полный спектр сетевого оборудования для нужд различных стран и отраслей, может и должна иметь трехуровневую модель обучения [1]:

- специалист
- профессионал
- эксперт

Уровень специалиста обеспечивает базовые знания в области построения сетей, что соответствует навыкам и знаниям, необходимым для проектирования и внедрения сетей до 250 пользователей. Обучение занимает несколько недель.

На уровне Профессионал осуществляется переход от общего курса в более специализированную область, сфокусированную на определенном аспекте построения сетей передачи данных: Безопасность, Коммутация и Маршрутизация, Сети провайдеров связи и пр. Обучение не возможно без практической работы и рассчитано на несколько месяцев.

Уровень эксперта характеризуется высочайшим уровнем знаний и опыта в рамках областей определенных на уровне Профессионала. Для достижения такого уровня необходимо учиться и работать несколько лет.

Автору удалось накопить богатый многолетний опыт преподавания ИТ в рамках первого и второго уровня вышеописанной трехуровневой модели обучения как в России, так и за рубежом. В данном конкретном случае будет рассмотрено обучение специалистов в условиях академии AMES (Окленд, Новая Зеландия).

Характеристики потока обучаемых в академии сетевых технологий

За период преподавания с 2008 по 2010 в AMES IT Academy (Auckland, New Zealand) были собраны статистические данные примерно о 200 студентах. В состав статистических данных входит:

- Имя
- Фамилия
- Пол
- Возраст
- Национальность
- Образование
- Соответствие образования профилю курса
- Результаты тестирований за 4 семестра (в общей сложности более 70 различных тестов)

На основе вышеприведенных данных были полученные следующие результаты:

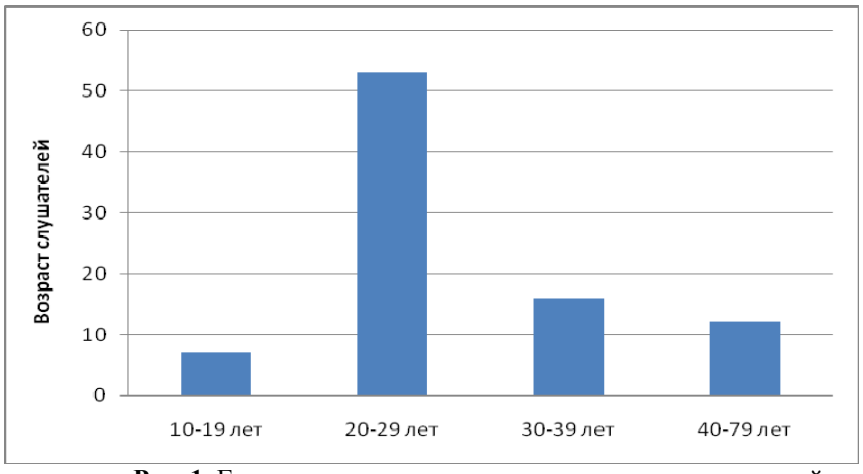


Рис. 1. Гистограмма возрастного распределения слушателей

Таблица 1. Возрастное распределение слушателей

Среднее	28.64773
Стандартная ошибка	0.91001
Медиана	27
Мода	28
Стандартное отклонение	8.536655
Дисперсия выборки	72.87448
Эксцесс	2.06222
Асимметричность	1.274141
Интервал	46
Минимум	16
Максимум	62
Сумма	2521
Счет	88
Наибольший(1)	62
Наименьший(1)	16
Уровень надежности (95.0%)	1.808744

Как видно из данной гистограммы и таблицы 52% слушателей составляют люди возрастной группы 20-29 лет

Таблица 2. Распределение результатов тестирования

Первый семестр		**Второй семестр**		**Третий семестр**		**Четвертый семестр**	
Среднее	90.96	Среднее	90.30	Среднее	87.48	Среднее	77.23
Стандартная ошибка	0.78	Стандартная ошибка	1.34	Стандартная ошибка	2.15	Стандартная ошибка	2.32
Медиана	92.3	Медиана	93.25	Медиана	92.35	Медиана	82.8
Мода	100	Мода	100	Мода	100	Мода	96
Стандартное отклонение	7.34	Стандартное отклонение	12.20	Стандартное отклонение	19.47	Стандартное отклонение	20.52
Дисперсия выборки	53.91	Дисперсия выборки	148.97	Дисперсия выборки	379.23	Дисперсия выборки	421.27
Эксцесс	7.49	Эксцесс	37.33	Эксцесс	13.59	Эксцесс	3.68

Асимметричность	-2.05	Асимметричность	-5.28	Асимметричность	-3.57	Асимметричность	-1.73
Интервал	45.2	Интервал	100	Интервал	100	Интервал	100
Минимум	54.8	Минимум	0	Минимум	0	Минимум	0
Максимум	100	Максимум	100	Максимум	100	Максимум	100
Сумма	7914	Сумма	7404.8	Сумма	7173.9	Сумма	6024.4
Счет	87	Счет	82	Счет	82	Счет	78
Наибольший(1)	100	Наибольший(1)	100	Наибольший(1)	100	Наибольший(1)	100
Наименьший(1)	54.8	Наименьший(1)	0	Наименьший(1)	0	Наименьший(1)	0
Уровень надежности(95.0 %)	1.56	Уровень надежности (95.0%)	2.68	Уровень надежности(95.0%)	4.27	Уровень надежности(95.0 %)	4.62

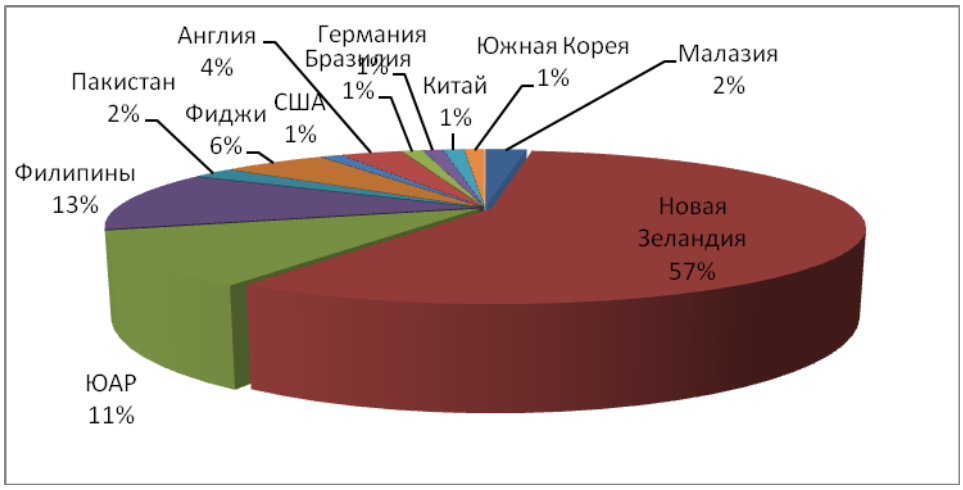


Рис. 2. Круговая диаграмма распределения слушателей по национальности

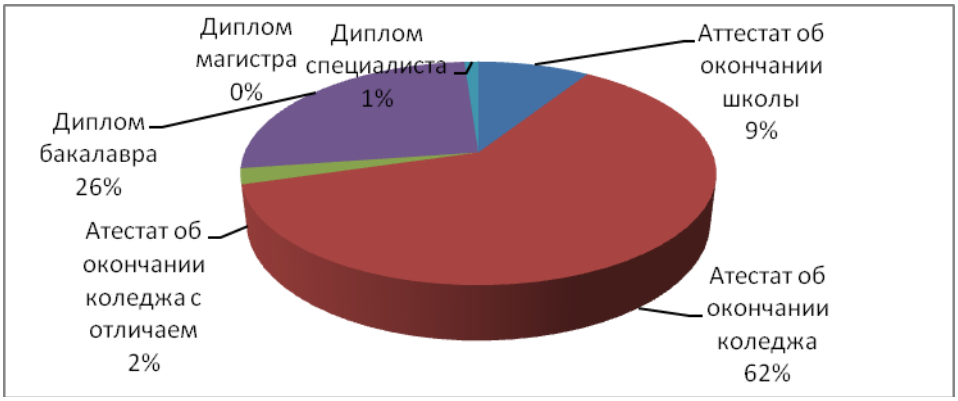


Рис. 3. Круговая диаграмма распределения слушателей по уровню образованию

Данная диаграмма позволяет увидеть, что преобладающий процент слушателей на момент поступления в учебное заведения уже имеют аттестат об окончания колледжа (среднее специальное образование).

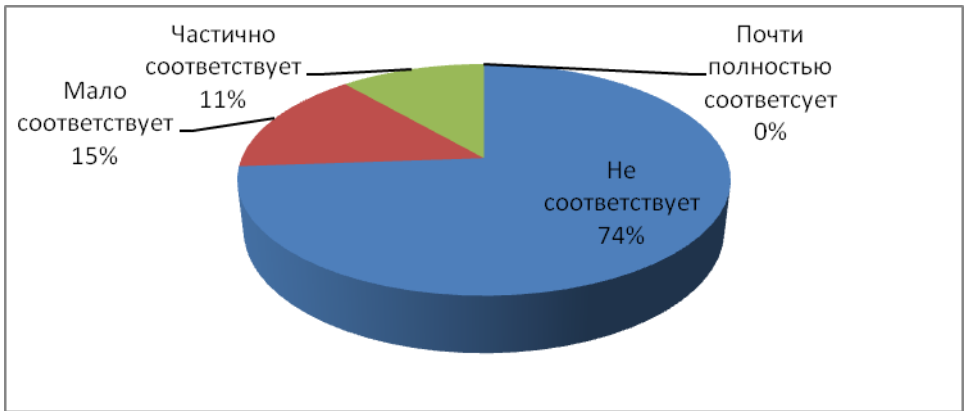


Рис. 4. Круговая диаграмма распределения слушателей по соответствию образования профилю курсов

Во многих случаях предыдущее образование не соответствует получаемому, потому что в преобладающем большинстве в колледжах Новой Зеландии изучают сельскохозяйственные специальности.

Общие выводы преподавания

В ходе преподавания курса сетевых технологий были сформированы следующие выводы:

- Сложность восприимчивости материала студентами возрастала с каждым днем занятий.
- Большинству студентов для закрепления материала необходимо давать большое количество работы на самостоятельную проработку.
- При соотношения 1 к 1 семинарных занятий к лекционным наблюдается наиболее лучший результат в успеваемости.

Общие замечания и выводы относительно успеваемости и влияющих на нее факторов

Анализ статистических данных привел к следующим выводам:

1. Процент иностранных слушателей достаточно велик (составляет 43%). Это связанно с тем, что специалист в области ИТ, а в частности с телекоммуникационном секторе, сильно востребован как на местном новозеландском рынке труда так и в мировом.
2. Процент слушателей женского пола мал (составляет 2%), что в свою очередь связанно со спецификой профиля. Как известно ИТ специализация более свойственна мужчинам чем женщинам.
3. Наиболее высокий процент не сдавших студентов экзамен это за итоговый экзамен 4. Это связанно с тем что, полный курс составляет 12 недель, а к концу обучения слушатели начинают забывать материал из первых 2-х семестров, что в свою очередь приводит к невысоким результатам за итоговый экзамен 4-го семестра, который опирается на знания всего курса.
4. ИТ специальность наиболее востребована среди слушателей в возрастной рамке между 20-29 лет

Литература

Горшков А.В. Обучение сетевым технологиям в сетевой академии: современное состояние и перспективы. / В сб: Современные информационные технологии в экономической деятельности. Материалы III международной научно-практической конференции. - Ростов-на-Дону: Эверест, 2010, с. 180-183

Автоматизированная система выявления индивидуального профиля студентов «деятельность-мотивация»

Е.А. Гресько
Самарский государственный архитектурно-строительный университет, Самара

Ведущей тенденцией развития современной цивилизации является возрастание роли общественного интеллекта в социальной эволюции. На фоне все более острых социальных, экономических и природных кризисов и катастроф происходит переход от стихийных регуляторов к механизмам устойчивого цивилизационного развития на базе эффективного общественного знания, позволяющего предвидеть и оптимальным образом организовать будущее в интересах общества и личности. Носителями же общественного интеллекта являются, в конечном счете, конкретные люди – только через них возможны его реализация и воспроизводство. Вот почему проблемы развития системы образования и повышения его эффективности являются актуальными как для отдельных стран, так и для всего мирового сообщества.

В настоящее время отсутствует программное обеспечение, выявляющее влияние деятельности на мотивацию.

Цель предстоящей работы состоит в том, чтобы описать взаимосвязанное нарастание во времени компонентов научной квалификации, вызванное собственной исследовательской деятельностью личности во взаимодействии с внешней средой. Взаимосвязь определяется тем, что функционирование личности при выполнении (и освоении ею) отдельных элементов научной деятельности различным образом влияет на такие факторы, как мотивацию (определяющую время, уделяемое научной деятельности), а также, возможно, на развитие интеллекта и креативности. С другой стороны, эти факторы задают темп освоения элементов деятельности. Роль внешней среды заключается в поддержке научной деятельности и ее стимулировании в зависимости от достигнутых результатов и квалификации.

Основной задачей данной работы является создание соответствующей математической модели и программного обеспечения для повышения эффективности и целесообразности образования.

При разработке программного обеспечения было принято, что необходимо разработать тест, выявляющий влияние различных видов исследовательской деятельности на мотивацию студентов.

Совместно с С.А. Пиявским были разработаны вопросы теста.

Коэффициенты являются количественной оценкой от 1 до 4 при выборе одного из ответов:

- верно;
- пожалуй, верно;
- пожалуй, неверно;
- неверно.

Ответу «верно» соответствует 4 балла, ответу «неверно» - 1 балл.

Также был разработан «ключ» влияния вопросов теста на эффективность влияния функций исследовательской деятельности на мотивацию студентов.

Для определения влияния конкретного вопроса на каждую из функций исследовательской деятельности были введены коэффициенты от 0 до3. То есть 0 означает минимальное влияние ответа на эффективность, 3 – сильное влияние.

В работе была использована следующая математическая модель. Ее структура:

1. *Математическая модель С.А. Пиявского*

i = 1..9 – номера основных функций исследовательской деятельности.

$j=1..4$ – номера научно-технических уровней исследовательской деятельности.

t – текущее время от начала развития личности в направлении исследовательской деятельности, месяцы.

I, K – текущая оценка интеллекта и креативности личности в психологическом понимании, баллы.

m_{ij} – текущее среднее время, уделяемое элементу научной деятельности (i, j), месяцы.

$\sum_{\substack{i=1..9 \\ j=1..4}} m_{ij} = M$ – суммарный уровень мотивации.

x_{ij} – текущая оценка квалификации личности по отдельным элементам научно-технической деятельности (i, j), измеряется в долях от полного овладения ими:

$0 \le x_{ij} \le 1$, $i=1..9$, $j=1..4$.

При описании суммарного уровня мотивации творческой личности соответствующее дифференциальное уравнение будет иметь вид:

$$\frac{dM}{dt} = \left(a_0 + \sum_{\substack{i=1..9 \\ j=1..4}} a_{ij} m_{ij} + a_x X_{prizn} \right) (M_{max} - M),$$ где:

a_0, a_{ij}, a_x – весовые коэффициенты, учитывающие относительную значимость соответствующую значимость соответствующих факторов;

M_{max} – физиологически предельный уровень мотивации;

X_{prizn} – признанная обществом квалификация личности:

$$X_{prizn} = \sum_{j=1..4} c_j^u x_{ij},$$

c_j^u – коэффициенты относительной значимости уровней творчества;

$(M_{max} - M)$ – «демпфирующий сомножитель», отражающий утомление, т.е. замедление темпов роста уровня мотивации при приближении к физиологическому пределу.

2. *Формирование оценок мотивации различных видов деятельности.*

Для формирования оценок мотивации различных видов исследовательской деятельности была сформирована таблица зависимостей функций исследовательской деятельности от ответов на вопросы разработанного теста (Приложение А, таблица 1).

Все ответы теста имеют коэффициент значимости от 1 до 4. На этой основе было принято формировать эффективность каждой из функций исследовательской деятельности следующим образом:

1. Расчет наихудшей мотивации всех функций. То есть если все ответы на тест были минимальными (1 балл).
2. Расчет наилучшей мотивации всех функций. То есть если все ответы на тест были максимальными (4 балла).
3. Расчет реальной мотивации всех функций. Принимались те ответы, которые указывал испытуемый.

На основе мною была сформирована следующая формула для расчета эффективности функций исследовательской деятельности:

$$F = \left(\frac{A - \frac{B-C}{3}}{C - \frac{B-C}{3}} \right) * 100\ \%$$

где A - реальная мотивация, B - наихудшая мотивация, C - наилучшая мотивация.

3. *Методы анализа данных*

По результатам исследования можно выделить следующие выводы:

1. Разработанный тест «деятельность-мотивация» полностью удовлетворяет целям работы. Тест уже внедрен в общую базу тестов ФИСТ СГАСУ и полностью работоспособен.

2. Были получены результаты ответов 54-х человек ФИСТ СГАСУ.

3. Разработанный тест удовлетворяет требованиям стандартизации с нормализацией **50 – 95%.** Такая нормализация связана с малым числом тестируемых студентов.

4. Тест отвечает требованиям повторяемости. Многие вопросы теста совпадают по своему содержанию с другими теста, такими как «Тест саморегуляции» и «Тест личностных характеристик».

5. При анализе результатов было выявлено, что:

a) многие студенты не полностью осознают назначения таких функций исследовательской деятельности, как «Выбор, освоение и реализация необходимого обеспечения» и «Реализация отдельных элементов исследования»;

b) самыми мотивируемыми функциями являются «Формирование ключевой идеи решения», «Выбор, освоение и реализация необходимого обеспечения» и «Синтез решения».

c) симметричному унимодальному распределению соответствуют «Постановка темы исследования», «Реализация отдельных элементов исследования», «Синтез решения» и «Ввод в научный обиход, защита и сопровождение решения»;

d) по анализу среднего квадратичного отклонений все функции имеют приблизительно одинаковый разброс. Это объясняется тем, что тестирование профодилось среди сравнительно одинакого контингента тестируемых (студенты первого-четвертого курсов ФИСТ СГАСУ).

6. Мотивация всех функций исследовательской деятельности в целом положительно влияет на разработку исследовательских проектов. Процент отрицательной мотивации функций исследовательской деятельности находится в пределах 5-31%. Самой неоднозначной является функция «Поиск тематики».

Разработка подсистема инвентаризации автоматизированного рабочего места инженера кафедры

Д.Н. Докшин, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Инвентаризация представляет собой проверку соответствия данных бухгалтерского учёта в организации реальному состоянию дел на текущий момент, она дает возможность получить объективную и детальную информацию о номенклатуре дел и актуальном состоянии инфраструктуры предприятия (оборудования, программного обеспечения и т.п.). В ходе инвентаризации проверяются и документально подтверждаются наличие имущества и обязательств, их состояние и правильность оценки. Все предприятия должны ежегодно проводить инвентаризацию, невыполнение этого требования может повлечь за собой серьёзные последствия, также как и неправильное оформление результатов инвентаризации.

Составление необходимых документов – важная составляющая инвентаризации, все результаты проверки заносятся в так называемую инвентаризационно-сличительную ведомость, которая отражает актуальное состояние и стоимость инфраструктуры предприятия. Как правило, форма этого документа и других сопроводительных документов очень редко меняется, в связи с чем целесообразно не только иметь электронные шаблоны документов, но и автоматизировать их заполнение, т.е. использовать так называемый «электронный документооборот».

На кафедре программных систем СГАУ разработано автоматизированное рабочее место (АРМ) инженера кафедры, представляющее собой WEB-приложение, в состав которого входит *подсистема инвентаризации*. Пользователями подсистемы могут быть:

– *администратор системы*, который имеет доступ ко всем функциям подсистемы инвентаризации;

– *инженер кафедры*, который отвечает за поддержку базы данных в актуальном состоянии, подготовку и генерацию отчетов и документов.

Подсистему инвентаризации можно разделить на две основных части: *подсистему ведения базы данных (БД)*, в которой ведется учет оборудования и программного обеспечения, закрепленного за кафедрой, и *подсистему ведения документации* (рис. 1).

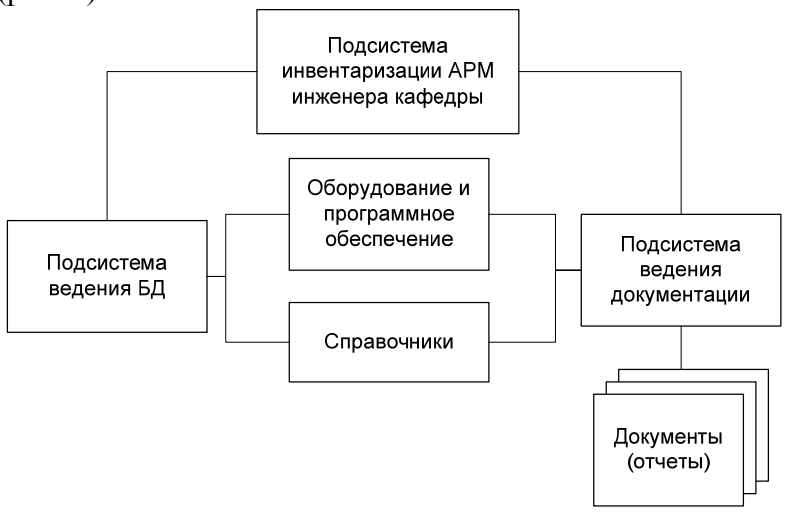


Рис. 1. Структура подсистемы инвентаризации

Подсистема ведения документации автоматически генерирует документы по заранее определенным шаблонам: инвентаризационные описи и ведомости, – которые записываются в БД и хранятся там по годам.

Центральным компонентом информационного обеспечения всей системы является БД, через которую осуществляется обмен данными между подсистемами, в ней хранится информация об оборудовании (позициях и категориях) и программном обеспечении, способах списывания, местах хранения оборудования и т.п., документы и вся связанная с ними информация. Локальная модель базы данных подсистемы инвентаризации представлена на рис. 2.

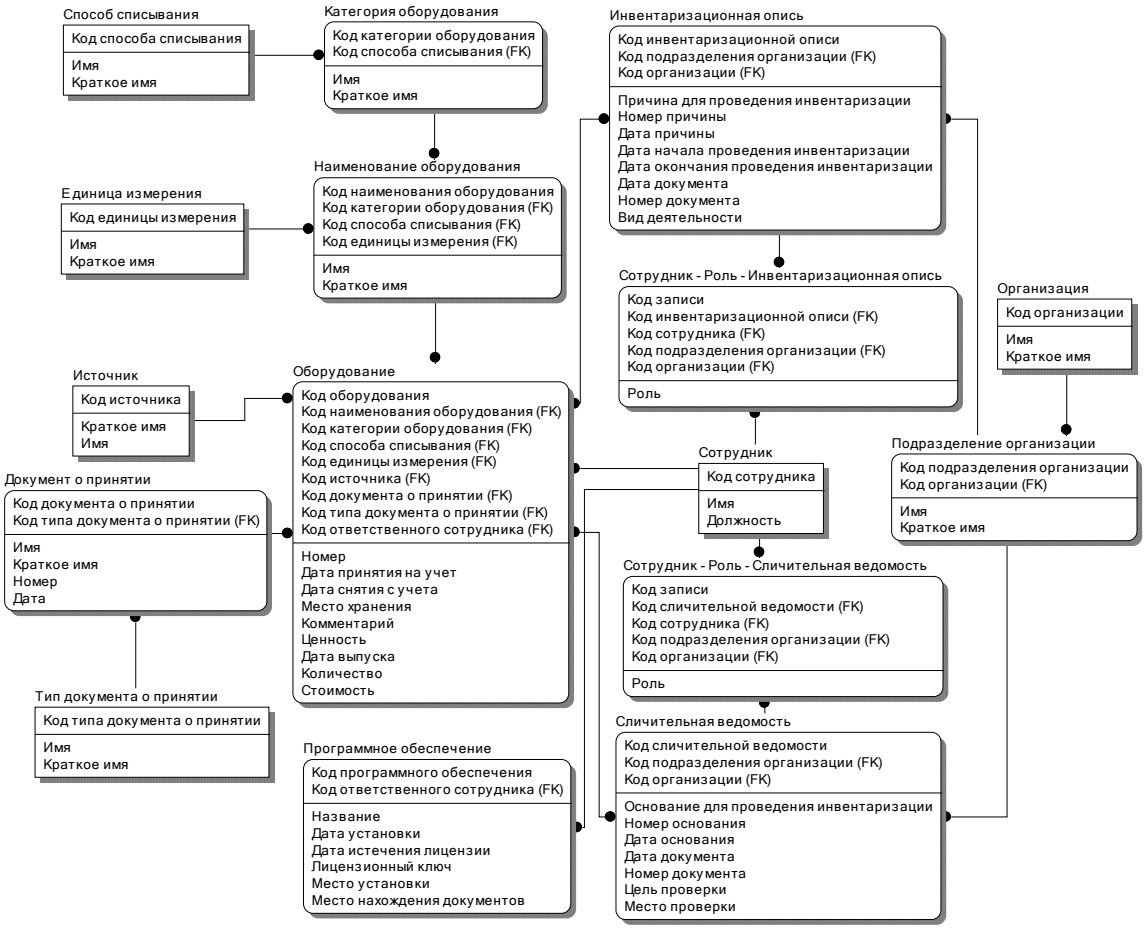


Рис. 2. Локальная логическая модель базы данных

В качестве среды разработки АРМ инженера кафедры выбрана Ruby on Rails – web-ориентированная среда с открытым кодом, обеспечивающая устойчивую работу пользователей с удаленной информацией и высокое быстродействие. Ruby – объектно-ориентированный язык, Rails – полноценный, многоуровневый фреймворк для построения веб-приложений, использующих базы данных, который основан на архитектуре MVC (Model-View-Controller, Модель-Представление-Контроллер) и обеспечивает гибкость разработки, отличную масштабируемость и читаемость приложений. Превосходный механизм работы с базой данных ActiveRecord позволяет организовать быстрый доступ к данным и работать с любой СУБД. Разработка базы данных подсистемы инвентаризации АРМ инженера кафедры проводилась с использованием CASE-средства AllFusion ERwin Data Modeler.

Автоматизированная система проведения онлайн-тестирования по иностранному языку «С.О.Т.А»

М.В. Ефремов, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

С развитием сети Интернет и технологий веб-программирования за последние несколько лет появились различные онлайн-сервисы, в том числе и системы проведения онлайн-тестирования.

С каждым годом во всех сферах жизнедеятельности человека, а особенно в образовании и науке, образуется все больший и больший пласт знаний, требующих освоения. Таким образом, очевидно, возникает проблема контроля качества освоения приобретенных знаний. А в связи с бурным развитием информационных технологий сегодня появляется возможность решить эту проблему путем автоматизации процесса проверки знаний, позволяя при этом преподавателю больше времени уделить совершенствованию лекционных, практических и тестовых материалов, а не самому процессу проверки.

Именно с этой целью кафедрой иностранных языков Самарского государственного аэрокосмического университета имени академика С.П. Королева было предложено разработать автоматизированную систему проведения онлайн-тестирования по иностранному языку. Разработка этой системы проводилась с учетом специфики данной кафедры и пожеланий, высказанных преподавателями.

Таким образом, создание автоматизированной системы проведения онлайн-тестирования является актуальной, сложной и интересной задачей веб-программирования, требующей понимания потребностей как тестируемых, так и людей, создающих тесты.

Описание возможностей системы

Основным пожеланием преподавателей кафедры иностранных языков являлось то, что система должна обладать интуитивно понятным и неперегруженным интерфейсом, чтобы с ней могли работать люди и студенты с разным уровнем владения навыками работы на компьютере, но при этом система должна иметь достаточную функциональность, чтобы полноценно проверять знания обучаемых. Эти и другие требования были учтены при разработке системы.

Структурная схема автоматизированной системы проведения онлайн-тестирования представлена на рисунке 1. Система состоит из серверной и клиентской частей.

Рассмотрим серверную часть. Подсистема идентификации пользователей решает задачу определения роли пользователя исходя из его логина и пароля, позволяя таким образом получать доступ к другим подсистемам. Подсистема администратора включает в себя подсистему работы с тестами, подсистему работы со списками студенческих групп, подсистему ведения журнала результатов тестируемого. Подсистема ведения журнала результатов получает данные от подсистемы проверки онлайн-теста и подсистемы работы со списками студенческих групп. Подсистема обмена сообщениями реализует взаимодействие серверной части с клиентской. Подсистемы проверки теста, администратора, подсистема идентификации пользователей взаимодействуют с базой данных.

Рассмотрим клиентскую часть. В нее входит подсистема проведения теста, которая передает данные в подсистему проверки, и справочная подсистема.

Справочная подсистема предоставляет доступ к пользовательскому руководству.

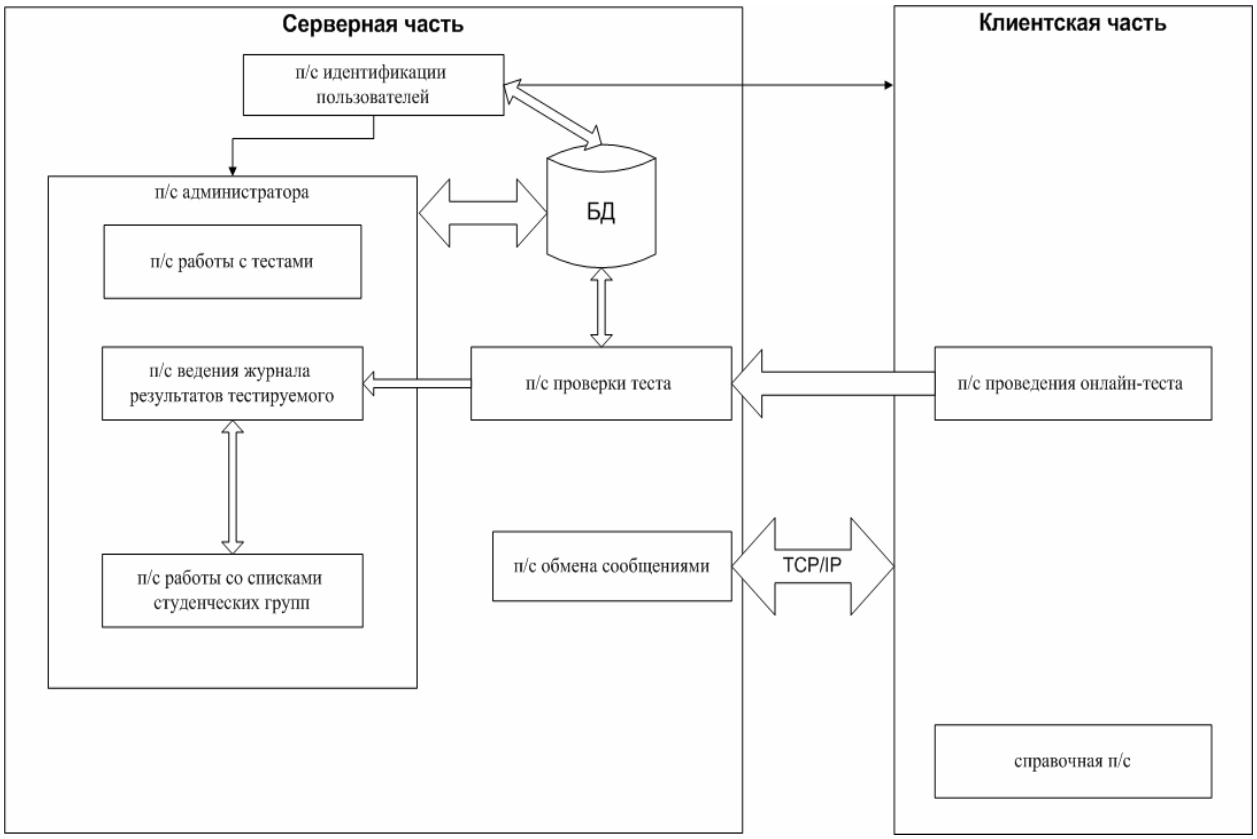


Рис.1. Структурная схема системы

В системе предусмотрены три роли пользователей: администратор, модератор и тестируемый. В начале работы с системой предусмотрена процедура прохождения авторизации, после которой в зависимости своих прав доступа пользователь получает доступ к тем или иным разделам системы.

В режиме администратора пользователь имеет следующие функциональные возможности:

- Работа с темами тестов: создание новой темы, редактирование, удаление.
- Работа с тестами: создание нового теста, редактирование, удаление теста.
- Работа со списком тестируемых: создание нового профиля тестируемого, редактирование, удаление профиля.
- Работа со списком модераторов: создание нового профиля модератора, редактирование, удаление профиля.
- Работа с журналом результатов тестирования.

Режим модератора – это режим, схожий с административным, но обладающий урезанным набором функций. Модератору доступны только те функции, которые связаны с тестами, и не доступны функции работы со списками тестируемых и журналов результатов.

Страницы добавления нового теста представлены на рисунках 2 и 3.

Из рисунка 2 видно, что система дает возможность создавать тесты с различным количеством вопросов открытой и закрытой форм.

На рисунке 3 представлена форма с добавлением вопроса закрытой формы в тест. Пользователь выбирает тему теста из ранее созданных, присваивает номер тесту, задает время, отведенное тестируемому на выполнение теста. Далее предлагается ввести текст вопроса, его вес, задать варианты ответа на вопрос и указать правильный ответ.

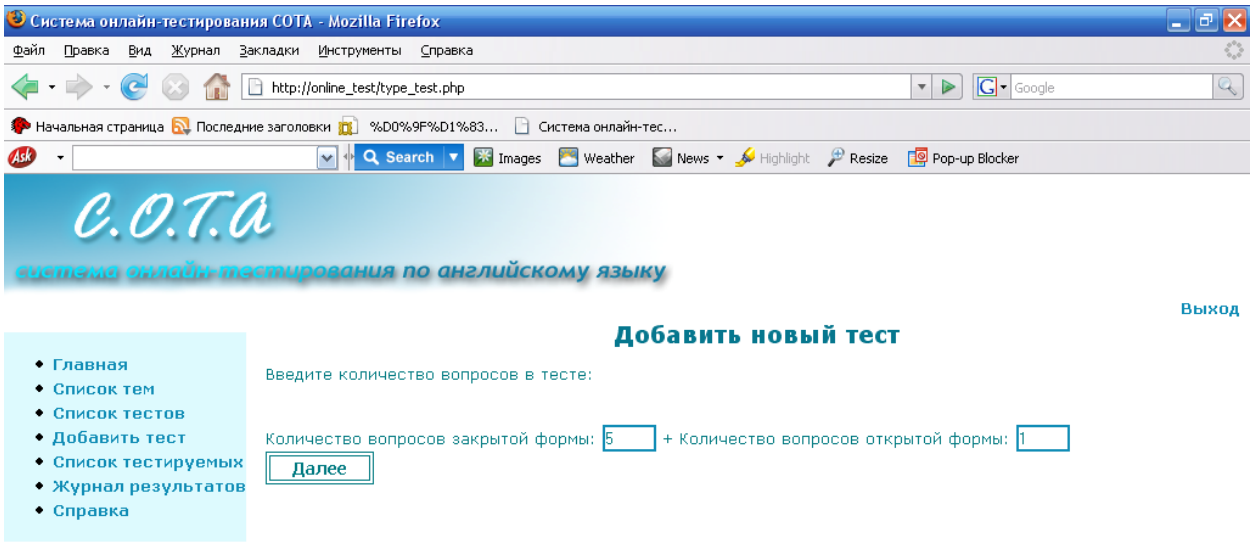


Рис. 2. Ввод количества вопросов в тесте

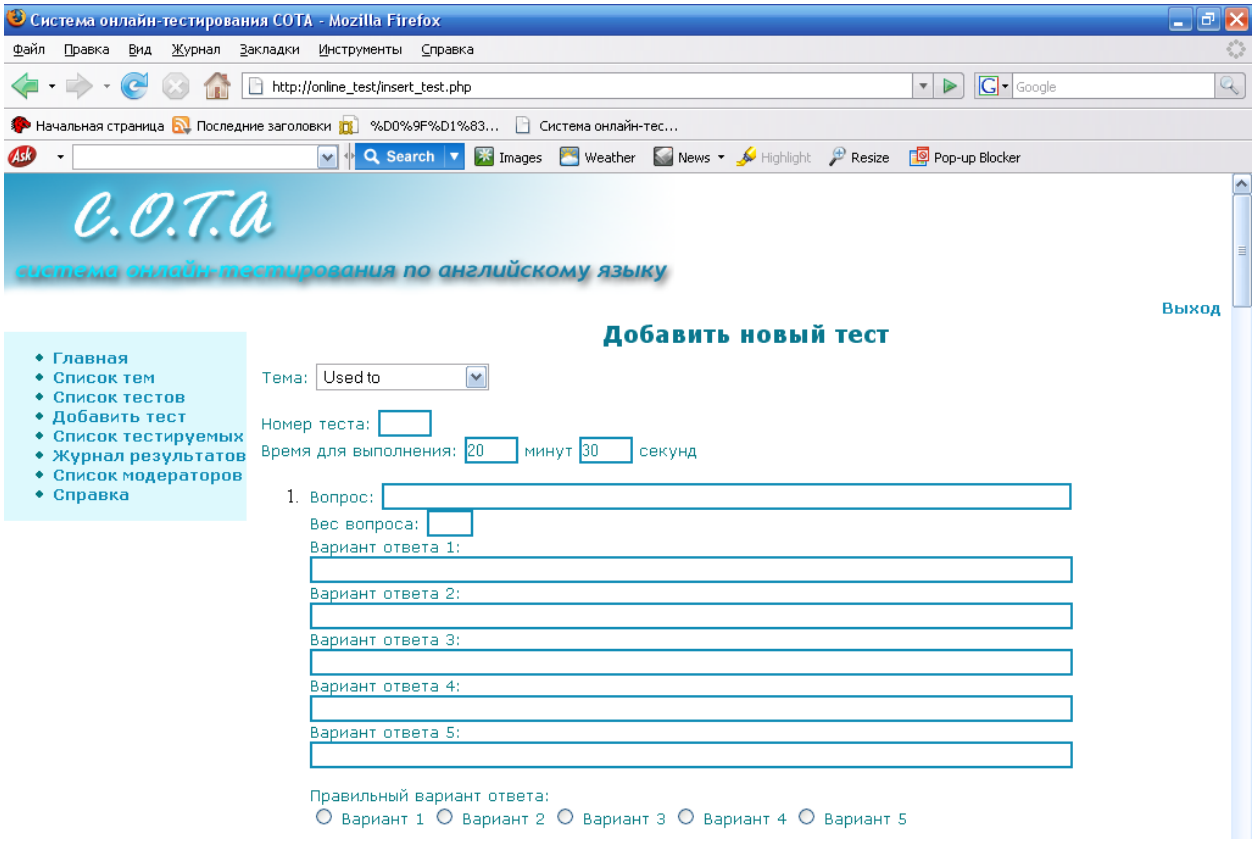


Рис.3. Страница добавления тестового вопроса

Далее созданные тесты могут подвергаться редактированию и в случае надобности удалению в режимах администратора и модератора.

Администратор также создает списки тестируемых с персональным логином и паролем, с помощью которых тестируемые выполняют вход в свою часть системы. Логином является номер студенческого билета, он уникален, а пароль генерируется системой случайным образом их букв различного регистра и цифр.

В режиме тестируемого пользователю доступен список тестов, которые он может выполнить, и справочное руководство по системе. После выбора тестируемым

определенного теста начинается его выполнение. Система следит за временем выполнения теста таким образом, чтобы оно не превысило время, установленное администратором или модератором. Если тестируемый не успевает выполнить тест до конца за отведенное время, система заносит в журнал результатов те баллы, которые он успел набрать.

После проведения теста администратор может просмотреть результаты по журналу в системе, где выводятся следующие данные:

- ФИО тестируемого, выполнившего тест.
- Номер группы тестируемого.
- Тема выполненного теста.
- Номер выполненного теста.
- Дата и время сдачи.
- Набранный бал.
- Процент правильно выполненных заданий.

Система онлайн-тестирования была написана на языке PHP 5.0 с использованием системы управления базами данных MySQL. С учетом популярности данной технологии такая реализация системы позволила добиться совместимости и легкой развертываемости на сервере в сети Интернет.

Выводы

Разработанная система удовлетворяет всем поставленным требованиям, имеет удобный и интуитивно понятный интерфейс, требуемую функциональность, организованную информационную поддержку в виде справочного руководства, а также ведет контроль ссылочной целостности в базе данных.

Система позволяет проводить онлайн-тестирование, просматривать результаты тестирования, работать с ними; работать с тестами, формировать списки студенческих групп для тестирования.

Данная система используется преподавателями кафедры иностранных языков Самарского государственного аэрокосмического университета имени академика С.П. Королева для контроля успеваемости студентов по иностранному языку. Предполагается дальнейшее развитие и совершенствование системы.

В качестве перспектив развития этого проекта можно выделить следующие направления:

- Расширение функциональности путем добавления новых типов тестовых заданий.
- Дальнейшее совершенствование подсистемы контроля ввода.
- Совершенствование информационных отчетов по результатам тестирования.

Концептуальные основы построения дистанционных обучающих систем

Л.С. Зеленко, Т.И. Михеева
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

Динамичное развитие информационных технологий требует внедрения новых подходов к обучению, обеспечивающих наравне с фундаментальностью образования развитие творческих и профессиональных компетенций подготавливаемых специалистов, потребностей их в непрерывном самообразовании. Применение инновационных технологий обучения, к числу которых относятся дистанционные образовательные технологии, дает возможность вузам выдерживать конкуренцию на рынке труда, обеспечивая качественную основу для подготовки специалистов высокой квалификации.

Дистанционное обучение или, точнее, электронное (от термина e-learning), позволяет в корне изменить процесс передачи знаний, сделать его более гибким, насыщенным и удобным для обучаемого и предлагает ресурсы, которые открывают новые возможности его активного участия в образовательном процессе. К числу таких ресурсов в первую очередь необходимо отнести компьютерные обучающие системы или учебные мультимедиа комплексы, которые дают возможность предоставлять учебно-методические материалы в электронной форме, постоянно обновлять контент и наращивать его потенциал по мере изменения структуры учебного курса.

Дистанционное обучение можно рассматривать и как *обучение*, при котором все или большая часть учебных процедур осуществляется с использованием современных информационных и телекоммуникационных технологий (ИКТ) при территориальной разобщенности преподавателя и студентов, и как *совокупность методов и средств*, обеспечивающих такую возможность [1]. Обучающие системы, включаемые в системы электронного дистанционного образования (СЭДО), – это сложные, многокомпонентные системы, которые динамично развиваются и непрерывно обновляются. При их проектировании необходимо применять и *системный подход*, и совместно с инновационными педагогическими технологиями информационные технологии, и *новые методологии*.

Построение методологии, позволяющей проектировать дистанционные обучающие системы, дающие максимальный учебный эффект при разумных затратах на разработку, можно считать ***актуальной задачей***. Данная методология должна использовать все новейшие достижения (методики) таких дисциплин, как педагогика, психология (в числе когнитивная психология), семиотика, нейролингвистика и др. Кроме того, в нее должен быть включен математический аппарат для представления и обработки информации, служащей объектом изучения вышеперечисленных наук: текст, таблицы, мультимедиа-объекты (графика, звук, видео и т.д.), сенсорно-моторные импульсы и т.п. [2]. И наконец, современная компьютерная обучающая система должна активно использовать последние достижения в области ИКТ (Internet-технологии, технологии виртуальной реальности и т.п.).

Концепция дистанционной обучающей системы

В основу комплекса образовательных методик, применяемых в СЭДО, положена как самостоятельная интерактивная работа обучаемого со специальными электронными учебными материалами (ЭУМ) – *режим самообразования*, так взаимодействие с преподавателем – *режим обучения*.

Эффективное взаимодействие с обучаемым требует построения в рамках обучающей системы адекватной математической модели как обучаемого, так и обучающего. При этом должны быть соблюдены принципы научности и системности, а также принципы развивающего обучения, наглядности, дифференциации и индивидуализации обучения.

Все эти требования дают основания к использованию в обучающей системе *когнитивной компьютерной графики* и *метода научной визуализации*. Значение компьютерной графики при стимуляции когнитивных процессов явственно следует из названия. При когнитивной визуализации знаний не ставится задача построить некое визуальное отображение, тождественное образному представлению знаний в головном мозге человека, задача когнитивной визуализации – стимулировать познавательные механизмы, творческое мышление, а не давать однозначную интерпретацию знаний. Следует отметить, что немаловажной особенностью когнитивной визуализации является *латентность* самого *обучения*: обучаемый, рассматривая предлагаемые зрительные образы, в идеале и не подозревает о том, что он на самом деле обучается, процесс обучения превращается в некое подобие игры. В таком случае обучаемый, со своей стороны, полностью раскрывает свои возможности, а обучающая система, со своей стороны, может с высокой степенью надежности использовать психолингвистически и математически обоснованные методики для достижения наилучшего результата обучения [3].

Метод научной визуализации предполагает использование визуальных компьютерных моделей, которые позволяют перевести набор формальных определений сложных объектов в наглядные геометрические свойства, за которыми можно наблюдать, динамически меняя параметры, и дает возможность непосредственно увидеть свойства и взаимосвязи отдельных частей, далеко не очевидные без компьютерного эксперимента и геометрической визуализации [2, 4]. Не подвергается сомнению и тот факт, что применение мультимедиа технологий в обучении позволяет поднять на качественно новый уровень и по существу изменить процесс передачи и усвоения знаний, их применение в учебном процессе может оказать революционное воздействие на всю индустрию компьютерных обучающих систем.

Проектирование ЭУМ требует серьезного научного подхода [5] и представляет собой весьма непростую задачу, требующую от обучающего (преподавателя) больших усилий и серьезной переработки «бумажных» учебных материалов. Содержание ЭУМ и форма подачи материала должны учитывать не только интерактивный характер взаимодействия обучаемого и обучающего, но и специфику конкретной предметной области, дистанционной формы обучения, возможности информационной среды, в первую очередь, Интернет. Кроме того, проектирование ЭУМ должно вестись с позиций *педагогической психологии*, необходимо учитывать требования, диктуемые психологическими особенностями восприятия информации, и эргономические требования.

Восприятие когнитивной информации человеком

Факт асимметрии функций человеческого мозга в современной нейрофизиологии считается общепринятым, что можно кратко охарактеризовать следующей цитатой: «правое полушарие связывают с пространственно-синтетической деятельностью, левое – с речевой и аналитической» [2]. Каждое из полушарий человеческого мозга является самостоятельной системой восприятия внешнего мира, переработки информации о нем и планирования поведения в этом мире. Согласно [2], «…правополушарный способ основан на целостном отображении объектов… Под целостным отображением понимается восприятие объекта как нерасчлененной совокупности его атрибутов… Это – существенно многомерное отображение, при

котором атрибуты объекта запечатлеваются в памяти параллельно, безусловно и одномоментно. Поэтому память в правом полушарии – эпизодическая, *данная в конкретном контексте*». Единственным видом отношений при правополушарном отображении объектов является *отношение сходства*, т.е. насколько один объект похож на другой в том или ином смысле, контексте, окружении. Знания, хранимые в правополушарной памяти, являются *экстенсиональными*. Важно отметить, что функции правого полушария правомерно также называть функциями обработки *невербальной* информации, поскольку под «невербальным» в современной психологии интеллекта понимается «имеющий отношение к реальным предметам, наглядным впечатлениям, жестам, а также к процессам оперирования этим материалом» [2]. Левое полушарие выполняет функции, комплементарные функциям правого. Исследователи зачастую сравнивают деятельность этого полушария с деятельностью компьютера: например, тьюринговская модель автомата, переписывающего символы на ленте, на самом деле есть модель левополушарной мыслительной деятельности. Левополушарный способ отображения действительности основан на «знаниях о статических и динамических закономерностях структуры воспринимаемой информации. Левополушарная память обнаруживает себя как хранилище этих знаний, выраженных словами, символами, значениями и отношениями между ними» [2]. Знания в левом полушарии – *интенсиональные*, это знания о структуре, законах. Процесс левостороннего мышления, в отличие от правостороннего, осуществляется *последовательно*, пошагово, причем каждый шаг этого процесса есть переход между узлами дерева, ребра которого помечены предикатными функциями, выполнение которых означает переход к нижестоящему узлу, а невыполнение – возврат к вышестоящему узлу. Левое полушарие занимается обработкой вербальной – структурированной, упорядоченной – информации. Оно рассматривает не конкретные предметы, а их абстрактные модели (классы), объединяя все объекты, имеющие одинаковые имманентные признаки, в один класс. Очевидна параллель с таксономической спецификацией межклассовых отношений предметной области [3].

Концепция дистанционной системы обучения информатике и программированию

Анализ существующих обучающих систем показал, что наибольшее распространение получили обучающие программы в области точных наук. Это, безусловно, связано с тем, что в области точных наук легче формализовать знания, т. е. выделить объекты, классы объектов, их ассоциации, установить регламент межклассовых отношений и структурировать все это в рамках заданной предметной области. При разработке компьютерной обучающей системы по основам информатики и программирования авторы попытались решить проблему формализации знаний и структурирования учебного материала следующим образом.

Предлагается весь учебный материал курса разбить на временные отрезки и разделы, структурированные в n–мерный массив, а единицей (объектом) формализации выбрать учебный элемент (УЭ). «Верхняя» грань этого массива – матрица, в строках которой расположены разделы учебного курса (например, «Системы счисления», «Алгебра логики», «Базовые алгоритмы обработки одномерных массивов» и т.п.), в столбцах – метки временной шкалы, представленные либо в абсолютных единицах (например, количество дней или недель, необходимых для усвоения материала), либо в относительных (например, в процентном отношении к длительности изучения всего курса). На пересечении строк и столбцов находятся один или несколько учебных элементов. Название «учебный элемент» условно, так как в него могут быть включены как элементарные алгоритмы, например, алгоритм обмена данными «Пустое ведро», так и сложные алгоритмы и процессы (обобщения), например, «Сортировки», «Поиск»

и т.п. Для каждого УЭ разработан паспорт, содержащий название УЭ; название раздела, к которому он относится; названия разделов, в которых УЭ будет использоваться. N-мерность массив приобретает за счет «разрастания» гипертекстового дерева, вершиной (корнем) которого является сам УЭ, расположенный в ячейке матрицы, а узлами и ветвями – мультимедийная информация, снабженная гипертекстовыми ссылками. Такая формализация позволяет проследить основные взаимосвязи УЭ, а сам изучаемый учебный курс представлен сложным переплетением составляющих его элементов.

Предлагаемая методика структурирования учебного материала не противоречит базовым концепциям SCORM (The Sharable Content Object Reference Model), а именно: составлению электронных образовательных ресурсов из блоков учебного материала, называемых совместно используемыми объектами содержания (Sharable Content Objects - SCOs). В качестве SCOs в обучающей системе используются отдельные фрагменты текста, графические иллюстрации, компьютерные анимации, самостоятельные компьютерные программы, видео- и аудио- файлы, элементы гипермедиа или их комбинации. SCORM не накладывает ограничений на размер SCOs и учебное время работы с ними, однако предполагает, что объект представляет относительно небольшую часть содержания изучаемого учебного материала.

Одной из наиболее перспективных мультимедиа технологий является технология виртуальной реальности или виртуальных миров. *Виртуальная реальность* - построенная компьютерными средствами трехмерная модель реальности, которая создает эффект присутствия человека в ней, позволяет взаимодействовать с ней объектами, включая новые способы взаимодействия: изменение формы объекта, свободное перемещение между микро- и макроуровнями пространства, перемещение самого пространства и т.п. Технология виртуальной реальности позволяет эффективно стимулировать правополушарные когнитивные механизмы с помощью невербальной пространственной интерпретации экстенсиональных знаний, и, тем самым, к активизации творческого индуктивного мышления обучаемого. Погружение обучаемого в виртуальную трехмерную среду позволяет более эффективно воздействовать на его визуально-пространственные рецепторы и, как следствие, дают ему возможность более точно представить предмет изучения (по сравнению с двумерными динамическими или статическими слайдами): пользователь сам может выбрать ракурс отображаемого объекта, приблизить или удалить его, поменять параметры освещенности или закраски [6]. Технология виртуальной реальности превращает процесс обучения в некое подобие игры (причем, совсем не обязательно подчеркнуто интеллектуальной). По мере продвижения по виртуальному пространству обучаемый получает необходимую информацию о текущем УЭ из его паспорта и список УЭ, связанных с текущим. Определив правильно цепочку связей за заданное число шагов, обучаемый заканчивает игру. В случае неудачи у него есть возможность сделать возврат на любое количество шагов и продолжить игру с этой точки. Во время игры можно при необходимости получить развернутую информацию о любом УЭ, вернувшись к исходной матрице. В таком случае обучаемый, со своей стороны, полностью раскрывает свои возможности, а обучающая система, со своей стороны, может с высокой степенью надежности использовать психолингвистически и математически обоснованные методики для достижения наилучшего результата обучения. Использование такой игры в обучающей системе позволяет решать фундаментальную образовательную задачу - систематизацию учебного материала, обеспечивает понимание информационного пространства как совокупности сложных и многообразных взаимосвязей.

Особо следует подчеркнуть следующее важнейшее требование: математическая адекватность требует от рассматриваемой системы способности *автоматической*

генерации учебного курса на основе заданной базы знаний в предметной области и базы знаний о способностях и потребностях обучаемого.

Выводы

Собственно учебный курс в обучающей системе представлен в виде рекомендуемых направлений перемещения в виртуальном пространстве (по построенной *n*–мерной матрице УЭ), при этом обучаемый может не следовать этому курсу, а самостоятельно исследовать все доступные в текущем сеансе обучения уголки виртуального мира. Процесс обучения делится на условные этапы: изучение текстового материала, просмотр медиаматериалов, выполнение индивидуальных заданий, прохождение тестов. Завершающий этап обучения - это выполнение различных «поручений», возлагаемых системой на обучаемого. Основные действия пользователя, количество пройденных заданий и их результаты сохраняются в базе данных обучаемого. На их основании система будет вычислять показатели эффективности обучения и строить схему изучения учебного материала при следующем сеансе: определять направления учебного курса, вид и содержание тестовых заданий, при этом акцент делается на тех направлениях и видах учебной информации, которые были менее всего изучены или были неверно усвоены.

Литература

1. Зеленко, Л.С. Дистанционное обучение как инновационная стратегия педагогического процесса [Текст] // Труды XV Всероссийской научно-методической конференции «Телематика-2008». - СПб: ИТМО. - 2008. - Т.2. – С. 461-463.
2. Дюк, В.А. Компьютерная психодиагностика [Текст]. – СПб.: «Братство», 1994. – 364 с.
3. Михеева, Т.И. Программная таксономия – основа для создания гипермедийных обучающих программ [Текст]/ Т.И. Михеева, И.Е. Михеенков // Информационные технологии. - 1998. - №8. - С. 40–43.
4. Зенкин, А.А. Когнитивная компьютерная графика и научное творчество [Текст]// В сб. «Будущее искусственного интеллекта». - М.: Наука, 1991. - С. 269–289.
5. Соловов, А.В. Электронное обучение: проблематика, дидактика, технология [Текст]. – Самара: «Новая техника», 2006. – 464 с.: ил.
6. Зеленко, Л.С. Интерактивная интеллектуальная обучающая система, построенная на основе технологии виртуальных миров, как средство активизации учебно-познавательной деятельности учащихся [Текст]/ Л.С. Зеленко, А.В. Топунов, Д.А. Загуменнов Д.А. // Труды XVII Всероссийской научно-методической конференции «Телематика-2010». - СПб: ИТМО. - 2010. - Т.2. - С.335-336.

Интернет-портал «Авторский Комикс»

А.А. Козлов, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Подавляющее большинство людей со словом «комикс» ассоциирует исключительно образы классических американских супергероев: Супермен, Бэтмен, Человек-Паук, Люди Икс и многие другие. Некоторые знают о существовании японской индустрии комиксов: так называемая «манга». И уж совсем мало людей представляют, что где-то в непосредственной близости, а именно: в Мировой Паутине Интернета, - зарождается альтернативная и самодостаточная индустрия комиксов: Веб-комиксы.

Впрочем, если быть до конца честным, слово «зарождается» применимо к русскоязычному сегменту Интернета: веб-комиксы на английском языке уже существуют в большом количестве и многие годы. Sinfest, Ctrl+Alt+Del, Penny Arcade, Cyanide&Happiness, Nemi, xkcd – эти франчайзы мало известны у нас, но на западе они очень популярны, а их авторы зарабатывают немалые деньги. Комикс Penny Arcade, к примеру, не только был издан в печатном варианте, но и стал весьма неплохой компьютерной игрой, мало того, коллектив авторов проводит ежегодный компьютерно-игровой фестиваль Penny Arcade Expo. Крупнейший иностранный рейтинг веб-комиксов TopWebComics.com сейчас насчитывает уже больше 7000 участников.

Отечественный же «Рейтинг Комиксов» *top.a-comics.ru* в начале этого года зарегистрировал только сотого участника. Эта цифра кажется ничтожно малой на фоне англоязычного рейтинга, и на самом деле так оно и есть. Одно но: всего 2 года назад, когда рейтинг только создавался, в нем было зарегистрировано всего 20 веб-комиксов, и каждый последующий год их число росло минимум на 130%, а такому росту можно только позавидовать.

Цели и задачи проекта «Авторский Комикс»

Чтобы лучше понять цели и задачи, преследуемые данным проектом, следует поподробнее рассказать о том, что такое Веб-комикс.

Веб-комикс – это одна из самостоятельных ветвей развития комиксов, которое отличается от классического бумажного вариант не только местом публикации, но и множеством косвенных признаков, характерных исключительно для этого вида творчества:

– они не «привязаны» к газетному или страничному формату: от выпуска к выпуску формат может варьироваться;

– они в общем смысле являются самиздатом (в лучшем смысле этого слова): автор не ограничен сроками, цензурой и пожеланиями заказчиков;

– они распространяются свободно, авторы не получают прямой прибыли от продажи своего творчество, хотя это не исключает побочных источников дохода:

– веб-комиксы – это оперативность: только что нарисованный и выложенный вами выпуск уже через секунды могут лицезреть ваши читатели.

Все это позволяет нам говорить о веб-комиксах, как о самостоятельном элементе индустрии развлечений информационного сообщества. Проект «Авторский Комикс» преследует цель не просто предоставить начинающим авторам площадку для публикации, но и объединять все публикующиеся на отдельных сайтах веб-комиксы со всего Рунета.

Проблемы и решения

Портал «Авторский Комикс» (*http://a-comics.ru)* является первым и единственным на сегодняшний день сервисом в Рунете, предоставляющим художникам площадку для публикации своих комиксов.

Первая проблема, с которой сталкиваются художники, желающие опубликовать свой веб-комикс в Интернете – место для публикации. Портал «Авторский Комикс» решает эту проблему: он предоставляет авторам-комиксистам собственную страницу для публикации,

интерфейс для обновления выпусков, ленту комментариев читателей и даже возможность размещения баннера в виртуальном магазине, где продаются товары по комиксу. Вместе с этим автор получает тысячи потенциальных читателей – постоянных посетителей сайта, которые могут заинтересоваться новым комиксом. Все сделано для того, чтобы художник мог заниматься любимым делом – создавать комикс, не занимал голову техническими вопросами публикации в интернете.

Другая проблема, с которой пришлось столкнуться порталу «Авторский Комикс», - чрезмерная нагрузка на сервер. Любой массовый наплыв посетителей на «Рейтинг» до оптимизации SQL-запросов мог вывести сайт из строя. Что касается самого сайта, то оказалось, что большую нагрузку на сервер оказывает… форум на стороннем движке Invision Power Board. Сейчас эта проблема решена.

Если взглянуть на сайт с точки зрения читателя, то мы увидим десятки разнообразных веб-комиксов на любой вкус. Многие из них обновляется на регулярной основе, но большинство комиксов не имеют постоянного графика обновления, поэтому для читателей было проблематично отслеживать процесс обновления ряда избранных комиксов. На помощь пришел протокол RSS: был создан настраиваемый RSS-канал, который позволяет пользователю выбирать комиксы, на которые он хотел бы подписаться.

Концепция сайта тоже имела некоторые нюансы. Однажды стало очевидно, что индустрия веб-комиксов в России доросла до такого уровня, что происходящие в ней события стоят того, чтобы их освещать более широко. Поэтому было решено открыть дочерний проект WebComUnity, который уже стал исключительно информационным порталом, частично взявшим на себя функции и форума, и блога, и новостной ленты.

В рамках проекта функционирует ранее упомянутый «Рейтинг Комиксов». Это первый и единственный на настоящий момент рейтинг веб-комиксов на русском языке. На западе есть, по крайней мере, три крупных рейтинга, в России – один, и в этом году число зарегистрированных в нем комиксов перевалило за сотню.

Выводы

Что самое главное в веб-сайте? Кто-то ответит, что контент (содержание), для кого-то главным будет дизайн, неплохой догадкой будет «посещаемость». И лишь после того, как внешний вид сайта будет отшлифован, контент создан, посещаемость появится, приходит понимание того, что чего-то не хватает. Это что-то называют иностранным словом «Комьюнити» или «Сообщество».

Комьюнити – это не число на счетчика посещений сайта, это группа людей, реально заинтересованных в теме (проблеме), которой посвящен сайт; которые участвуют в дискуссиях, которые готовы словом и делом помочь сайту. В конце концов, люди – конечная цель создания любого сайта. Сайт всегда будет популярен и успешен, если он формирует вокруг себя группу заинтересованных людей, независимо от того, какие сложные технологии были заложены в его основу. Именно поэтому администрация «Авторского Комикса» уделяет большое внимание взаимоотношению между авторами комиксов и читателями, продвижению идей веб-комиксов в массы.

Модель кластеризации студенческой группы на основе непрерывного мониторинга успеваемости

В.В. Козлов
Самарский государственный архитектурно-строительный университет, Самара

В ходе непосредственного обучения в семестре требуется наличие хорошо отлаженной обратной связи, обеспечивающей оперативный учет индивидуальных особенностей восприятия учебного материала отдельными студентами. Далеко не все студенты одинаково осваивают учебные дисциплины, так как поток студентов изначально неоднороден. В данной статье в качестве инструмента, обеспечивающим обратную связь с преподавателем, предлагается использовать систему непрерывного рейтингового мониторинга успеваемости учащихся. Важное задачей преподавателя является необходимость в ходе учебного процесса в оперативном режиме выявлять отстающие группы студентов и своевременно ликвидировать причины этого отставание [1]. Решить эту задачу можно на основе математического моделирования путем кластеризации студенческой группы в ходе ее обучения на основе непрерывного мониторинга успеваемости [2, 3]. Таким образом, непрерывный индивидуальный мониторинг успеваемости имеет три цели [4]:

- на основе индивидуального компьютерного тестирования (или иным способом) дать информацию об уровне успеваемости студента в оперативном режиме;
- определять рейтинг студента в оперативном режиме и осуществлять динамическую кластеризацию студенческой группы;
- прогнозировать предположительную (рекомендуемую) оценку студента при заключительной аттестации по дисциплине.

В «стандартной» схеме обучения в вузе: лектор читает лекции, а студенты независимо от этого выполняют лабораторный практикум с разбиением на подгруппы (обычно по алфавиту). Контроль при таком подходе проводится обычно в середине семестра и только в этот момент возникает обратная связь. Индивидуализированная же технология обучения предполагает групповое обучение в однородных по своему составу группах. В статье предлагается индивидуализированная коррекция преподавания учебной дисциплины путем динамической кластеризации студенческой группы на основе данных непрерывного мониторинга учебного процесса. Доступность любой момент времени получения списка в котором студенты разделены на несколько подгрупп (кластеров) в зависимости от выявленной степени их освоения учебной дисциплины позволяет преподавателю индивидуализировать свою работу по отношению к студентам при проведении занятия, например, уделяя больше внимания подгруппе, состоящей из «слабых» студентов, в то время как более «сильные» студенты могут заниматься самостоятельно.

За счет использования математического аппарата на уровне непосредственного обучения предусматривается индивидуализация хода изучения отдельных разделов и тем внутри каждой из дисциплин. При этом средствами мониторинга обеспечивается оперативный контроль за ходом учебного процесса, ведется индивидуальный рейтинг студента и довольно сильно меняется сам ход учебного процесса. Вся дисциплина разбивается на разделы и темы. Каждый раздел состоит из нескольких тем с оценкой трудоемкости их освоения в учебных часах. На раздел в целом и каждую тему в отдельности составлены рекомендации и методические указания, выложенные в сети Интернет. По каждой теме также предусмотрен тест или иной способ контроля с указанием минимально допустимого уровня ее освоения. Для всех тем в дисциплине задана матрица взаимозависимостей, определяющая возможные траектории изучения отдельных дисциплин. Таким образом, на этом уровне мы имеем индивидуальные траектории изучения дисциплины для каждого из студентов, что, однако, довольно сложно реализовать на практике и в сложившейся образовательной среде вуза целесообразнее перейти к групповому индивидуализированному обучению.

В рамках модели индивидуального мониторинга рассмотрим ход изучения одной дисциплины одним студентом. Дисциплина разбивается на N тем. Темы обозначаются индексом k, который изменяется от *1* до N. Вектор G_k задает трудоемкость изучения k-той темы в занятиях в академических часах. Множество

занятий (тем) обозначим через G_{kt}, где через *t* обозначен момент времени, когда произошло занятие по *i*-той теме.

Вектор A_k определяет текущий балл студента по *k*-той теме, A_{kt} – балл студента по *k*-той теме в момент времени *t*. Зададимся вектором пороговых значений отдельных тем из дисциплины H_k. Тема считается зачтенной, если выполняется неравенство:

$A_k \le H_k$.

Введем оператор *F(x,y)*:

$$F(x,y) = \begin{cases} 1, \textit{если } x \le y \\ 0, \textit{иначе} \end{cases}$$

Тогда к моменту времени *τ* будет прочитано некоторое количество зачтенных часов, а именно:

$$G_\tau = \sum_{t=0}^{\tau} \sum_{k=1}^{N} G_{kt} F(t,\tau)$$

При этом студент будет иметь некоторое количество зачтенных часов, определяемое выражением:

$$V_\tau = \sum_{k=1}^{N} G_{kt} F(A_{k\tau}, H_{k\tau})$$

Нормированную успеваемость студента можно подсчитать как отношение зачтенных часов к прочитанным часам:

$$X_\tau = \sum_{k=1}^{N} \frac{G_k F(A_{k\tau}, H_k)}{\sum_{t=0}^{\tau} G_{kt} F(t,\tau)}$$

Рейтинг студента, выраженный в балах, с учетом трудоемкости изучения отдельных тем как весовых коэффициентов будет:

$$Y_\tau = \sum_{k=1}^{N} \frac{G_k F(A_{k\tau}, H_k) A_{k\tau}}{G_k}$$

Очевидно, что на момент окончания изучения дисциплины (сдачи дисциплины) величина *X* примет значение *100 %* то есть студент освоил все темы. Величину *Y* можно использовать как предварительную (рекомендованную) оценку студента на экзамене или зачете.

На основе непрерывного мониторинга производится динамическая кластеризация студенческой группы.

Пусть величина A_{ik} определяет уровень освоения *i*-тым обучаемым (*i=1..N*) *k*-того раздела (*k=1..M*) дисциплины.

Введем величину, характеризующую степень общности *i*-того и *j*-того обучаемого (дистанция между ними),

$$\rho_{ij} = \frac{1}{M} \sum_{k=1}^{M} \left| A_{ik} - A_{jk} \right|$$

Пусть f_{zap} – радиус кластера

Введем таблицу признаков того, что дистанция между двумя обучаемым менее радиуса кластера:

$$b_{ij} = \begin{cases} 1, \textit{если } \rho_{ij} \le f_{zap} \\ 0, \textit{иначе} \end{cases}$$

Пусть u_i – признак того, что *i*-тый обучаемый является центром кластера

$$u_i = \begin{cases} 1, \textit{если i – тый обучаемый является центром кластера} \\ 0, \textit{иначе} \end{cases}$$

Пусть v_i – признак исключения обучаемого из процесса кластеризации

$$v_i = \begin{cases} 1, \textit{если i – тый обучаемый входит какой – либо кластер} \\ 0, \textit{иначе} \end{cases}$$

Пусть h_i – вектор принадлежности хотя бы одному из кластеров

$h_i = \sum_{j=1}^{N} b_{ij} u_j$

Условие охвата системой обучения всех обучаемых

$h_i + v_i \geq 1, для\ \forall i = 1..N$

Ограничим число студентов, исключенных из процесса кластеризации

$\sum_{i=1}^{N} v_i \leq V_{\max}$

Целевая функция Y определяет число кластеров

$Y = \sum_{i=1}^{N} u_i \to \min$

Подводя итоги, получим формальную постановку задачи кластеризации студенческой группы:

Исходные данные: Уровень освоения тем - A_{ik}

Переменные оптимизации: Признак центра кластера - u_i,
Признак исключения обучаемого из кластеризации - v_i

Ограничения: Предельная дистанция между обучаемыми - f_{zap}
Максимальное количество обучаемых вне кластеров – V_{max}

Целевая функция:

$$Y = \sum_{i=1}^{N} u_i$$

Или:

Целевая функция:

$$Y = \sum_{i=1}^{N} u_i \to \min$$

Ограничения:

$u_i, v_i = двоичное$

$$\sum_{i=1}^{N} v_i \leq V_{\max}$$

$$h_i + v_i \geq 1, для\ \forall i = 1..N$$

Переменные оптимизации: u_i, v_i

Литература

1. Козлов В.В. Организация индивидуализированного сопровождения учебной деятельности студентов в семестре / Актуальные проблемы в строительстве и архитектуре. Образование. Наука. Практика. Материалы 66-й Всероссийской научно-технической конференции по итогам НИР за 2008г. / Под редакцией д.т.н., профессора Чумаченко Н.Г.Самара: СГАСУ, 2009. 288 с.

2. Пиявский С.А. Математическое моделирование при оптимизации сложных систем, монография, Самара, 2008 – 180 с.

3. Козлов В.В. Модель индивидуализированной подготовки специалистов в инфокоммуникационной среде вуза / «Инфокоммуникационные технологии». Периодический научно-технический и информационно-аналитический журнал, том 7, №3, 2009 – с.18 –23 / С.А. Пиявский

4. Козлов В.В. Мониторинг учебной деятельности студента по дисциплине / Естественнонаучное образование в вузе: проблемы и перспективы: сборник II Всероссийской научно-методической конференции (13-14 ноября 2008г.) / Самарск. Гос. Арх.-строит. Ун-т. – Самара, 2008 – 140с.

Интерактивное обучение информатике и основам программирования с использованием когнитивной компьютерной графики

Р.В. Малышев, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

Исходным аргументом для всех рассуждений об эффективности когнитивной визуализации в сфере обучения служит тот факт, что, по мнению психологов, более 80% всей информации, используемой человеком при решении задач, получается им в виде образной зрительной информации. Поэтому использование компьютерной графики очень перспективно в различных областях человеческой деятельности, она широко используется в архитектуре, моделировании и производстве различных изделий, в моделировании физических явлений; в медицине и т.п. Кроме того, компьютерная графика используется для реализации пользовательского интерфейса, что улучшает взаимодействие пользователя с компьютерной системой.

Развитие электронных средств мультимедиа открывает для сферы обучения принципиально новые дидактические возможности: системы интерактивной графики и анимации позволяют в процессе анализа изображений управлять их содержанием, формой, размерами, цветом и другими параметрами для достижения наибольшей наглядности. Применение мультимедиа в электронном, в том числе дистанционном, обучении не только увеличивает скорость передачи информации обучаемым и повышает уровень ее понимания, но и способствует развитию таких важных качеств, как интуиция, профессиональное «чутье», образное мышление. Воздействие интерактивной компьютерной графики на интуитивное, образное мышление привело к возникновению нового направления в проблематике искусственного интеллекта – когнитивной (то есть способствующей познанию) компьютерной графике.

Объекты когнитивной компьютерной графики

Человеческое познание пользуется двумя механизмами мышления. Один из них – возможность работать с абстрактными цепочками символов, с которыми связаны некоторые семантические и прагматические представления. Это – умение работать с текстами в самом широком смысле этого слова. Такое мышление можно было бы назвать символическим или *алгебраическим*. Другой – способность работать с чувственными образами и представлениями об этих образах, которые обладают куда большей конкретностью и интегрированностью, чем символические представления. Но они и значительно более «расплывчаты», «менее логичны», чем то, что скрывается за элементами, с которыми оперирует алгебраическое мышление. Способность работать с чувственными образами (и, прежде всего, со зрительными образами) определяет то, что можно было бы назвать *геометрическим* мышлением.

Возникает необходимость появления специальных *средств работы* со зрительными представлениями и способы перехода от них к текстовым представлениям и обратного перехода, а также таких *моделей представления знаний*, в которых была бы заложена возможность однообразными средствами представлять как объекты, характерные для алгебраического мышления, так и образы-картины, с которыми оперирует геометрическое мышление. Эти комбинированные структуры – основные объекты *когнитивной графики*.

Задачи и требования когнитивной компьютерной графики

Известный специалист в области искусственного интеллекта Д.А. Поспелов сформулировал три основных задачи когнитивной компьютерной графики. Первой задачей является создание *моделей представления знаний*, о которых говорилось ранее. Вторая задача - визуализация тех человеческих знаний, для которых пока невозможно подобрать текстовые описания. Третья - поиск путей перехода от наблюдаемых образов-картин к формулировке некоторой гипотезы о тех механизмах и процессах, которые скрыты за динамикой наблюдаемых картин. С позиций информационных технологий обучения к данным трем

задачам необходимо добавить четвертую: создание условий для развития у обучаемых профессионально-ориентированных интуиции и творческих способностей.

При разработке компьютерных систем инженерного анализа, проектирования и обучения обычно исходят из первых двух задач когнитивной графики, когда знания о техническом объекте, полученные в ходе исследований на многомерных математических моделях и представленные в обычной символьно-цифровой форме, становятся недоступными для анализа человеком из-за большого объема информации.

Четкое осознание третьей и четвертой задач когнитивной графики позволяет формулировать дополнительные требования как к собственно графическим изображениям, так и к соответствующему программно-методическому обеспечению. Среди них можно выделить: адекватность изучаемым объектам или процессам, используемым инженерным методам и методикам обучения; естественность и доступность для восприятия слабо подготовленными или даже неподготовленными пользователями; удобство для анализа качественных закономерностей распределения параметров; эстетическую привлекательность, быстроту формирования изображения.

Применение когнитивной компьютерной графики в интерактивном обучении

В рамках проекта «Школа информатики СГАУ» разработана интерактивная гипертекстовая среда для обучения основам информатики и программирования, в состав которой включены анимационные демонстрации и виртуальные тренажеры, позволяющие в интерактивном режиме настраиваться на решение конкретных задач и «отрабатывать» технологические навыки, при этом на любое действие обучаемого наблюдается «реакция» со стороны обучающей среды. Для разработки был выбран графический пакет Macromedia Flash, который позволяет интегрировать различные виды информации: графические, текстовые и звуковые данные, предлагает широкие возможности по созданию анимации. Кроме того, современные Flash-технологии позволяют создавать небольшие по объему анимационные файлы, что немаловажно для систем дистанционного обучения, и обеспечивают независимость от разрешающей способности экрана монитора пользователя.

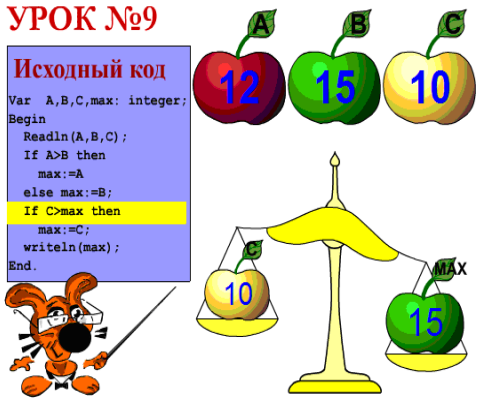


Рис. 1. Пример урока с применением интерактивного тренажера

Программа каждого урока содержит необходимый текстовый материал, а также мультимедиа объекты с интуитивно понятными образами, которые позволяют легко запоминать и воспринимать материал (рис. 1). На данный момент разработаны уроки по темам: «Системы счисления», «Двоичная арифметика», «Базовые алгоритмы программирования». В будущем планируется дополнить программу обучения проверочными и тестовыми заданиями, что позволит закрепить полученные знания по курсу обучения.

В заключение хотелось бы отметить, что когнитивная компьютерная графика позволяет изменить процесс передачи знаний, сделать его более интересным и удобным для обучаемого.

Программный комплекс формирования региональных образовательных целевых программ

М.А. Мартынов
Самарский государственный архитектурно-строительный университет, Самара

В настоящее время актуальна проблема образования и развития одаренной молодежи. Это, прежде всего, связано с потребностью общества в неординарной творческой личности. Неопределенность современной окружающей среды требует не только высокую активность человека, но и его умения, способности нестандартного поведения. Исходя из этих предпосылок, мною был создан макет целевой комплексной программы «Одаренная молодежь».

Любая целевая программа, представляет собой увязанный по ресурсам, исполнителям и срокам осуществления комплекс заданий и мероприятий, направленных на эффективное решение определенной проблемы. Макет целевой комплексной программы (ЦКП) – это модель, отражающая цели и возможные пути их достижения, которая позволяет получить представление о реальной ЦКП.

Залог успеха любого исследования – его *системность и комплексный подход* к решению поставленных задач. В работе используется комбинация математических и психологических методов исследования, для наилучшего достижения целей. Для реализации этих методов был разработан программный комплекс (ПК) формирования и оптимизации региональных ЦКП (ФИОРЦКП). Разработанный ПК состоит из двух блоков. Первый блок позволяет сформировать структуру ЦКП и внести все данные программы в базу данных. Второй блок является оптимизационным, который по итогам расчета формируется набор мероприятий, наиболее удовлетворяющий заданным целям. Интерфейс позволяет пользователю влиять на результаты с помощью модулей ввода ограничений на затраты по подпрограммам, целям, принципам, функциональным задачам, показателям эффективности, охвату населения, а также с помощью расстановки целевых коэффициентов и предельных затрат.

На основе выходных параметров макета ЦКП ОМ был проведен анализ путем перебора значений целевых коэффициентов по следующим данным:

1. Целевая функция
2. Выделенные денежные средства
3. Удовлетворение каждой из задач
4. Общая удовлетворенность
5. Список мероприятий вошедших в программу

По значениям целевой функции и общей удовлетворенности целей программы были выбраны четыре наиболее оптимальных варианта. Выбранные модели практически индеитичны, что указывает на высокую результативность работы программы.

Так с помощью исследований была доказана практическая значимость использования целевых коэффициентов и ограничений, что является подтверждением «активности» программно-целевого метода – он позволяет не только наблюдать ситуацию, но и влиять на ее возможные последствия.

Данный комплекс позволяет формировать региональные ЦКП за короткий промежуток времени (от нескольких минут до одного часа), в то время как в организациях это делается в ручную (несколько дней).

Разработка виртуального тренажёра для дистанционного обучения основам информатики и программирования

А.В. Невинский, М.А. Шершнев, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Результаты единого государственного экзамена (ЕГЭ) по информатике в Самарской области выявили многочисленные пробелы в знаниях по данному предмету. К сожалению, школьная подготовка по информатике не соответствует тому уровню требований, которые предъявляет современный вуз к абитуриенту, многие старшеклассники не знают базовых понятий информатики и информационных технологий, особенно это характерно для выпускников сельских школ.

Известно, что дети лучше всего учатся в процессе деятельности, удовлетворяющей присущую ребенку любознательность: игры-исследования, творческое моделирование и конструирование и т.п. Поэтому необходима разработка виртуальных тренажеров, которые позволят в интерактивном режиме настраиваться на решение конкретных задач и «отрабатывать» требуемые технологические навыки. Применение виртуальных тренажеров в учебном процессе позволит повысить эффективность усвоения знаний, понимание материала. Виртуальные тренажеры могут быть использованы в дистанционном обучении, электронных учебниках, автоматизированных системах тестирования знаний и других учебных программных средствах с различной степенью интерактивности.

Роль виртуальных тренажеров в обучении

Наиболее перспективное применение виртуальных тренажеров – использование их при дистанционном обучении, т. к. они позволяют: освоить предметную область на разных уровнях глубины и детальности; проверить знания в игровой форме; получить навыки решения типовых практических задач. Виртуальные тренажеры являются дополнением к основному учебному материалу. В системах дистанционного обучения все учебные материалы представлены в электронном виде: базовый теоретический материал – обычно гипертекстовый документ, дополненный перекрестными ссылками, позволяющими перейти к ранее изученным разделам; тематические глоссарии, справочники; дополнительная учебная и методическая литература. Учебный материал может сопровождаться интерактивной анимацией по технологии flash и 3D-анимацией и т.п.

Система управления обучением LMS Moodle

Наиболее популярной учебной средой при дистанционном обучении является система управления обучением LMS Moodle, которая предназначена для организации обучения с использованием технологий Интернет. К достоинствам системы можно отнести:

– поддержку многообразных процедур обучения, комбинированием которых может быть организовано эффективное обучение в образовательном учреждении;

– предоставление возможности инсталляции образовательных ресурсов (учебных материалов) и обеспечение средствами доступа к ресурсам и управления ими;

– обеспечение коммуникационного взаимодействия участников образовательного процесса, реализуемое в форме интернет-конференций – форумов, дискуссий, а также обмена посланиями, содержащими, в том числе, задания обучаемым, решения заданий и комментарии;

– модульный принцип построения и открытый код, что позволяет расширять функциональность системы за счет включения авторских модулей.

LMS Moodle переведена на десятки языков, в том числе и русский, и используется почти в 50 тысячах организаций из более чем 200 стран мира. В РФ зарегистрировано более 400 инсталляций, количество пользователей в некоторых инсталляциях достигает 40 тысяч человек. LMS Moodle написана на языке PHP с использованием SQL-базы данных, может работать с объектами SCO и отвечает стандарту SCORM.

Проект «Школа информатики СГАУ»

Одно из важнейших направлений учебного процесса в старшем звене средней школы – подготовка к централизованному тестированию по обязательным и профилирующим предметам, в том числе и по информатике. Обеспечить преемственность процесса обучения – одна из основных задач вуза, в связи с этим на факультете информатики была создана школа информатики СГАУ, цель которой – подготовить школьников по информатике и программированию для поступления в высшие учебные заведения на специальности, связанные с изучением информационных технологий, в первую очередь на факультет информатики СГАУ.

Занятия в школе информатики пока проводятся в традиционной (очной) форме, но предполагается внедрение дистанционной формы обучения. В рамках проекта разработана и проходит опытную эксплуатацию система дистанционного обучения школы информатики (distance.itschool.ssau.ru), построенная на базе LMS Moodle. В состав системы дистанционного обучения входят:

– гипертекстовая среда обучения основам информатики и программирования, включающая в себя учебный материал, представленный в текстовом, табличном, графическом виде, мультимедиа-объекты (фото, звук, анимации и т.п.), и тематические глоссарии и словари;

– гибкий комплект сервисов: форумы, рабочие тетради, тесты, ресурсы, опросы, анкеты, задания, чаты и практикумы;

– система тестирования знаний, которая гибко конфигурируется в зависимости от поставленной задачи и позволяет представить конечный тест практически в любом желаемом виде.

Несмотря на то, что система тестирования поддерживает около десяти различных видов тестов, в ней не предусмотрено выполнение творческих заданий по программированию, которые можно было бы выполнить в режиме on-line тестирования. В связи с этим появилась необходимость реализации авторского модуля «Виртуальный тренажер по программированию», который можно было бы интегрировать в среду Moodle, и который позволял бы преподавателю (разработчику учебного курса) создавать тестовые задания с развернутым ответом для проверки знаний блока С ЕГЭ по информатике.

Описание возможностей модуля «Виртуальный тренажер по программированию»

Экзаменационная работа ЕГЭ по информатике состоит из трех частей, каждый из которых имеет несколько уровней сложности:

– блок А – задания с выбором ответа;

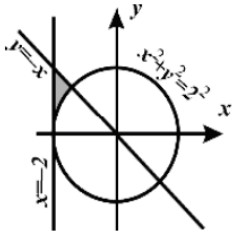
– блок В – задания с кратким ответом;

– блок С – задания с развернутым ответом.

Как показала практика, наиболее сложными для школьников является задачи блока С, которые призваны проверить знания по основам алгоритмизации и программированию. В блок С входят четыре задания:

– С1 - умение прочесть фрагмент программы на языке программирования и исправить допущенные ошибки (рис. 1);

С1



Требовалось написать программу, при выполнении которой с клавиатуры считываются координаты точки на плоскости (x,y – действительные числа) и определяется принадлежность этой точки заданной заштрихованной области (включая границы). Программист торопился и написал программу неправильно.

ПРОГРАММА НА ПАСКАЛЕ	ПРОГРАММА НА БЕЙСИКЕ	ПРОГРАММА НА СИ
var x,y: real; begin readln(x,y); if x*x+y*y>=4 then if x>= -2 then if y<= -x then write('принадлежит') else write('не принадлежит') end.	INPUT x, y IF x*x+y*y>=4 THEN IF x>= -2 THEN IF y<= -x THEN PRINT "принадлежит" ELSE PRINT "не принадлежит" ENDIF ENDIF ENDIF END	void main(void) { float x,y; scanf("%f%f",&x,&y); if (x*x+y*y>=4) if (x>= -2) if (y<= -x) printf("принадлежит"); else printf("не принадлежит"); }

Последовательно выполните следующее:
1) Приведите пример таких чисел x, y, при которых программа неверно решает поставленную задачу.
2) Укажите, как нужно доработать программу, чтобы не было случаев ее неправильной работы. (Это можно сделать несколькими способами, достаточно указать любой способ доработки исходной программы).

Рис. 1. Пример задания С1

– С2 - умение написать короткую (10 – 15 строк) простую программу (например, обработки массива) на языке программирования или записать алгоритм на естественном языке;

– С3 - умение построить дерево игры по заданному алгоритму и обосновать выигрышную стратегию;

– С4 - умение создавать собственные программы (30 – 50 строк) для решения задач средней сложности.

За подготовку заданий блока С отвечает пользователь системы, которому приписана роль «*Разработчик учебного курса*». Система предоставляют ему элементы ввода и управления для выполнения следующих действий:

1. Выбор категории вопроса и задание названия.
2. Ввод условия задания.
3. Ввод эталонного и неправильного кода.
4. Ввод количества баллов за выполнение каждого из заданий.
5. Ввод массива тестовых данных.

Все задания сохраняются в базе данных и доступны для дальнейшего использования и редактирования.

Тестовые задания блока С могут выполнить только зарегистрированные и подписанные на учебные курс пользователи, им присваивается роль «Обучаемый». Система предоставляет *обучаемому* текст условия задания и элементы для ввода данных, необходимых для решения. Вводимый код может быть написан на любом из

трех языков программирования: C, Pascal, Basic, в зависимости от выбора пользователя (рис. 2). Оценка работы обучаемого производится автоматически.

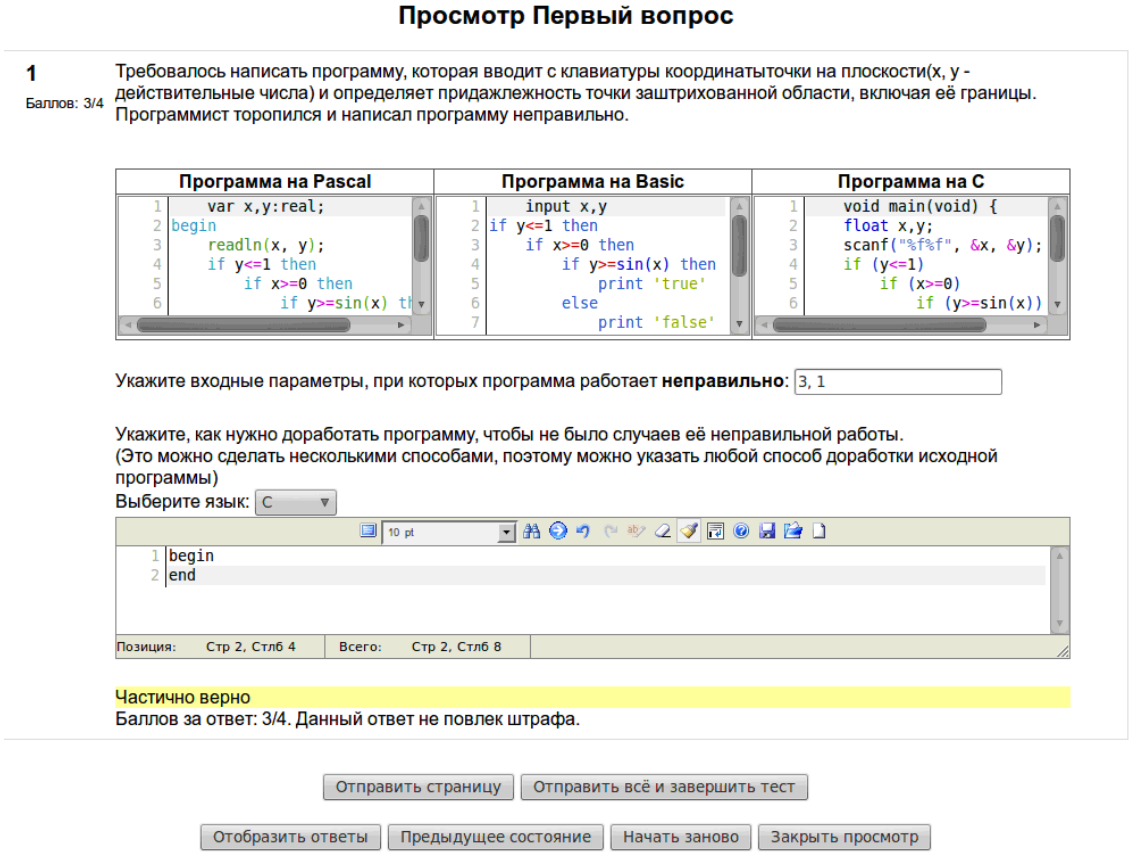


Рис. 2. Экранная форма решения вопроса типа C1

Выводы

Виртуальный тренажер по программированию окажет неоценимую помощь как обучаемому (он позволяет повторять упражнения до полного понимания и закрепления изучаемого материала, приобрести навыки при решении типовых задач и устранить пробелы в знаниях), так и преподавателю (не требуется его постоянного присутствия, тренажер беспристрастно оценивает полученное решение, методические рекомендации выдаются автоматически, программа сама указывает на допущенные ошибки).

Рейтинговая оценка текущей успеваемости студентов СГАУ

В.В. Никишкин, Д.С. Оплачко, Е.И. Чигарина
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В настоящее время в Самарском государственном аэрокосмическом университете (СГАУ) разработана автоматизированная интернет-система контроля текущей успеваемости и посещаемости студентов. Основная цель разработанной системы – постепенный переход к объективным методам оценки уровня знаний студентов. Система предназначена для заполнения преподавателями и просмотра студентами и родителями через интернет сведений о текущей успеваемости студентов факультета информатики. Текущая успеваемость и посещаемость в течение семестра контролируется на определенных неделях, а именно 6, 10 и 14 до начала сессии. Для решения этих задач в системе реализован доступ к базе данных, учитывающий рабочий учебный план на семестр и распределение нагрузки по преподавателям кафедр факультета, а также позволяет учесть все виды занятий по каждому предмету семестра.

Интернет-система имеет удобную систему ссылок на справочную информацию по специальностям и кафедрам факультета СГАУ, возможность поиска сведений о группе и специальности студента по указанной фамилии и в наглядной форме получения сведений об успеваемости и посещаемости студентов в течение семестра.

Кроме того, интернет-система учета текущей успеваемости студентов СГАУ позволяет получить отчеты об успеваемости в виде файлов в формате Excel по группам, отчеты по студентам, имеющим задолженности по контрольным точкам по конкретному предмету, а также вывести список студентов заданной группы или специальности, имеющих задолженности более чем по трем предметам семестра, что облегчает работу учебного отдела и деканатов вуза по контролю за учебным процессом.

Программа обеспечивает выполнение следующих функций:

- Доступ к базе данных системы.
- Возможность заполнения сведений об успеваемости и посещаемости студентов преподавателями.
- Разграничение прав доступа пользователей и контроль целостности данных.
- Возможность просмотра по группам сведений о текущей успеваемости студентов.
- Возможность получения справочной информации по специальностям и кафедрам факультета.
- Формирует следующие отчеты:
  - просмотр 6 формы по группе;
  - должники по группе;
  - должники по предмету;
  - должники по преподавателю.

Программная реализация итнернет-системы выполнена с помощью средств языка программирования PHP5 под управлением CMS Drupal 6. База данных системы реализована под управлением СУБД MySQL 5, логическая модель которой представлена на рис. 1.

Слово «рейтинг» происходит от английского «to rate» (оценивать) и «rating» (оценка, оценивание).

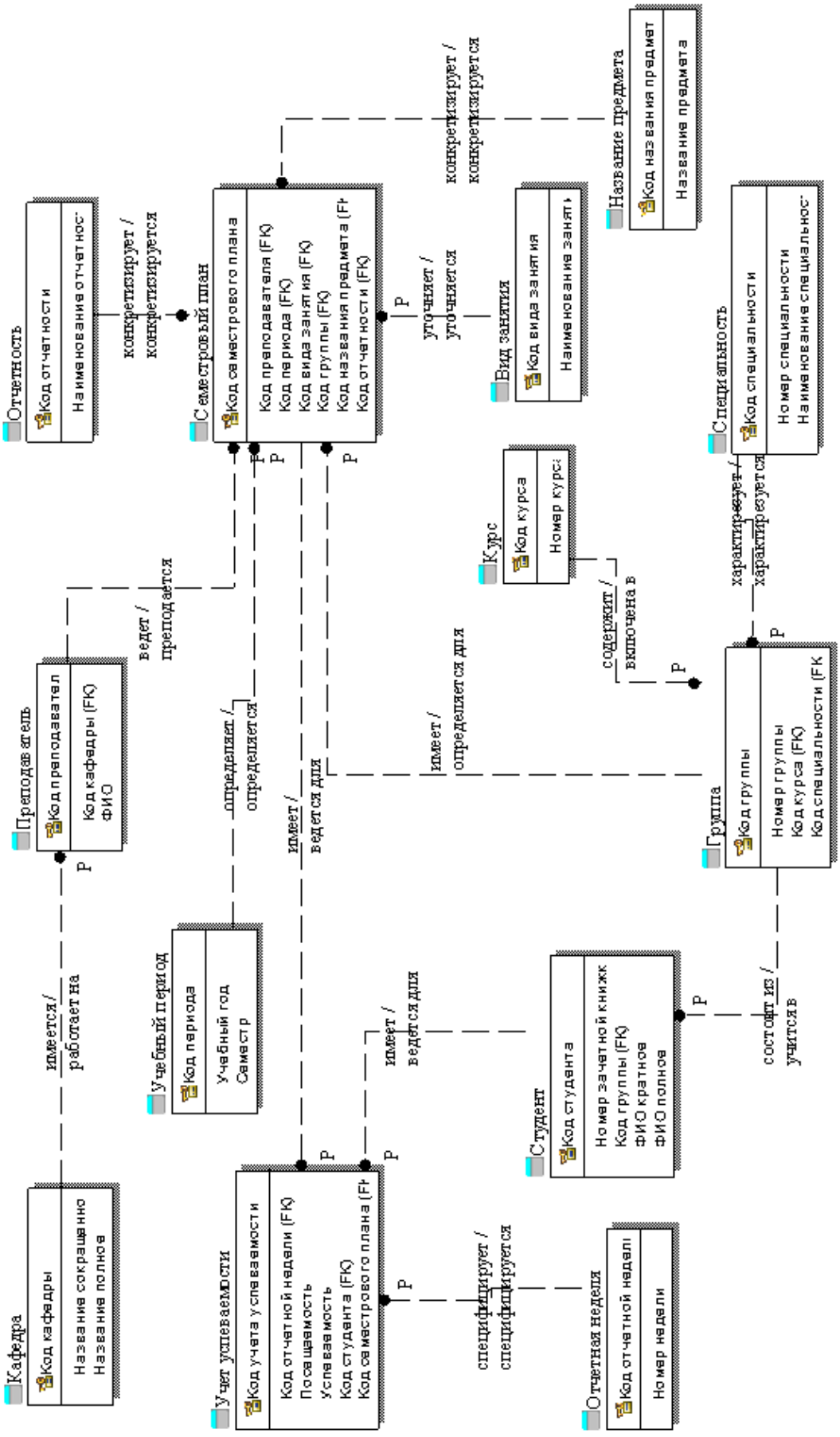


Рис. 1. Логическая модель базы данных интернет-системы

Рейтинговая технология оценивания результатов обучения студентов по некоторой дисциплине основана на учете накапливаемых ими оценок в баллах за выполнение текущих работ (лабораторных, курсовых, практических и других видов) и регулярно проводимых контрольных мероприятий. Так же оценка рейтинга учитывает показатель посещаемости студентом занятий. В отличие от традиционного способа оценивания, рейтинговая технология предполагает последовательное суммирование оценок студента по данной дисциплине в течение некоторого периода времени. Текущая рейтинговая оценка по дисциплине складывается из оценок всех без исключения практикуемых в вузе видов учебной работы и контроля знаний за 6-ю, 10-ю и 14-ю отчетную неделю.

Целью введения рейтинговой системы оценки успеваемости студентов является комплексная оценка качества учебной работы студентов в процессе обучения по программам высшего профессионального образования.

Главные задачи рейтинговой системы заключаются в:

- повышении мотивации студентов к освоению образовательных программ путем более высокой дифференциации оценки их учебной работы;
- повышении уровня организации образовательного процесса в ВУЗе.

Рейтинговая оценка может быть вычислена за любой контрольный период или за весь семестр в целом, по конкретному предмету или по всем предметам, преподаваемым в данном семестре по специальности. Полученные данные для наглядности отображаются на диаграмме, где для каждого студента выбранной группы указан его рейтинг. Таким образом, визуально можно получить информацию, как об относительной, так и об абсолютной успеваемости студентов.

Формула, для вычисления рейтинга студента, имеет следующий вид:

$$Р= Кт \cdot Кв \cdot 100 \cdot (1 – Кп \cdot (1 – П) / 100)) \cdot (1 – Ку \cdot (1 - У / 100)),$$

где

Р – рейтинг студента;
Кт – коэффициент вида занятий;
Кв – коэффициент важности предмета;
Кп – коэффициент посещаемости, П – посещаемость в процентах;
Ку – коэффициент успеваемости, У – успеваемость;

Коэффициент вида занятий - является численной оценкой важности вида занятий (лекции, практики, курсового) и определяется деканатом. По умолчанию коэффициент равен единице.

Коэффициент важности предмета – является численной оценкой важности предмета в данном семестре у группы по специальности, определяется деканатом. По умолчанию коэффициент равен единице.

Коэффициент посещаемости – коэффициент, учитывающий посещаемость предмета и влияющий на итоговый рейтинг студента, определяется преподавателем.

Коэффициент успеваемости – коэффициент, учитывающий успеваемость по предмету и влияющий на итоговый рейтинг студента, определяется преподавателем.

Из формулы видно, что максимальных рейтинг студента, при обычных условиях, будет находиться в районе 100 баллов, но при этом для каждого предмета преподаватель может установить свои правила для набора максимального количества баллов. Так же, коэффициент важности предмета позволяет установить максимальный балл выше для наиболее сложных предметов в семестре.

Система работает через web-интерфейс, что позволяет получить информацию об успеваемости студента или группы любому желающему. Студент и родители могут просмотреть показатели рейтинга для контроля успеваемости, как по предмету, так и за семестр в целом.

Информационная технология формирования исследовательских компетенций студентов НИТУ

С.А. Пиявский
Самарский государственный архитектурно-строительный университет, Самара

Переход высшей школы на Федеральные образовательные стандарты третьего поколения, реализующие компетентностную парадигму образования, требует реорганизации всего учебно-воспитательного процесса в вузе. Особенно это справедливо в отношении национальных исследовательских технических университетов (НИТУ). Многолетний опыт Линии элитной подготовки, а впоследствии факультета информационных систем и технологий СГАСУ позволяет предложить целостную систему формирования универсальных (исследовательских, общекультурных, социально-личностных) компетенций студентов на основе специально разработанной информационной технологии.

Чтобы осуществить полноценный переход от «знаниевой» к «компетентностной» парадигме образования, необходимо перейти от линейной к матричной структуре студенческого коллектива и выстроить достаточно мощную организацию внеучебной деятельности студентов. Созданная система должна охватывать каждого студента в течение всего периода обучения в вузе, вести мониторинг формирования у него универсальных компетенций и организовывать его деятельность в этом направлении.

Организационным стержнем такой системы, придающий взаимодействию со студентами регулярный, целенаправленный и контролируемый характер, является наддисциплинарный учебный курс, охватывающий весь период обучения. Каждый обучаемый входит не только в свою студенческую группу, вместе с которой он изучает учебные дисциплины. Он входит также в какую-либо бригаду научной группы, занятой решением конкретной научной проблемы. Научную группу возглавляет ученый – специалист в соответствующей научной области. Решаемая научной группой проблема разбита на ряд задач, решение которых занимаются отдельные бригады. Бригада – это разновозрастный студенческий коллектив. Его возглавляет магистрант или пятикурсник. Этот коллектив включает цепочку студентов с первого по четвертый – пятый курсы, в которой старшие руководят младшими.

Технологической основой системы являются автоматизированные информационные системы (АИС) мониторинга внеучебной и, в первую очередь, научно-исследовательской деятельности студентов, объединенные с АИС мониторинга учебной деятельности.

В общих чертах, организация образовательного процесса на факультете в условиях компетентностной парадигмы представляется следующим образом.

Профессиональные компетенции вырабатываются в процессе учебной деятельности студента и оцениваются по принятой в вузе системе оценки академической успеваемости по результатам сдачи зачетов и экзаменов в сессию. Поддержку учебного процесса осуществляет соответствующая АИС. В ней на основе промежуточного и итогового тестирования на Интернет-портале ведутся электронные журналы студенческих групп. В журналах старостами с ведома преподавателей отражаются посещение студентами занятий и результаты прохождения ими регулярного промежуточного контроля на усвоение отдельных дидактических единиц учебных дисциплин. Это позволяет системе еженедельно вычислять учебные рейтинги каждого студента, студенческой группы, курса, предмета и строить вокруг этих данных работу со студентами, их родителями (на младших курсах) и преподавателями.

Общенаучные и инструментальные компетенции, в основном, вырабатываются в процессе выполнения студентами индивидуальных исследовательских работ под руководством ученых вуза и старшекурсников. Предусмотрено выполнение каждым студентом в каждом семестре одной курсовой работы творческого характера по индивидуальной тематике. При этом методическая структура работ меняется от семестра к семестру по научно разработанной системе, позволяющей последовательно и гармонично развивать различные составляющие компетенций. Комиссионная оценка работ в конце каждого семестра по специально разработанной скользящей системе критериев дает количественную оценку различных компонентов творческих (общенаучных и инструментальных) компетенций и позволяет рассчитать творческий рейтинг каждого студента. Они также оцениваются по результатам работы (публикациям, докладам на конкурсах и конференциях, особенно на ежегодной студенческой научной конференции вуза, допускающей введение единой методически обоснованной системы критериев оценки).

Сами работы размещаются на личных сайтах студентов, что создает за период обучения подробное электронное «портфолио», отражающее их творческий потенциал. АИС мониторинга формирования научных компетенций студентов создает технологическую основу для этой деятельности. Ее целевое назначение состоит, в частности, в том, чтобы скоординировать и сделать максимально нетрудоемкой и эффективной деятельность научных руководителей студенческих НИР, а также увлечь студентов наглядно наблюдаемым процессом «вызревания» их способностей.

Социально-личностные и общекультурные компетенции вырабатываются в процессе внеаудиторной активности студентов: в трудовой деятельности, спорте, науке и технике, искусстве, общественной деятельности и пр. Они оцениваются по определенной шкале на основании сведений, вносимых самими студентами или осведомленными лицами в АИС мониторинга формирования личностных компетенций студентов. Она позволяет существенно усилить воздействие организаторов учебно-воспитательного процесса на эту деятельность за счет придания ей измеримого, соревновательного, в хорошем смысле, характера. С этой целью на соответствующем Интернет-портале ведется еженедельно обновляемая виртуальная доска почета, связанная с общедоступной базой личных достижений. Любой член коллектива может через Интернет записать информацию о своих или чужих достижениях в различных сферах жизни. Информация немедленно становится видной всем. После того, как она оценивается по значимости ответственным членом студенческого совета, она автоматически учитывается при расчете рейтинга активности соответствующего студента. Значимость достижения со временем плавно убывает, таким образом, давние достижения имеют меньший «вес».

Непрерывный автоматический расчет в АИС для каждого студента трех рейтингов: учебного, творческого и личностного позволяет показать наглядно достигнутый им уровень развития в трехмерном пространстве на фоне выбираемой им референтной группы других студентов. Такой подход, зримо и объективно отражающий формирующуюся компетентность студентов, оказывает определенное влияние на их самооценку, способствует тому, что они с интересом наблюдают за динамикой собственного развития, а интерес, естественно, вызывает и желание «стать лучше». Еще большее влияние этот подход оказывает на преподавателей и организаторов учебно-воспитательного процесса.

Функционирование АИС позволяет сформировать новое приложение к диплому, выдаваемому выпускнику вуза, так называемое портфолио. В нем указывается, в какую часть выпуска (первую десятку, первые 25%, вторые 25% и т.д.) входит выпускник по своему учебному, творческому и личностному рейтингам, подсчитываемым

информационно-коммуникационной системой вуза в течение всего периода его обучения, а также адрес его веб-страницы на официальном сервере вуза. Кроме того, в портфолио перечисляются основные достижения студента в различных областях деятельности за период его обучения. Мы полагаем, что без этого компетентностный подход никогда не станет живым делом, непосредственно ориентирующим студента на повседневную всестороннюю активность в вузе. Ведь в современном динамичном обществе выпускнику предстоит работать в разных странах, у разных работодателей. Указанное приложение к диплому откроет им выпускника вуза с максимальной полнотой.

Почему в основу формирования компетенций студентов на ФИСТ положена исследовательская деятельность?

Исследовательская деятельность в максимальной степени аккумулирует в себе компетенции, входящие в группы общенаучных и инструментальных, а на старших курсах – и профессиональных компетенций.

Новые виды образовательной деятельности, порожденные переходом к компетентностной парадигме образования, не могут вырасти «на пустом месте». Они должны быть результатом развития уже существующих массовых форм внеучебной работы с молодежью в вузе и их переосмысления в сторону расширения спектра задач. Из них наиболее развитой и охватывающей максимальное число компетенций является научная исследовательская работа студентов (НИРС). Конечно, в существующем виде система НИРС не решает задач управляемого развития универсальных компетенций всех студентов, такая задача перед ней и не ставилась. Да и сама по себе она не свободна от недостатков. Однако она может стать основой для новой системы управляемого формирования универсальных компетенций студентов, причем в рамках этой системы более эффективно решить и свои изначальные задачи – поиск и развитие научных способностей наиболее одаренной в творческом отношении молодежи, подготовки кадрового резерва науки.

Каждый преподаватель высшей школы в той или иной мере занимается научными исследованиями. Следовательно, для ведения подобного курса уже существует готовый преподавательский коллектив.

Существенным является вопрос: возможно ли обеспечить индивидуальное руководство ВСЕМИ студентами в течение ВСЕГО периода обучения при том, что в технических вузах штатный коэффициент ППС составляет 10-12 студентов на одного преподавателя? При этом часть их из них, например, преподаватели кафедр физвоспитания, не могут вести руководство НИРС. Таким образом, на одного преподавателя, который может участвовать в руководстве НИРС в вузе приходится около 15 студентов, а то и больше.

Вопрос решается положительно, если учесть, что органической частью предлагаемого подхода к компетентностной парадигме вузовского образования является АКТИВНОЕ вовлечение в организацию образовательного процесса всего студенческого коллектива. Именно этот деятельностный подход обеспечивает достижение целей образования. В соответствии с этим подходом, на 1 – 3 курсах индивидуальное руководство (скорее, шефство) студентами осуществляют старшекурсники. Преподаватели, аспиранты, магистранты при этом руководят лишь «продвинутыми» студентами 2-3 курсов и всеми студентами 4-5 курсов.

Конечно, руководство студентов студентами нуждается в контроле и помощи со стороны квалифицированных преподавателей. Поэтому организационно-методически исследовательская деятельность оформлена в виде специального учебного курса «Технология исследовательской профессиональной деятельности». Он является наддисциплинарным (то-есть интегрирует знания, умения и навыки, приобретаемые

студентами при освоении многих других дисциплин) и сквозным (то-есть преподается в течение всего периода обучения).

Введения наддисциплинарного учебного курса имеет три основные цели:

1. Обеспечить в течение всего периода обучения студента в вузе индивидуальное целенаправленное оперативное управление его компетентностным развитием, имея в качестве субъектов управления организаторов учебно-воспитательного процесса, преподавательский персонал, а также самого студента.

2. Обеспечить высокий уровень сформированности универсальных компетенций обучаемых.

3. Расширить объем и повысить уровень научно-исследовательской работы выпускающей кафедры за счет вовлечения в научные коллективы талантливой мотивированной молодежи.

Курс читается в течение всего периода обучения по 2 часа в неделю. Половину его составляют лекции, половину – семинары-консультации. В каждом семестре студент выполняет курсовую исследовательскую работу и защищает ее перед комиссией.

Результаты обучения по дисциплине оцениваются в каждом семестре зачетом с оценкой, который проставляется коллегиально на основе оценки выполненных курсовых работ и проектов по комплексной системе критериев. При постановке темы работы она в предположительном плане оценивается по этой системе критериев, так что становится ясно, на какую оценку она «потянет»; затем в середине семестра руководитель снова совместно со студентами производит такую «прикидку».

В студенческом коллективе в начале каждого семестра выделяется группа «продвинутых» студентов, проявляющих признаки одаренности в старшем школьном и студенческом возрасте и ориентированных на развитие своих творческих способностей. Это мотивированные еще в школьный период участники и победители различных профильных олимпиад и конференций исследовательских работ молодежи, а также отлично и хорошо успевающие студенты, перед которыми преподаватели обязаны развернуть перспективы занятий творческой деятельностью. Им, помимо общего плана занятий, предлагаются темы курсовых работ и проектов исследовательского характера, с ними проводятся еженедельные индивидуальные консультации.

«Исследовательская составляющая» является как бы «несущей частотой» наддисциплинарного курса, но ею не исчерпывается его содержание.

Ведение курса обеспечивает комплексный преподавательско-студенческий коллектив, сформированный на базе выпускающей кафедры. В дальнейшем будем называть его организационным коллективом. В него входят:

1) научный руководитель направления подготовки студентов,
2) преподаватели потоков,
3) руководители научных групп,
4) индивидуальные научные руководители студентов,
5) руководитель внеучебной деятельности студентов,
6) руководители научных бригад (студенты)
7) старосты, учебные и внеучебные аналитики учебных групп (студенты).

Опишем проведение наддисциплинарного курса в течение семестра.

Первым этапом является распределение студентов по научным руководителям. Это тонкий процесс, непосредственное участие в котором принимает научный руководитель направления подготовки. Он совместно с преподавателями потоков распределяет научных руководителей по студентам. Затем преподаватели потоков в предварительном порядке согласовывают с научными руководителями темы индивидуальных исследовательских работ. Эта информация вводится преподавателями

потоков в АИС и на занятии доводится до сведения студентов, которым даются координаты научных руководителей. В течение недели студенты должны обсудить со своими руководителями тематику и содержание предполагаемых исследовательских работ, составить и ввести в АИС аннотации работ.

Второй этап состоит в предварительной оценке творческого уровня предлагаемых работ по специальной системе критериев с помощью АИС. Эта оценка производится на занятии с использованием проектора и обсуждением со всеми студентами. Анализируется, насколько выполнение работы повысит творческий рейтинг студента и гармонизирует структуру его творческих компетенций. В процессе обсуждения содержание работы и, соответственно, аннотация корректируются.

Третий этап, занимающий весь семестр, включает два параллельных сюжета.

Во-первых, выполнение студентом запланированной работы с еженедельными консультациями научного руководителя и преподавателя потока. При этом научный руководитель и преподаватель записывают соответственно комментарии и задания в «электронный дневник» студента в АИС, а сам студент непосредственно до очередного аудиторного занятия заносит в АИС краткий отчет о ходе работы. Такая система побуждает студента регулярно заниматься выполнением исследования, дает возможность своевременно его направлять. Кроме того, важно, что анализ хода работы каждого студента буквально «на глазах» его товарищей (с использованием проектора) позволяет заслуженно поощрить или пристыдить его перед одногруппниками, мнением которых он дорожит, возможно, больше, чем преподавательским.

Во-вторых, привлечение внимания студентов к содержанию и интенсивности их учебной и внеучебной деятельности, отражаемой соответствующими АИС. Эту функцию выполняет преподаватель потока на аудиторных занятиях в контакте с ответственными за воспитательную работу на факультете. Плодотворность такой работы зависит от того, имеется ли в вузе достаточная система стимулов успешной деятельности студентов в семестре, за которые молодым людям было бы интересно бороться.

Четвертый этап проходит в середине семестра на аудиторном занятии с участием научного руководителя направления подготовки и, по возможности, индивидуальных научных руководителей. Он предусматривает коллективное (с помощью проектора) обсуждение выполняемой работы каждого студента по специальной системе критериев с помощью АИС с позиций ее перспектив при защите в конце семестра. По итогам обсуждения уточняется аннотация работы, а, возможно, и сама ее тематика.

Наконец, пятый, заключительный, этап состоит в гласной защите на аудиторном занятии каждым студентом выполненной исследовательской работы перед комиссией, состоящей из преподавателя потока и научного руководителя направления подготовки. Ими производится оценка работы по специальной системе критериев с помощью АИС, тем самым измеряется продемонстрированный студентом уровень сформированности его общенаучных, инструментальных и, частично, профессиональных компетенций. На этом же занятии подводится итог сформированности компетенций, проявляющихся во внеаудиторной трудовой, культурной, спортивной и общественной сферах деятельности.

Управление учебной и внеучебной деятельностью студента в вузе на основе информационно-коммуникационных технологий

С.А. Пиявский, З.Ф. Камальдинова
Самарский государственный архитектурно-строительный университет, Самара

Введение

На факультете ИСТ СГАСУ под руководством профессора, д.т.н. С.А. Пиявского разработана и внедрена система управления учебной и внеучебной деятельностью студентов на основе информационно-коммуникационных технологий. Система размещена на портале факультета по адресу: www.sciyouth.ru. Рассмотрим положения обеспечивающие работоспособность системы.

Управление учебной деятельностью студентов

Данные для управления учебной деятельностью студентов предоставляются системой мониторинга процесса обучения и вычислением учебного рейтинга. Учебный рейтинг рассчитывается как среднее арифметическое двух рейтингов: посещаемости занятий и регулярной отчетности (результатов освоения разделов дисциплин по контрольным точкам с регулярностью, предусмотренной учебным графиком).

Контрольные точки служат для фиксации регулярности и качества работы студента в семестре. По каждой дисциплине предусматривается **не менее шести** контрольных точек в семестре. Имеется возможность использовать при оценке результатов работы студента разработанную на ФИСТ тестовую систему с ориентацией на федеральный Интернет-экзамен по профессиональному образованию. Результаты аттестации студента по контрольной точке впоследствии изменению не подлежат, т.к. основное назначение контрольной точки – именно контроль регулярности работы студента в семестре.

Данные для расчета учебного рейтинга представитель деканата получает у преподавателя каждой дисциплины, посетив его на плановом учебном занятии, совпадающем с очередной контрольной точкой. Преподаватель обязан при этом проставить в ведомость, которую ему передает представитель деканата, число пропусков занятий каждым студентом с начала семестра и аттестовать студента в выполнении им учебного плана на текущий момент. Аттестация проводится по 3-х бальной системе: 0- не аттестован, 1 – аттестован, 2 – аттестован с отличием.

Сведения о пропусках занятий по уважительным причинам вводятся в систему мониторинга на основании документов, представленных студентом в деканат не позднее чем через неделю после прекращения действия причины, вызвавшей пропуски. После этого никакие документы во внимание не принимаются.

Допуская к занятию опоздавшего студента, преподаватель отмечает ему один час пропуска занятия по неуважительной причине.

В исключительных случаях положительно зарекомендовавший себя студент может подать декану заявление на обучение по индивидуальному графику, предусматривающее сочетание различных форм обучения (п.3.1 Устава СГАСУ). В случае согласия преподавателя и разрешения декана учет посещаемости по ряду дисциплин может не вестись.

Управленческие воздействия на основе учебного рейтинга

Студенту, аттестованному по всем контрольным точкам по дисциплине, **автоматически выставляется зачет или экзамен по дисциплине с оценкой как минимум «удовлетворительно»**. *(Это, в частности, повышает ответственность преподавателя за качество аттестации по отдельным контрольным точкам).* При этом по итогам обучения в семестре, по усмотрению преподавателя, студенту может быть выставлена и более высокая оценка. Студент при желании может сдавать зачет или экзамен на более высокую оценку в обычном порядке.

Студент, имеющий неаттестованные контрольные точки по дисциплине, допускается к сдаче зачета или экзамена по этой дисциплине в обычном порядке **без дополнительных условий.**

Студент может быть представлен к отчислению из вуза за систематические пропуски аудиторных занятий и невыполнение в срок учебного графика без уважительных причин, что является грубым нарушением учебной дисциплины («систематическое нарушение … учебной дисциплины (в том числе систематические пропуски без уважительной причины аудиторных занятий)», пункт 3.23 Устава СГАСУ). Мерой такого нарушения является коэффициент отставания студента.

Коэффициент отставания студента состоит из базовой части и поправки на внеаудиторную активность.

Базовая часть равна сумме процента часов пропусков занятий и процента количества неаттестованных контрольных точек.

Поправка на внеаудиторную активность рассчитывается более сложно. Как показывает статистика 2008-2009г.г, текущий суммарный рейтинг внеучебной активности лучших 10% студентов факультета равен примерно 30 единицам. Учитывая, что в течение одного года оценка любого достижения снижается до нуля, можно представить себе уровень успешности таких студентов на следующем примере. Чтобы иметь к концу некоторого года такой рейтинг, необходимо в течение этого года добиться в совокупности пяти достижений, оцененных оценкой 6 каждое, т.е, например, очно участвовать в Международной научной выставке и получить там медаль, получить награду на Международных спортивных соревнованиях, награду на областном и выше художественном конкурсе, быть профоргом факультета и активно участвовать в профсоюзных мероприятиях, - и все это по придирчивой оценке студсовета факультета. Принимаем, что такой студент заслуживает права на определенное, но не безграничное, отставание от учебного графика. Полагаем целесообразным принять это отставание также равным 30%.

В таком случае, коэффициент отставания студента рассчитывается как базовая часть, из которой вычитается рейтинг внеаудиторной активности этого студента. Комплексный учет успеваемости, посещаемости и позитивной внеаудиторной деятельности позволяет успевающим студентам более гибко распоряжаться своим временем с учетом индивидуальных особенностей и интересов в рамках выполнения требований Государственного образовательного стандарта, предусматривающего гармоничное развитие профессиональных и универсальных компетенций обучаемых.

Начиная с 5-й недели семестра, студент, имеющий 30-процентный коэффициент отставания **представляется к отчислению из вуза** за систематическое нарушение учебной дисциплины.

Рейтинг внеучебной деятельности и управленческие воздействия на его основе

Рейтинг внеучебной деятельности рассчитывается по записям, введенным в систему мониторинга достижений на портале ФИСТ любым зарегистрированным пользователем и оцененным заместителем декана по воспитательной работе. На его основе рассчитывается комплексный рейтинг, который является средним арифметическим учебного и внеучебного рейтингов.

Значения рейтинга и его составляющих «снимаются» еженедельно и используются для управления учебно-воспитательным процессом на факультете.

Важнейшей формой управления является проведение конкурса на звание **«Лучший студент факультета информационных систем и технологий»**. В соответствии с Положением о конкурсе, утвержденном ректором СГАСУ 18.05.08, это звание присваивается по итогам каждого семестра студенту, стабильно в течение семестра имеющему наивысшие по факультету показатели комплексного рейтинга. При определении победителя конкурса учитываются результаты студентов, хотя бы на одной неделе занявших по результатам комплексного рейтинга первые три места в своей студенческой группе. Еженедельно в течение семестра для этих студентов суммируются комплексные рейтинги. Победителем признается студент, получивший максимальный суммарный рейтинг в течение первых пятнадцати недель семестра.

Победитель конкурса определяется на 16-й неделе каждого семестра. Победитель конкурса представляется к единовременной стипендии Ученого совета (стипендии ректора) за успехи в учебной и внеучебной деятельности. **Его портрет на весь период существования факультета вывешивается в кабинете декана.**

Делается соответствующая запись в его зачетную книжку. Он имеет право досрочной сдачи всех экзаменов и зачетов в течение двух ближайших сессий.

На 16 неделе определяются **10% студентов факультета, имеющих наивысший комплексный рейтинг**. В их зачетные книжки делается запись «по итогам семестра вошел в число 10% лучших студентов факультета».

При вручении диплома специалиста студенту вручается его **портфолио**, заверенное деканатом, отражающее наивысшие достижения студента в период обучения в вузе.

Еще одной формой управления является проведение конкурса на звание **«Лучшая студенческая группа факультета информационных систем и технологий»**. Ею признается группа, имеющая на 16-й неделе наивысший средний (по студентам группы) комплексный рейтинг. Староста, профорг, учебный и внеучебный аналитики группы награждаются почетными грамотами факультета. Студенты лучшей группы имеют приоритет при распределении различных форм стимулирования студентов на факультете.

Система отображения внеучебной деятельности студентов

1. Достижения студентов в различных видах деятельности заносятся самими студентами или другими заинтересованными лицами на факультетский портал www.sciyouth.ru в раздел "Мониторинг учебно-воспитательного процесса".
2. Эта информация хранится в портале и полностью отображается в портфолио студента.
3. Для объективной оценки степени формирования социально-личностной характеристики студентов значимость их достижений оценивается в порядковой шкале (1- 6 баллов). Приведенные в ней уровни носят рекомендательный и опорный характер. Деятельность студента нужно оценивать, сопоставляя ее объем и значимость с приведенными ниже вариантами оценки.
4. Оценка производится на основе коллективно разработанной системы критериев внеучебными аналитика студенческих групп поочередно (по графику) совместно с уполномоченным сотрудником деканата.
5. С целью учета "устаревания" значимости достижений их первоначальное значение уменьшается ежедневно на 1 / 400 часть, а по истечении 400 дней с момента введения оценки она перестает учитываться при подведении итогов рейтинга внеучебной деятельности, но остается в портфолио студента.

Критерии оценки внеучебной деятельности студентов

I. Научная деятельность

1 – за достижение в школе/вузе
2 – за I и II место в школе/вузе (в школе – только для 1-го курса)

Заочное участие

2 – в региональном конкурсе (конференции) (I тур)
3 – на Всероссийских конференциях (выставках)
4 – на Международных конференциях (выставках)

3 – за **прохождение** на II-й тур (очный) в региональном конкурсе

Очное участие

4 – на Всероссийских конференциях (выставках)
5 – на Международных конференциях (выставках)

Награды

5 – за I и II место и лауреатство на Всероссийских конференциях (выставках)
6 – за любые награды на Международных конференциях (выставках)
6 – за получение медали!!!

Статьи

2 - отправка статьи в сборник вуза
3 - отправка статьи в сборник Всероссийской конференции
3 - печать статьи в сборнике вуза
4 - печать статьи в сборнике Всероссийской конференции
5 - публикация статьи в центральной печати

6 - публикация статьи в журнале ВАК

II. Спортивная деятельность

1 – за выступление в школе/вузе

2 – за I и II место в школе/вузе (в школе – только для 1-го курса)

1 – за постоянные занятия в спортивных секциях в течение семестра

3 – за участие региональных соревнованиях

4 – за участие Всероссийских соревнованиях

5 - за участие Международных соревнованиях

4 – за I и II место и на региональных соревнованиях

5 – за I и II место и лауреатство на всероссийских соревнованиях

6 - за любые награды на международных соревнованиях

III. Культурная деятельность

1-3 – участие в культурной работе вне вуза

1 – участие в конкурсах-смотрах в группе или факультете

2 – участие в конкурсах-смотрах в вузе

3 – участие в конкурсах-смотрах вне вуза

4 – участие в конкурсах-смотрах российского характера

5 – участие в конкурсах-смотрах международного характера

2 – награды в конкурсах-смотрах в группе или факультете

3 – награды в конкурсах-смотрах в вузе

6 – награды на областных, российских и международных художественных конкурсах

4 – награды за помощь в организации и проведении СТЭМа

5 – За постоянные занятия в СТЭМе на протяжении семестра

IV. Общественная деятельность

1 – за помощь в хозяйственных, трудовых делах факультета

1-3 – за конкретную работу внутри общественной организации (партии и т.д.)

2 – за семестровую работу профорга группы, старосты группы, редколлегии или аналитика (учебного или внеучебного)

3 – за организацию мероприятий на факультете

5 – за семестровую работу профорга факультета и помощь в делах профсоюза

V. Трудовая деятельность

1 – за сообщение «работаю» без указания места и должности работы

2 – за указание места и специальности не по профессии

3 – за указание места и специальности по профессии

1-3 – за одноразовую работу не по профессии

2-5 – за одноразовую работу по профессии

1, 2 – за посещение каких-либо курсов, только при получении сертификата, в котором указано не менее 15 часов

3 – за получение профессиональных сертификатов корпоративного уровня

4 – за получение профессиональных сертификатов российского уровня

6 – за получение профессиональных сертификатов международного уровня

4 – за получение 2-го высшего образования

Результаты опроса студентов ФИСТ об отношении студентов к действующей на факультете системе мониторинга учебно-воспитательного процесса

В феврале 2010г. был проведен анонимный опрос 50-ти студентов факультета ФИСТ (с 1 по 4-й курсы). Его результаты сравнили с результатами анонимного опроса проведенного в 2007г.

Опрос показал, что степень одобрения студентами системы, с некоторыми замечаниями, вырос с 65% до 74%. Процент студентов регулярно интересующимися результатами мониторинга через Интернет и на информационной доске возрос 27% до 84%. Доля студентов регулярно интересующихся результатами мониторинга через Интернет или на информационной доске близка к 100% (выросла с 90% до 96%).

По-прежнему результатами мониторинга интересуются порядка 50% родителей, причем доля родителей самостоятельно использующих для этого Интернет выросла с 9,5% до 24%.

72% студентов считают, что рейтинг стимулирует вовремя сдавать лабораторные и практические работы (в 2007г. их было 47,6%).

Ниже в таблице1 приведены подробные результаты опроса студентов ФИСТ СГАСУ.

Таблица 1. Результаты анонимного анкетирования студентов

№ п.п	**Вопрос**	Варианты ответа и результаты в % в 2007г. и в 2010г.					
1	Смотрите ли вы учебный рейтинг, и если да, то как часто?	Каждую неделю	Несколько раз в неделю	Раз в месяц	Раз в семестр	Никогда	Другое
		57	33	8	нет	нет	2
		62	34	нет	4	нет	Нет
2	Где вы смотрите информацию о рейтинге?	Только в Интернете	Только распечатку на информационной доске	И в Интернете и на доске			
		51	22	27			
		6	10	84			
3	Показываете ли вы дома родителям рейтинг?	Да	Нет	Родители смотрят сами			
		36,5	54	9,5			
		26	50	24			
4	Считаете ли вы, что рейтинг стимулирует вас вовремя сдавать лабораторные и практические (чтобы не допустить отставания)?	Да	Нет	Другое			
		47,6	41,3	11,1			
		72	24	4			
5	Выскажите, пожалуйста, ваше отношение к системе рейтинга в целом.	В целом одобряют	Не одобряют	Ничего не сказали			
		65	25	10			
		74	4	22			

Об автоматизированной информационной системе управления компетентностным развитием студенческого коллектива

А.А. Тимофеев, Д.С. Шишков
Самарский государственный архитектурно-строительный университет, Самара

Кроме знаний высшие учебные заведения дают студентам жизненные навыки, которые, как правило, и формируют его как специалиста, на которого можно полагаться. Вы сами знаете, что работодатели кроме оценок в аттестате обращают свое внимание на способности человека, которые ученик приобретает в течение обучения в вузе. В настоящее время российская система высшего образования переходит на государственные стандарты нового поколения. Они предполагают не только освоение студентом учебного плана, но и формирование целого комплекса личностных характеристик у студентов, таких как: качества лидера, работа в коллективе, умение принимать решения, оправданный риск, и т.д.

Цель работы – разработать структуру управленческой технологии управления компетентностным развитием студентов на факультете вуза на основе результатов функционирования соответствующей автоматизированной информационной системы. Разделить студентов на группы (произвести кластеризацию студенческого коллектива), для того, чтобы упростить процесс управления компетенциями.

Практическая значимость работы заключается в создании эффективного элемента, позволяющего каждому ВУЗу корректировать содержание обучения с учетом особенностей студента, повышая тем самым потенциал подготавливаемых специалистов.

Автоматизированная информационная система управления компетентностным развитием студентов представляет собой 3 блока: блок входных данных, оптимизационный блок и блок вывода результата.

Работа представляет собой часть общей большой системы АИС управления компетентностного развития студенческого коллектива. В рамках данной работы была разработана база данных с использованием СУБД MS Access, в которой содержатся данные о факультете, группе, студентах, виде деятельности, периоде обучения, компетенции, ресурсу вузу и т.д. Предрасположенности студентов к видам внеаудиторной активности и их уровень начальной сформированности компетенций выявляются при помощи интернет-тестов собственной разработки.

Разработанная система служит для ведения учета информации о состоянии и динамике компетентностного развития студенческого коллектива. На основании информационной картины, создаваемой системой, руководители различного звена могут принимать решения об управляющих воздействиях с целью оптимизации развития компетенций у студентов. Программный комплекс разработан при помощи языка программирования сценариев HTML-страниц JavaScript.

На данный момент система охватывает 5 студенческих групп, в которых обучается 96 студентов. Проведены два теста, которые позволили осуществить сбор данных по 14-ти компетенциям и 12-ти видам внеаудиторной деятельности.

Исследование проводилось с целью выявления кластерной структуры студенческого коллектива и статистической обработке полученных результатов. Все результаты, полученные в ходе исследования можно разделить на несколько этапов: выявление кластерной структуры студенческого коллектива; анализ полученных кластеров; статистическая обработка исходных данных, выявления зависимостей.

Полученные результаты кластеризации неоднозначны, поэтому кластеры должны подвергаться дополнительному анализу эксперта, который сможет из полученных кластеров распределить студентов так, чтобы общие виды деятельности отражались только на увеличении компетенций студенческого коллектива.

Дистанционная интеллектуальная система, реализующая интерактивное обучение в виртуальном пространстве

А.В. Топунов, Д.А. Загуменнов, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Формирование творческой личности будущего специалиста является актуальной проблемой не только для средней и высшей школы, но и важнейшей социально-экономической задачей всего общества. Процесс многостороннего формирования знаний специалиста является сложной задачей и требует использования различных приемов и методик подготовки и передачи знаний. Традиционные образовательные методики в условиях постоянно возрастающего информационного потока уже не обеспечивают требуемого уровня подготовки специалистов. Преподавателям любых образовательных учреждений (школ, колледжей, лицеев, вузов) приходится постоянно искать новые методы и формы образовательной деятельности, совершенствовать методику обучения, внедрять в учебный процесс более эффективные методы и средства с тем, чтобы активизировать процессы усвоения знаний, формирования умений и навыков. Думаем, что все согласятся с утверждением, что для развития творческих способностей учащихся на всех этапах обучения необходим комплексный подход, при этом в процессе обучения наравне с традиционными методиками должны применяться новейшие технические средства обучения и последние достижения в области информационных технологий.

На сегодняшний день одним из наиболее продуктивных направлений в разработке интерактивных обучающих систем является внедрение в их состав мультимедийных ресурсов: аудиовизуальной информации любых форм (текст, графика, анимация и др.), реализующей интерактивный диалог пользователя с системой и разнообразные формы самостоятельной деятельности по обработке информации. Одной из наиболее перспективных, по мнению авторов, мультимедиа технологий является технология виртуальной реальности или виртуальных миров (Virtual Reality).

В последние время ведущие мировые ИТ-компании активно ведут разработку средств построения и поддержки трехмерных объектов и виртуальных пространств, наработки в этой сфере можно использовать в процессах исследования и обучения. Технология виртуальной реальности широко использует когнитивную компьютерную графику, основная задача которой – стимулировать познавательные механизмы, творческое мышление, а не давать однозначную интерпретацию знаний. Погружение обучаемого (исследователя) в виртуальную трехмерную среду позволяет более эффективно воздействовать на его визуально-пространственные рецепторы и, как следствие, дают ему возможность более точно представить объект изучения: измененять форму объекта, свободно перемещаться между микро- и макроуровнями пространства, перемещать само пространство и т.п.

В связи с этим возрастает необходимость построения инструментария, которые позволили бы создавать обучающие системы, легко встраиваемые в традиционный учебный процесс, охватывающие весь курс обучения и дающие максимальный учебный эффект при разумных затратах на разработку. Для того чтобы процесс обучения давал хорошие результаты, а система эффективно функционировала, необходимо, чтобы выполнялись следующие требования:

- должны использоваться современные математические способы представления и обработки информации;

- активно применяться последние достижения в области информационных технологий: Internet-технологии позволяют организовать дистанционные интеллектуальные обучающие системы с применением языков Java, JavaScript и средств визуализации виртуальных миров, использующих язык X3D [1];

- учебные курсы должны автоматически генерироваться на основе заданной базы знаний предметной области и базы данных о способностях и потребностях обучаемого;

- должен обеспечиваться принцип моделирования содержания: в зависимости от цели обучения и содержания курса необходимо отобрать тот объем знаний, который будет достаточен, чтобы точно представить объект познания. При этом также необходимо учитывать познавательные потребности отдельных обучаемых, связанные с их индивидуальными интересами и т.п.

- каждый учащийся, как индивид, обладает определенными способностями, поэтому обучение должно быть направлено на выявление их исходного уровня и дальнейшего их развития - принцип индивидуализации обучения.

На кафедре программных систем СГАУ в рамках проекта «Школа информатики СГАУ» разрабатывается интеллектуальная система для обучения основам информатики и программирования. Она удовлетворяет всем перечисленным выше требованиям и представляет собой виртуальное пространство обучаемого, в котором он может передвигаться и воздействовать на окружающий его виртуальный мир.

Сам процесс обучения делится на условные этапы: изучение текстового материала, просмотр медиа-материалов с применением компьютерной графики, прохождение тестов. Завершающий этап обучения - это выполнение различных «поручений», возлагаемых системой на обучаемого. Основные действия пользователя, количество пройденных заданий и их результаты сохраняются в базе данных обучаемого, на их основании система вычисляет показатели эффективности обучения и строит схему изучения учебного материала при следующем сеансе: определяются направления учебного курса, вид и содержание тестовых заданий, при этом акцент делается на тех направлениях и видах учебной информации, которые были менее всего изучены или были неверно усвоены.

Важной особенностью системы является то что, виртуальный мир создаётся динамически на стороне сервера для каждого конкретного пользователя в зависимости от выбранного курса обучения и предустановленных настроек, и передаётся клиенту в виде текстового файла описания мира на языке X3D. Для того чтобы данный файл отобразился в окне Интернет-браузера обучаемого, на стороне клиента должен быть установлен специальный плагин.

Литература

1. Титтел Э., Сандерс К., Скотт Ч., Вольф П. Создание VRML-миров: пер. с англ. – К.: Издательская группа «BHV», 1997. 320 с.

Реализация дистанционного образования на базе системы дистанционного обучения кафедры программных систем СГАУ

Р.Д. Фахрутдинов, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

Современные тенденции развития образовательного процесса и образовательных технологий настоятельно требуют совершенствования системы информатизации обучения, текущего и промежуточного контроля результатов обучения и внедрения компьютерных систем тестирования знаний студентов. Современное состояние вычислительной техники, имеющейся в высших учебных заведениях, и использование инновационных подходов к ИТ-образованию позволяют эффективно использовать информационные технологии для совершенствования обучения студентов, анализа их профессиональной компетенции.

Применение дистанционных образовательных технологий может служить хорошим «подспорьем» к основным формам образования за счет [1]:

- обеспечения удобного доступа студентов к материалам лекционных курсов и практических занятий;
- возможностей эффективной организации общения между студентами и преподавателями, возникающих в процессе выполнения практических заданий и изучении материалов учебного курса;
- реализации автоматизированного приема работ по практическим заданиям, включающего систему информирования студентов об оценках и статусах работ;
- возможности организации дистанционного тестирования для оценки знаний студентов.

Немаловажное значение имеет и тот факт, что применение дистанционных технологий позволяет эффективно использовать образовательные ресурсы в рамках кафедры, факультета, университета, они существенно изменяют роль преподавателя и студента в учебном процессе. Преподаватель все больше выступает в роли консультанта, советчика, он координирует познавательный процесс студентов. Его работа по развитию и совершенствованию электронных информационно-образовательных ресурсов должна иметь системный и непрерывный характер. Студент повышает свой творческий и интеллектуальный потенциал за счет самоорганизации, стремления к знаниям, умения взаимодействовать с компьютерной техникой и самостоятельно принимать ответственные решения.

На кафедре программных систем ведутся работы по созданию и наполнению информационных ресурсов, основанных на использовании технологии дистанционного обучения.

Организация системы дистанционного обучения

Реализация дистанционных образовательных технологий осуществляется через системы дистанционного обучения (СДО), которые используют специальные инструментальные средства и оболочки. Это имеет ряд преимуществ, таких как, снижение времени и затрат на разработку курсов, широкое привлечение к разработке курсов преподавателей, обеспечение современного уровня функциональных и коммуникационных возможностей и пользовательского графического интерфейса курсов. Одной из наиболее популярных оболочек для разработки СДО является

программная среда LMS Moodle, на основе которой разработан сайт кафедры. LMS Moodle построена по модульному принципу и может расширяться за счет включения в нее авторских программных модулей, реализующих дополнительные функциональные возможности.

К достоинствам LMS Moodle можно отнести:

- реализацию коллаборативной технологии обучения, основанной на взаимообмене знаний в процессе совместного решения учебных задач преподавателя и студента;
- предоставление широких возможностей для коммуникаций: поддерживается обмен файлами любых форматов - как между преподавателем и студентом, так и между самими студентами;
- возможность оперативного информирования всех участников курса или отдельных групп о текущих событиях (сервис рассылки);
- возможность организации обсуждения учебных проблем, в том числе в режиме реального времени, при этом обсуждение можно проводить индивидуально или по группам (форум, чат);
- гибкую систему тестирования, которая легко конфигурируется в зависимости от поставленной задачи и позволяет представить конечный тест практически в любом желаемом виде.

В Moodle в качестве электронного информационного ресурса может выступать любой материал для самостоятельного изучения, проведения исследования, обсуждения: текст, иллюстрация, web-страница, аудио или видео файл и др. Для создания web-страниц в систему встроен визуальный редактор, который позволяет преподавателю, не знающему языка разметки HTML, с легкостью создавать web-страницы, включающие элементы форматирования, иллюстрации, таблицы.

Для поддержания СДО в актуальном состоянии большинство разделов сайта должны быть динамическими, за обновление информации отвечают администратор и разработчики курсов.

Развитие системы дистанционного обучения

Любое образовательное учреждение должно заботиться о престиже своего бренда, создавать условия для обеспечения приемлемого качества обучения, для совершенствования образовательных ресурсов, увеличения их внутреннего потенциала [2]. Реализация инновационных проектов в сфере образования, в первую очередь дистанционных образовательных технологий, позволит обеспечить качественную основу для подготовки специалистов высокой квалификации, которые будут востребованы на рынке труда. Растет число ВУЗов, ведущих подготовку с помощью дистанционных образовательных технологий. Отставание в этом направлении может существенно снизить статус нашего университета.

В качестве концептуальной модели информационного обеспечения СДО предлагается использовать гомогенную сетевую модель, когда совокупная информация, необходимая для функционирования системы дистанционного обучения, определенным образом распределена либо по территориальному, либо логическому принципу на сети серверов (рис. 1). Базовый сервер университета хранит всю совокупную информацию, распределенную по сети серверов, это позволяет в случае необходимости восстанавливать потерянную информацию или заменить временно вышедший из строя узел. Следующим уровнем иерархии является факультетский сервер, который с одной стороны обслуживает рабочие станции пользователей, с другой – связан с кафедральными серверами. Кафедральные серверы связаны с дисплейными классами, где и происходит образовательный процесс.

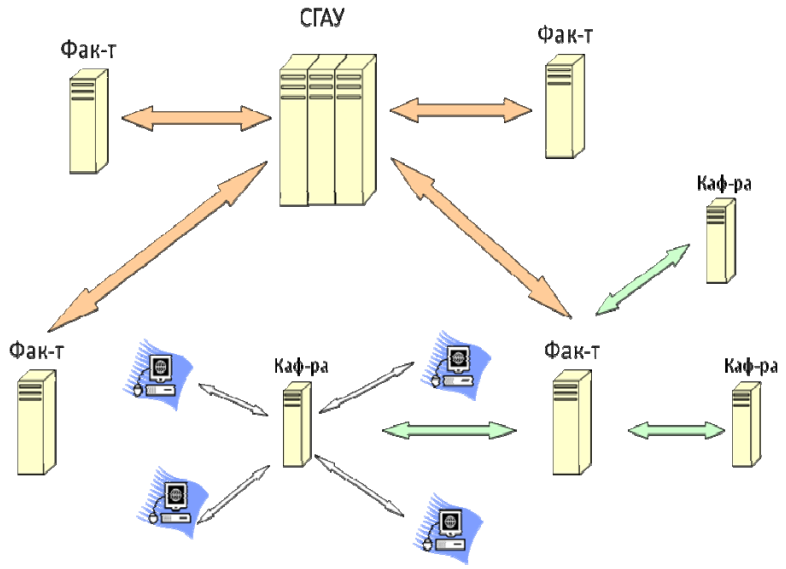


Рис. 1. Модель информационного обеспечения СДО

Отчасти такое решение вызвано естественными причинами, например, географической удаленностью рабочих площадок университета, отчасти - соображениями мобильности и надежности организации учебного процесса, происходящего в рамках дистанционной обучающей системы.

Литература

1. Зеленко, Л.С. Проблемы и перспективы реализации компетентностного подхода в сфере дистанционного образования [Текст]// Труды международной научно-практической интерент-конференции «Информационные технологии в науке и образовании» и IV Всерос.семинара «Применение MOODLE в сетевом обучении»/ редкол.: А.Э.Попов и [др.]. / Л.С. Зеленко, Р.Д.Фахрутдинов/ – Шахты: ГОУ ВПО «ЮРГУЭС», 2010. С. 205-207.

2. Зеленко, Л. С. Реализация инновационных подходов к образованию на основе дистанционного обучения [Текст] /Л.С. Зеленко Р.Д. Фахрутдинов, О.М. Лисневская // Сборник материалов межрегион. научно-методич. конф. «Актуальные проблемы развития университетского технического образования в России». Самара: изд-во СГАУ, 2009. – С. 89-90.

Подсистема электронного документооборота в составе автоматизированного рабочего места инженера кафедры

О.В. Федюнев, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Крупные общеобразовательные учреждения не могут обходиться без систем организации и хранения информации. Это обусловлено тем, что они состоят из большого числа взаимодействующих друг с другом подразделений, единственным приемлемым средством коммуникации между которыми остаются различного рода документы (приказы, распоряжения, отчеты и т.д.), составление которых обычно отнимает много времени у сотрудников.

Основным подразделением, которое отвечает за проведение учебного процесса в вузе, является кафедра. Организация учебного процесса требует подготовки большого количества плановых и отчетных документов как внутри кафедры, так и на уровне деканатов и ректората, большинство из которых являются типовыми и не изменяются из года в год. К таким документам в первую очередь можно отнести индивидуальный план преподавателя, сводный план работы кафедры, табель учета рабочего времени, различного рода служебные записки и сопроводительные документы. Поэтому *актуальной* остается задача автоматизации документооборота внутри кафедры.

На кафедре программных систем разработано автоматизированное рабочее место (АРМ) инженера кафедры, в состав которого включена подсистема электронного документооборота. Она позволит каталогизировать составленные документы, получать к ним простой доступ, а также заранее напомнит о приближении срока написания очередного документа. Архивация устаревших данных позволит экономить место для хранения более важной информации.

Описание подсистемы электронного документооборота

При работе с подсистемой электронного документооборота предусмотрено распределение ролей: администратор, преподаватель и лаборант. *Администратор кафедры* имеет возможность манипулировать общекафедральными документами и документами любого зарегистрированного в системе пользователя (рис. 1); *преподаватель* же, напротив, может управлять только своими собственными документами, *лаборант* может работать только с типовыми справочниками.



Рис. 1. Структура типовых документов кафедры

Наиболее трудоемкими и сложными документами, которые должны ежегодно подготавливать сотрудники кафедры, являются индивидуальный план работы преподавателя и сводный план работы кафедры. Они включают в себя распределение учебной нагрузки и учебных поручений, наименования всех выполняемых типов работ

(учебная, учебно-методическая, научно-методическая, научно-исследовательская работы) и сроках их выполнения, данные о повышении квалификации сотрудников кафедры и т.п. Подсистема электронного документооборота автоматически генерирует требуемые документы по заданным шаблонам и сохраняет их в базе (по годам). В подсистеме предусмотрена возможность архивирования устаревших и неиспользуемых документов и резервное копирование новых документов.

Локальная модель базы данных, используемая при генерации документа, представлена на рис. 2. Она позволяет хранить информацию об учебных планах преподавателей кафедры, кафедры в целом, имеющихся приказов и заявлений, сведений о работе ГЭК.

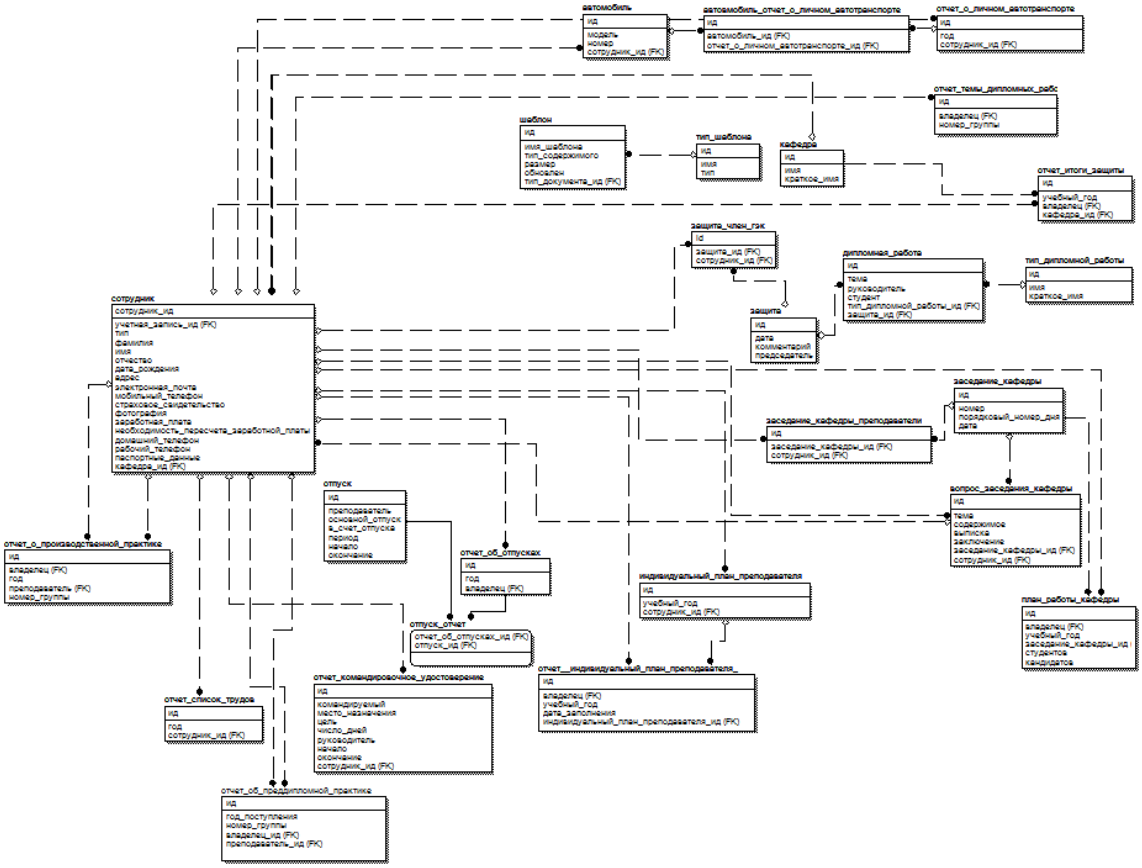


Рис. 2. Локальная логическая модель базы данных

Описание технологий разработки приложения

Проектирование программных интерфейсов последнее время шагнуло далеко вперед, появление новых языков программирования и технологий разработки привело к тому, что некоторые разработчики заговорили о тупике развития нынешних графических оболочек, а другие о переносе разработки приложений в WEB. В настоящее время уже имеются примеры удачных интерактивных интерфейсов на базе WEB (Flickr, Google Maps), которые позволяют организовать работу пользователей с удаленной информацией и обеспечивают высокое быстродействие. Разработка WEB-приложений - достаточно рутинное занятие, поэтому появление Ruby On Rails (RoR) было воспринято разработчиками с большим энтузиазмом.

RoR не является традиционной средой разработки приложений, а представляет собой полноценный, многоуровневый фреймворк, интегрированный с объектно-ориентированным языком Ruby для разработки Интернет-приложений, использующих базы данных. Он обеспечивает гибкость разработки, отличную масштабируемость и читаемость приложений. В среде Rails реализованы строгие философские решения, позволяющие существенно упростить разработку сложных WEB-приложений и сократить время разработки от нескольких месяцев до нескольких недель, с ее

помощью можно быстро и эффективно разработать приложение практически любой степени сложности.

Одними из основных решений, лежащих в основе Rails, являются:

– *Плавная интеграция.* Rails блестяще использует наилучшие свойства языка программирования Ruby и расширяет его возможности, отличная интеграция Active Record (механизм персистенции Rails) и среды Модель-Представление-Контроллер (model-view-controller - MVC) позволяет легко создавать модели и автоматически генерировать их пользовательский интерфейс.

– *Соглашение по конфигурации.* Rails отказывался от поддержки объемных, всеобъемлющих конфигурационных файлов (как например, в Java-системах). Он допускает обычную структуру каталогов проекта и простые соглашения по именованию методов, классов, таблиц и столбцов, в результате этого Rails-приложения нуждаются только в части конфигурационного кода, часто уменьшая его объем в 10 и более раз.

– *Низкая повторяемость.* В среде Rails реализована возможность абстрагироваться от повторяющихся задач.

– *Немедленная обратная реакция.* Изменение кода становится активным при следующей загрузке Web-страницы. Миграции могут стать видимыми немедленно после обновления вашей базы данных.

Учитывая достоинства Ruby On Rails и необходимость дистанционного доступа к документам, на кафедре программных систем СГАУ было принято решение использовать данную технологию при разработке автоматизированного рабочего места инженера кафедры, это позволит пользователям оперативно решать производственные задачи и обеспечит гибкое планирование учебного процесса.

Выводы

Для разрабатываемой системы «АРМ инженера кафедры» подсистема электронного документооборота имеет большую значимость. Она позволяет генерировать определенный набор документов из шаблонов, поддерживает напоминания о создании определенных документов, ведение базы созданных документов, архивирование документов, изменение базы данных кафедры. Это позволяет существенно сократить время на составление типовых документов, их поиск и организацию. В настоящее время подсистема электронного документооборота проходит эксплуатацию на кафедре программных систем СГАУ, дорабатывается модель данных, рассматриваются пути повышения удобства графического интерфейса.

Роль системы дистанционного обучения в проекте «Школа информатики СГАУ»

А.А. Флигин, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

Для того чтобы качество подготовки специалистов не отставало от требуемого на рынке труда, необходимо расширять образовательные технологии. Учебно-методическая база, образовательно-информационные технологии любого вуза таковы, что они вообще не зависят от формы обучения. Если знания, весь учебный материал (в том числе его дидактическая составляющая) оформлены и находятся в формализованном виде (в виде файлов или баз данных) и хранятся на компьютерах, то все равно, куда подать эти знания: в аудиторию (соответственно оснащенную) или находящемуся за пределами учебного заведения пользователю [1].

Цели проекта

Требования к системе образования, в первую очередь к базовой (школьной) ступени, повышаются год от года. Одно из важнейших направлений учебного процесса в старшем звене средней школы – подготовка к централизованному тестированию (к сдаче единого государственного экзамена (ЕГЭ) по обязательным и профилирующим предметам, в том числе и по информатике).

Главные цели проекта «Школа информатики СГАУ» - привить школьникам навыки информационной культуры, необходимые при дальнейшем обучении в вузе; дать представления о сети Интернет как о средстве самообучения и самовыражения; всесторонне подготовить их к сдаче ЕГЭ по информатике, а также дать более глубокие знания по программированию школьникам, планирующим поступать на специальности, связанные с информационными технологиями. Другая цель проекта - создать в сети Интернет общедоступный высококачественный контент по дисциплинам «Информатика» и «Основы программирования» [2].

Современные информационные и телекоммуникационные технологии позволяют более гибко построить учебный процесс и организовать взаимодействие его участников в режиме реального времени. Дистанционное обучение с использованием Интернет сегодня востребовано и быстро развивается, оно становится важнейшей частью системы образования.

Система дистанционного обучения школы информатики

На кафедре программных систем СГАУ разработана и проходит опытную эксплуатацию система дистанционного обучения школы информатики СГАУ, построенная на базе LMS Moodle. Выбор данной LMS обусловлен следующими причинами: она построена по модульному принципу; имеет простой, эффективный, совместимый для разных браузеров web-интерфейс; позволяет гибко добавлять или удалять элементы на различных уровнях иерархии, имеет мощную и гибкую систему тестирования знаний.

Авторами разработаны логическая структура сайта, интерфейс системы; статические разделы системы наполнены конкретным содержанием; разработаны контенты учебных дисциплин «Информатика» и «Основы программирования», в них добавлен большой объем тестовых заданий.

Наиболее важные этапы проектирования - разработка логической структуры сайта и интерфейса пользователя, так как именно интерфейс определяет набор функций, доступных пользователю, механизм взаимодействия пользователя и системы,

как следствие, скорость получения результата. На рис. 1 представлена главная страница сайта.

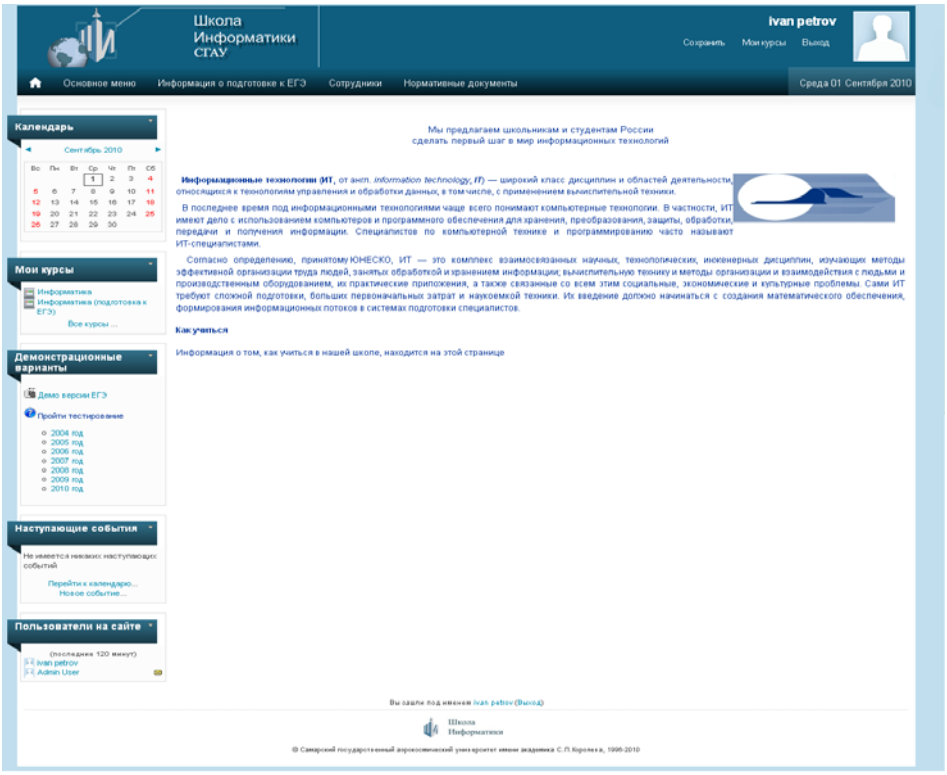


Рис. 1. Главная страница сайта дистанционного обучения школы информатики

Сайт разбит на разделы. Панель меню под баннером содержит ссылки на общую информацию, доступную для всех пользователей сайта. Разделы «Календарь», «Мои курсы», «Наступающие события» отображают информацию, специфичную для каждого пользователя, например, раздел «Календарь», изображенный на рис. 2, отображает события, которые могут быть индивидуальными как для каждого пользователя, так и для группы пользователей, подписанных на определенный курс.

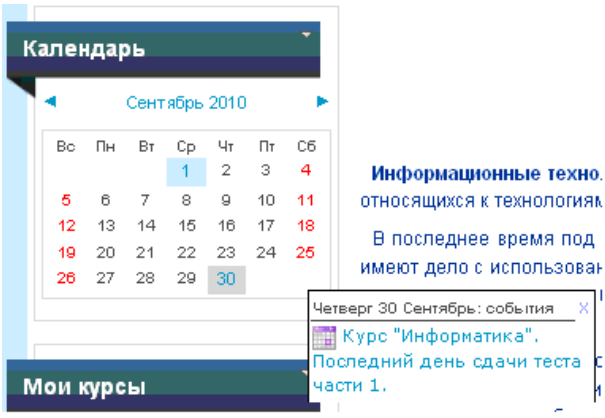


Рис. 2. Календарь событий

Раздел «Мои курсы» содержит список доступных для пользователя курсов. При выборе курса видимая структура сайта меняется. Появляются новые разделы «Люди», «Элементы курса», «Поиск по форумам», «Управление», «Новостной форум», «Последние действия». На рис. 3 приведена главная страница учебного мультимедиа комплекса «Информатика», в состав которого входит: хорошо структурированный текстовый материал, снабженный большим количеством перекрестных ссылок; компьютерные (анимационные) демонстрации различных алгоритмов; виртуальные тренажеры, позволяющие усилить подготовку в области программирования, приобрести навыки при решении типовых задач блока С ЕГЭ; тесты по всем разделам учебного курса и другие учебно-методические материалы.

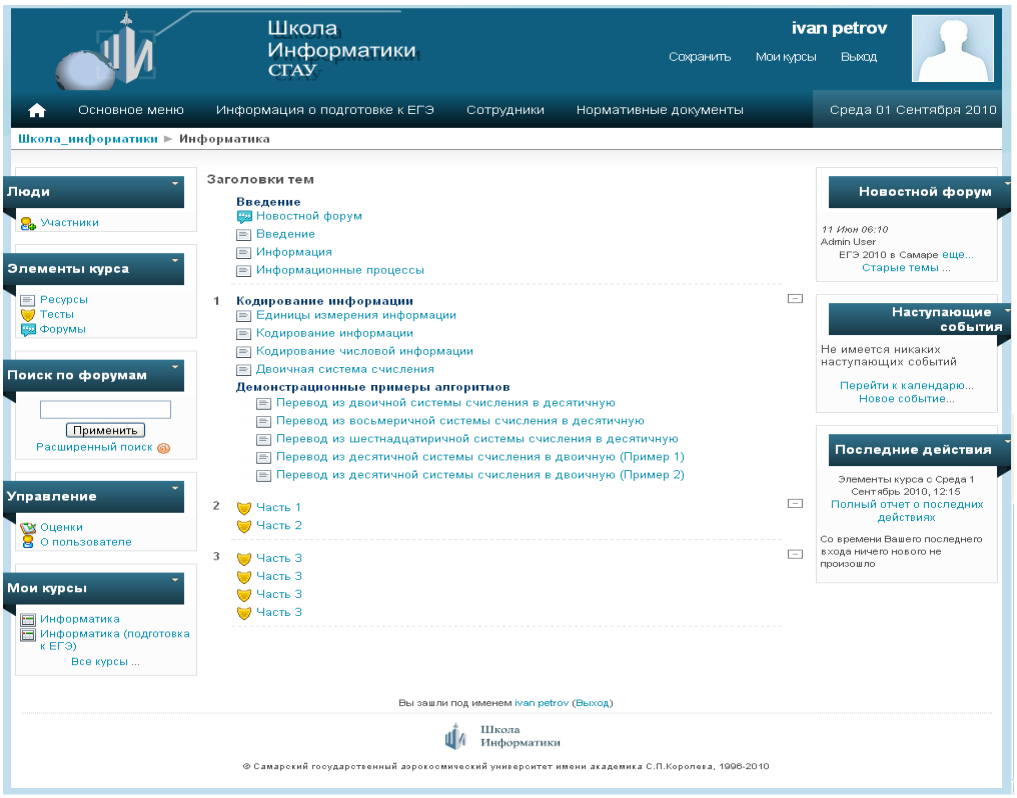


Рис. 3. Главная страница курса «Информатика»

При дистанционном обучении роль тестовой аттестации возрастает, поскольку тест является единственно возможным средством проверки знаний обучаемых, территориально удаленных от образовательного учреждения. В LMS Moodle имеется встроенная система тестирования, которая очень гибко конфигурируется в зависимости от поставленной задачи и позволяет представить конечный тест практически в любом желаемом виде. В системе имеются следующие виды тестов: вопрос-задача, вопрос с развернутым ответом, вопрос на соответствие, вопрос с множественным выбором, вопрос с кратким ответом, вопрос с числовым ответом, вопросы, на которые нужно ответить однозначно: истина или ложь. Каждый ответ вопрос несет в себе процентную долю корректности (правильности), а также весовой коэффициент, учитывающий сложность вопроса. В случае неправильного ответа на вопрос имеется возможность повторного ответа, обучаемый имеет возможность обдумать, где и почему он ошибся. Он думает и анализирует, следовательно, он обучается.

Выводы

Учитывая актуальность и перспективность направления создания электронных информационно-образовательных ресурсов, работа по их развитию и совершенствованию в системе дистанционного обучения должна иметь системный и непрерывный характер. Совмещение очного и дистанционного обучения позволит школьникам лучше усвоить учебный материал; дистанционное тестирование поможет выявить типовые ошибки и устранить пробелы в знаниях; индивидуальная работа с тренажерами позволит изучить основы алгоритмизации и программирования.

Литература

1. Зеленко, Л.С. Дистанционное обучение как инновационная стратегия педагогического процесса// Труды XV Всероссийской научно-методической конференции «Телематика-2008». - СПб: Изд-во ИТМО, 2008. - Т. 2. - С.461-463.

2. Зеленко, Л.С. Методика преподавания информатики в школе информатики СГАУ// Труды XVII Международной конференции «Математика, компьютер, образование»/ Л.С.Зеленко. Москва-Ижевск: «R&C Dynamics». - выпуск 17. - С. 413.

Интерактивный тренажёр «Конструктор алгоритмов»

И.В. Юдин, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Для преодоления противоречий между стремительно нарастающим объемом знаний, подлежащих усвоению, и реальными сроками подготовки специалистов, необходимо определить новые организационные формы обучения, способы получения знаний студентами, закрепления умений и навыков [1]. Для школьников при изложении нового, трудного для понимания материала, требуется наглядное его пояснение на простых примерах. Включение мультимедиа объектов в состав учебного курса способствует более быстрому запоминанию учебного материала, позволяет дать более емкое представление об изучаемых явлениях и предметах. Как показывает опрос школьников, лучше всего запоминаются те анимации, которые имеют не строго учебный, а скорее развлекательный характер, «живые» элементы мультимедиа позволяют вспомнить материал при тестировании и/или на экзамене. Технологии мультимедиа позволяют в полной мере использовать принцип: человек получает 80 % информации из окружающего мира с помощью зрения, при этом люди запоминают 20 % того, что они видят, 40 % того, что они видят и слышат, и 70 % того, что они видят, слышат и делают [2]. Это дает основание для утверждения тезиса об эффективности визуализации в сфере обучения.

Один из наиболее простых и понятных способов записи алгоритма – графический, он построен на использовании универсальных примитивов, позволяющих отразить последовательность выполняемых действий и не «привязанных» к конкретному языку программирования. В связи с этим было принято решение о разработке интерактивного тренажера «Конструктор алгоритмов», который поможет старшекласникам и первокурсникам, изучающим информатику, закрепить такие базовые понятия, как «алгоритм», «цикл», «условие» и т.п. Используя «Конструктор алгоритмов», обучаемый сможет самостоятельно разрабатывать простейшие алгоритмы и соответствующие им программы. Процедура построения схемы алгоритма достаточно проста: необходимо выбрать нужный функциональный блок (процесс-вычисление, цикл, решение-условие и т.д.) из заданного набора инструментов, расположенных на панели; перетащить его на рабочую область; определить связи между блоками, задав тем самым последовательность действий; и, наконец, ввести в соответствующие блоки операнды и выражения, оперирующие ими. После завершения процесса конструирования алгоритма обучаемый может проверить его работу, вводя необходимые данные в режиме диалога. В тренажере предусмотрена возможность сохранения полученной схемы алгоритма или его части (блока) в файл. Отлаженные и изученные алгоритмы обучаемый может в дальнейшем использовать при построении более сложных алгоритмических конструкций.

Данный тренажер будет интегрирован в систему дистанционного обучения школы информатики СГАУ. Его применение на практике позволит: сделать процесс обучения более творческим и интересным; повысить эффективность освоения учебного материала; вовлечь обучаемого в активную когнитивную деятельность по осмыслению и закреплению учебного материала и поднять уровень подготовки учащихся.

Литература

1. Бабенко, Т.А. Применение средств мультимедиа в процессе обучения будущих учителей информационным технологиям [Текст]: Дис. ... канд. пед. наук: 13.00.08/ Бабенко Татьяна Арутюновна. Армавир, 2003. - 201 с. - 61:03-13/1442-8.

2. http://www.sike.ru/products/multimedia/staty/teach-virt [Электронный ресурс]. – Обучение на основе Виртуальной реальности.

Подсекция 4
«Автоматизированные системы реального времени»

Многоуровневая модель разработки ПО АСРВ средствами операционной системы реального времени

А.В. Баландин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

При проектировании и разработке программного обеспечения автоматизированных систем реального времени (ПО АСРВ) приходится рассматривать различные уровни поведения системы и решать комплекс разноплановых, но взаимоувязанных задач функционирования системы, что относит ПО АСРВ к классу сложных программных систем. Согласно теории сложных систем их анализ и синтез должен осуществляться в соответствии с принципом *иерархической организации* системы. Согласно этому принципу ПО АСРВ следует представлять в виде многоуровневой системы, структурное моделирование которой необходимо осуществлять на различных, взаимосвязанных уровнях абстракции. Наиболее последовательно иерархический принцип для многоуровневого моделирования *вычислительных многокомпьютерных управляющих систем*, работающих в режиме реального времени (т.е. АСРВ), был применён и описан Гертенбахом (Gertenbach W.P.) в [1]. В этой работе, на основе обобщения опыта построения и использования вычислительных управляющих систем в области ядерных исследований, введено в рассмотрение семейство иерархических моделей АСРВ, описывающих функциональную структуру АСРВ с целью обоснования состава и структуры комплекса технических средств и последующего распределения выделенных функций по компонентам (модулям) технических средств АСРВ. При этом полагается, что функции, возложенные на процессоры, специфицируют семантику функционирования ПО АСРВ и должны быть программно реализованы. Однако в работе никаких принципов моделирования структуры ПО АСРВ для программной реализации отнесённых к процессорам функций не предлагается. Учитывая, что программная составляющая АСРВ в свою очередь сама является сложной (программной) системой, то её проектирование и разработку также необходимо осуществлять на основе многоуровневого моделирования структуры ПО АСРВ в виде взаимосвязанных моделей различных уровней абстракции, которые и предлагаются в данной работе.

Общая структура ПО АСРВ

Предложенная в [1] модель АСРВ не указывает границ между программной частью системы и программно-управляемым комплексом технических средств АСРВ. Очевидно, что аппаратная составляющая обеспечивает непосредственное или опосредованное взаимодействие программной части с объектом управления (ОУ), субъектом управления (СУ) и хранилищем данных (архивом). Будем рассматривать её как *окружение* программной системы. В качестве элементов программной системы, предназначенных для взаимодействия со своим окружением введем в рассмотрение абстрактные объекты памяти – *истоки*/*стоки*, осуществляющие коммуникацию данных

между ОУ и программной системой, между программной системой и СУ, а также между программной системой и хранилищем данных. Выделение и спецификация истоков и стоков является предметом моделирования внешних данных программной системы. В итоге спецификация истоков и стоков составит границу между программной системой и её окружением. Вычленение программной системы из АСРВ вместе с истоками и стоками представим в виде, показанном на рис. 1.

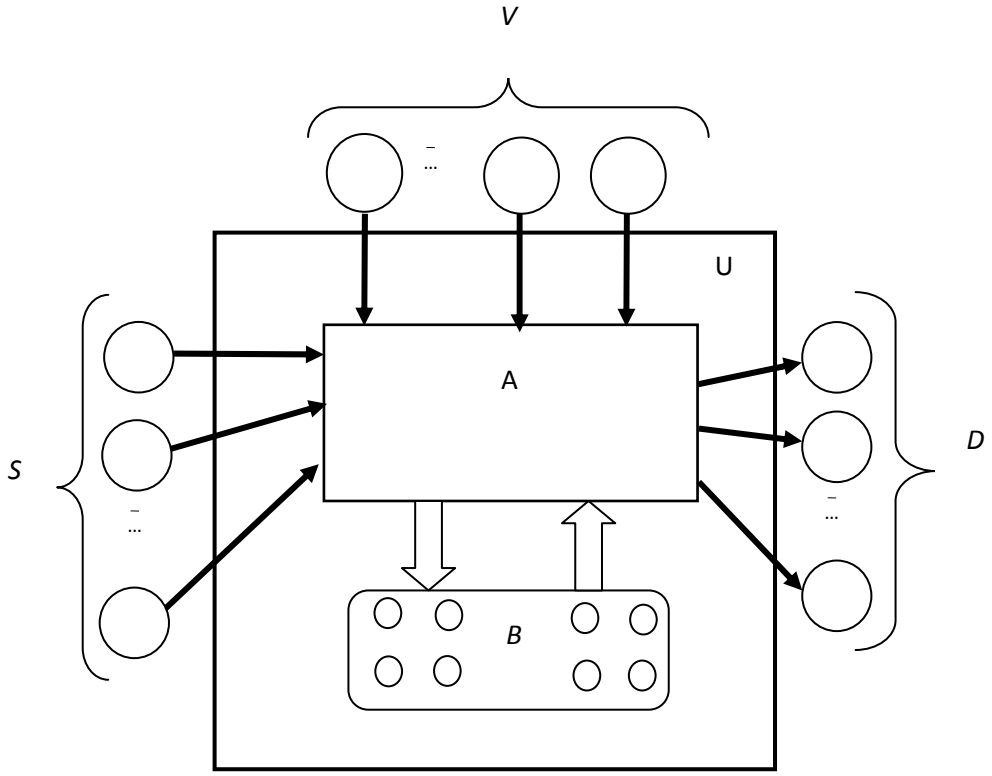


Рис. 1. Вычленение программной системы

Формально программную систему, как вычленение АСРВ, опишем парой U=<M,A>, где M=$S\cup V\cup D\cup B$ – конечное множество абстрактных объектов памяти программной системы U, A - программный агрегат, получающий входные данные из истоков, и помещающий выходные данные в стоки. Подмножество $(S\cup V\cup D)\subset$M – *внешняя память* (окружение) программного агрегата A, $S=\{s_1,s_2,\ldots,s_{N_S}\}$ и $V=\{v_1,v_2,\ldots,v_{N_v}\}$ – множества истоков (N_S – мощность множества S, N_v – мощность множества V), $D=\{d_1,d_2,\ldots,d_{N_D}\}$ – множество стоков (N_D – мощность множества D). Подмножество $B\subset$M - внутренняя память программного агрегата A. Разница между истоками из S и истоками из V в том, что истоки из S ассоциируются с параметрами состояния ОУ, а истоки из V, ассоциируются с установочными параметрами программного агрегата A (служебная информация). Стоки в D ассоциируются с регулировками исполнительных механизмов ОУ, элементами отображения информации для СУ и помещением данных в архив.

Существуют различные инструментальные средства и методы разработки программных систем. Их выбор во многом определяет парадигму структуризации при проектировании и разработке программной системы. Если в качестве инструментальных средств программирования используется SCADA-система (специализированное средство для разработки программных систем компьютерного мониторинга), то структура программной системы на этапе проектирования формируется средствами выбранной SCADA-системы, а программный код затем в

значительной части генерируются автоматически. На долю разработчика выпадает в основном конфигурирование и настройка программной системы, используя инструментальные средства SCADA-системы, последующая её адаптация к специфическим особенностям используемого УСО, разработка графического пользовательского интерфейса и параметрическая настройка системы в целом.

Иное дело, если в качестве программных инструментов используются универсальные средства программирования операционной системы реального времени (ОСРВ). В этом случае разработка ПО АСРВ начинается с выбора принципа структуризации программной системы, который в итоге определяет выбор методов структурного анализа, моделирования и макетирования программной системы. В дальнейшем будем полагать, что разработка ПО АСРВ осуществляется в следующих условиях:

− программная система реализуется на базе распределенного вычислительного комплекса (РВК), представляющего собой компьютерную или промышленную сеть, включающую в свой состав устройства ввода/вывода для связи с объектом управления и оператором. Одномашинный вариант рассматривается как частный случай РВК;

− в качестве инструментальных средств разработки программной системы используются: многопроцессная операционная система реального времени с многонитиевой организацией вычислений внутри процессов и с клиент-серверной организацией взаимодействия процессов посредством сообщений; специализированные драйверы УСО и файлы устройств ввода/вывода; библиотека функций математической обработки; библиотека классов графических объектов для изображения мнемосхем оператора; средства архивирования данных (файловая система ОС, СУБД).

Многоуровневая модель ПО АСРВ

Программная система U=<M,A>, является сложной системой. Поэтому согласно принципу построения сложных систем в самом общем виде структуру программного агрегата А необходимо представить в виде иерархии уровней абстрагирования программного агрегата А с соответствующими каждому уровню задачами проектирования и разработки ПО АСРВ. На каждом уровне программную систему будем представлять в виде программного агрегата, извлекающего данные из истоков и отправляющего результаты работы в стоки.

Представим многоуровневую модель ПО АСРВ в виде следующих уровней иерархии (рис. 2):

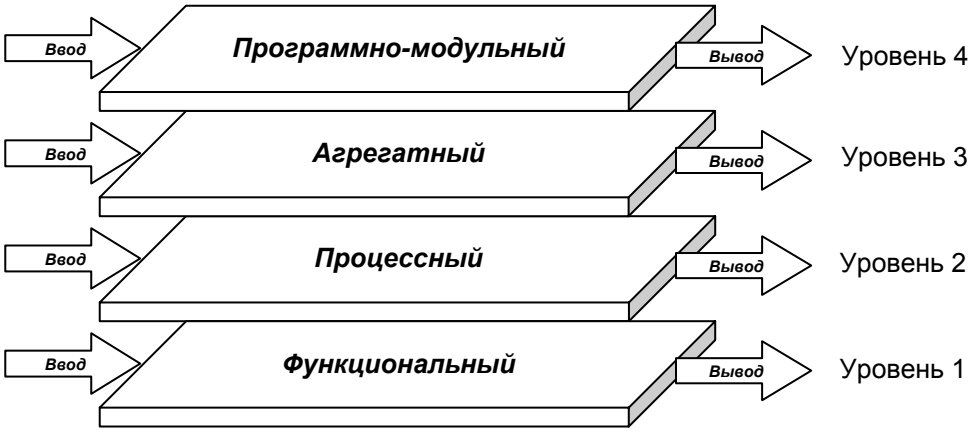


Рис. 2. Стратумная модель ПО АСРВ

- функциональный,
- процессный,

- агрегатный,
- программно-модульный.

Функциональный уровень предназначен для моделирования функциональной структуры программной системы на рассматриваемом уровне АСРВ. Для этого осуществляется анализ требований к функционированию АСРВ, изложенных в техническом задании на разработку системы.

На функциональном уровне целевая функция уровня декомпозируется в структуру взаимосвязанных по данным неделимых функций, отражающих семантику обработки данных, принятия решений и выдачи управляющих воздействий на объект управления, а также отображения информации или занесения результатов обработки в архив на соответствующем уровне АСРВ. Для структуризации ПО АСРВ на функциональном уровне используется язык схем асинхронных потоков данных (АСПД-схем) [3].

Процессный уровень предназначен для моделирования структуры ПО АСРВ в виде совокупности взаимодействующих параллельных процессов, обеспечивающих выполнение функций, представленных на функциональном уровне. Структуризация ПО АСРВ на процессном уровне осуществляется в виде сети виртуальных потоковых машин (ВПМ-сети) [3].

Агрегатный уровень предназначен для моделирования ПО АСРВ в виде структуры взаимодействующих абстрактных программных агрегатов (А-сети), интегрирующих в себе компоненты структуры процессного уровня [3]. Программные агрегаты рассматриваются в качестве неделимых программных элементов ПО АСРВ, загружаемых в процессорные компоненты РВК.

Программно-модульный уровень предназначен для представления ПО АСРВ в виде структуры каталогов с файлами, содержащими программные модули ПО АСРВ: исполняемые модули, объектные модули, библиотечные файлы и т.п. На программно-модульном уровне решается задача размещения программных модулей ПО АСРВ на внешних носителях РВК по компонентам сетевой файловой системы.

В соответствии с введёнными уровнями абстракции процесс проектирования программной системы предполагает решение следующих задач:

Структурный анализ и моделирование функциональной структуры программной системы компьютерного мониторинга на рассматриваемом уровне АСРВ в виде АСПД-схемы.

Синтез на основе АСПД-схемы процессной структуры программной системы в виде ВПМ-сети, реализующей функции функциональной структуры.

Синтез на основе ВПМ-сети агрегатной структуры программной системы и спецификация размещения программных агрегатов по процессорным компонентам РВК.

Распределение программных файловых объектов программной системы в каталогах файловой системы распределенного вычислительного комплекса АСРВ - формирование программно-модульной структуры ПО АСРВ.

Литература

1. Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем. – М.: Мир, 1973. – 344с.

2. Gertenbach W.P. A Hierarchical Organization of Real-Time Multicomputer Process Control Systems // Real-Time Data Handling & Process Control: Europ.Symp. 1-st, Berlin/West/, 1979, Brussels. – 1980. – P. 443-446.

3. Баландин А.В., Николаев А.В. Метод структуризации и РВ-верификации приложений реального времени для систем промышленной автоматизации // Надежность и качество. - Труды международного симпозиума. – Пенза: Изд-во Пенз. гос. ун-та. 2003. – С.378-380.

К проблеме построения оптимизирующих преобразований спецификации управляющих алгоритмов реального времени

А.Ю. Богатов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В последнее время при разработке программного обеспечения всё более широкое применение находит трансформационный подход к программированию. В первую очередь это связано с возрастающей сложностью, а также с высокими требованиями, предъявляемыми к надёжности и стабильности работы программных систем. Данное замечание особенно актуально для систем жёсткого реального времени (HRT-системы). Корректность работы HRT-системы определяется двумя условиями:

- корректность по выходным данным, то есть логическая правильность результатов;
- корректность по времени, то есть система переходит из одного запланированного состояния в другое в заранее заданные сроки.

При этом невыполнение второго условия потенциально может привести к катастрофическому отказу системы.

В работе приводятся подходы к проведению оптимизирующих преобразований для спецификации HRT-систем, основанные на исчислении управляющих алгоритмов реального времени.

Алгебра управляющих алгоритмов реального времени

Управляющие алгоритмы реального времени (УА РВ) можно представить в виде четверок объектов: $УА\ РВ = \{< f_i, t_i, \tau_i, \bar{l}_i >\}, i = \overline{1, N}$, где f_i – функциональная задача (действие); t_i — момент начала выполнения действия (целое неотрицательное число); τ_i — длительность действия (целое неотрицательное число); $\bar{l}_i$— логический вектор, обуславливающий выполнение функциональной задачи.

Множество УА РВ в смысле наборов четвёрок объектов U, множество логических условий L вместе с набором операций Ω образуют универсальную двухосновную алгебру $< U, L; \Omega >$. Система $< U, L; \Omega >$ называется алгеброй управляющих алгоритмов реального времени. Сигнатура Ω состоит из следующих операций: $\Omega = \{СН, СК, \rightarrow, Н, ЗА, @, +, \overset{\alpha}{\rightarrow}\}$, которые определяют координацию задач между потоками и процессами.

Операция $СН$ — совпадение по началу, задачи начинают своё выполнение в один и тот же момент времени.

Операция $СК$ — совпадение по концу, задачи заканчивают своё выполнение в один и тот же момент времени.

Операция $\rightarrow$ — непосредственное следование, задача В начинает своё выполнение после окончания выполнения задачи А.

Операция $Н$ — непосредственное следование, задача В начинает своё выполнение через промежуток времени t после начала выполнения задачи А.

Операция $ЗА$ — следование со сдвигом, задача В начинает своё выполнение через промежуток времени t после окончания выполнения задачи А.

Операция $@$ — привязка начала выполнения задачи к абсолютному значению времени t.

Операция $+$ — операция выбора динамического объекта.

Операция $\overset{\alpha}{\rightarrow}$ — запуск задачи на выполнение, если условие α истинно.

В алгебре управляющих алгоритмов реального времени выполняются следующие тождества:

1) $T_1 \; CH \; T_2 = T_2 \; CH \; T_1$,
2) $T_1 \; CK \; T_2 = T_2 \; CK \; T_1$,
3) $T_1 + T_2 = T_2 + T_1$,
4) $(T_1 \rightarrow T_2) \rightarrow T_3 = T_1 \rightarrow (T_2 \rightarrow T_3)$,
5) $(T_1 \; CH \; T_2) \; CH \; T_3 = T_1 \; CH \; (T_2 \; CH \; T_3)$,
6) $(T_1 \; CK \; T_2) \; CK \; T_3 = T_1 \; CK \; (T_2 \; CK \; T_3)$,
7) $(T_1 \rightarrow T_2) CH (T_1 \rightarrow T_3) = T_1 \rightarrow (T_2 \; CH \; T_3)$,
8) $(T_1 \rightarrow T_2) CK (T_3 \rightarrow T_2) = (T_1 \; CK \; T_3) \rightarrow T_2$,
9) $(T_1 \rightarrow T_2) + (T_1 \rightarrow T_3) = T_1 \rightarrow (T_2 + T_3)$,
10) $(T_1 \rightarrow T_2) + (T_3 \rightarrow T_2) = (T_1 + T_3) \rightarrow T_2$,
11) $(T_1 \; CH \; T_2) + (T_1 \; CH \; T_3) = T_1 \; CH \; (T_2 + T_3)$,
12) $(T_1 \; CH \; T_2) + (T_3 \; CH \; T_2) = (T_1 + T_3) \; CH \; T_2$,
13) $(T_1 \; CK \; T_2) + (T_1 \; CK \; T_3) = T_1 \; CK \; (T_2 + T_3)$,
14) $(T_1 \; CK \; T_2) + (T_3 \; CK \; T_2) = (T_1 + T_3) \; CK \; T_2$,
15) $(\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} (T_1 \rightarrow T_2) = ((\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} T_1) \rightarrow ((\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} T_2)$,
16) $(\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} (T_1 \; CH \; T_2) = ((\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} T_1) CH ((\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} T_2)$,
17) $(\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} (T_1 \; CK \; T_2) = ((\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} T_1) \; CK ((\alpha_1 = 1) \overset{\alpha_1}{\rightarrow} T_2)$,
18) $T_1 \; CH \; T_1 = T_1$,
19) $T_1 \; CK \; T_1 = T_1$,
20) $(\alpha \overset{\alpha}{\rightarrow} T) + (\neg \alpha \overset{\alpha}{\rightarrow} T) = T$,
21) $(\alpha \overset{\alpha}{\rightarrow} (\beta \overset{\beta}{\rightarrow} T)) = (\alpha \wedge \beta) \overset{\alpha \wedge \beta}{\rightarrow} T$.

Оптимизирующие преобразования спецификации УА РВ

Спецификация является формальным описанием задач, решаемых проектируемой системой и представляет собой систему уравнений в алгебре УА РВ.

С точки зрения модели пред- и постусловий Хоара, критерий корректности работы HRT-системы будет выглядеть следующим образом: $(A(D_0, t_0)) \pi (B(D_1(t_1), D_2(t_2), ..., D_k(t_k))) \equiv 1$. Здесь $(A(D_0, t_0))$ означает корректное задание начальных условий на момент времени t_0, а $(B(D_1(t_1), D_2(t_2), ..., D_k(t_k)))$ означает, что в результате выполнения программы π были корректно выполнены целевые задачи в моменты времени $t_1, ..., t_k$. Данное соотношение является основополагающим и должно присутствовать в спецификации каждой программной системы.

Ниже приведены подходы, использованные для построения алгоритмов оптимизирующих преобразований спецификации УА РВ, критерием оптимальности является минимум длины формулы спецификации.

1) Формальные преобразования спецификации. По набору тождеств 1) ─ 22) строится каноническая система переписывания термов с отношением редукции, определяемым выбранным критерием оптимальности. На основании полученной системы переписывания термов задаётся конечный автомат, определяющий алгоритм оптимизирующих преобразований.

2) Оптимизирующие преобразования спецификации, основанные на представлении термов конечными автоматами.
Переход от терма к конечному автомату выполняется следующим образом: по дереву синтаксического разбора терма строится граф функции перехода искомого автомата, в соответствии с заранее определённым алгоритмом нумерации элементов из алгебры $< U, L; \Omega >$, находятся состояния конечного автомата. Далее к заданному таким образом автомату применяется алгоритм минимизации.

3) Оптимизирующие преобразования, основанные на представлении спецификации в виде системы линейных неравенств.
Любое отношение в алгебре УА РВ однозначно определяется отношениями <, >, = на декартовом произведении $N \times N \times \{0,1\}^J$, где J ─ размерность логического вектора, обуславливающего выполнение функциональных задач данного множества. Так же любая операция в алгебре УА РВ однозначно определяется стандартными операциями умножения и сложения на множестве натуральных чисел. В соответствии с этим утверждением выполняется переход к системе линейных неравенств. Решение полученной системы задаёт оптимальное представление спецификации, если система несовместна, значит, спецификация содержит противоречивые условия.

Измерительная система для прочностных испытаний элементов авиационных конструкций

Л.Э. Вилоп
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Опыт разработки летательных аппаратов показал, что перед стендовыми прочностными испытаниями летательного аппарата в целом целесообразно проводить прочностные испытания каждого отдельного элемента конструкции. Такой подход позволяет более тщательно испытать каждый элемент и сократить впоследствии время стендового испытания летательного аппарата в целом. При этом удается получить более полную информацию о прочностных характеристиках элемента, проверить правильность расчетов, подыскать лучшие варианты конструкции, обладающие статической прочностью и требуемым ресурсом при наименьшей массе.

Отличительной особенностью прочностных испытаний элементов конструкции является небольшое число точек, в которых необходимо производить измерение относительных деформаций, что делает нецелесообразным применение для этих целей больших измерительных систем. Вместе с тем большая номенклатура элементов, подлежащих испытаниям, и большая продолжительность ресурсных испытаний предопределяет необходимость применения для этих целей недорогой автоматизированной системы на основе персонального компьютера.

Для дистанционного измерения деформаций используются тензорезисторные датчики, особенностью которых является очень малый диапазон относительного изменения сопротивления не превышающий 0,6 %. В связи с этим на результат преобразования относительного изменения сопротивления в цифровой код большое влияние оказывает нестабильность параметров элементов канала преобразования, например, температурные изменения сопротивлений проводников линии связи, э.д.с. смещения усилителя нормализатора и т.д.

Целью данной работы являлась разработка несложной измерительной системы, удобной в эксплуатации и обладающей точностью преобразования относительного изменения сопротивления в цифровой код, необходимой для проведения экспериментальных исследований.

Функциональная схема подсистемы аналого–цифрового преобразования разработанной измерительной системы приведена на рис. 1.

При разработке системы учитывалось следующее:

–прочностные испытания элементов конструкции, как правило, производятся в лабораторных условиях, длина линий связи между датчиками и электронной аппаратурой не превышает нескольких метров;

– перед очередным статическим нагружением и в периодических перерывах ресурсных испытаний возможна аддитивная автокалибровка всего канала преобразования, включая датчик;

– практически всегда в таких измерительных системах используются полумостовые измерительные цепи с одним рабочим (R_{T_i}) и одним компенсационным тензорезистором ($R_{T_{комп}}$), который может быть общим для всех рабочих тензорезисторов системы;

– необходимость трудоёмкой и ненадёжной начальной аналоговой балансировки измерительных цепей, обусловленной большим разбросом значений начальных сопротивлений $\Delta R_{оном}$ тензорезисторов, достигающем единиц процентов, устраняется

в современных тензоизмерительных системах, так называемой цифровой балансировкой, которая возможна при использовании в системе АЦП повышенной разрядности.

В системе реализовано поочерёдное импульсное питание измерительных цепей посредством коммутатора напряжения питания датчиков с активной компенсацией влияния остаточных параметров бесконтактных ключевых элементов (K_{KT} , K_i).

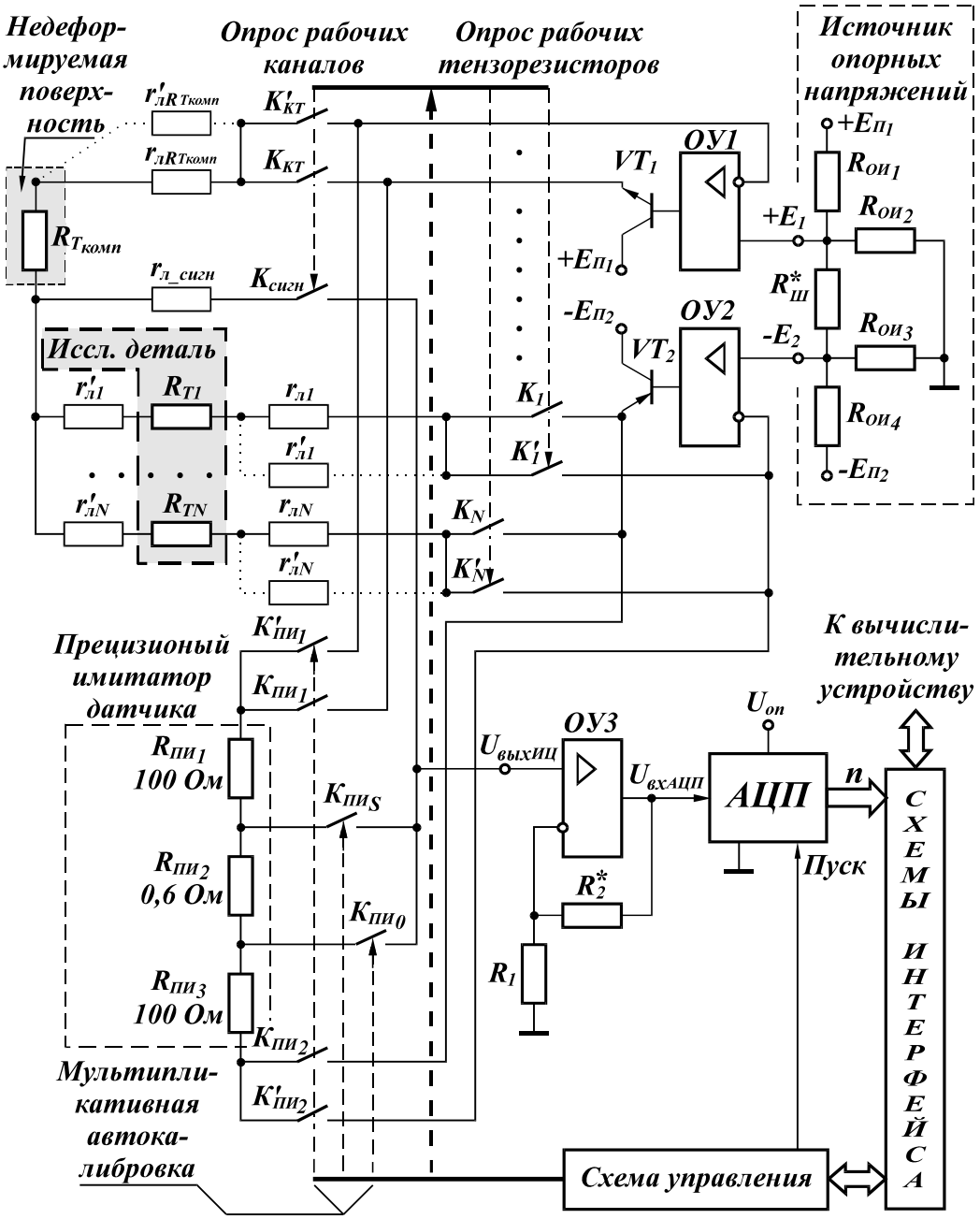


Рис. 1. Функциональная схема подсистемы аналого-цифрового преобразования измерительной системы

Напряжения отрицательной обратной связи на инвертирующие входы операционных усилителей ***ОУ1*** и ***ОУ2*** могут подаваться как с выходов ключевых элементов (K_{KT} , K_i) коммутатора напряжения питания, так и непосредственно с датчиков через дополнительные проводники линии связи ($r'_{лR_{Tкомп}}$, $r'_{лi}$). В последнем случае влияние сопротивлений проводников ($r_{лR_{Tкомп}}$, $r_{лi}$) исключается практически полностью.

Выходные сигналы измерительных цепей с опрашиваемыми датчиками через ключевой элемент $K_{сигн}$ на МОП–транзисторе подаются на вход неинвертирующего усилителя ($ОУ3, R_1, R_2$) для усиления и последующего аналого-цифрового преобразования.

Несложно показать, что для реализации возможности цифровой балансировки должно выполняться соотношение:

$$\Delta U_{вх\ АЦП_i} = \frac{\varepsilon_{макс}}{\Delta R_{оном} / R_{оном}} \Delta U_{вх\ АЦП_{макс}},$$

где: $\Delta U_{вх\ АЦП_i}$ – диапазон изменения измерительного сигнала на входе АЦП, устанавливаемый при настройке канала преобразования коэффициентом усиления неинвертирующего усилителя (резистор R_2);

$\Delta U_{вх\ АЦП_{макс}}$ – паспортное значение диапазона входного сигнала АЦП;

$\Delta R_{оном} / R_{оном}$ – относительный "разброс" начальных сопротивлений в партии тензорезисторов;

$\varepsilon_{макс} = \Delta R_{макс} / R_{оном}$ – диапазон относительного изменения сопротивления тензорезистора при его деформации (для металлических тензорезисторов $\varepsilon_{макс} \le (6-7) \cdot 10^{-3}$).

При этом диапазон изменения выходного кода АЦП (Δn_i) при опросе любого из датчиков определяется аналогичным выражением:

$$\Delta n_{i\,макс} = \frac{\varepsilon_{макс}}{\Delta R_{оном} / R_{оном}} \Delta n_{макс},$$

где $\Delta n_{макс}$ – диапазон изменения выходного кода АЦП, определяемый его разрядностью.

Для исключения из результирующей погрешности измерения мультипликативной погрешности, вносимой каналом преобразования электрических величин, в измерительной системе используется периодическая мультипликативная автокалибровка по прецизионному имитатору датчика (резисторы $R_{ПИ1}$, $R_{ПИ2}$, $R_{ПИ3}$), на которую одновременно возложена функция градуировки измерительной системы.

Схема на указанных резисторах имитирует полумост с двумя рабочими тензорезисторами. Значение калибровочного относительного изменения сопротивления $\varepsilon_к$, воспроизводимого таким прецизионным имитатором при его опросе (ключи $K_{ПИ0}$ и $K_{ПИS}$), определяется выражением

$$\varepsilon_к = 2 \frac{R_{ПИ2}}{R_{ПИ1} + R_{ПИ3}}.$$

При этом на выходе АЦП получаются коды $n_{к2}$ (при замкнутом ключе $K_{ПИS}$) и $n_{к1}$ (при замкнутом ключе $K_{ПИ0}$). Разность этих кодов прямо пропорциональна коэффициенту преобразования относительного изменения сопротивления в цифровой код $\frac{n_{к2} - n_{к1}}{\varepsilon_к}$ и "отслеживает" все его изменения.

Напряжение на выходе полумостовой измерительной цепи с одним рабочим тензорезистором определяется известным выражением: $U_{вых\,ИЦ} = \frac{1}{2}\varepsilon E$. Каналом преобразования это напряжение преобразуется в цифровой код на выходе АЦП:

$$n = \frac{1}{2}\varepsilon \frac{n_{к2} - n_{к1}}{\varepsilon_к}.$$

С учётом кода n_{0_i} на выходе АЦП, получаемого при нулевой нагрузке и устраняющего влияние на результат измерения начального разбаланса полумоста и аддитивных погрешностей (за исключением погрешности квантования), приведённое выше равенство можно записать как

$$n - n_{0_i} = \frac{1}{2}\varepsilon \frac{n_{к2} - n_{к1}}{\varepsilon_к}.$$

Откуда получим выражение, определяющее значение ε:

$$\varepsilon = 2\,(n - n_{0_i})\frac{\varepsilon_к}{n_{к2} - n_{к1}}.$$

Из этого выражения следует, что погрешность определения ε определяется только отклонением действительного значения $\varepsilon_к$ от значения, используемого в вычислениях. Кроме этого, при таком построении системы обеспечивается инвариантность к начальным сопротивлениям используемых в системе тензорезисторов. То есть без всякой перестройки в системе могут использоваться тензорезисторы из одной партии с любым значением $R_{ном}$. Естественно, что значение тока через тензорезистор (с поправкой на импульсное питание) при этом не должно превышать допустимого.

Если учесть, что относительная деформация тензорезистора $\varepsilon_l = \varepsilon / K_T$, где K_T – коэффициент тензочувствительности тензорезистора, то можно записать окончательное выражение для вычисления значения измеряемой деформации в следующем виде:

$$\varepsilon_l = \frac{2}{K_T}(n - n_{0_i})\frac{\varepsilon_к}{n_{к2} - n_{к1}}.$$

Таким образом, всё, что нужно при эксплуатации разработанной системы, это:

– проконтролировать соответствие допустимого тока через тензорезистор току, создаваемому в измерительных цепях схемой питания ($I_{R_0} = \frac{E_1 + E_2}{2R_{0ном}} \le I_{R_{0доп}}$, где E_1 и E_2 – напряжения на выходах источника опорных напряжений, $I_{R_{0доп}}$ – паспортное значение допустимого тока через тензорезистор), и при необходимости изменить значения E_1 и E_2 шунтирующим резистором $R_ш$;

– ввести в память вычислительного устройства значение $\varepsilon_к = 2\frac{R_{ПИ2}}{R_{ПИ1} + R_{ПИ3}}$;

– ввести в память вычислительного устройства значение коэффициента тензочувствительности K_T для используемой при испытаниях партии тензорезисторов;

– опросить перед началом испытаний измерительные каналы, для получения и внесения в память кодов n_{0_i}.

Формализованное описание производственного процесса предприятия

А.П. Зеленин
Оренбургский государственный университет, Оренбург

Работа выполняется при финансовой поддержке гранта № 613 «Исследование механизма и закономерностей перехода от технического задания к техническому предложению на создание гибких производственных ячеек» в рамках аналитической ведомственной целевой программы «Развитие научного потенциала высшей школы (2009-2010 годы).

При разработке и интеграции программных продуктов, предназначенных для автоматизации разработки продукции и подготовки производства, встает вопрос формализованного описания производственной системы предприятия.

Предприятие занято изготовлением множества машиностроительных изделий $P = (p_1, p_2, ..., p_j, ..., p_n)$, где, p_i – тип изделия, n – ассортимент.

Каждое изделие изготавливается партиями Or размерами num_i штук

$$Or = \{P = (p_1, p_2, ..., p_j, ..., p_n) \times Num = (num_1, num_2, ..., num_j, ..., num_n)\}.$$

Соответственно, множество изделий предприятия образует пространство

$$P' = \{ p_1 \times num_1 \in (Or_1, ..., Or_j, ..., Or_n), ..., p_i \times num_i \in (Or_1, ..., Or_j, ..., Or_n), ...,$$

$$p_k \times num_k \in (Or_1, ..., Or_j, ..., Or_n)\}.$$

С другой стороны, каждое изделие p_i может состоять из неограниченного количества сборочных единиц *au*, комплектов *s*, комплексов *c*, деталей *d*, стандартных изделий *sp*, прочих изделий *op* и материалов *m*. При этом все элементы, кроме материалов, являются самостоятельными изделиями:

$$p_i = \{au, s, c, d, sp, op, m\}.$$

Соответственно, каждое изделие предприятия представимо в виде:

$$p_i = \begin{Bmatrix} au_1 \times num_1^{au}, ..., au_j \times num_j^{au}, ..., au_n \times num_n^{au} \\ s_1 \times num_1^{s}, ..., s_k \times num_k^{s}, ..., s_m \times num_m^{s} \\ c_1 \times num_1^{c}, ..., c_a \times num_a^{c}, ..., c_l \times num_l^{c} \\ d_1 \times num_1^{d}, ..., d_b \times num_b^{d}, ..., d_z \times num_z^{d} \\ sp_1 \times num_1^{sp}, ..., sp_e \times num_e^{sp}, ..., sp_x \times num_x^{sp} \\ op_1 \times num_1^{op}, ..., op_q \times num_q^{op}, ..., op_y \times num_y^{op} \\ m_1 \times num_1^{m}, ..., m_t \times num_t^{m}, ..., m_w \times num_w^{m} \end{Bmatrix}$$

Состав изделия можно представить в виде графа (см. рис. 1).

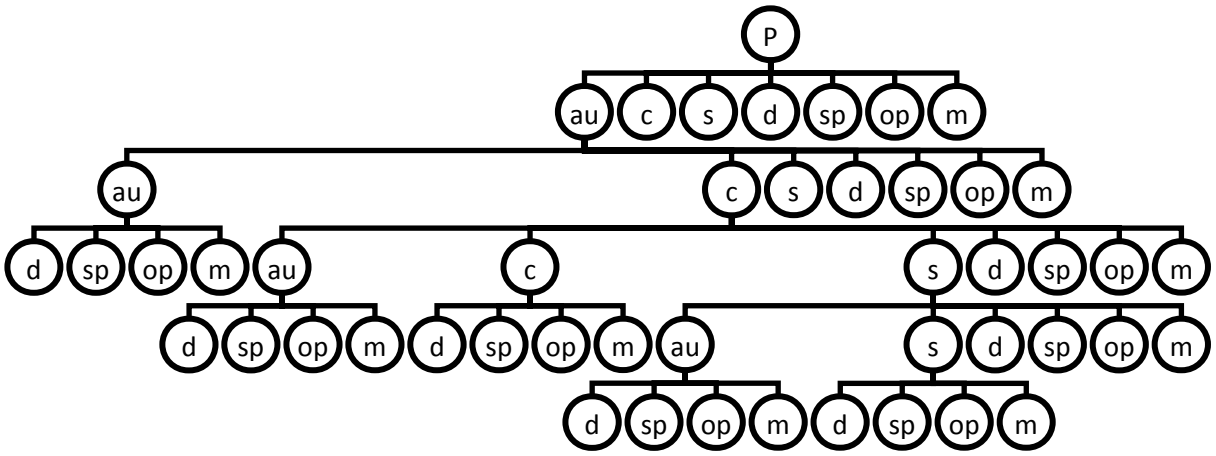


Рис. 1. Состав изделия

На нижних уровнях графа всегда будут детали, стандартные изделия, прочие изделия и материалы. Как правило, в иерархических системах физический смысл имеют только элементы нижнего уровня. Высшие уровни носят организационный или информационный характер. Однако при описании состава изделия и производственной системы физический смысл имеют все элементы, т.к. каждому элементу, вплоть до высшего, сопоставлен уникальный технологический процесс.

Кроме того, у каждого изделия может быть конструкторский и производственный состав. В этом случае верхний элемент и набор элементов нижнего уровня одинаковы, а элементы промежуточных уровней и связи между ними могут значительно отличаться.

Все составы всех изделий можно описать одним типом бинарной связи «входит в… - состоит из….», связывая объекты попарно. Таким образом, мы имеем множество связей для пар объектов

$$VHV = \{Vhv_1(p_i, p_j),...,Vhv_k(p_i, p_j),...,Vhv_n(p_i, p_j)\}.$$

Все элементы могут иметь как уникальные, так и общие свойства. Свойства могут суммироваться в соответствии с иерархией (масса, время обработки) и могут появляться в виде синергетического эффекта свойств элементов нижних уровней. Возможно, определить общие свойства (атрибуты) для определенных типов объектов, а так же указать какие из свойств будут обязательными $p_i(ob_i, name_i, mass_i,...)$.

На предприятии используется множество материалов $M = (m_1, m_2,..., m_j,..., m_n)$.

Каждая деталь, стандартное и прочее изделие может изготавливаться из одного материала m_i, при этом существуют возможные замены $Mzam = (m_1,..., m_k) \in M$.

Соответственно существует множество бинарных связей «изготавливается из … - для изготовления …» для пары деталь-материал

$$PI = \{Pi_1(d_i, m_j),..., Pi_k(d_i, m_j),..., Pi_n(d_i, m_j)\}.$$

Любому изделию может быть поставлен в соответствие один или несколько технологических процессов $T_k(p_i)$.

Соответственно существует множество бинарных связей «изготавливается по … - для изделия …»

$$Pb = \{Pb_1(p_i, t_j),..., Pb_k(p_i, t_j),..., Pb_n(p_i, t_j)\}.$$

Каждый технологический процесс представляет собой определенную последовательность операций

$$T_i = \left\langle op_1 \prec op_2 \prec ... \prec op_j \prec ... \prec op_m \right\rangle = \langle op, Ro \rangle,$$

где $op = \{op_i\}$, Ro - порядок следования операций.

Каждая операция может быть представлена в виде

$$op = (Eq, Ri(Toll, Ad, Rm, Me), Maux, PRo, t_{шт}, t_{п.з.}),$$

где Eq – применяемое оборудование; Ri – оснастка; $Tool$ – инструмент основной и вспомогательный; Ad – приспособления; Rm – средства защиты; Me – средства измерения; $Maux$ – используемый вспомогательный материал; PRo – профессия рабочего, выполняющего операцию; $t_{шт}$ – штучное время; $t_{п.з.}$ – подготовительно-заключительное время.

Если все элементы графа производственного состава изделий заменить на технологические процессы с подчиненными им операциями, то на нижних уровнях графа получим множество операций.

$$Op = (op_1, ..., op_i, ..., op_n).$$

Для производственного процесса имеет значение только этот нижний уровень графа. Все остальные уровни имеют только организационный характер.

Последовательность операций можно описать множеством связей

$$Link = \{Link_1(op_i \prec op_j), ..., Link_k(op_i \prec op_j), ..., Link_n(op_i \prec op_j)\}.$$

Если конечную операцию, элемента стоящего выше в иерархии связать с начальной операцией нижестоящего элемента, то можно построить граф стоящий только из значимых элементов – операций. Таким образом, мы получим операционный граф (диаграмму предшествования) производственной системы предприятия (см. рис. 2).

Соответственно всю производственную систему можно представить в виде двух множеств – операций и связей операций.

$$Op = (op_1, ..., op_i, ..., op_n)$$

$$Link = \{Link_1(op_i \prec op_j), ..., Link_k(op_i \prec op_j), ..., Link_n(op_i \prec op_j)\}$$

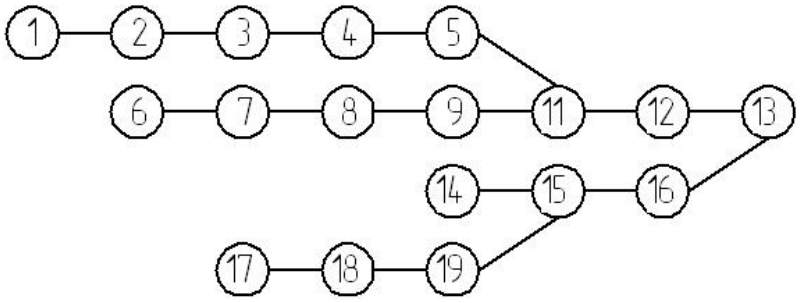


Рис. 2. Диаграмма предшествования

Применение метода экспресс-анализа для архивирования трендов АСУТП

А.Ю. Зенов, М.П. Строганов
Пензенский государственный университет, Пенза

В последнее десятилетие отмечается интенсивный процесс внедрения SCADA-систем. Он обусловлен возможностями современной измерительной базы, позволяющей не только выполнять измерения в полном объеме и с высокой точностью, но и передавать их в компьютер с помощью разнообразных телекоммуникационных средств. Информационной базой для работы SCADA-систем являются результаты измерений контрольно-измерительного оборудования, установленного на технологических объектах. Современные SCADA-системы обеспечивают сбор и долгосрочное хранение трендов всех технологических параметров для последующего анализа в случае необходимости. Тренд представляет собой массив точек переменных, каждая из которых записывается в реальной системе в память компьютера через определенные интервалы времени. Анализ трендов необходим для повышения надежности работы объекта, увеличения его КПД и т.д.

Хотя данные такого типа хорошо сжимаются архиваторами, имеется возможность сократить объем исходных трендов, а, следовательно, и сжатых данных за счет предварительной обработки.

Для этого предлагается:

– исходный тренд, за определенный временной промежуток подвергнуть низкочастотной фильтрации;

– из отфильтрованного сигнала выделить экстремумы;

– сохранить в преобразованном тренде только экстремальные значения переменных $t_{iэ}$, $x_{iэ}$;

– впоследствии разархивировать тренд и восстановить по предлагаемой далее методике.

На рис. 1 показан пример тренда. Покажем возможность применения предложенной методики. На рис. 2 показан переходной процесс в системе при изменении задания, а на рис. 3 показано применение метода предварительной обработки исторических данных путем экстремальной фильтрации – фильтрации, выделения экстремумов и восстановление на основе интерполяции.

Рассмотрим метод подробнее. Предлагается аппроксимация реализации гауссовыми кривыми, при этом аппроксимирующая функция принимает вид:

$$f(t)=\sum_{i=1}^{l}\sum_{k=1}^{m_i} x_{ki}\, e^{-\beta_{ki}^2\left(t-c_{ki}\right)^2}.$$

Процедура аппроксимации связана с выделением из сигнала знакопеременных составляющих, каждая из которых может быть отнесена к определенной полосе частот.

Значения экстремумов и интервалы между ними определяют параметры колокольных составляющих для сигнала вида $e^{-x^2/2\beta^2}$. Параметр β_i определяется следующим образом:

$\beta_i=\min((c_i-c_{i-1}),(c_{i+1}-c_i))/2{,}2$ или $\beta_i=(c_{i+1}-c_{i-1})/4{,}4$.

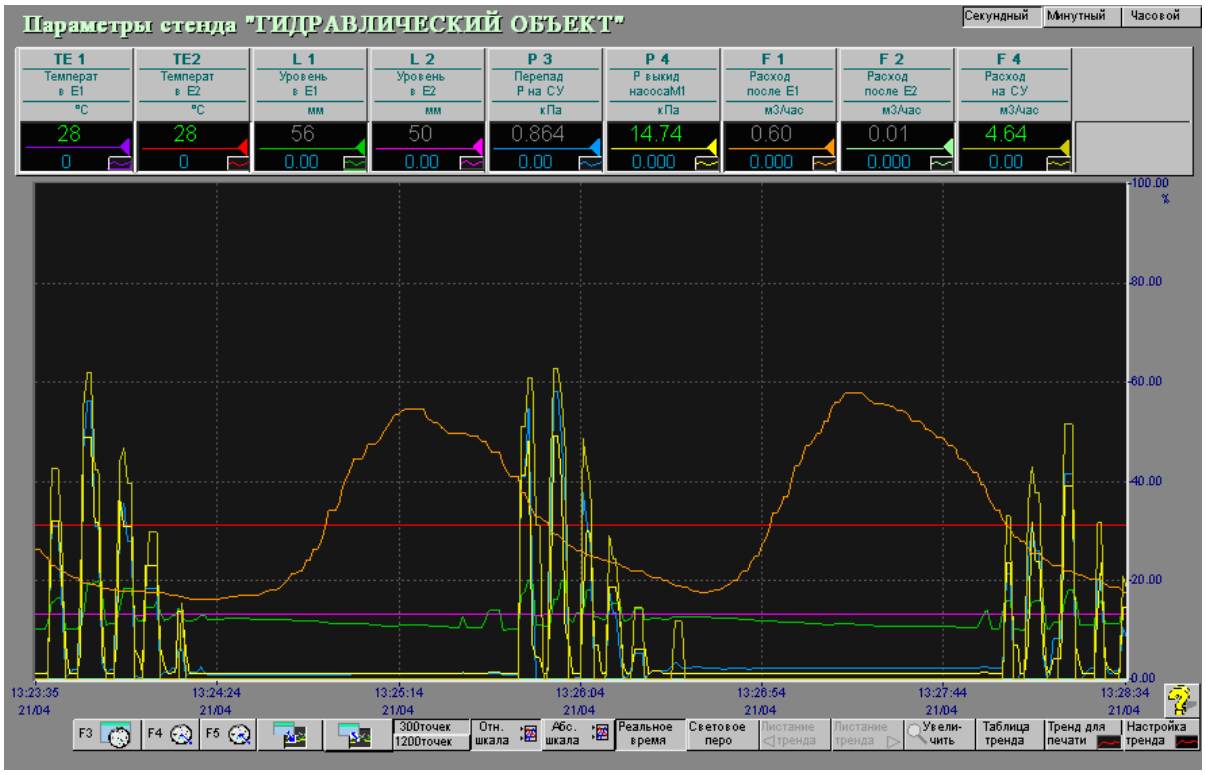


Рис. 1. Тренд учебного стенда «Гидравлический объект»

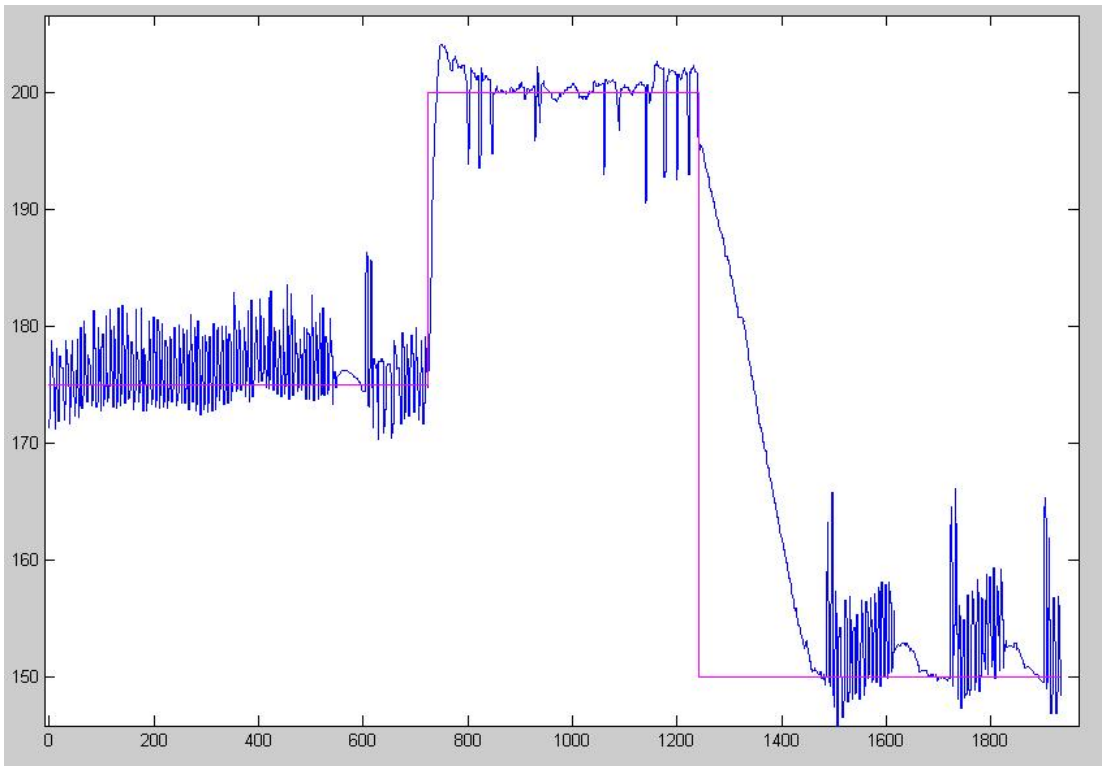


Рис. 2. Изменение регулируемого параметра при изменении задания

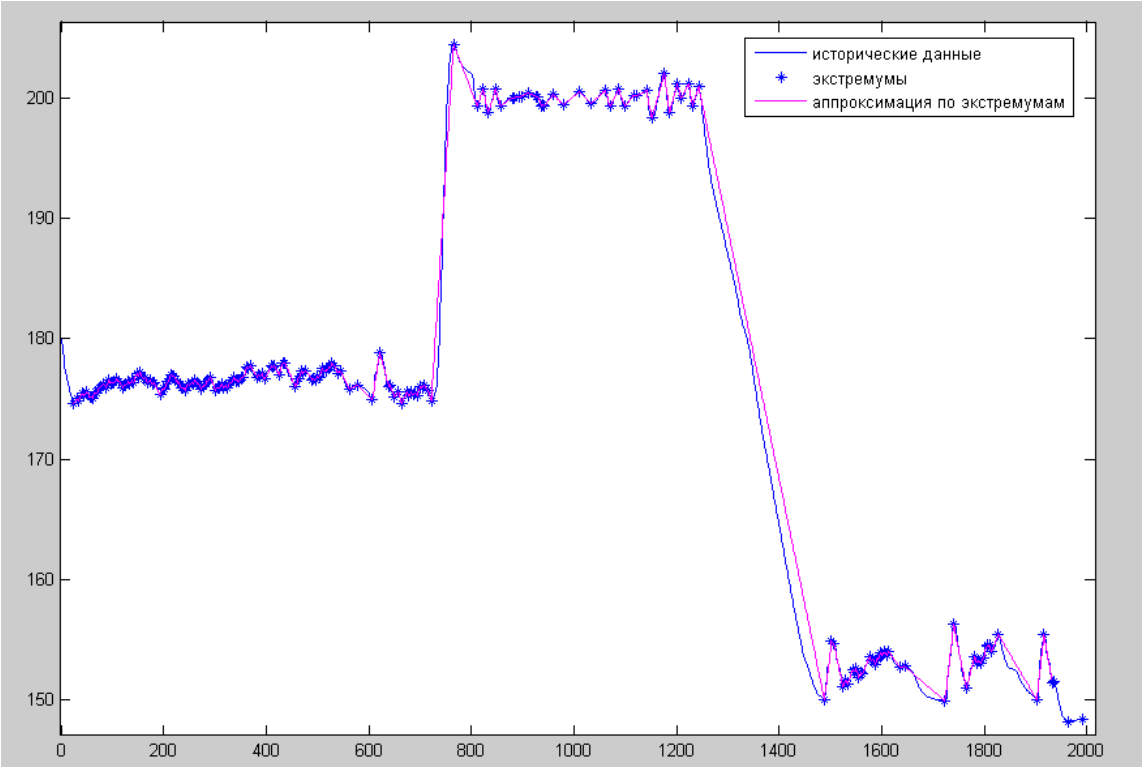


Рис. 3. Предварительная обработка исторических данных путем экстремальной фильтрации: синяя кривая – отфильтрованный исходный сигнал; * – выделенные экстремумы; розовая кривая – интерполяция по экстремумам

Преимущества метода по сравнению с другими методами аппроксимации в том, что он не требует оптимизации: выбор параметров аппроксимирующего сигнала осуществляется автоматически, при этом обеспечивается точность, приемлемая для практических нужд.

В результате проведенных экспериментов была доказана возможность применения метода экстремальной фильтрации и получена степень сжатия в 4 раза, а если не хранить моменты появления экстремумов, то в 8 раз по сравнению со стандартным способом обработки. Однако в реальности этот показатель будет ниже, т.к. степень сжатия зависит от количества параметров объекта.

Выводы

В рассмотренной статье показана возможность применения метода предварительной обработки исторических данных путем экстремальной фильтрации с возможностью восстановления, однако необходимы дополнительные исследования с разнообразными параметрами, а также их количеством.

Литература

1. Строганов М.П., Берестень М.П., Мясникова Н.В. Обработка сигналов в системах диагностики/ Под редакцией Е.П. Осадчего: Монография. - Пенза: Изд-во Пенз. гос. техн. ун-та, 1997.- 119 с.

2. Мясникова Н.В., Берестень М.П. Экстремальная фильтрация и ее приложения// Датчики и системы, №4, 2004. - С. 8-11.

3. Экспресс-анализ сигналов - Автоматика и телемеханика [Электронный ресурс] / Пензенский государственный университет, Автоматика и телемеханика; - Электрон. дан. – Пенза: ПГУ, 2008 - Режим доступа: http://www.pnzgu.ru/dep/k_ait/?q=express, свободный. — Загл. с экрана.

Разработка автоматизированной системы прогнозирования термодеформационного состояния технологического оборудования

К.В. Марусич
Оренбургский государственный университет, Оренбург

Способом оценки точности технологического оборудования служат производственные испытания, проводимые при изготовлении и после каждого ремонта. Реально на практике провести испытание с одновременным получением информации обо всей номенклатуре выходных параметров невозможно, прежде всего, из-за необходимости размещения всей измерительно-диагностической аппаратуры на испытательном стенде. Поэтому испытания рекомендуется проводить поэтапно, включая в каждый этап полные испытания по отдельным параметрам. Одним из таких параметров выделяют температурные деформации. Достоверно известно, что они составляют от 30 до 70% от общей погрешности обработки при изготовлении деталей на металлорежущих станках. Это связано с тем, что тепловые процессы в станках носят в течение всей рабочей смены ярко выраженный нестационарный характер, так как изменяются режимы резания, режущий инструмент, происходит чередование рабочих и холостых ходов, смена обрабатываемых поверхностей и т.п.

Температурное поле деталей станков может быть установлено опытным или теоретическим путем. Экспериментальные методы отличаются достоверностью результатов исследования, но вместе с этим и большой трудоёмкостью, ограниченной возможностью изменения исходных данных. Поэтому предпочтение обычно отдается аналитическим способам, которые позволяют определить температуру в любой точке детали. Однако детали при этом должны быть простой конфигурации с тем, чтобы их, с небольшим удалением от действительности, можно было представить в виде тел элементарных геометрических форм (стержня, пластины, цилиндра и т.д.) с простыми условиями теплообмена и теплообразования на поверхности.

Создание автоматизированной системы прогнозирования термодеформационного состояния технологического оборудования на этапе их эксплуатации позволит принимать решения и создавать автоматизированные системы компенсации температурных погрешностей. На данный момент существуют автоматизированные системы прогнозирования теплового состояния станков работающих в условиях непрерывного режима их работы, что на практике не реализуется в условиях реальной механической обработки. Поэтому разработка современных методов и средств оценки качества и надежности технологического оборудования является актуальной задачей.

Решением задачи повышения теплоустойчивости станков занимаются на кафедре «Технологии машиностроения, металлообрабатывающие станки и комплексы», Аэрокосмического института, Оренбургского государственного университета. В эту работу также вовлекаются и студенты, а результаты ее и мировые достижения в этой области они изучают на соответствующих дисциплинах и используют при выполнении курсовых и дипломных работ. В настоящее время на кафедре разрабатывается методика и автоматизированная система тепловых испытаний станков в условиях повторно-кратковременных режимов работы. Данный комплекс позволит в режиме реального времени прогнозировать сведения о термодеформационном состоянии станка.

Для решения поставленной научной задачи работы были проведены следующие натурные испытания станка высокой точности Deckel FP3:

– непрерывный режим работы станка на частотах вращения 800, 1000, 1250, 1600 и 2000 об/мин;

– повторно-кратковременный режим с хаотичным изменением частот вращения для различных типов циклограмм;

– повторно-кратковременный режим с регулярным изменением частот вращения для различных типов циклограмм;

– режим нагрев-остывание.

В качестве измерительной аппаратуры используются многоканальный измеритель температуры МИТ-12ТП-11 и термометр типа «Замер» для контроля технической среды. Цифровой прибор оснащен интерфейсом RS-232, поэтому все данные в режиме реального времени передавались в компьютер. Датчики для измерения температуры устанавливаются в различных точках станины, шпиндельной головки и на электродвигателе привода главного движения. Измерения температурных перемещений осуществляются с помощью трех многооборотных индикаторных головок типа МИГ и МИГП с ценой деления 1 мкм, которые фиксируют перемещения вдоль трех осей шпинделя станка.

На сегодняшний день общий объем выполненных экспериментальных тепловых испытаний станков превысил 150 часов. В основном, длительность каждого испытания станка составила от 300 до 480 минут. На рис. 1 представлены экспериментальные тепловые характеристики станка на четырех разных режимах работы:

1) на рисунке 1.а показаны температурные характеристики для датчика установленного на гильзе шпинделя станка;

2) на рисунке 1.б – характеристики температурных перемещений датчика расположенного вдоль оси шпинделя.

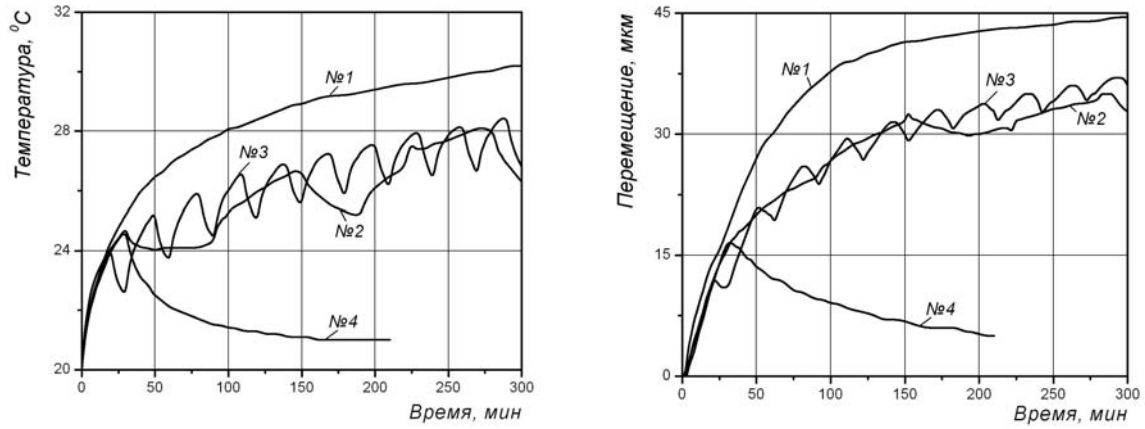


Рис. 1. Экспериментальные тепловые характеристики станка

Полученные данные натурных испытаний обрабатывались с помощью разработанного ранее и постоянно совершенствуемого авторского программного обеспечения на основе экспериментального модального анализа. Математическую основу используемого экспериментального модального анализа составляет решение для уравнения теплопроводности вида:

$$T_{ij,м}(t_j) = \sum_{k=1,4...,3m-2} x_{k,i}(1-e^{-t_j/x_{k+1,i}}) + \sum_{k=1,4,...,3m-2} x_{k+2,i}e^{-t_j/x_{k+1,i}},\ i=1,...,N,\ j=1,...,L \qquad (1)$$

где $x_{k,i}, x_{k+1,i}, x_{k+2,i}$ – модальные параметры; m – число температурных мод; i – номер датчика температуры; t_j – отдельный интервал времени, в который проводились измерения; N – общее число точек несущей системы станка, в которых измеряется

температура; L – число интервалов времени, составляющих рассматриваемый диапазон времени $t \in [t_1, t_L]$.

В настоящее время проводится обработка результатов натурных испытаний с использованием зависимости (1) для уточнения математической модели, применяемой в условиях повторно–кратковременном режиме работы станка. Рабочей гипотезой, лежащей в основе проводимых исследований, является физическое представление о суперпозиции различных тепловых процессов, протекающих в несущей системы станка, согласующихся с изменениями эксплуатационных режимов его работы. Фактически, в условиях повторно–кратковременного режима работы станка модальные параметры в уравнении (1) являются локальными в пределах каждого участка циклограммы. Есть два пути решения поставленной задачи: либо определение модальных параметров на каждом временном участке циклограммы станка, либо применение некоторой эквивалентной частоты вращения станка.

Модель оценивания достоверности значений параметров АСУТП

А.А. Сидоров, В.Е. Захарченко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Настоящая диссертационная работа посвящена разработке модели для оценивания достоверности (МОД) значений параметров АСУТП в реальном времени без остановки технологического процесса.

Поскольку АСУТП принимает управляющие решения на основе поступающей информации, то она должна быть достоверной. Проблема определения достоверности значений параметров АСУТП включает множество аспектов: корректность показаний источника информации, правильность передачи, вычислений, работоспособности связанных элементов системы, каналов передачи данных и многими другими.

Анализируя современные промышленные системы, необходимо отметить, что существует широкий спектр методов проверки достоверности входной информации, большинство из которых рассчитано на определение явной неисправности. Каждый метод формирует независимую оценку по некоторому критерию достоверности, однако по этой оценке нельзя судить о достоверности параметра в целом. Применительно к АСУТП методы определения достоверности параметров должны быть реализованы с учетом ограничений на время обработку информации и на ресурсы системы управления. В этой связи рассматриваемые вопросы представляется актуальными, и имеют важное научно-практическое значение.

В представленной работе описание математической модели для оценки достоверности параметров проводится на основе сравнения измеренной величины и некоторого диапазона, характерного для этого параметра в рассматриваемый момент времени. Для оценивания достоверности значений параметров в реальном времени предлагается синтезировать предложенную математическую модель оценивания достоверности с имитационной моделью технологического процесса, формирующей характерные диапазоны значений для параметров АСУТП в режиме реального времени на основе исторических и экспертных данных.

Для описания модели технологического процесса используются исторически сложившееся описание в виде установившихся и переходных режимов. Установившиеся режимы характеризуются стабильным изменением параметров по некоторому легко формализуемому закону. Эксперт разделяет весь процесс на участки, которые логично связать с технологическими режимами объекта управления (ОУ). В каждом режиме предполагается поведение параметра, качественно отличное от поведения на других участках. Формально описание режимов можно выполнить в виде карты состояний. Переход из одного режима в другой осуществляется событийно, то есть мгновенно по выполнению некоторого условия. По этой теме есть немало работ, выполненных Д. Форрестером, Д. Харелом, Д. Бучем, Якобсоном, Д. Раумбахом, Б. Селиком , А. С. Устенко, Н. П. Бусленко, А.Н. Мелиховым, Ю.Г. Карповым, Ю.Б. Сениченковым, Ю.Б. Колесовым и другими.

Картой состояний будем называть детерминированный гибридный автомат:

$$H = \{YC, SC, Q_0, F, Pr, Alg\}, \quad (1)$$

где *YC* - множество входных и выходных параметров, *SC = {Q, E}* – граф с множеством режимов $Q = \{q_1, \dots, q_z\}$ и множеством ориентированных дуг $E = \{(q_i, q_j),\ i, j = \overline{1, z}\}$, каждому режиму поставлена в соответствие одна из функций $f_i: T \times Y \to Y$, а каждой дуге – пара (pr_i, alg_i), состоящая из предиката и алгоритма, Q_0

– определяет множество начальных режимов. Причём каждый узел графа (режим) может быть и гиперузлом, кроме того в каждый момент непрерывного времени T на дугах, исходящих из текущего состояния, может стать истинным только один предикат.

Пусть параметр $Y \in YC$ измеряется и передаётся в систему управления, тогда в некоторый момент времени значение параметра Y в АСУТП основывается на измеренном значении, с учетом ошибок измерения и преобразования, передачи данных, вычисления: $y = y_{изм} + \varepsilon$, причем $\varepsilon = y - y_{изм} \neq const$. При пуско-наладочных работах АСУТП проводится метрологическая экспертиза по каждому параметру $Y \in YC$, характеризующая точность. Точность зависит от средств измерения, каналов передачи данных, внешних воздействий, ошибок округления, вычисления и прочими. После метрологической экспертизы можно упрощенно представить значение параметра в виде интервала:

$$Y = [y - a,\ y + a], \tag{2}$$

где $a = \max(\varepsilon) > 0$ – величина, характеризующая неопределённость параметра в рассматриваемый момент времени таким образом, что истинное значение параметра технологического процесса y^* непременно принадлежит интервалу $[y - a,\ y + a]$ в каждый момент времени технологического процесса. В общем случае величина a может изменяться во времени вместе с технологическим процессом. Существует достаточно большое количество методов оценивания величины a, описанных у Сквайрса Дж., Тюрина Н.И., Новицкого П.В. и многих других.

Оценивание достоверности

Достоверность значения параметра y есть степень соответствия значения в АСУТП y истинному значению параметра технологического процесса y^*, то есть достоверность D обратно пропорциональна отклонению $|y \quad y^*|$. Однако на практике истинное значение y^* неизвестно, поэтому достоверность можно оценить путём сравнения с другой величиной: характерным диапазоном значений параметра, полученным, путём отличным от (2).

Пусть задан характерный диапазон параметра Y как экспертная интервальная оценка:

$$Y_{хар} = [y_{min},\ y_{max}], \tag{3}$$

Оценка достоверности формируется в результате сравнения значения параметра в настоящий момент времени Y с границами характерного диапазона этого параметра $Y_{хар}$ в этот же момент времени. Достоверность значения параметра y оценивается мерой вхождения интервала Y в интервал $Y_{хар}$, определённой на интервале [0; 1]:

$$D(y) = \begin{cases} 0, & y - a > y_{max} \text{ или } y + a < y_{min} \\ 1, & y - a > y_{min} \quad и\ y + a < y_{max} \\ \dfrac{y + a - y_{min}}{2a}, & y - a \leq y_{min} \\ \dfrac{y_{max} - (y - a)}{2a}, & y + a \geq y_{max} \end{cases} \tag{4}$$

Рассуждая таким образом, можно заметить, что значение параметра, заданное с функцией $D(y)$ образует трапецеидальную нечёткую величину $\overline{y}$ заданную на множестве вещественных чисел $\Re$, с множеством функций определяющих степень принадлежности $D(y)$ значения параметра Y к достоверным (4). Иначе выражение (4) можно записать в виде $\overline{y} = (y_1; y_2; y_3; y_4) = (y_{min} - (2a);\ y_{min};\ y_{max};\ y_{max} + (2a))$.

Поскольку в общем случае для некоторых параметров Y системы можно сформировать не одну, а несколько интервальных оценок $Y_1..Y_K$ вида (3), то для каждого Y_k можно определить нечёткую величину $\overline{y_k}$ с достоверностью, определяемой выражением (4), $1 \leq k \leq K$, где K – общее число интервальных оценок параметра Y.

Объединенную оценку достоверности можно получить на основе нечёткой арифметики несколькими способами: объединением нечётких величин, анализом минимумов и максимумов функций принадлежности и другими. Выбор способа расчёта должен обеспечивать необходимое быстродействие для систем АСУТП, а также отвечать особенностям технологического процесса и знаниям эксперта. В этой связи для расчёта объединённой оценки достоверности предлагается способ с низкой вычислительной сложностью, а именно – линейная комбинация нечётких величин, где линейные коэффициенты определяются экспертом:

$$D_{общ} = \sum_{k=1}^{K} \gamma_k D(y_k), \quad (5)$$

где $D_{общ}$ – объединённая оценка достоверности, $D(y_k)$ - достоверность параметра Y по k-й интервальной оценке, $\gamma_k \in [0; 1]$ – весовой коэффициент такой, что $\sum_{k=1}^{K} \gamma_k = 1$, γ_k – задаётся экспертом исходя из анализа приоритетов интервальных оценок. Если все характерные диапазоны эквивалентны (имеют равный приоритет, по мнению эксперта), то $D_{общ} = \frac{1}{K} \sum_{k=1}^{K} D(y_k)$.

Для интерпретации результата введём две величины: порог доверия d_{trust} и порог сомнения d_{doubt} такие, что $d_{trust}, d_{doubt} \in [0,1]$ и $d_{trust} > d_{doubt}$. Значение параметра Y в момент времени t (см. Рис. 2):

$$y_t = \begin{cases} достоверно, & D_{общ} \geq d_{trust} \\ недостоверно, & D_{общ} < d_{doubt} \\ сомнительно, & иначе. \end{cases}, \quad (6)$$

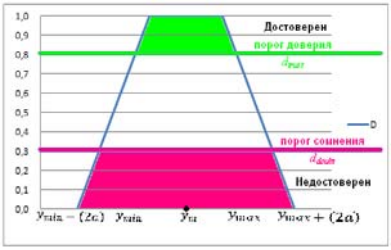


Рис. 2. Интерпретация достоверности

На основе представленной математической МОД возможно формирование модельного значения. В случае отказа источника данных или канала система управления может ориентироваться на модельное значение, вычисляемое как нечёткое среднее по всем нечетким величинам y_k вида $\overline{y_k} = (y_1; y_2; y_3; y_4)$:

$$y_m = \frac{1}{4}(y_1 + y_2 + y_3 + y_4) \quad (7)$$

Модельное значение параметра может участвовать в алгоритмах управления, как минимум, для безаварийной приостановки технологического процесса. Кроме того модельное значение может быть использовано для проверки алгоритмов системы управления без ОУ, а также для целей обучения персонала. По описанной методике производится формирование модельных значений параметров в АСУ стендами наземных и высотных испытаний авиационных двигателей ТС12 и ДГ4.

Формирование характерного диапазона

В представленной работе предлагается алгоритм формирования характерного диапазона параметра, основанный на имитационной модели технологического процесса и учитывающий историю функционирования АСУТП, опыт экспертов и функциональные зависимости между параметрами процесса. Модель описывает ожидаемое развитие технологического процесса и исключает основные источники недостоверности параметров большинства современных АСУТП. Основные трудности при создании имитационной модели для проверки достоверности связаны с формальным описанием истории функционирования системы и функциональных зависимостей.

Оценка достоверности на основе исторических данных предполагает сравнение значения параметра в настоящий момент времени (2) с аналогичными значениями в прошлом. Учитывая, что число предшествующих реализаций постоянно увеличивается,

удобно выделить диапазон *характерных* значений параметра в каждом из режимов, и производить сравнение текущего значения параметра с границами диапазона.

Рассмотрим формирование характерного диапазона, основанное на исторической информации поведения параметра АСУТП, в одном режиме. Пусть имеется N реализаций некоторого режима, тогда значение параметра $Y \in YC$ в каждый момент времени можно с некоторой уверенностью заключить в интервал, определённый предыдущими реализациями $y_1,\ldots,y_N$, которые, в общем случае, имеют различную длительность режима $T_1 \ldots T_N$. Если отмерять время от начала режима $t \in [0, max(T_1 \ldots T_N)]$, то значение параметра y в некоторый момент времени t предположительно будет лежать в интервале между наименьшим значением параметра из всех реализаций на данный момент времени t и наибольшим значением параметра по всем реализациям в тот же момент времени:

$$\min\left(y_1(t),\ldots,y_N(t)\right) = y_{min}(t) \leq y(t) \leq y_{max}(t) = \max\left(y_1(t),\ldots,y_N(t)\right) \qquad (8)$$

Объединяя наименьшие значения параметра во всех реализациях режима одной функцией, а наибольшие – другой, получается такая пара функций $Y_{хар}$ вида (3), которая в каждый момент времени образует интервал, содержащий значения всех предшествующих реализаций параметра. В каждом из режимов на основе данных архива системы управления (SCADA-системы) можно сформировать характерный диапазон изменения параметра.

Очевидно, недостаток ретроспективных данных может восполнить эксперт технологического процесса: Эксперт на основе своего опыта и знаний определит величину, на которую следует увеличить характерный диапазон в каждом режиме. Для уменьшения количества режимов, а также для формализации поправок эксперта могут быть использованы различные способы описания границ характерного диапазона, например на основе аппроксимации с использованием нелинейных функций.

Кроме исторических данных описание характерного диапазона для оценивания достоверности параметров АСУТП может быть выполнено на основе вычисления значения одного параметра системы через значения других с использованием формализованных закономерностей, описанных в виде функций из предметной области технологического процесса. В случае отсутствия строгих равенств или сложности их определения, на производстве часто используют эмпирические зависимости. Таким образом, один параметр системы может быть определён расчётным путём через другие.

Оценка достоверности на основе функциональных зависимостей предполагает сравнение измеренного значения параметра в настоящий момент времени (2) с расчётным значением этого же параметра в этот же момент времени, которое можно представить в виде (3). Во множество функций F карты состояний (1) включаются функциональные зависимости вида:

$$Y_{хар} = Y_{расчётное} = f(Y_1 .. Y_n) = [y_{min}; y_{max}], \qquad (9)$$

где $Y_{расчётное}$ – интервальное вычисленное значение параметра Y, f – интервальная функция, заданная для параметра Y, $Y_1..Y_n$ – интервальные значения параметров в рассматриваемый момент времени, n – число параметров, от которых зависит Y в функциональной зависимости. В общем случае для некоторых параметров Y системы можно сформировать не одно, а систему интервальных равенств вида (9). Вычисление функций (9) может быть организовано с использованием аппарата достоверных вычислений, описанных у У. Кулиша, интервального анализа, описанного у И.Ю. Шокина, С.П. Шарого и др., или недоопределённых моделей А.С. Нариньяни.

Описание имитационной модели представляется картой состояний (1), где множество функций сформировано в виде (8) и (9).

Исполнение такой имитационной модели синхронно с АСУТП позволит формировать в каждый момент времени характерные диапазоны по параметрам

АСУТП, и, как следствие, проводить системное оценивание каждого параметра АСУТП по математической модели (4) и (5), при этом оценивание не зависит от помех, шума, вибраций, неисправностей измерительного, приёмо-передающего оборудования и прочего.

Испытания модели оценивания достоверности

Испытания МОД проводились на основе реальных данных системы управления. Была смоделирована функция, для имитации различного рода ошибок и наложена на реальный тренд. Сначала модель работала только на основе исторических данных, затем добавили анализ одной функциональной зависимости. Результаты экспериментов отражены в таблице. Проводя эксперимент на 500000 имитированных ошибочных точках (100%) без модели, используя только самые распространённые методы определения ошибок: методы сглаживания, контроля скорости изменения и контроля диапазона измерения параметра, - удалось определить лишь 17% ошибок, при количестве грубых ошибок 30%. При тех же условиях МОД, построенная только на основе исторических данных, выявила все грубые, явные ошибки. Если же к историческим данным добавить функциональные зависимости, то результат обнаружения ошибок достигает 97%.(см. Рис. 3)

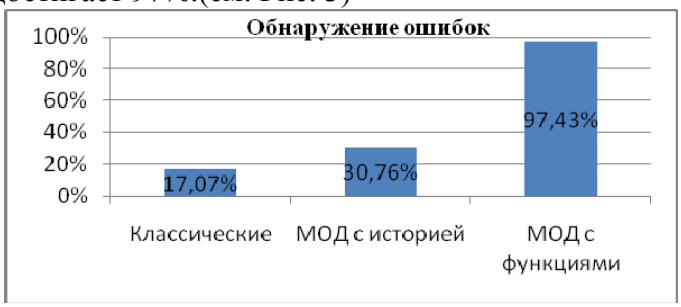


Рис. 3. Диаграмма обнаружения ошибок МОД

Выводы

Использование АСУТП совместно с МОД значений параметров на основе представляет ряд значительных преимуществ:

Управляющие решения принимаются на основе проверенных, достоверных данных. Как следствие, уменьшается число ложных срабатываний АСУТП, а также снижается время бесполезного простоя оборудования.

Сокращается время на поиск и устранение неисправности оборудования.

Улучшается надёжность системы управления в случае неопределенности: при недостоверности значения параметра АСУТП может воспользоваться его модельной оценкой или определить резервный и исправный, и осуществлять управление по его показаниям. Кроме того, реализация модели предоставляет следующие возможности: автоматическое прогнозирование поведения параметра в заданном режиме, проверка алгоритмов управления без объекта управления, обучение персонала на основе модельных значений.

Итак, предлагаемое решение обладает низкой вычислительной сложностью, основывается на ретроспективных данных и на опыте экспертов технологического процесса, что позволяет оценивать достоверность параметров АСУТП в реальном времени.

Направления дальнейших исследований: поиск быстродействующих и качественных алгоритмов параметризации модели, создание быстродействующих алгоритмов вычислений над нечёткими и интервальными числами, исследование применимости моделей с использованием прогнозирования совместно со сверхбыстрыми АСУТП (с малым периодом обработки данных), разработка методов и средств оценивания и оптимизации действий эксперта.

ЛОГИЧЕСКИЙ ВЫВОД В СИСТЕМЕ ВЕРИФИКАЦИИ ТРЕБОВАНИЙ К УПРАВЛЯЮЩИМ АЛГОРИТМАМ РЕАЛЬНОГО ВРЕМЕНИ

А.А. Тюгашев, А.В. Шулындин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Ключевую роль при управлении современными космическими аппаратами (КА) играют бортовые вычислительные системы (БВС), в состав которых входят одна или несколько бортовых цифровых вычислительных машин (БЦВМ). На них возлагаются задачи контроля работоспособности бортовой аппаратуры (БА), управления движением КА и навигации, выдачи управляющих воздействий на БА при решении КА целевых задач. Функции управления реализуются при этом бортовым программным обеспечением (БПО). Среди ошибок БПО значительное количество приходится на сбои синхронизации и согласования логики управления БА при одновременном функционировании ряда бортовых систем и программ БПО в рамках решения КА целевых задач (ошибки в управляющих алгоритмах реального времени - УА РВ).

Так как полное тестирование сформированной программы будет очень дорогостоящим и займет много времени, то для повышения уровня надежности УА РВ для БВС КА могут быть использованы аналитические методы верификации, которые на основе логически строгого доказательства способны определять наличие или отсутствие у управляющего алгоритма желаемых свойств.

Модель семантики УА РВ может быть построена как набор кортежей (четверок) Φ_i. Φ_i обычно подразумевает работу какого-то прибора или агрегата, входящего в состав БА, или выполнение функциональной программы из комплекса БПО.

Как известно, верификация - это процесс доказательства соответствия между программной реализацией задачи и спецификацией задачи. Цель верификации – демонстрация свойства корректности программы. Для программы управления КА важна верификация требований реального времени, потому что от правильно заданных требований зависит успешное выполнение функций, возложенных на управляющую программу.

Решение систем уравнений исчисления УА становится возможным путем перехода от формул теорий УА РВ к уравнениям и неравенствам на множестве целых чисел.

Так как математическая модель УАРВ предполагает создание связок между функциональными задачами, входящими в спецификацию, а так же формулировку правил вывода, позволяющих переходить от одних формул, специфицирующих требования, к другим, решение задачи верификации УА РВ может быть «погружено» в некоторую среду логического вывода. В частности – среду языка логического программирования Пролог, основанную на логике предикатов первого порядка.

Подсекция 5
«Сетевые технологии»

Перспективы внедрения протокола IPv6 и исследование параметров сетей нового поколения

С.С. Болдырев, П.В. Веселовский
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В последние годы проблема перехода на IPv6 является основной темой обсуждения в органах технического управления глобальной сетью. Например, на двух последних конференциях Европейского Центра Управления Интернет (RIPE NCC) более 70% докладов были непосредственно связаны с IPv6 [1]. И хотя в настоящий момент темпы перехода невелики, задача освоения и исследования IPv6 актуальна в меру неизбежного истощения пула свободных адресов IPv4 [2] и недостатков этого протокола [4].

Первый раздел тезисов акцентирует внимание на причинах внедрения нового протокола, связанных с ключевыми проблемами IPv4, решения которых предлагаются в протоколе IPv6; а также на взаимодействии двух протоколов.

Во втором разделе описывается предложенная нами измерительная инфраструктура для изучения качества параметров сети нового поколения.

Раздел 1. Причины и последствия внедрения протокола IPv6.

Благодаря удачному исходному дизайну, IPv4 выдержал испытание на масштабируемость, которому сопутствовал рост обслуживаемых им сетей. Однако, в протоколе IPv4 было изначально заложено ограничение глобального адресного пространства. Это сыграло свою роль по мере роста масштабов сетей в последние два десятилетия. На рис. 1 приведены данные, описывающие статистику распределения IPv4-адресов за последние 11 лет:

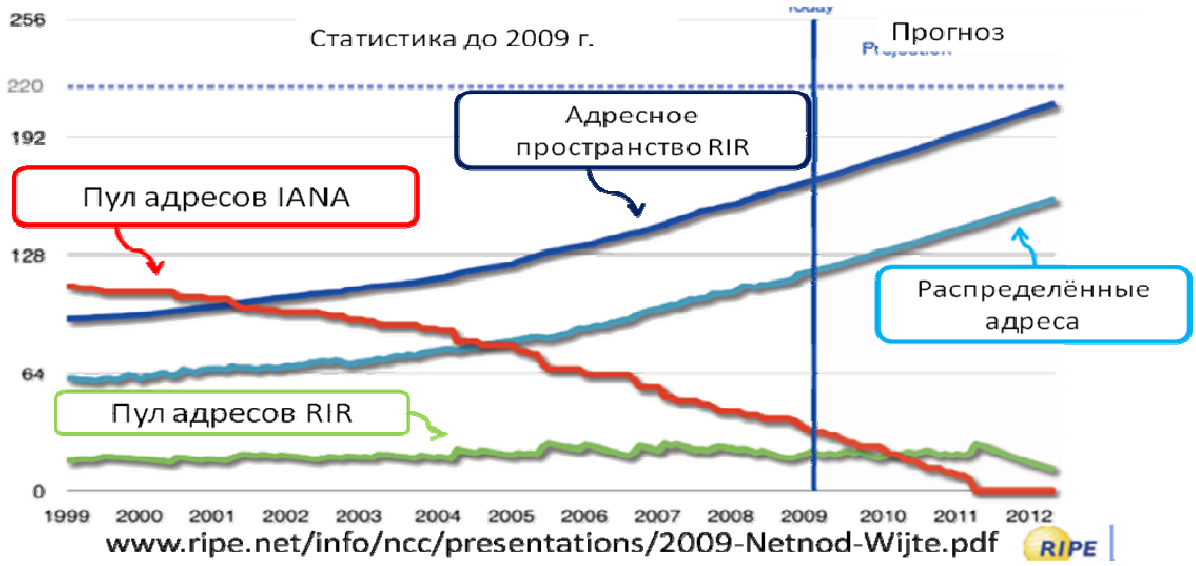


Рис. 1.

Появление в 1994 году технологии Network Address Translation (NAT) помогло значительно снизить темпы уменьшения числа доступных адресов. Однако, технология NAT не решила, а лишь отложила проблему истощения адресного пространства. IPv4 имеет также ряд прочих слабых сторон, устранение которых не входит в возможности NAT или иных технологий, но успешно производится в IPv6. Перечислим основные из них.

- 32 битный формат IPv4-адреса привёл к трудностям развития рынка интернет-услуг в его современных темпах из-за сокращения ресурса внешних адресов. 128 битный формат в IPv6 позволяет составить не менее $3{,}8 \cdot 10^{38}$ различных адресов. Становятся необязательными технологии преобразования адресов (например, NAT), и существенно упрощается маршрутизация.
- В IPv4 имеет место проблема роста таблиц маршрутизации на основных роутерах интернета. В IPv6 она решается формированием эффективной иерархической инфраструктуры маршрутизации в новом адресном пространстве.
- Невозможность гибкой настройки multicast-передачи пакетов в IPv4 значительно загружает сети и не способствует распространению мультимедиа-приложений. В IPv6 multicast-адреса введены на уровне стандарта, что существенно снижает нагрузку на магистральные линии и оптимизирует работу мультимедийных приложений.

При всей важности описанных проблем, основная причина глобальной смены протокола — истощение самого адресного пространства, что критически затрудняет работу региональных (RIR) и локальных (LIR) интернет-регистраторов.

В течение времени, которое займет процесс перехода на IPv6, будет играть особую роль вопрос его взаимодействия с протоколом IPv4. Оно осуществляется посредством нескольких методов:

- трансляторы протоколов,
- двухпротокольные стеки (dual stacks) и
- туннелирование (tunnels).

Из них особую роль играет последний, поскольку реализация связности по протоколу IPv6 будет осуществляться именно методом туннелирования, пока большая часть Интернета работает по протоколу IPv4 – на ринних этапах внедрения нового протокола [6]. На рис. 2 приведена схема туннеля "IPv6 поверх IPv4":

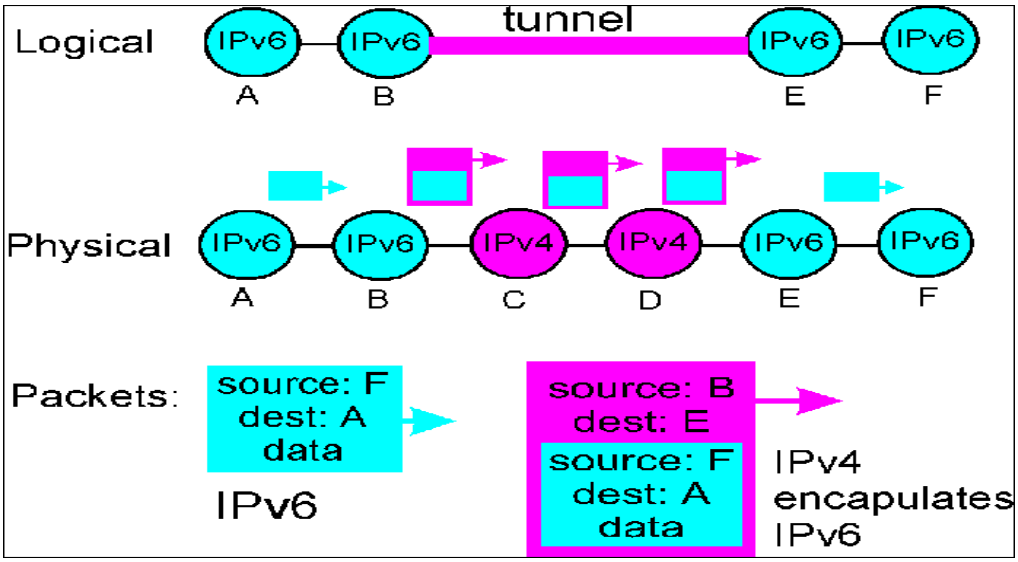


Рис. 2

Раздел 2. Базовые принципы системы измерения параметров сети в протоколе IPv6

Практическая часть данной работы заключается в создании системы, способной измерять параметры качества сетевых соединений по протоколу IPv6.

Основа этой системы — измерительный механизм RIPE Test Box. Каждая точка (Test Box) включает в себя BSD-сервер с подключенной GPS-антенной, служащей для высокоточной синхронизации (точность измерений для задержки пакетов составляет 2 мкс). Для измерения параметров сети производится обмен тестовыми пакетами с почти сотней аналогичных устройств по всему миру, работающих в рамках проекта RIPE TTM [3]. В России на сегодняшний момент есть лишь 4 таких точки:

- Самарский государственный аэрокосмический университет имени академика С.П. Королёва, Самара
- Институт органической химии им. Н.Д. Зелинского РАН, Москва
- Ростовская точка обмена интернет-трафиком (RND-IX), г. Ростов-на-Дону
- Институт теоретической физики им. Л.Д. Ландау РАН, г. Черноголовка

Из них в первых трёх система установлена и настроена в рамках гранта РФФИ 06-0789074а командой, в состав которой мы входим.

Важно подчеркнуть, что в Европе в рамках проекта RIPE TTM уже успешно проводятся измерения и в сетях протокола IPv6 [3], тогда как в России эта возможность еще не реализована. Предлагаемый нами проект не влечет больших дополнительных затрат при реализации [5].

Ключевым этапом для нас здесь является получение блока внешних (реальных) IPv6-адресов для нашей экспериментальной локальной сети и настройка маршрутизации, что и позволит исследовать параметры IPv6-сети. Затем последует ряд экспериментов и анализ данных, полученных с RIPE Test Box. Здесь мы выделяем два ключевых направления исследований:

- Определение типа распределения задержки пакетов в глобальной сети нового поколения. А также проверка применимости существующих для протокола IPv4 моделей распределения [7].
- Определение доступной пропускной способности сети на протоколе IPv6 по данным о значениях задержки пакетов разных размеров. Проверка пригодности существующих для IPv4 моделей в сети на протоколе IPv6 [8].

На данном этапе проводится анализ данных IPv6 трафика с измерительных точек RIPE (другой сети).

Предполагаемые сроки запуска опытной площадки на базе СГАУ им. С.П. Королёва и получения первых экспериментальных результатов — до конца 2010 года. Конечный этап работы — обобщение результатов о качестве связи (значении ключевых параметров) в IPv6-сети и взаимодействии протоколов IPv4 и IPv6.

Выводы

В работе представлен подготовительный проект в рамках смены протокола IP с версии 4 на версию 6. Прогнозируемые сроки истощения запаса IPv4-адресов и повсеместного перехода на IPv6 — ближайшие 2-3 года [2].

Поставлена задача об исследования параметров трафика в IPv6-сети. Кратко описаны проводимые для этого разработки и планы их развития. Предлагается внедрение в России передовой сетевой технологии, уже успешно функционирующей в Европе. Ближайшая задача — получить свой блок внешних IPv6 адресов и договориться о данных для анализа с другой измерительной точки, пока не будет полностью готова площадка в СГАУ.

Анализ статистики позволит судить об эффективности использования протокола IPv6, а также о его взаимодействии с IPv4. Приобретенные опыт и знания облегчат и ускорят освоение Интернета нового поколения.

Литература

1. http://www.ripe.net/ripe/meetings/archive/index.html
2. http://www.ipv6actnow.org/info/what-is-ipv4/
3. Georgatos F., Gruber F., Karrenberg D., Santcroos M., Susanj F., Uijterwaal H., Wilhelm R., Providing Active Measurements as a Regular Service for ISP's, PAM2001, http://www.ripe.net/projects/ttm/Documents/PAM2001.pdf
4. http://www.ripn.net/articles/IPv6_transition/
5. Платонов А.П., Сидельников Д.И., Стрижов М.В., Сухов А.М., Измерительная инфраструктура для изучения качества соединений в российском сегменте Интернет, Телекоммуникации, 2009, № 1, с.11-16
6. Мендкович А.С., Сидельников Д.И. IPV6 — новый этап развития Интернета, Информационное общество, №4, с. 11-22 (2000).
7. A.M. Sukhov, N Kuznetsova, What type of distribution for packet delay in a global network should be used in the control theory? 2009; arXiv: 0907.4468.
8. A.M. Sukhov, T.G. Sultanov, M.V. Strizhov, A.P. Platonov, Throughput metrics and packet delay in TCP/IP networks, RIPE59 Meeting, Lisbon, 2009; arXiv:0907.3710.

Методы построения оптимальной структуры HRBF нейронной сети для решения задачи прогнозирования

В.Г. Литвинов, О.П. Солдатова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Прогнозирование – это ключевой момент при принятии решений. Конечная эффективность любого решения зависит от последовательности событий, возникающих уже после принятия решения. Возможность предсказать неуправляемые аспекты этих событий перед принятием решения позволяет сделать наилучший выбор, который, в противном случае, мог бы быть не таким удачным.

Для построения прогнозов существует большое количество классических методов, но они не являются эффективными (высока погрешность прогноза), если пытаться исследовать зависимость прогнозируемой величины от независимых переменных, а при использовании нейронных сетей данная задача легко решаема в виду их хорошей обобщающей способности.

В настоящее время, в нейросетевых технологиях, существует нерешенная проблема автоматического подбора оптимального, с точки зрения решения задачи, количества базисных функций. При этом каждой функции соответствует один скрытый нейрон. Слишком малое количество нейронов не позволяет уменьшить в достаточной степени погрешность обучения, тогда как слишком большое их число увеличивает погрешность решения задачи.

Вследствие невозможности априорного определения точного количества скрытых нейронов применяются адаптивные методы, которые позволяют добавлять и удалять их в процессе обучения.

Сам процесс обучения нейронной сети состоит из двух основных этапов:

1. Подбор начальных параметров сети;
2. Оптимизация параметров сети в зависимости от входных данных.

В рамках проведенных исследований, был использован алгоритм самоорганизации для подбора начальных параметров сети. Этот процесс организации обучающих данных автоматически разделяет пространство на так называемые области Вороного, определяющие различающиеся группы данных [1]. В виду того что на основе радиально-базисные нейронные сети сами по себе производят локальную аппроксимацию, данный метод показывает отличные результаты на практике.

Пример такого разделения двухмерного пространства представлен на рис. 1 [1]. Данные, сгруппированные внутри кластера, представляются центральной точкой, определяющей среднее значение всех его элементов. Центр кластера в дальнейшем отождествляется с центром соответствующей радиально-базисной функции.

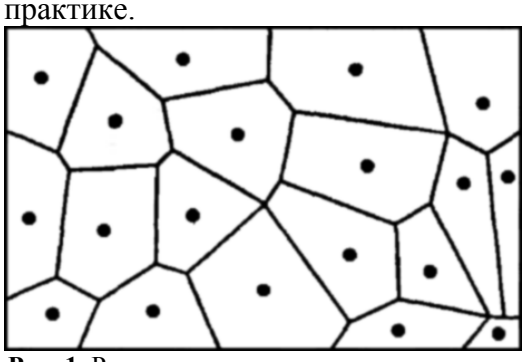


Рис. 1. Разделение двумерного пространства

Разделение данных на кластеры можно выполнить с использованием алгоритма *k*-усреднений. Уточнение центров производится после предъявления каждого очередного вектора $\mathbf{x}_p$ из множества обучающих данных.

Такой алгоритм уже предлагался во многих работах по нейросетевым технологиям, но он лишь определял расположение нейронов скрытого слоя в пространстве данных, и не определял их количество. Авторами была предложена модификация алгоритма, который на этапе подбора начальных параметров нейронов подбирает оптимальную структуру скрытого слоя нейронной сети (количество базисных функций и их расположение в пространстве обучающих входных данных). Модификация заключается в том, что при предъявлении нового входного вектора $\mathbf{x}_p$, с использованием евклидовой метрики определяется расстояние от него до всех существующих кластеров данных. Если минимальное расстояние не удовлетворяет параметру точности кластеризации обучения заданному предварительно, то создается новый кластер с центром в точке $\mathbf{x}_p$.

Второй алгоритм, предложенный авторами, позволяет подобрать оптимальную структуру нейронной сети уже на втором этапе обучения. В нем, для определения стратегии формирования структуры сети был применен подход каскадного наращивания сети, при котором центр радиально-базисной функции нового нейрона размещается в месторасположение выборки данных $\mathbf{x}_p$, на которой возникает максимальная погрешность. Подбор структуры сети при таком подходе происходит параллельно с оптимизацией ее параметров путем добавления на каждой итерации обучения одного скрытого нейрона [2].

Таким образом, определение структуры сети и реализацию алгоритма ее обучения можно трактовать как выполнение подбора оптимальной архитектуры искусственной нейронной сети.

В процессе обучения каждый нейрон старается так адаптировать веса своих связей, чтобы быть востребованным для выполняемого сетью отображения данных. На начальном этапе формируется сеть, состоящая только из входных узлов и выходного нейрона и двух скрытых нейронов. Количество входов зависит исключительно от специфики решаемой задачи и не подлежит модификации. Скрытые нейроны добавляются в сеть по одному (рисунок 2). Каждый добавляемый нейрон подключается ко всем входам сети и его центр устанавливается в месторасположение выборки данных $\mathbf{x}_p$, на которой возникает максимальная погрешность, а диагональные элементы матрицы радиусов уточняются. Далее производится уточнение коэффициентов сети гибридным алгоритмом [3].

Если результат функционирования сети считается удовлетворительным с точки зрения ожидаемой или допустимой погрешности, либо достигнуто ограничение по количеству нейронов, процесс обучения и формирования структуры сети завершается. В противном случае структура сети повторно расширяется и производится уточнение коэффициентов.

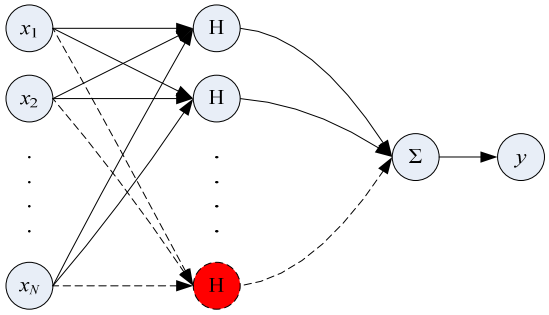


Рис. 2. Добавление нейрона

Кроме подбора оптимальной структуры нейронной сети, одним из основных факторов успеха прогнозирования нейронной сетью является правильный выбор методики подачи входных данных (подбор размерности входного вектора данных обучения с задержками). Чем она больше, тем большими степенями свободы обладает нейронная сеть, так как количество компонент входного вектора равно количеству входов. При этом обобщающие способности нейронной сети возрастают. Но следует

учесть, что при увеличении размерности входного вектора, растет и время, затрачиваемое на обучение и функционирование сети.

Авторами предложен метод подачи входных данных с задержками. При этом нейронная сеть на основе радиально-базисных функций по своему функционалу становится похожа на рекуррентную сеть. Заключается он в том, что на вход подаются векторы (можно подавать только часть компонент вектора) $\mathbf{x}_p, \mathbf{x}_{p-1}, \mathbf{x}_{p-2} \ldots \mathbf{x}_{p-n+1}$, где ***n*** – размерность входного вектора, ***p*** – номер выборки из обучающих данных.

Все предложенные выше алгоритмы метода были реализованы в программной модели HRBF сети. Проведены экспериментальные исследования на данных котировок акций ОАО «Газпром», подтвердившие работоспособность данного метода (результаты исследования алгоритма наращивания сети на этапе оптимизации приведены на рис. 3 - 4).

Если результат функционирования сети считается удовлетворительным с точки зрения ожидаемой или допустимой погрешности, либо достигнуто ограничение по количеству нейронов, процесс обучения и формирования структуры сети завершается.

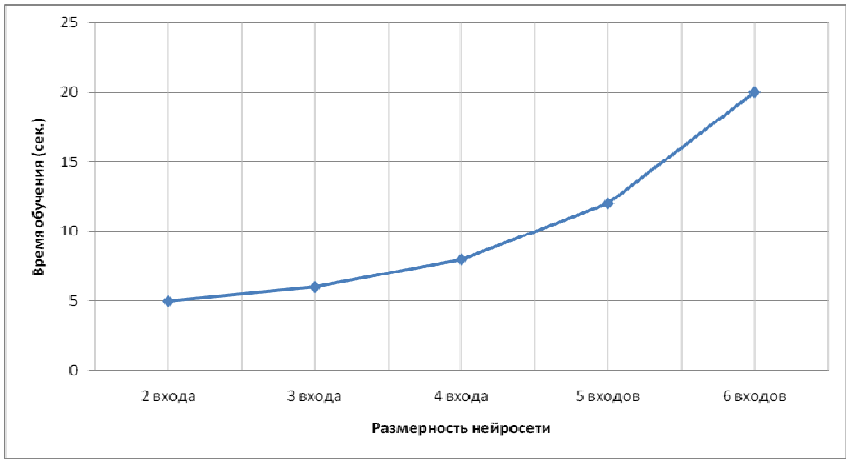


Рис. 3. Зависимость времени обучения от размерности нейронной сети

В противном случае структура сети повторно расширяется и производится уточнение коэффициентов.

На основании проведенных исследований можно сделать вывод, что использование нейронных сетей на основе радиально-базисных функций является целесообразным в задачах прогнозирования (сравнивая с классическими методами).

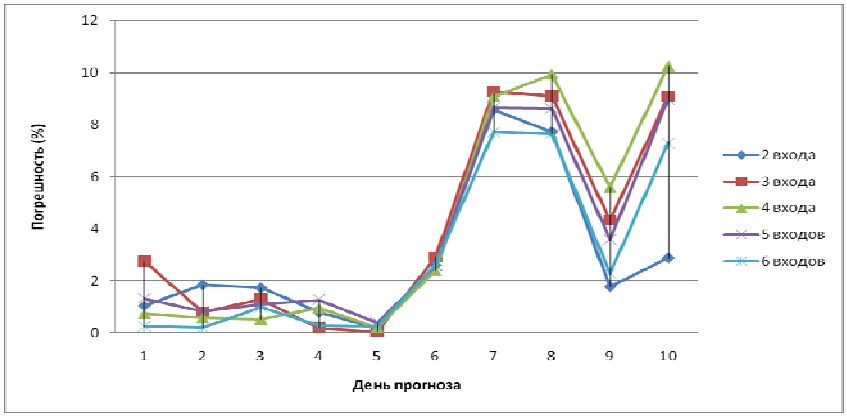


Рис. 4. Относительная погрешность прогнозирования при различных размерностях нейронной сети

Эксперименты показали, что разработанный метод позволяет производить краткосрочные и среднесрочные прогнозы с хорошей достоверностью. Полученные прогнозы с точностью до 2% были достоверны в течение 5-ти дней.

Литература

1. Осовский С. Нейронные сети для обработки информации [Текст] / Осовский С. – М.: Финансы и статистика, 2002 г. – 344 с.

2. Каюкова А.С., Солдатова О.П. Аппроксимативный анализ с использованием нейронной сети HRBF с произвольными функциями активации Электронный журнал «Исследовано в России», 37, стр. 343-349, 2006 г., http://zhurnal.ape.relarn.ru/articles/2006/037.pdf.

3. Хайкин, С. Нейронные сети: полный курс [Текст] / С. Хайкин, – М.: ООО “И. Д. Вильямс”. 2006. – 1104 с.

Информационная система структуризации знаний из Интернета

А.С. Лубышев, А.Х. Галеев
Самарский государственный архитектурно-строительный университет, Самара

На сегодняшний день Интернет является одним из основных источников информации. С каждым годом в Глобальной сети появляется все большее количество информации. Причем каждые три года объем данных Интернета увеличивается вдвое. Число пользователей сети так же растет. С каждым днем ритм жизни ускоряется, и у пользователей сети просто нет времени на поиск нужной информации среди всего объема Интернета, большую часть данных которого составляет информационный шум.

Таким образом, на сегодняшний день одной из наиболее актуальных проблем становится проблема структуризации знаний из Интернета. В рамках данной работы рассматривается один из вариантов решения этой проблемы.

Была предложена технология извлечения и дезинтеграции фактографической табличной информации из Интернета на элементарные единицы знаний – «атомы знаний», а также последующего формирования кластеров знаний из таких «атомов», описывающих определенную предметную область.

Под фактографической информацией понимается наиболее ценная информация. Когда такой информации много ее удобнее всего представлять в структурированном виде. Такая информация находится в огромном количестве в таблицах. В результате исследования было выявлено, что около 40% текстовой информации является таблицами. Следовательно, есть смысл извлекать фактографическую информацию из таблиц.

В данной работе были решены основные проблемы дезинтеграции: сложная структура таблиц, включающая в себя вложенность таблиц друг в друга и проблему нахождения различий между фактографическими таблицами и табличной версткой – оформлением ресурсов при помощи таблиц. Были разработаны способы приведения сложно структурированных таблиц к более простому, удобному для дезинтеграции виду. Явных способов «распознавания» фактографических таблиц не существует, поэтому был разработан собственный способ «распознавания», работающий на основе многокритериального анализа характеристик таблиц. Так же были предложены способы качественного взаимодействия с существующими поисковыми машинами Интернета.

На основе данной технологии была создана программа, которая ищет таблицы в сети Интернет с помощью поисковых машин, проверяет таблицы на предмет фактографической информации в них, затем дезинтегрирует таблицы, получая «атомы знаний» и прикрепленные к ним метаданные, при помощи специально разработанного алгоритма – парсинга.

Результаты дезинтеграции записываются в базу данных в четко структурированном виде знаний. Такая таблица является основой для последующего формирования кластеров знаний. При этом, благодаря хранению метаданных, появляется возможность для формирования кластеров знаний, описывающих ту или иную область. Область поиска может задаваться пользователем. Таким образом, вся система работает по аналогии с информационно-поисковыми системами, вместе с тем преодолевая основные недостатки ИПС, в частности получение актуальной и полезной информации, релевантной запросу пользователя.

В рамках данной работы рассматривался упрощенный вариант кластера знаний. Под ним понимается двумерная таблица, в строках и столбцах которой располагаются некие элементарные «лексемы», а в соответствующих им ячейках некая значащая информация, в рамках данной работы под такой информацией понималась числовая информация.

Были рассмотрены два варианта кластеров знаний:

1) Полностью заполненный кластер – такая подтаблица исходной матрицы знаний, на пересечении любых строк/столбцов которой находятся только единицы, т.е. есть информация, характеризующая любые пары терминов, входящих в кластер.

2) Частично заполненный кластер – такая подтаблица исходной матрицы знаний, на пересечении хотя бы одной строки/столбца которой находится ноль, т.е. отсутствует информация, характеризующая зависимость соответствующих пар терминов.

Опираясь на данные варианты кластеров, были предложены следующие виды кластеризации:

1) Выделение из заданной таблицы полностью заполненных кластеров знаний максимального размера

2) Выделение из заданной таблицы частично заполненного кластера знаний заданного размера.

Были предложены математические модели для осуществления предложенных видов кластеризации. В данных моделях используются методы целочисленного линейного программирования.

Реализован программный комплекс, использующий для вычислений предложенные математические модели. Данный комплекс позволяет формулировать предметную область с использованием тезауруса терминов и получать отчеты, содержащие найденные кластеры для выбранной области.

В рамках исследования работоспособности всего комплекса были обработаны несколько исходных таблиц из Интернета в области «IT в образовании» и получены реальные плотные кластеры знаний.

Данная работа может являться основой для создания фактографической информационно-поисковой системы Интернета.

Нейроимитатор нейронных сетей на основе нечеткого многослойного персептрона

В.В. Новиков, Е.А. Расщепкин, О.П. Солдатова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Разработан нейроимитатор, позволяющий создавать и изменять структуру сети, отображать структуру созданной сети, обучать сеть и решать задачи прогнозирования и классификации. Нейроимитатор позволяет также проводить исследования качества обучения сети. В нейроимитаторе представлены следующие типы сетей – многослойный персептрон, нечёткий многослойный персептрон и рекуррентная сеть Эльмана.

Среда разработки: Visual Studio 2008, язык программирования: C#, операционная система: MS Windows XP Professional и выше.

Введение

Нейронные сети могут решать широкий круг задач обработки и анализа данных − распознавание и классификация образов, прогнозирование, управление и т.д. Конкурентами являются классические методы анализа данных, статистики, идентификации систем и управления.

На практике чаще всего используются программы-нейроимитаторы - они работают на обычных компьютерах, а современные нейроалгоритмы позволяют быстро обрабатывать значительные объемы информации, т.е. использовать нейронные сети для решения сложных реальных задач. Программы включают в себя как набор базовых операций по созданию, обучению и манипулированию нейронными сетями, исходными данными, параметрами нейросетей и нейрорешений, так и автоматизированные процедуры выполнения наиболее рутинных цепочек действий, например, для определения оптимальных настроек сети и алгоритма обучения.

После создания нейросетевых моделей реализуются либо автоматическая генерация описания обученной нейросети на некотором языке программирования для того, чтобы полученный модуль можно было вставить в программу пользователя, либо поддерживаемые нейроимитаторами стандартные средства межпрограммной коммуникации (DDE, OLE, COM в среде Windows) для возможности вызова этих нейропрограмм из программ пользователя.

На специфические вычислительные архитектуры (например, сигнальные процессоры) обычно переносят сгенерированное программой-нейроимитатором описание обученной нейронной сети на языке программирования. Для ПЛИС/PLM/FPGA-кристаллов существуют и CAD-системы, в которых номенклатуру нейроэлементов можно описать в виде набора модулей и потом из этих модулей путем выбора и "перетаскивания" мышкой собрать нужную топологию-архитектуру нейросети, затем проект скомпилировать и "прошить" кристалл. В настоящее время можно создавать нейрокомпьютеры и на базе видеокарт (графических процессоров)[1].

Так что существует широкий спектр достаточно универсальных способов организации инструментальных средств и собственно процесса применения нейронных сетей на различной программно-аппаратной базе. Всегда можно подобрать наиболее оптимальный для некоторой задачи способ − всё определяется свойствами задачи и требованиями к решению, а также компетентностью пользователя.

1. Теоретическая часть

Многослойный персептрон. Персептрон представляет собой сеть, состоящую из нескольких последовательно соединенных слоев формальных нейронов (как прави-

ло, сигмоидальных нейронов). На низшем уровне иерархии находится входной слой, состоящий из сенсорных элементов, задачей которого является только прием и распространение по сети входной информации. Далее имеются один или, реже, несколько скрытых слоев. Каждый нейрон из скрытого слоя имеет несколько входов, соединенных с выходами нейронов предыдущего слоя или непосредственно с входными сенсорами *x1..xn*, и один выход. Нейрон характеризуется уникальным вектором весовых коэффициентов *w*. Веса всех нейронов слоя формируют матрицу, которую мы будем обозначать *W*. Функция нейрона состоит в вычислении взвешенной суммы его входов с дальнейшим нелинейным преобразованием ее в выходной сигнал:

$$y = \frac{1}{\left(1 + \exp\left(-\left[\sum_i W_i x_i - \Theta\right]\right)\right)}. \quad (1)$$

Выходы нейронов последнего, выходного слоя выдают результат решения задачи *Y=Y(X)*. Особенности работы персептрона состоят в следующем. Каждый нейрон суммирует поступающие к нему сигналы от нейронов предыдущего уровня иерархии с весами, определяемыми состояниями входов, и формирует ответный сигнал (переходит в возбужденное состояние), если полученная сумма выше порогового значения. Персептрон переводит входной образ, определяющий степени возбуждения нейронов самого нижнего уровня иерархии, в выходной образ, определяемый нейронами самого верхнего уровня. Число последних, обычно, сравнительно невелико. Состояние возбуждения нейрона на верхнем уровне говорит о принадлежности входного образа к той или иной категории [2, 3, 4].

Нечёткий персептрон. Его структура отличается от структуры обычного персептрона наличием на входе многослойного персептрона нечёткого слоя, состоящего из радиально-базисных нейронов с гауссовой функцией активации. Это позволяет работать с неполными данными или с зашумленными данными.

Сеть Эльмана. Она характеризуется частичной рекуррентностью в форме обратной связи между скрытым и входным слоем, реализуемой с помощью единичных элементов запаздывания. Каждый нейрон скрытого слоя имеет свой «собственный» элемент запаздывания. Выходной слой состоит из нейронов, однонаправлено связанных с нейронами скрытого слоя, подобно многослойному персептрону.

2. Описание программы

Особенностью программы является специально разработанное ядро, которое позволяет создавать новую сеть путём описания только её параметров и типов соединения нейронов в сети. Это возможно благодаря разработанной в нейроимитаторе архитектуре построения сети (см. рисунок 1).

В нейроимитаторе реализованы следующие возможности:

- выбор типа сети:
- многослойный персептрон (MLP);
- нечёткий многослойный персептрон (FMLP);
- рекуррентная сеть Эльмана;
- выбор алгоритма обучения:
- обратного распространения ошибки и наискорейшего спуска [2, 3,4];
- c-means [5].
- выбор функции активации нейрона:
- пороговая;
- линейная;
- униполярная сигмоида;
- биполярная сигмоида;

– функция Гаусса;
– синусоидальная;
– генерация сети, выбранной из ранее приведенного перечня, с заданной размерностью и параметрами;
– редактирование созданной сети;
– просмотр структуры сети вплоть до значения весовых коэффициентов отдельно взятого нейрона;
– визуальное отображение построенной сети и графическое конструирование сети;
– редактирование выборок данных при помощи встроенного редактора;
– отображение результатов хода обучения и тестирования сети в виде графиков;
– проверка качества обучения сети.

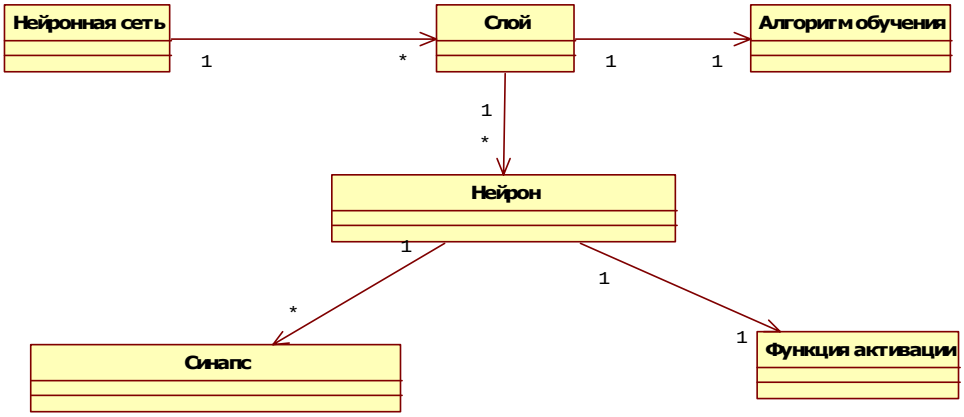


Рис. 1. Упрощённая диаграмма классов сети

Пример созданной сети типа «Нечёткий многослойный персептрон» приведен на рис. 2. На нём также отражены основные интерфейсные решения программы.

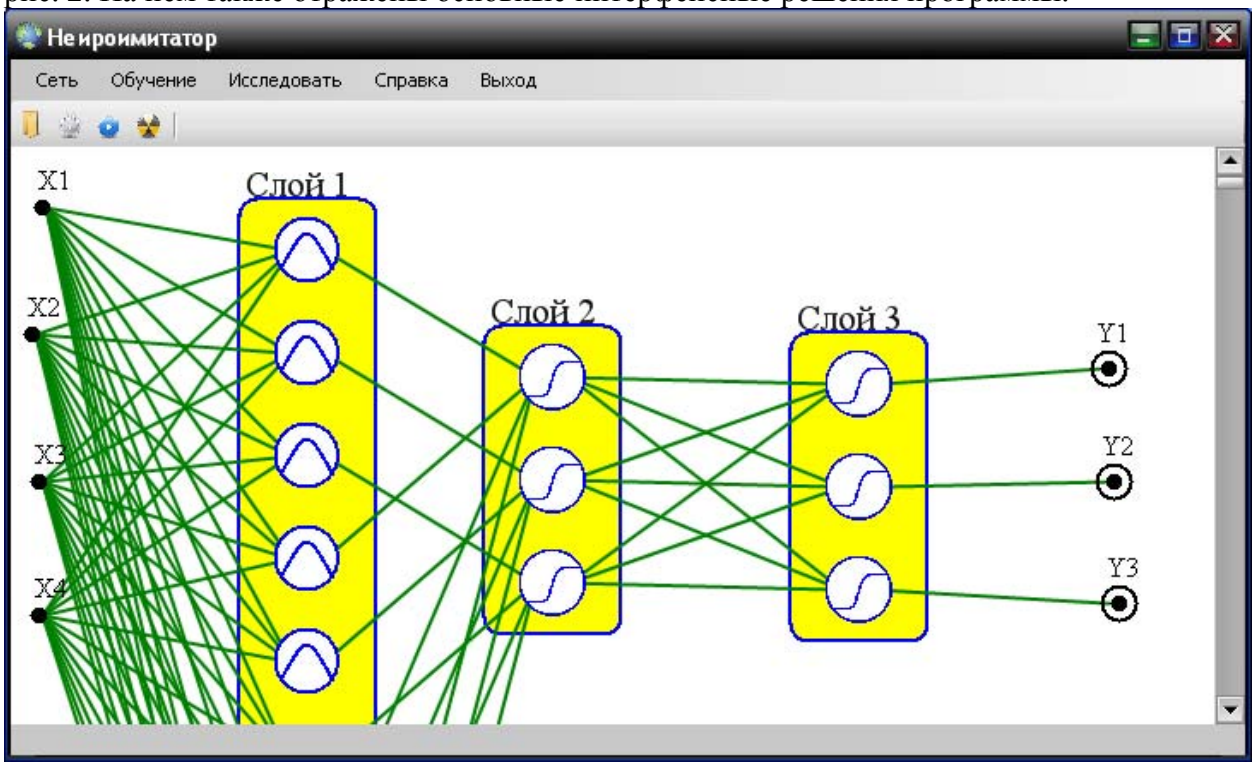


Рис. 2. Структура сети «Нечёткий многослойный персептрон»

3. Описание экспериментов

Для примера используем «Нейроимитатор» в качестве нейросетевого классификатора. Для примера возьмем задачу классификации ирисов Фишера (это набор данных для задачи классификации, на примере которого Рональд Фишер в 1936 году продемонстрировал работу разработанного им метода дискриминантного анализа). Имеются данные измерений для 150 экземпляров ирисов, в равных частях (по 50 штук) принадлежащих к трем видам (iris setosa, iris versicolor, iris virginica). Для каждого экземпляра ириса известны 4 величины:

– длина чашелистика (Sepal Length);
– ширина чашелистика (Sepal Width);
– длина лепестка (Petal Length);
– ширина лепестка (Petal Width).

Входной файл состоит из 150 строк (по 50 для каждого сорта).

Для эксперимента использовался многослойный персептрон со следующей архитектурой:

– входной слой с размерностью 4;
– один скрытый слой с 4 нейронами (функция активации - униполярная сигмоида);
– выходной слой с 3 нейронами (функция активации - униполярная сигмоида);
– алгоритм обучения – наискорейшего спуска и метод обратного распространения ошибки с шагом обучения равным 0,1.

Нечеткий многослойный персептрон с архитектурой:

– нечеткий входной слой с 12 нейронами (функция активации - гауссова);
– один скрытый слой с 3 нейронами (функция активации - униполярная сигмоида);
– выходной слой с 3 нейронами (функция активации - униполярная сигмоида);
– алгоритм обучения нечеткого слоя - c-means, алгоритм обучения четкого слоя – наискорейшего спуска и метод обратного распространения ошибки с шагом обучения равным 0,1.

Сеть Эльмана:

– число входов – 3 (функция активации - униполярная сигмоида);
– число нейронов в скрытом слое – 10 (функция активации - униполярная сигмоида).
– алгоритм обучения – наискорейшего спуска и метод обратного распространения ошибки с шагом обучения равным 0,1.

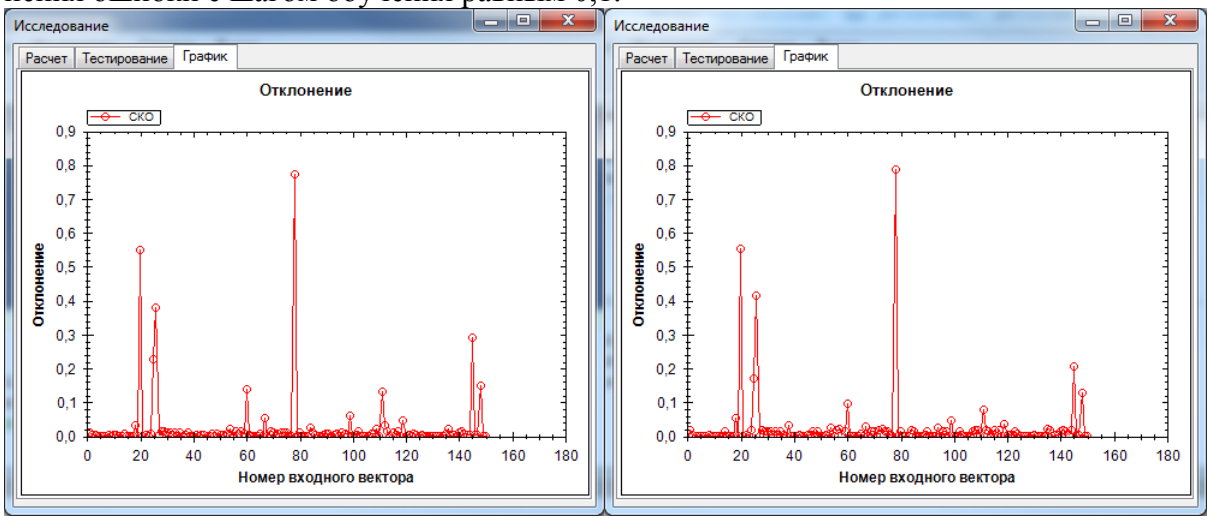


Рис. 3. Результаты экспериментов (слева – MLP, справа – FMLP)

Результаты экспериментов представлены на рисунках 3 и 4. На графиках показано СКО выходных векторов от истинного результата (по оси X номер вектора). В итоге многослойный персептрон неправильно классифицировал 2 ириса, сеть Эльмана неправильно классифицировала 2 ириса, нечёткий персептрон неправильно классифицировал 2 ириса.

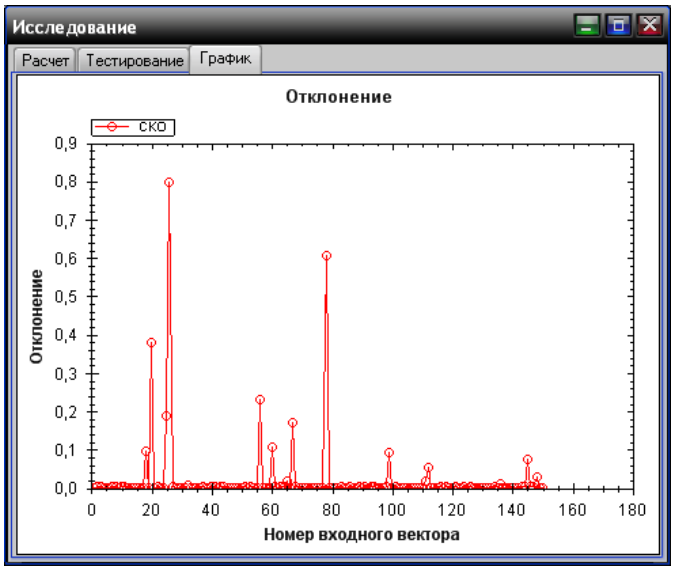


Рис. 4. Результаты экспериментов (сеть Эльмана)

Выводы

Была разработано программное средство для моделирования и работы с нейронными сетями, основанными на многослойном персептроне (многослойный персептрон, нечёткий многослойный персептрон, рекуррентная сеть Эльмана). Для каждой сети были реализованы алгоритмы обучения, а также проведён анализ работы сетей на примере решения задачи классификации. Анализ полученных результатов показал, что все сети показывают схожие результаты классификации. Тем не менее, можно видеть, что минимальное СКО на данных выборках показал нечеткий многослойный персептрон.

Литература

1. Царегородцев В. Г. // Искусственные нейронные сети [Электронный ресурс]. – Режим доступа: http://neuropro.ru/neu4.shtml.
2. Осовский С. Нейронные сети для обработки информации / Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344с.:ил.
3. Методы нейроинформатики / Под ред. А.Н. Горбаня, КГТУ, Красноярск, 1998. 205с.
4. Хайкин, Саймон Нейронные сети: полный курс, 2-е издание.: Пер. с англ. - М.: Издательский дом "Вильямс", 2006. - 1104с.: ил. - Парал. тит. англ.
5. Palit, Ajoy K. Computational intelligence in time series forecasting: theory and engineering application – Bremen, Germany: Springer, 2005. – 381p.:il.

Математическое моделирование динамического последовательного заполнения сетей с кластерной топологией

И.И. Курочкин, Я.Р. Гринберг
ИСА РАН, Москва

Введение

Пусть есть некоторая телекоммуникационная сеть, которая соединяет множество абонентов посредством каналов связи, где каждый канал соединяет пару абонентов. В данном случае телекоммуникационная сеть представляется графом, где абоненты (полюса) – узлы, а каналы связи - дуги. Попытку проложить канал между парой абонентов (полюсов) будем называть заявкой. А последовательный процесс удовлетворения заявок будем называть последовательным процессом заполнения сети или заполнением сети. Основанная задача при заполнении сети – удовлетворение заявок на прокладку каналов так, чтобы было удовлетворено как можно большее число заявок. Поскольку время жизни заявки не оговорено и считается бесконечно долгим, то будем считать, что рассматривается статический вариант последовательного заполнения сети. В связи с рассмотрением различных топологий телекоммуникационных сетей было замечено, что результаты заполнения различных множеств сетей имеют различия и зависят, в том числе от топологии телекоммуникационной сети.

Рассмотрим топологию «связные кластеры». Кластером назовем некоторый граф. Кластеры связаны между собой дугами. Дуги, связывающие кластеры, соединяют узлы, принадлежащие этим кластерам. В кластере может быть любое количество узлов и дуг, связывающих данный кластер с другими. Узлов сети вне кластеров не существует. Исходная сеть может быть разделена на кластеры по некоторому признаку: по силе связности, по территориальному (географическому) расположению узлов, по пропускной способности дуг между узлами, экспертное разделение (классификация). Топология связных кластеров довольно часто встречается при конструировании телекоммуникационных сетей. Так как не определена топология для каждого кластера и топология соединения кластеров, то можно сказать, что рассматриваемая топология является общей топологией или метатопологией для телекоммуникационных сетей.

Использование топологии связных кластеров позволяет применять декомпозицию для вычисления кратчайших путей. Так как по теореме об оптимальном пути считается, что отрезок оптимального пути также является оптимальным путем. Из этого следует, что кратчайший (оптимальный) путь можно проложить от кластера-источника до кластера-стока, а потом уже прокладывать путь внутри транзитных кластеров и конечных кластеров. Использование этого метода позволяет исключить из рассмотрения кластеры, которые не участвуют в прокладке пути. Применяемый метод декомпозиции позволяет сократить вычислительную сложность задач нахождения минимальных разрезов и прокладки кратчайших путей в сети.

Динамическое заполнение сети

Динамическое заполнение телекоммуникационной сети предполагает, что заявки через определенное время освобождают ресурсы сети, то есть имеют некоторое время жизни, по истечении которого пропускные способности дуг, принадлежащие проведенному потоку заявки, увеличиваются на величину проведенного потока.

В зависимости от распределения времени жизни заявок заполнение сети может протекать в нескольких режимах:

ненасыщенный режим – одновременное количество заявок в сети небольшое, отказы никогда не происходят, величина максимального потока не достигается, величина меры неравномерности мала вне зависимости от числа проведенных заявок;

стационарный режим – одновременное количество заявок в сети большое, иногда происходят отказы, величина максимального потока иногда достигается, но массовых отказов нет; величина меры неравномерности в зависимости от проведенных заявок – величина постоянная;

насыщенный режим – величина максимального потока достигается быстро, величина меры неравномерности достигает максимальных значений, отказы в обслуживании массовые;

режим полного заполнения – равноценен статическому заполнению сети заявками, когда время жизни достаточно велико по сравнению с временем моделирования, полный отказ происходит в 100% случаев.

Более всего интересен стационарный режим, потому что режим полного заполнения и насыщенный режим равноценны статическому заполнению, а ненасыщенный режим не дает информации о поведении сети при достижении значений максимального потока и использовании большей части ресурсов.

Алгоритмы последовательного заполнения

В эксперименте принимали участие следующие алгоритмы:

- Простой алгоритм;
- Равномерный по дугам алгоритм;
- Субоптимальный по дугам алгоритм;
- Субоптимальный минимально-разрезный алгоритм;
- Аддитивный минимально-разрезный алгоритм;
- Гибридный минимально-разрезный алгоритм.

Расчет путей по всем этим алгоритмам производится как определение пути между двумя полюсами в сети, имеющего минимальную стоимость из всех возможных путей. Алгоритмы различаются разными способами определения этих дополнительных весов для дуг сети.

Параметры математического моделирования

Рассматриваются сети с топологией связные кластеры, с несколькими парами полюсов (парами источник-сток). Заполнение производится единичными потоками.

Модель сети определяется с помощью следующих параметров:

- количество кластеров;
- количество вершин в каждом кластере;
- максимальное количество дуг между двумя кластерами;
- количество (часть от общего числа) нулевых дуг;
- максимальная пропускная способность дуг в сети;
- граф связности, в котором указываются пропускные способности дуг сети;
- вектор пар полюсов (пар источник-сток);
- избыточный вектор заявок на проведение единичного потока (Заявка определяется как номер пары полюсов, между которыми следует провести единичный поток);
- идентификационный уникальный номер сети.

Математическое моделирование процесса заполнения сети происходит до наступления следующих критериев:

- критерий первого отказа (проведение очередной заявки невозможно, одна из пар источник-сток несвязна; минимальный из множества минимальных разрезов между парами полюсов равен 0);
- критерий полного отказа (стоки не достижимы из источников, для всех пар полюсов; все минимальные разрезы между парами полюсов равны 0);
- выход на стационарный режим, когда критерий полного отказа не достижим, но уже обработано достаточно большое число заявок.

Общие результаты численного эксперимента и некоторые выводы

Используя данные по отказам можно сказать, что субоптимальный минимально-разрезный алгоритм не так хорош, как при одной оценке общего проведенного потока, так как число отказов в несколько раз больше, чем при использовании гибридного и дугового субоптимального алгоритма.

В таблицах 1 и 2 приведены обобщенные данные по отказам раздельно по каждому алгоритму для всех сетей.

Таблица 1. Данные по отказам

Количество отказов						**Процент отказов на отрезке первый отказ – полный отказ**					
П	**Д**	**ДС**	**РС**	**РА**	**РГ**	**П**	**Д**	**ДС**	**РС**	**РА**	**РГ**
572860	453462	592350	571378	462269	747543	57	53	57	54	57	63

Таблица 2. Процент превышения общего количества отказов

Процент превышения общего количества отказов					
П	**Д**	**ДС**	**РС**	**РА**	**РГ**
0	-20,8	3,4	-0,3	-19,3	30,5

Все отказы располагаются после выполнения критерия (1) (до первого отказа) и до выполнения критерия (2) (до полного отказа). Лучшими алгоритмами по количеству отказов являются субоптимальные алгоритмы дугового и минимально-разрезных подходов.

По проценту превышения проведенного потока лучшими алгоритмами являются гибридный алгоритм и субоптимальные алгоритмы минимально-разрезного и дугового подхода.

Несмотря на то обстоятельство, что общий проведенный поток не может показать большинство преимуществ алгоритмов при динамическом последовательном заполнении, приведение суммарных результатов по всем сетям позволяет выявить алгоритмы, которые одинаково хорошо работают в статическом и в динамическом режиме.

Таблица 3. Процент превышения общего проведенного потока

Проведенный поток до 1-ого отказа (1)						**Проведенный поток до разрыва всех пар полюсов (2)**					
П	**Д**	**ДС**	**РС**	**РА**	**РГ**	**П**	**Д**	**ДС**	**РС**	**РА**	**РГ**
0	2,1	2,9	3,0	2,3	3,0	0	-8,5	2,1	2,0	-6,1	1,8

Так как при генерации сетей была поставлена задача создания сетей, которые бы при заполнении были бы на границе между «стационарным » и насыщенным режимом, то наличие временного стационарного режима может говорить о некотором стабильном режиме работы сети под нагрузкой.

Выводы

Наличие и протяженность стационарного режима является важной характеристикой сетей при динамическом заполнения; существенно здесь также и то, что свойство стационарности проявляется в сетях, которые в начальный период времени были «пустыми»; это свойство, качественно выявленное на основе экспериментальных данных, и его зависимость от параметров задачи также требуют пристального изучения;

Для некоторых сетей выбор того или иного алгоритма заполнения оказывается очень существенным и значения проведенного потока могут различаться на 50%-70%.

В среднем по всем сетям и по нескольким критериям алгоритмы можно ранжировать (от лучшего к худшему) следующим образом:

- гибридный алгоритм,
- субоптимальный дуговой алгоритм,
- субоптимальный минимально-разрезный алгоритм,
- аддитивный минимально-разрезный алгоритм,
- простой алгоритм,
- дуговой алгоритм.

Литература

1. Афанасьев А.П., Гринберг Я.Р., Курочкин И.И.; Равномерное заполнение телекоммуникационной сети каналами связи.(с.118-123) Прикладные проблемы управления макросистемами (Апатиты, 5-9 апреля 2004 года). Сборник докладов. Том 8. Труды ИСА РАН. – М.: Едиториал УРСС, 2004, 288с.
2. Афанасьев А.П., Гринберг Я.Р., Курочкин И.И.; «Равномерные» алгоритмы последовательного заполнения потоковой сети потоками продуктов.(с.118-140) Проблемы вычислений в распределенной среде. Модели обработки и представления данных. Динамические системы. Том 14. Труды ИСА РАН. – М.: КомКнига, 2005, 224с.
3. Гринберг Я.Р., Курочкин И.И.; Анализ результатов численного эксперимента по последовательному заполнению сетей со стохастической топологией. (с.99-128) Проблемы вычислений в распределенной среде: распределенные приложения, коммуникационные системы, математические модели и оптимизация. Том 25. Труды ИСА РАН. – М.: КомКнига, 2006, 224с.
4. Гринберг Я.Р., Курочкин И.И. «Методы оценки математического моделирования по динамическому последовательному заполнению телекоммуникационных сетей потоками связи.» Материалы XVI Международной конференции по механике и современным прикладным программным системам (ВМСППС'2009), 25-31 мая 2009, г.Алушта, Украина. – М.: МАИ-ПРИНТ, 2009г. стр.231-233.
5. Герасимов А.И.. Аналитические методы исследования и оптимизации вычислительных систем и сетей на основе сетевых моделей массового обслуживания. – М.: Радио и связь, 2001.
6. Крылов В.В., Самохвалова С.С.. Теория телетрафика и ее приложения. – Санкт-Петербург: БХВ-Петербург, 2005.
7. Остерлох Х.. Маршрутизация в IP-сетях. Принципы, протоколы, настройка. – СПб.: ДиаСофтЮП, 2002. – 512 с.
8. Шмалько А.В.. Цифровые сети связи: основы планирования и построения. – М.: Эко-Трендз, 2001.
9. Ху Т. Целочисленное программирование и потоки в сетях. - М.: Мир, 1973.
10. Форд Л.Р., Фалкерсон Д.Р. Потоки в сетях. М.: Мир, 1962.
11. Вишневский В.М. - Теоретические основы проектирования компьютерных сетей. – М.: Техносфера, 2003.

Повышение безопасности передачи данных в сетях с временной синхронизацией через NTP протокол

М.А. Макаров, Е.А. Симановский
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Во время передачи пакетов существует возможность их перехвата злоумышленником, а так же изменения и ретрансляции в качестве подлинного. Для повышения безопасности передаваемых по сети данных предлагается использовать механизм контроля сетевых задержек в сетях с временной синхронизацией по протоколу сетевого времени NTP (Network Timing Protocol). Функциональная схема, иллюстрирующая работу системы безопасности на основе NTP серверов представлена на рисунке ниже (рис. 1).

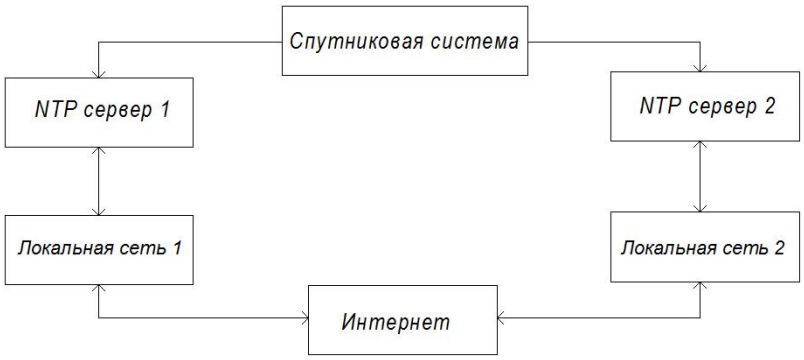


Рис. 1. Функциональная схема работы системы безопасности на основе NTP серверов

Архитектура системы представляет собой несколько сетей, внутри которых осуществляется временная синхронизация с точностью порядка 1-2 мс. Такая точность позволяет высчитывать отклонение времени задержки передачи пакета от допустимого диапазона задержки. Величина этого временного коридора является определяющим критерием для формирования решения о подлинности передаваемого сообщения.

При передаче информационных пакетов через сети TCP/IP, клиент, находящийся на приемной стороне, получает их с задержкой D(W), являющейся функцией от величины пакета информации. Задержка представляет собой совокупность постоянной и переменной составляющей, D_{fixed} и d_{var}. Нашей задачей является определение такого d_{var}, при котором принятый пакет можно было бы считать подлинным.

$$D(W) = D_{fixed}(W) + d_{var}. \quad (1)$$

На рисунке ниже представлены экспериментальные результаты по передаче пакетов по сети Интернет. График отображают зависимость функции распределения (CDF - cumulative distribution functions) от временной задержки при передаче пакета через сеть Интернет.

CDF показывает какое количество пакетов (в процентном отношении), принятых на другом конце канала, имеет задержку ниже или равную задержке, соответствующей интересующему нас значению CDF. Например, значению CDF равному 0,8 соответствует задержка 44,35 мс. Это говорит о том, что задержкой 44,35 мс и меньше обладают 80% полученных пакетов.

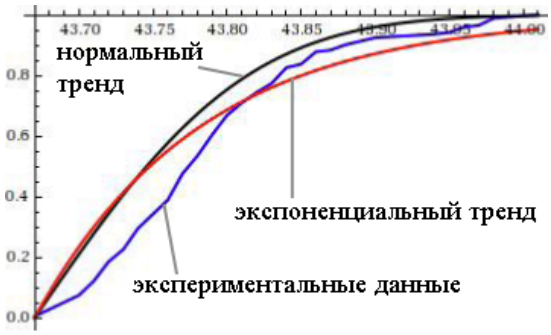


Рис. 2. Зависимость CDF от задержки пакетов при использовании 100 байтных пакетов

Экспоненциальное распределение значительно превосходит нормальное по корреляции с экспериментом. Поэтому в дальнейшем предполагался экспоненциальный характер распределения задержки.

На основе полученных данных была выведена формула для переменной задержки:

$$d_{var} = -(D_{av} - D_{min})\, ln\{1 - F(D,W)\}. \qquad (2)$$

где D_{av} – усредненное значение задержки;

D_{min} – минимальное значение задержки;

F(D,W) – функция совокупного распределения (изменяется от 0 до 1).

Совокупная функция распределения является зависимой от двух переменных: от задержки и от размера передаваемых пакетов. Последнюю переменную мы исключаем, используя для системы трафик с фиксированной длиной посылок. Используя эти зависимости можно найти однозначное соответствие каждого значения задержки, своему значению CDF и наоборот. Для Интернет трафика удовлетворительным значением потери (слишком продолжительной задержки) являются потери 0,5% пакетов. Для увеличения надежности соединения зададимся критерием потери 1 пакета из 1000. Это говорит о том, что значение совокупной функции распределения должно равняться 0,999. Разница D_{av} и D_{min} установлена экспериментально и составляет порядка 1 – 5 мс. Подставив имеющиеся данные, получим максимально допустимую задержку (абсолютную погрешность):

$$d_{var} = 5 \cdot 3 ln 10 = 34{,}5\, мс.$$

Принимая пакет от источника, сервер аутентификации проверяет метку времени, содержащуюся в его заголовке и сверяет ее со своим локальным временем, которое синхронизировано с помощью NTP сервера. Принятие решения об обработке пакета или же об отвержении получается в следствии алгоритма:

Считывается временная метка (значение времени T_1)

Проверяется на правильность условие:

$$T_2 < T_1 + \frac{RTT}{2} \le T_2 + d_{\text{var}}, \qquad (3)$$

где RTT (Round time trip) – время передачи пакета туда и обратно.

Принимается решение: если (3) выполняется – пакет принимается, если условие ложно – то пакет уничтожается.

Литература

Choi, B.Y. Analysis of Point-To-Point Packet Delay In an Operational Network [Текст]/B.Y. Choi, S. Moon, Z.L. Zhang, K. Papagiannaki, C. Diot. - Hong Kong: Infocom, 2004. – C. 1797-1807.

Модель сверточной нейронной сети для извлечения паттернов признаков на основе рецептивных полей

Е. Пудикова, О.П. Солдатова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Центральным объектом обсуждений в данной статье является метод распознавания рукописных цифр и букв с помощью нейронных сетей. Рассмотрены уже существующие методы распознавания в сверточных сетях. Предложена модель архитектуры нейронной сети, предполагающая подход извлечения конкретных геометрических признаков, соответствие которым и будет определять принадлежность паттерна к тому или иному классу.

Сверточные сети для распознавания рукописных цифр, букв и других паттернов изображений

Сверточные нейронные сети - это класс сетей, построенных на модели многослойного персептрона и использующие соответствующие методы обучения, например, алгоритм обратного распространения ошибки. Принципиальное отличие сверточной сети от многослойного персептрона в том, что каждый сверточный слой имеет общий паттерн разделяемых весовых коэффициентов и состоит из нескольких карт, каждая из которых хранит представление входного изображения. Что характерно, число нейронов в каждой карте с увеличением порядкового номера слоя уменьшается, при этом одновременно количество карт в слое возрастает.

Особенностью архитектурного проектирования сетей класса CNN (Convolutional Neural Network) является то, что каждая сеть проектируется под конкретную задачу путем подбора числа карт в каждом слое. Критерием работоспособности сети и успеха подобранной архитектуры является: динамика изменения качества и скорости распознавания. Такой подход разработки сети влечет за собой то, что аналитически обосновать причину выбора определенной конфигурации сети сегодня однозначно определить нельзя, что осложняет проектирование сложных архитектур данного класса сетей для решения других задач.

Однако все же определить, количество нейронов в карте можно обосновать. Ведь размер каждой карты зависит от размера рецептивного поля сверточного нейрона в конкретном слое сети. При этом каждое рецептивное поле конкретного нейрона должно пересекаться с соседним рецептивным полем на 75%, это позволяет добиться хорошего обобщения при извлечении признаков [5]. Учитывая это, можно определить количество нейронов, удовлетворяющих данному условию. Требование того, чтобы рецептивные поля всех нейронов слоя умещались полностью в предыдущей карте или слое, определяет количество сверточных нейронов.

Сверточные сети – это гибридные сети, они состоят из слоев, извлекающих признаки, и слоев классифицирующих данные, полученных слоями признаков. Качество классификации можно корректировать выбором классификатора или изменением качества извлекаемых признаков.

Центральной задачей обучения в данных сетях является обучение слоев признаков извлечению конкретных паттернов закономерностей из группы паттернов множества обучения.

Наиболее известными сверточными сетями, успешно решающими задачу распознавания рукописных цифр, являются: LeNet4, Boosted LeNet4, LeNet5, созданные профессором Яном ЛеКуином [1]. Также известны примеры сверточных сетей применяемых для распознавания изображений объемных объектов: сверточные сети с машиной опорных векторов, выступающей в качестве классификатора [2].

Модель сверточной сети

Рассмотрим самую простую задачу распознавания рукописных символов: распознавание рукописных цифр.

Была выбрана следующая архитектура, представленная в [6]. Сеть состоит из четырех слоев: два слоя свертки и два слоя нейронов, имеющих входы со всех нейронов предыдущего слоя и выполняющих классификацию.

Подробно архитектура описана в [6].

Важно понять, что отображает каждая карта слоев признаков в подобной сети. Так исходное изображение интерпретируется по-разному за счет различных весовых коэффициентов рецептивных полей нейронов последующих 6 карт. Это позволяет создать несколько альтернативных представлений внутри одной сети, что увеличивает обобщающую способность сети. Процесс же обучения приводит к тому, что слои свертки способны выделить некоторые признаки, общие для входного множества обучения.

К сожалению, данный метод получения признаков, характеризующих изображение, не всегда позволяет извлечь достаточные данные для классификации. Это можно увидеть на ярком примере различия структур сетей решающих задачу распознавания рукописных цифр и сетей, классифицирующих объекты, и неэффективности вторых для классификации входных данных.

Если же рассматривать принцип обработки сигнала в биологических системах, можно заметить, что изображение претерпевает прохождение через группы нейронов, которые вносят новые информационные составляющие о нем - признаки. Определение извлекаемого из изображения признака, которого было бы достаточно для классификации. При этом вид признака изначально может быть не определен, и сеть должна обучиться извлекать его. Т.о., предлагается, процессу обучения обеспечить сходимость к конкретной конфигурации соотношения коэффициентов, для распознавания признака.

Модель сети извлекающей признаки

На пути решения поставленного вопроса, можно обратиться к существующим изученным системам обработки визуальных сигналов, а именно моделям зрительных систем млекопитающих (кошек, обезьян, человека) описываемых в [3].

Так в статье «Computational predictions of receptive fields and organization of V2 for shape processing»[3] была разработана модель, описывающая самоорганизующиеся ритинотопические карты, находящиеся в зоне V1 и V2 зрительной коры (рис. 1). В [3] было показано, что рассматриваемую систему нейронных слоев можно привести к состоянию, достаточному для извлечения первичных паттернов изображения. На рис. 1 представлена модель такой структуры: (early visual pathway – первичный путь обработки визуального сигнала, Afferent – афферентные связи, Lateral Excitatory – латеральное усиление сигнала, Lateral Inhibitory – латеральное торможение.)

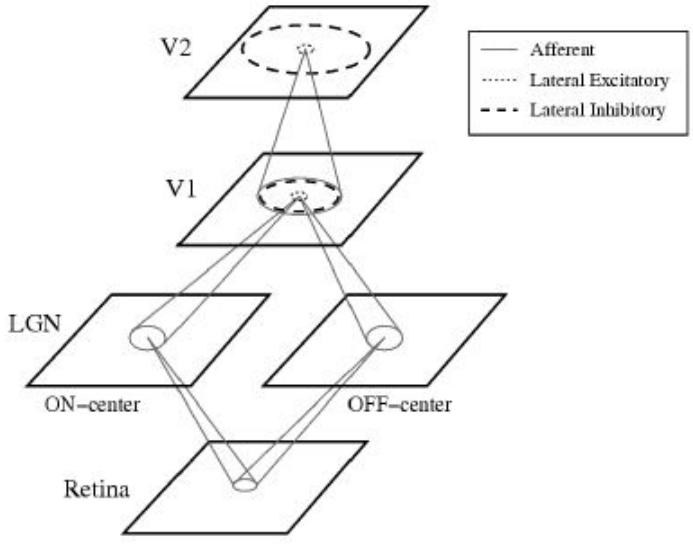


Рис. 1. Модель первичного зрительного пути

На данной схеме представлены несколько слоев нейронной структуры эмитирующих соответственно: ретину, латеральное коленчатое тело, слой V1 и V2. Стоит отметить, что в классических работах Хьюбеля и Визеля [4] было показано, что в зоне V1 клетки извлекают элементарные паттерны изображения, такие как: информацию о направлении, контрасте, и размере объектов.

Исследователям удалось создать систему близкую по функциональным свойствам к V1 обезьяны [3]. На рис. 2 представлены карты сымитированного слоя V1 (а) и реальной зоны зрительной коры обезьяны (б). Наблюдается хорошее соответствие карт модели и органической структуры.

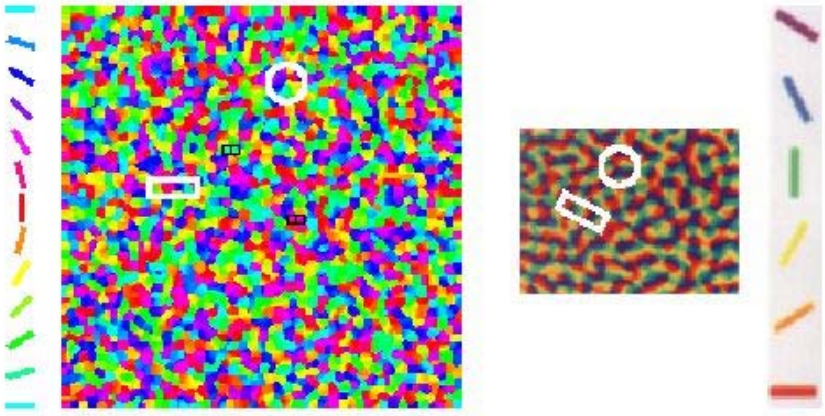


Рис. 2. Карты ориентации в модели V1(a) и реальной зоны зрительной коры макаки(b)

При моделировании V2 удалось показать, что наблюдается увеличение кластеров, отвечающих за анализ смежных паттернов. Так же было показано, что генерируемые рецептивные поля в ходе процесса самоорганизации, становиться чувствительны к примитивам соответствующим различным углам и прямым различных направлений, это представлено на рис. 3.

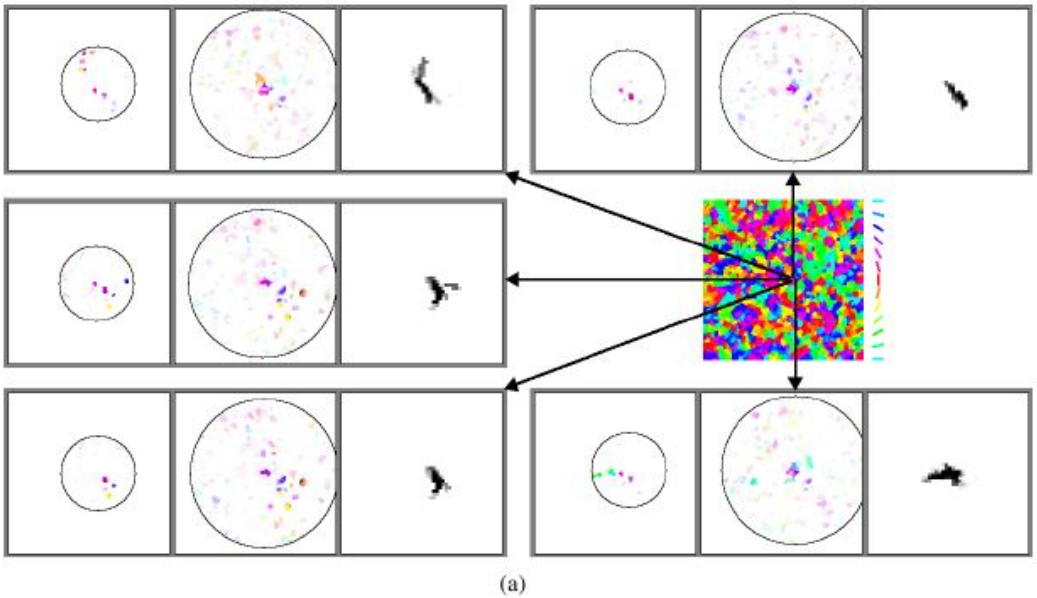


Рис. 3. Связи и предпочитаемые паттерны в зоне V2 в области одного домена ориентации.

Преимущество модели состоит в том, что можно сформировать так организованные нейронные структуры, которые будут способны извлекать строго определенные геометрические паттерны, что сделает в целом работу системы предсказуемой. Необходимо математически описать представление сигнала после его обработки слоем.

Можно представить математическую модель, позволяющую построить нейронную сеть с заложенными правилами извлечения паттернов признаков на основе рецептивных полей.

Представим модель первичного зрительного пути, изображенного на рисунке 1.

Пусть фоторецептор ретинного слоя - это пиксель входного изображения. И пусть входные изображения представлены в градации серого цвета. Все связи от ретинного слоя до слоя Латерального Коленчатого Тела (ЛКТ) фиксированы и выбраны так, чтобы представить его рецептивные поля клеток, используя при этом стандартную модель difference-of-Gaussians (разница гауссианов)[7]. При этом каждый нейрон ЛКТ связан с ретинными нейронами по

принципу, установленному в сверточных сетях. ЛКТ представлено двумя картами отображения рецептивных полей клеток с ON и OFF центрами. Можно определить веса связей клеток ЛКТ и ретинного слоя следующим образом. Обозначим вес фоторецептора (x,y) рецептивного поля клети ЛКТ (a,b) с ON-центром как $L_{xy,ab}$, с центральным нейроном с координатами X_c, Y_c , $L_{xy,ab}$ вычисляется так:

$$L_{xy,ab} = \frac{\exp\left(-\frac{(x-x_c)^2+(y-y_c)^2}{\sigma_c^2}\right)}{\sum_{uv}\exp\left(-\frac{(u-x_c)^2+(v-y_c)^2}{\sigma_c^2}\right)} - \frac{\exp\left(-\frac{(x-x_c)^2+(y-y_c)^2}{\sigma_s^2}\right)}{\sum_{uv}\exp\left(-\frac{(u-x_c)^2+(v-y_c)^2}{\sigma_s^2}\right)},$$

где σ_c определяет ширину ON-центра рецептиваного поля и σ_s определяет окружающее рецептивное поле OFF-области. Веса рецептивных полей с OFF-центром отрицательны по отношению к весам с ON-центром.

Отклики нейронов с ON и OFF каналов в ЛКТ рассчитываются через взвешенную сумму весов их рецептивных полей. Т.е. отклик ζ_{ab} рассчитывается следующим образом:

$$\xi_{ab} = \sigma\left(\sum_{xy}\chi_{xy}L_{xy,ab}\right),$$

где χ_{xy} выходной сигнал фоторецептора (x,y), принадлежащего нейрону (a,b). $L_{xy,ab}$ - представляет латеральные (связи данного нейрона с нейронами этого же слоя, важны для самооргонизации) связи нейрона (a,b). А σ – это функция активации данного нейрона.

Следующий слой за ЛКТ это зона V1. Результирующий выходной сигнал каждого нейрона в котором вычисляется с учетом латеральных и афферентных (связей данного нейрона, с нейронами предыдущего слоя, входящими в его рецептивное поле) связей. Сначала определим афферентные связи $s_{1i,j}$ нейронов слоя V1 и нейронов ЛКТ, входящих в их рецептивные поля.

$$s_{1ij} = \gamma_A\left(\sum_{ab\in ON}\xi_{ab}A_{1ab,ij} + \sum_{ab\in OFF}\xi_{ab}A_{1ab,ij}\right),$$

Здесь ζ_{ab} выходной сигнал нейрона (a,b) слоя ЛКТ, входящего в рецептивное поле (i,j) нейрона зоны V1 соответственно из ON и OFF каналов. $A_{1ab,ij}$ соответствующие веса афферентных связей и γ_A – постоянный фактор масштабирования. Начальный отклик (i,j) нейрона определяется через:

$$\eta_{1ij}(0) = \sigma(s_{1ij}).$$

После активации отклики нейронов определяются близкодействующими связями возбуждения и дальнодействующими связями торможения, что можно представить в виде:

$$\eta_{1ij}(t) = \sigma\left(s_{1ij} + \gamma_E\sum_{kl}\eta_{1kl}(t-1)E_{1kl,ij} - \gamma_I\sum_{kl}\eta_{1kl}(t-1)I_{1kl,ij}\right),$$

где $\eta_{1kl\,(t-1)}$ отклик другого нейрона V1 на предыдущем шаге. $E_{1kl,ij}$ это матрица весов, определяющих возбуждающие латеральные связи, а $I_{1kl,ij}$ определяет матрицу весов тормозящих латеральных связей. Все значения весов положительные. γ_E и γ_I это константы масштабирования, определяющие отношение между степенью влияния латеральных связей возбуждения и торможения соответственно.

Аналогично подсчитываются отклики нейронов в следующем слое – зоне V2:

$$\eta_{2mn}(t) = \sigma\left(\gamma_A\sum_{kl}\eta_{1kl}(t')A_{2kl,mn} + \gamma_E\sum_{pq}\eta_{2pq}(t-1)E_{2pq,mn} - \gamma_I\sum_{pq}\eta_{2pq}(t-1)I_{2pq,mn}\right),$$

Соответствующие константы характеристик и величины определены аналогичным образом, что и в V1, за тем исключением, что индексы матриц были заменены на соответствующие для V2.

Обучение же нейронов, производится по правилу Хебба, с постсинаптической нормализацией:

$$w'_{xy,ij} = \frac{w_{xy,ij} + \alpha X_{xy} \eta_{kij}}{\sum_{uv}(w_{uv,ij} + \alpha X_{uv} \eta_{kij})},$$

где $w_{xy,ij}$ – текущее значение афферентной или латеральной матрицы весов. α - это соответствующая определенному типу связей коэффициент обучения. X_{xy} и η_{kij} досинаптический и постсинаптический сигналы на конкретном нейроне.

Использование подобной модели позволило получить путем самоорганизации приведенное на рис. 3 ретинотопическое представление откликов нейронов на рецептивных полях конкретной конфигурации, что показано в [3]. Автором была построена сеть аналогичной архитектуры и обучающаяся по указанным правилам. Использование подобной структуры в рамках гибридной сети может позволить классифицировать объекты, лишь по наличию у них строго определенных, общих геометрических признаков и их взаимной конфигурации, безотносительно к природе этих объектов. На основе этого предположения в дальнейшем планируется, доработать архитектуру уже созданной сети, добавив к ней классификатор, использующий данные со слоев признаков.

Литература

1. Y. LeCun, L. Jeckel, L. Bottou. Comparison of learning algorithms for handwritten digit reorganization. // *Bell Laboratories. – 1996.*

2. F.U. Huang, Y. LeCun. Large-scale learning with SVM and Convolutional nets for Generic object categorization..// - *2006.*

3. Yiu Fai Sit, Risto Miikkulainen. Computational predictions of receptive fields and organization of V2 for shape processing.*// Universety of Texas at Austin. – 2009.*

4. John G. Nicholls, A. R. Martin, B.G. Wallace, P.A. Fuchs. From neuron to brain (fourth edition).// *Sinauer Associates, Inc.*Publishers. Sunderland, Massachusetts, USA. – 2001.*

5. Y. LeCun, B. Boser, J.S. Denker. Handwritten digit reorganization with a Back-Propagation network.// *Bell Laboratories. – 1992.*

6. Гаршин А.А. Применение сверточной нейронной сети для распознавания рукописных цифр /. А.А.Гаршин, О.П.Солдатова // *Компьютерная оптика. – 2010. – Том 34,№2. – с. 252-260.*

7. Sohail F. Tavazoie and R. Clay Reid. Diverse receptive fields in the lateral geniculate nucleus during thalamocortical development. *// Department of Neurobiology, Harvard Medical School, Massachusetts.-2000.*

Пути повышения качества передачи видео в беспроводных сетях

Е.С. Сагатов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Мобильные решения в области телекоммуникаций все более настойчиво входят в повседневную жизнь, но существует серьезное препятствие для повседневного использования видео сервисов в беспроводных сетях: соответствующее качество связи оставляет желать лучшего. Большой процент потерь пакетов и вариация задержки (сетевой джиттер) делают беспроводные сети малопригодными для передачи данных в режиме реального времени, так как мультимедийный трафик чрезвычайно чувствителен к подобным искажениям. Пакеты потокового видео теряются при передаче по сети, изменяют порядок следования из-за значительной вариации задержек пакетов. На получаемом видеоизображении появляются множественные артефакты, происходит рассинхронизация потока, что приводит к искажениям изображения, а иногда и к полной остановке воспроизведения видео.

Методика паспортизации пакетных сетей в рекомендации RFC-2544 [3] определяет следующие основные параметры качества сети: пропускная способность, задержка при передаче пакета, пакетный джиттер, количество потерянных пакетов, количество пакетов с ошибками.

Идея субъективного тестирования (MOS) состоит в том, что видео, закодированное одним из кодеков MPEG-2, MPEG-4 или Windows Media Video 9, передается через сеть Интернет и беспроводную сеть стандартов WiFi, 3G или WiMAX. Полученное после передачи видео, демонстрируется группе экспертов, которые выставляют оценки, основываясь на своих впечатлениях от качества. Существует много методов демонстрации последовательностей и сбора оценок, некоторые из них описаны в рекомендациях ITU [5]. К сожалению, в основном они рассчитаны на сравнение видео в телевизионном формате, и не очень удобны для проведения тестирования на PC.

В настоящей работе рассматривается проблема адаптации современных алгоритмов кодирования и передачи видео для беспроводных сетей, таких как 3G, WiFi и WiMAX, а также для других сетей с плохими характеристиками качества. В работах [6, 7] было показано, что субъективная оценка качества видео, равно как и сетевых параметров, имеет градацию "хорошо" (Good), "приемлемо" (Acceptable) и "плохо" (Poor) (GAP). Такое описание помогает понять качественные зависимости, возникающие при сетевой трансляции видео и сделать первичные численные приближения.

В настоящей работе сделана попытка найти численную зависимость качества видеоизображения от сетевых параметров. Отличительной особенностью нашего подхода является то, что указанная зависимость описывается простой математической моделью, что позволяет нам сравнить численные значения коэффициентов. На основании подобного сравнения возможно найти не только наиболее существенные факторы, влияющие на качество видео, но также и сопоставить между собой различные кодеки.

В работе учитывается различие между искажениями, которые вносят повреждения ключевых и обычных кадров. Ключевым называется кадр, который несет в себе полную информацию о видеоизображении и может быть восстановлен без привлечения дополнительных данных, а обычным – кадр, который кодирует разницу между предыдущим кадром и текущим. Степень сжатия ключевого кадра меньше, чем обычного, а размер в несколько раз больше. В настоящей работе будет дано численное сравнение влияния ошибок, произошедших на ключевых кадрах, для кодеков MPEG-2 и MPEG-4.

Предпосылки для моделирования

При передаче видео по сети качество связи ухудшается в зависимости от характеристик сетевого соединения. Для того чтобы описать качество передаваемого видео в зависимости от сетевых параметров была рассмотрена универсальная функция $Q_{mos}(p, j, D, B)$, описывающая качество видео по шкале MOS. Эта функция может быть разложена в степенной ряд по сетевым

параметрам, при этом можно ограничиться линейными членами и только одним вкладом второго порядка, описывающим передачу видео с различными скоростями потока.

На конференции TERENA 2005 [4] было показано, что для фиксированной скорости видео потока достаточно рассмотреть только члены разложения, описывающие линейную зависимость от потерь пакетов и сетевого джиттера:

$$Q_{mos}=Q_{ideal} - \alpha p - \beta j \quad (1)$$

где

- Q_{ideal} - максимальное качество видео для данного кодека, баллы от одного до пяти;
- p - процент потерь пакетов, %;
- j - сетевой джиттер (вариация задержки) в момент ошибки, сек.;
- Q_{MOS} - качество видео на приемной стороне, баллы от одного до пяти;
- α, β - коэффициенты модели, которые могут быть найдены экспериментально.

Для проведения исследований была разработана единая видео последовательность, которая обрабатывалась кодеками MPEG-4 (DivX), MPEG-2 и Windows Media video 9 с постоянным битрейтом 256 кбит/с.

Основная задача, поставленная в настоящей работе, это выявление влияния ключевых кадров на качество получаемого видео изображения. Поэтому отдельно вычисляются коэффициенты α^k и β^k для потока с повреждением ключевых кадров, а также α^w и β^w для последовательностей без повреждения ключевых кадров. Например, $\alpha^{\mathrm{k}}_{DivX}$ будет обозначать усредненный коэффициент, который характеризует ухудшение качества видео, закодированного кодеком MPEG-4 (DivX), если повреждение пришлось на ключевой кадр. Далее в работе будут найдены и сопоставлены $\alpha^{\mathrm{k}}_{DivX}$ и $\alpha^{\mathrm{w}}_{DivX}$, а также $\alpha^{\mathrm{k}}_{Mpeg2}$ и $\alpha^{\mathrm{w}}_{Mpeg2}$, чтобы определить в численном виде меру ухудшения качества видеоизображения.

Планирование эксперимента

Для нахождения коэффициентов уравнения (1) нами был разработан и проведен ряд экспериментов. Закодированные кодеками MPEG-4 (DivX), MPEG-2 и Windows Media Video 9 видео файлы, пересылались на ноутбук, подключенный к беспроводной сети WiFi, WiMAX или 3G. На ноутбуке проводилась запись получаемого видео в файл, параллельно записывался сетевой трафик на уровне пакетов при помощи сетевого сниффера Wireshark [9]. Таким образом, по полученному видеоизображению можно установить качество видео по шкале MOS, а по сетевым логам – параметры сетевого соединения.

Все записанные в ходе экспериментов видеофрагменты и сетевые логи опубликованы на сайте компании НПЦ «Интернет ТВ» [2].

Для сетей Wi-Fi, WiMAX и 3G эксперименты будут выглядеть одинаково, меняется лишь сетевое оборудование. На рис. 1 представлена схема проведения эксперимента.

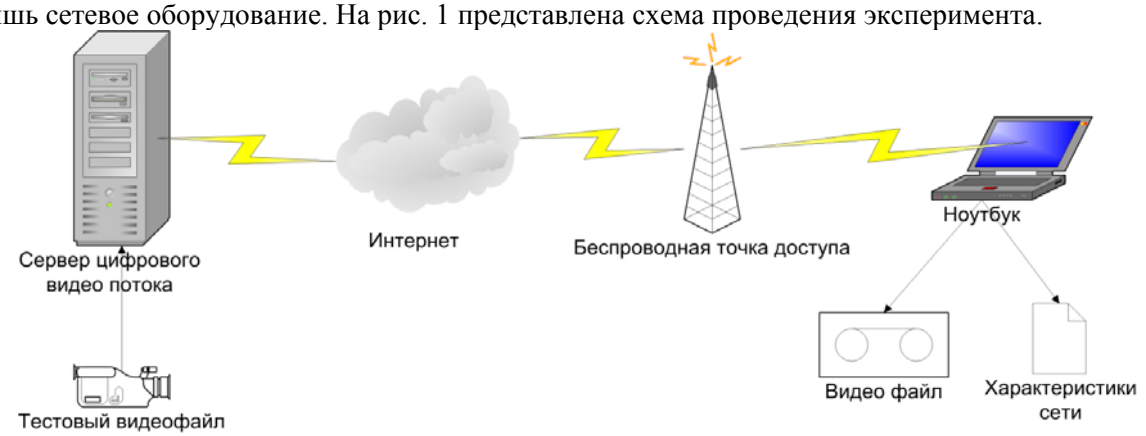


Рис. 1. Схема проведения эксперимента

Для эксперимента был подготовлен один видеоряд с различными типами изображения: статичное, со слабым движением, с быстрым движением, с изменением яркости. Затем видеоряд был закодирован с использованием кодеков MPEG-4 (DivX), MPEG-2 и Windows Media Video 9. При этом установлены следующие параметры видео:

- разрешение картинки - 320 x 240;
- частота кадров – 24 кадр/с.;
- битрейт 256 кбит/с;
- качество – максимальное.

Файлы исходного видео также опубликованы на сайте компании НПЦ «Интернет ТВ» [2].

В качестве программы для трансляции видео по сети использовался программный продукт VideoLan VLC [8]. Данная программа запускалась с сервера, подключенного к сети Интернет с реальным IP адресом.

На компьютере, который принимал видео, был установлен VLC media player, а также сетевой сниффер Wireshark, записывающий весь получаемый по сети трафик в момент передачи видео. VLC media player был настроен на одновременное отображение принимаемого видео на экран и сохранение его в видеофайл. Таким образом, при проведении каждого опыта записываются два файла для дальнейшего анализа: видеофайл и файл сетевых логов.

Обработка результатов экспериментов

Для сети стандарта Wi-Fi характерно периодическое ухудшение характеристик сети, соответственно изображение ухудшается именно в определенные моменты. Нами были проанализированы полученные ранее видеоизображения и соответствующие им сетевые логи, были даны оценки качества видео и найдены характеристики сети в моменты сбоев.

Оценка качества видео проводилась по шкале MOS. Алгоритм оценки следующий:

1. На первом искаженном кадре видео и последнем устанавливаются метки в программе VirtualDub [10].
2. Искаженное изображение можно просматривать несколько раз и сравнить его с оригиналом.
3. Визуально оценивается качество видеоизображения в момент ошибки по шкале MOS от 1 до 5.

В настоящей работе для каждой ошибки свою оценку от 2 до 4 выставляли 4 эксперта. Затем по их оценкам вычислялось среднеарифметическое значение.

При обработке файлов, закодированных при помощи кодека WMV9, возникли сложности. К сожалению, программа VirtualDub не отображает ключевых и неключевых кадров для видео, закодированного кодеком Windows Media Video 9. Поэтому эксперименты с ним не были нами обработаны.

Результаты экспериментов

Полученные в результате проведения экспериментов данные были обработаны по описанной в предыдущем разделе методике. Найдено субъективное качество видео Q_{MOS} в зависимости от процента потерь пакетов p и сетевого джиттера j. Полученные значения коэффициентов сведены в Таблицу 1.

Таблица 1. Значения коэффициентов модели для кодеков MPEG2, DivX в сети WiFi

№ пп	Кодек	$Q_{i\text{-}deal}$	α^k	β^k	α^w	β^w
1	MPEG2	4,2 ±0,2	0,15 ±0,03	0,011± 0,002	0,04 ±0,01	0,003± 0,001
2	DivX	4,7 ±0,2	0,27 ±0,05	0,013± 0,003	0,13 ±0,02	0,01±0 ,002

Было установлено, что качество видеоизображения зависит не только от процента потерь пакетов и сетевого джиттера, но и от типа кадра, на котором произошла ошибка. Поэтому в таблице 1 нами специально выделено два типа коэффициентов – с потерями на ключевых кадрах и без них.

В Таблице 1 α^k и β^k - коэффициенты модели с потерями пакетов на ключевых кадрах, α^w и β^w - коэффициенты для неповрежденных ключевых кадров, а Q_{ideal} – оценка по шкале MOS для исходного файла (до пересылки по сети).

Изначально у видеоизображения, закодированного кодеком MPEG-4 (DivX), качество выше, чем у видео, закодированного MPEG-2, но при ухудшении характеристик сети качество снижается ощутимее и при больших сетевых помехах становится схожим с качеством MPEG-2. В случае повреждения ключевого кадра качество изображения, закодированного MPEG-4 (DivX) упадет более чем в 2 раза при тех же характеристиках сети, а MPEG-2 в 3,75 раза.

Таким образом, для значительного повышения качества видеоизображения при передаче в беспроводной сети необходимо выполнить два обязательных пункта по модернизации схемы связи:

1. Модернизировать видео плейер на приемной стороне с тем, чтобы автоматически откидывать дублирующиеся RTP пакеты.
2. Сервер потокового видео должен дублировать пакеты, содержащие информацию ключевых кадров.
3. Период между ключевыми кадрами не может превышать 2 секунд (оптимально 1 секунда).

Эти простые меры приведут к тому, что даже плохие по оценке GAP сети будут транслировать видео с оценкой MOS большей 3,5. Стоит отметить, что для кодеков MPEG-4 (DivX) и MPEG-2 передаваемый объем трафика увеличится в среднем на 7%, а качество видео не менее чем в 2 раза. Некоторые исследованные видеоплееры при воспроизведении MPEG-TS MPEG-4 потоков уже автоматически отбрасывают повторяющиеся кадры, что дает принципиальную возможность повышения визуального качества передаваемого видео путем незначительной модификации алгоритма MPEG-TS инкапсуляции на сервере трансляций.

К сожалению, на данный момент эксперименты в беспроводных сетях третьего и четвертого поколений еще не обработаны. Эксперименты, проведенные на сети WiMAX, показали очень хорошую устойчивость данной сети к ухудшению качества связи. Сети WiMAX по своим характеристикам сопоставимы с фиксированными сетями Ethernet.

3G сети очень чувствительны к внешним помехам и даже при хорошем уровне сигнала присутствуют значительные потери пакетов (Poor) и вариация задержки порядка 40 мс (Acceptable). Также было установлено, что у одного из 3G провайдеров оборудование зачастую дублировало исходящие пакеты, что само по себе создавало на видео очень сильные артефакты. Устранить данные помехи поможет наше решение с дублированием, а значит и отсеиванием дублированных кадров.

Решение с дублированием ключевых кадров должно значительно повысить качество видео при передаче его в ненадежных из-за расстояния и шумов сетях, таких как WiFi и 3G. Также данное решение может быть применено в беспилотных разведывательных самолетах и спутниках, так как организовать для них надежный беспроводной канал связи для передачи видео очень сложно.

Выводы

В данной работе было проанализировано качество передачи видео в беспроводных сетях стандартов WiFi, 3G и WiMAX и исследовано влияние искажения ключевых кадров на качество видео. Оказалось, что сеть WiMAX достаточно надежна и по характеристикам соответствует проводной локальной сети. На сетях WiFi и 3G качество видео изображения значительно снижается при передаче. Для повышения качества видео в этих сетях нами предложен метод дублирования ключевых кадров сервером потокового видео при использовании кодека MPEG-4 (DivX). Это увеличит качество видео на сетях WiFi и 3G минимум в 2 раза при незначительном увеличении размера потока на 7%. Кроме того данный метод может повысить качество видео за счет модернизации воспроизводящего плейера функцией отсеивания дублирующихся кадров.

В работе были найдены коэффициенты математической модели, описывающей качество передачи видео [4] для случая беспроводной сети WiFi и кодеков MPEG-2 и MPEG-4 (DivX). Для расчета значений коэффициентов была разработана специальная экспериментальная методика, которая позволила собрать и обработать данные.

Литература

1. Cisco Systems, “Cisco Visual Networking Index: Forecast and Methodology, 2008–2013”, 9 июня 2009. (http://www.cisco.com/en/US/ solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf)

2. “Traces of video stream in wireless networks (WiMAX, 3G, WiFi)”. (http://www.ip4tv.ru/stati/aaa.html)

3. S. Bradner, J. McQuaid, RFC2544 - Benchmarking Methodology for Network Interconnect Devices, March 1999

4. A. Sukhov, P. Calyam, W. Daly, A. Ilin, Towards an analytical model for characterizing behavior of high-speed VVoIP applications, Computational Methods in Science and Technology, 11(2), 161-167 (2005)

5. International Telecommunication Union, “Methodology for the subjective assessment of the quality of television pictures (ITU-R BT.500-11)” (2002)

6. M. Claypool and J. Tanner, “The Effects of Jitter on the Perceptual Quality of Video,” in ACM Multimedia, 1999.

7. P. Calyam, M. Sridharan and et. al., “Performance Measurement and Analysis of H.323 Traffic,” in PAM Workshop, 2004.

8. VideoLAN team, “VideoLAN, Free streaming and multimedia solutions for all OS!”. (http://www.videolan.org/)

9. Wireshark Foundation, “Wireshark. Go deep”. (http://www.wireshark.org/)

10. Avery Lee, “Welcome to virtualdub.org! - virtualdub.org”. (http://www.virtualdub.org/)

Объектно-сегментированный анализ телекоммуникационных данных на основе апостериорных вероятностей

А.А. Салмин
Поволжский государственный университет телекоммуникаций и информатики, Самара

В телекоммуникационных компаниях, занимающихся предоставлением массовых услуг (сотовые операторы, Интернет-провайдеры или операторы дальней связи), в процессе деятельности накапливаются огромные объемы данных. Это и информация об абонентах, и всевозможная статистика по использованию ими предоставляемых услуг, размеры генерируемого трафика и многое другое. Эффективное использование всей накопленной информации – ключ к пониманию клиента. Причем следует отметить, что клиент для телекоммуникационных компаний и услуги, которыми он пользуется, воспринимаются как объекты, на которые компания-оператор направляет определенные действия с целью обеспечения более эффективной работы компании. Каждый клиент-объект должен обязательно характеризоваться наличием количественных и качественных атрибутов, объединенных в единое целое на основе инкапсуляции, например, клиент (К) с атрибутами К.А – возраст объекта К; К.S – пол объекта К и так далее. Кроме того, объект может иметь различные состояния, обладать некоторым определенным поведением и уникальной идентичностью, что обязательно должно быть учтено оператором. Состояние объекта может характеризоваться перечнем (обычно статическим) свойств данного объекта и текущими (обычно динамическими) значениями каждого из этих свойств. В качестве поведения воспринимается то, как объект действует и реагирует на возможные действия оператора, то есть поведение объекта можно интерпретировать как его наблюдаемую и проверяемую извне деятельность.

Следует также отметить, что современные телекоммуникационные компании уделяют большое внимание такому понятию, как «сегментация», заключающаяся в аналитическом разделении всех объектов на однородные группы. С помощью сегментации компании могут создавать «маркетинговые сегменты», объединяя объекты в группы по предпочтениям и запросам, используемым услугам, способам использования услуг и другим критериям. Причем каждый такой сегмент должен иметь четко определенные границы, связанные с множеством атрибутов объекта. Например, сегмент, связанный с платежеспособностью (П) объекта К (К.П), которая также является одним из его атрибутов, можно определить как:

$$Сегмент = \{S_i\} = \left\{ K \underbrace{\left| \underline{П}_i \le К.П \le \overline{П}_i \right.}_{C} \right\}, \qquad (1)$$

где $\underline{П}_i$ и $\overline{П}_i$ – соответственно нижняя и верхняя граница платежеспособности i-го клиента, принадлежащего сегменту S_i.

Интерпретация формулы (1) с одной стороны определяет множество клиентов, принадлежащих сегменту S_i, определенному условием С, а с другой стороны S_i – это интервал (фрагмент) шкалы платежеспособности. Шкала платежеспособности – это упорядоченный набор сегментов-интервалов $Ш = (S_1, S_2, ..., S_N)$. Порядок определяется отношением $\overline{П}_i = \underline{П}_{i+1}$, $i = \overline{1, N-1}$.

Определим мощность множества символом m. Тогда количество клиентов в i-ом сегменте (платежном диапазоне) будет определяться как: $m_i = m\left\{K \left| \underline{П}_i \le К.П \le \overline{П}_i \right.\right\}$, а

отношение величин $\frac{m_i}{\sum_{i=1}^{N} m_i}$ можно интерпретировать как вероятность «попадания» клиента К в i-й сегмент p_i (на основе соответствующей ему платежеспособности К.П). Значение N в данном случае – количество сегментов.

Оперирование с частостями $\left(m_i \Big/ \sum_{i=1}^{N} m_i \right)$ с использованием такой интерпретации сегмента дает возможность формализации вероятностных событий (отношений между объектами). Например, вероятность события (попадания молодого клиента в третий платежный диапазон):

$$\frac{m\left\{K\left|(K.A<20)\,\&\,\left(\underline{\Pi}_3 \le K.\Pi \le \overline{\Pi}_3\right)\right.\right\}}{\left\{K(K.A<20)\,\&\,\left(\underline{\Pi}_3 < K.\Pi < \overline{\Pi}_3\right)\right\}}. \qquad (2)$$

Логическая конъюнкция, задающая классообразующий признак, в выражениях для вероятностей событий должна интерпретироваться как условие для определения условной вероятности:

$$\frac{M\left\{K\left|(K.A<20)\,\&\,\left(\underline{\Pi}_3 \le K.\Pi \le \overline{\Pi}_3\right)\right.\right\}}{M\left\{K\left|\left(\underline{\Pi}_3 \le K.\Pi \le \overline{\Pi}_3\right)\right.\right\}} = p\left[\underbrace{(K.A<20)}_{\text{Событие}}\left|\underbrace{\left(\underline{\Pi}_3 \le K.\Pi \le \overline{\Pi}_3\right)}_{\textit{Условие}}\right.\right] \qquad (3)$$

Понятия события, условия и гипотезы сводятся к концепции отношения (равенства, неравенства, эквивалентности и т.п.).

Условие – это зарегистрированный (установленный) факт, событие, исчисляемое допускаемой априорной вероятностью.

Гипотеза – это ожидаемое, возможное планируемое событие (событие, исчисляемое апостериорной вероятностью).

Ожидаемое событие может и должно интерпретироваться как гипотеза, а условие как возможный исход некоторого опыта (или как наблюдаемый факт). При такой интерпретации $p\left[(K.A<20)\left|\left(\underline{\Pi}_3 \le K.\Pi \le \overline{\Pi}_3\right)\right.\right]$ рассматривается как гипотеза того, что клиент К окажется молодым, если он принадлежит третьему сегменту платежеспособности.

Формально и условие, и событие интерпретируются как некоторые отношения между объектами или их атрибутами. Пусть (К.А < 20) будет отношением R1, а $\left(\underline{\Pi}_3 \le K.\Pi \le \overline{\Pi}_3\right)$ отношением R2. Вероятность $p\left(R1\left|R2\right.\right)$ можно вычислить по формуле (3), но вычислить $p\left(R2\left|R1\right.\right)$ (апостериорную вероятность) на основе (3) уже невозможно. В общей постановке задача вычисления апостериорной вероятности гипотез на основе имеющихся наблюдений фактов заключается в следующем. Пусть имеются N гипотез относительно объекта, в качестве которого может выступать клиент телекоммуникационной компании, услуга и т.д. Множество гипотез $A_1, A_2, ..., A_N = \{A\}$ являются несовместными событиями, то есть они не могут произойти одновременно:

$$\sum_{i=1}^{N} p(A_i) = 1.$$

Каждая из гипотез A_i может иметь место при наличии совокупности условий $\{B\} = (B_1, B_2, ..., B_M)$.

Например, гипотезой A_i является принадлежность клиента к i-му платежному диапазону, а совокупность условий $\{B\}$ является набором конкретных характеристик

клиента, по которым его можно отнести к тому или иному сегменту. Так для первого платежного диапазона должен выполняться комплекс условий: B_1 – возраст клиента ($20 < B_1 \le 27$), B_2 – район проживания клиента ($B_2 = 3$), B_3 – семейное положение ($B_3 = 1$) и т. д. Причем некоторые условия могут иметь качественные характеристики, поэтому для их описания они должны быть переведены в количественные эквиваленты. Таким образом, клиент попадет в первый платежный диапазон, если будет выполнен комплекс условий $\{B\}_1$ для данного диапазона, то есть $\{ПД_1 | \{B\}_1\}$.

Предъявление такого рода комплекса условий, полученного на основе объектно-ориентированного анализа статистических данных компании, позволит формировать сегменты и выявлять принадлежности объектов к этим сегментам. Аппаратом сегментации при предъявлении априорных данных объекта, которыми являются частности выполнения комплекса условия в определенном сегменте, может являться формула Байеса, позволяющая на основе априорных данных получать апостериорные значения и имеющая в этом случае вид:

$$p(A_i | \{B\}) = \frac{p(\{B\} | A_i) \cdot p(A_i)}{\sum_{i=1}^{N} p(\{B\} | A_i) \cdot p(A_i)}.$$

Причем если условия B – независимы, тогда:

$$p(\{B\} | A_i) = p(B_1 | A_i) \cdot p(B_2 | A_i) \cdot \ldots \cdot p(B_M | A_i).$$

Если же условия B являются зависимыми между собой, то есть предъявление следующего условия возможно только при совершении предыдущих, вероятность $p(\{B\} | A_i)$ будет иметь вид:

$$p(\{B\} | A_i) = p(B_1 | A_i) \cdot p(\{B_2 | B_1\} | A_i) \cdot p(\{B_3 | B_1, B_2\} | A_i) \cdot \ldots \cdot p(\{B_M | B_1, B_2, \ldots, B_{M-1}\} | A_i).$$

Следует заметить, что зависимость условий может быть не всегда очевидна, причем не учитывание этого фактора может сильно сказаться на конечный результат. Например, зависимость между услугами, предоставляемыми компанией-оператором, на первый взгляд не очевидна, но выявление этой зависимости позволит оператору управлять своими ресурсами и проводить более точные маркетинговые кампании по привлечению и удержанию клиентов.

Определение апостериорных значений на основе формулы Байеса и, как следствие, отнесение объекта к определенному сегменту не несет простых объяснительных функций, оно лишь предлагает исследователю уточнить его знания на основе новых знаний, закодированных в новых порциях информации. В этом смысле наблюдения над многоэтапными экспериментами как процесс, который интегрирует в себя новые и новые знания, может оказаться значительно более информативным, чем любые альтернативные подходы работы с клиентами.

На основе объектно-сегментированного подхода, во главе которого стоит объект, появляется возможность создавать маркетинговые сегменты, учитывая все особенности деятельности компании-оператора и возможные взаимоотношения компании и ее объектов, будь то услуги или клиенты. Достоинство такого подхода сегментации в том, что могут создаваться группы для целевого воздействия, определяться атрибуты, нужды и потребности каждого сегмента, для которых вырабатываются определенные действия с целью более эффективного управления телекоммуникационной компанией.

Самоорганизующиеся нейронные сети в обработке пространственных данных

О.Н. Сапрыкин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В мировой практике организации дорожного движения все активнее используется управление транспортной инфраструктурой крупных городов с применением технологий интеллектуальных транспортных систем. На данный момент существует устойчивая тенденция дальнейшего совершенствования и внедрения подобных систем. В России реализованы отдельные элементы интеллектуальных транспортных систем.

Геоинформационные системы обладают инструментарием описания пространственных данных, однако его не достаточно для описания процессов, происходящих на улично-дорожной сети города. Это обусловлено большим количеством параметров, влияющих на поведение объектов подсистем, нелинейным характером процессов, динамичностью системы. Для создания модели с подобными характеристиками можно прибегнуть к использованию самоорганизующихся нейронных сетей. Самоорганизация, как метод адаптивного синтеза сложных систем, основана на предположении о том, что информация о взаимной корреляции параметров скрыта в экспериментальных значениях этих переменных [1].

По своим конфигурациям, сетевые парадигмы имеют много общего, несмотря на чрезвычайное разнообразие искусственных нейронных сетей. В статье рассмотрены дважды многорядные самоорганизующиеся нейронные сети с активными нейронами на основе метода группового учета аргументов (МГУА). Метод, реализующий процесс построения *дважды многорядной нейронной сети с активными нейронами,* позволяет синтезировать нелинейную структуру многопараметрической регрессионной модели. В основу метода положены принципы *неокончательности решений, внешнего дополнения, массовой селекции* [2].

Механизм самоорганизации нейронной сети с активными нейронами

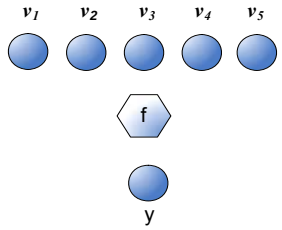


Рис. 1. Состояние нейронной сети перед началом процесса самоорганизации

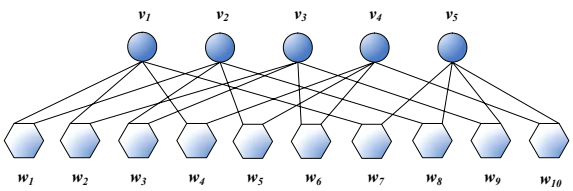


Рис. 2. Перебор возможных моделей и построение первого слоя нейронов

Алгоритм действует по принципам нейронной сети, с одним лишь отличием: структура сети заранее неизвестна и строится в процессе самоорганизации. При этом устанавливается только критерий оптимальности, а эффективные входы выбираются самими нейронами в процессе самоорганизации. Нейроны с подобным поведением называются *активными*.

На нулевом шаге итерации имеем: n входных нейронов $(v_1 \ldots v_n)$, равных количеству переменных; функцию f построения полиномиальной модели (нейрон); значение y, полученное после расчета по построенной модели (рис. 1) [3]. Если в качестве функции используется один из многорядных алгоритмов МГУА, генерирующий полиномиальную функцию, то в итоге будем иметь дело с дважды многорядной нейронной сетью (рис. 2).

На следующем шаге нейроны ранжиру-

ются в соответствии с внешним критерием выбора модели. Выбирается *n* моделей, как оптимальные, остальные отбрасываются (рис. 3).

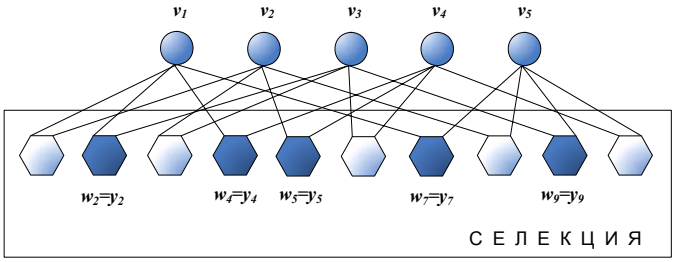


Рис. 3. Выбор нейронов по внешнему критерию в первом слое

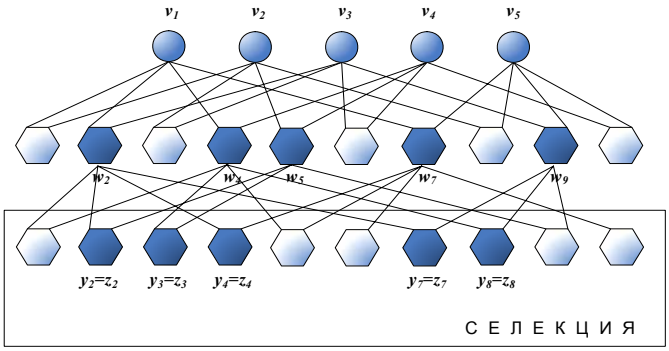


Рис. 4. Выбор нейронов по внешнему критерию во втором слое

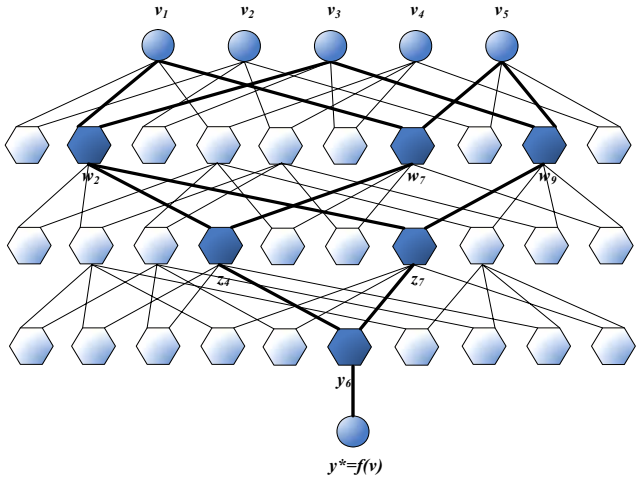


Рис. 5. Нейронная сеть, полученная в результате самоорганизации

Эти *n* моделей будут участвовать в дальнейшей организации модели. Они являются входными нейронами для вновь построенного слоя, который перебирает все возможные комбинации. На втором слое также проходит выборка оптимальных моделей, вычисление внешнего критерия оценки (рис. 4) и вычисление среднего значения критерия в слое ε_{cp}. Если ε_{cp} уменьшилось, значит на этом слое построена модель лучше и самоорганизация продолжится, иначе дальнейшее наращивание слоев не имеет смысла.

Процесс самоорганизации продолжается до тех пор, пока ε_{cp} в текущем слое не станет больше ε_{cp} из предыдущего слоя. В конце обучения выбирается один нейрон, имеющий наилучшее значение внешнего критерия оценки. Из предыдущего слоя удаляются все не связанные с ним, и так далее до входных нейронов (рис. 5).

В отличие от обычной нейронной сети, дважды многорядная нейронная сеть является суперпозицией более мелких нейронных сетей, которые живут параллельно во время самоорганизации. Только после полной организации остается одна нейронная сеть, тогда как другие погибают.

Адаптация методов самоорганизации для работы с пространственно-координированными объектами

Геоинформационные объекты интеллектуальной транспортной системы являются пространственно-координированными, т. к. расположены в пространстве, занимают в нем некоторый объем и привязаны к системе координат, находящейся в этом пространстве. Для обеспечения решения глобальных задач, направленных на повышение уровня безопасности транспортной инфраструктуры, необходимо построить новые модели геообъектов, используя методы самоорганизации. Перед процессом самоорганизации входные пространственно-координированные данные проходят предварительную обработку (препроцессинг). Задача препроцессинга в силу своей специфичности может

быть решена только в частных случаях и не имеет общего решения [4]. При проектировании геоинформационных аналитических систем следует учитывать данный факт и предоставлять большую библиотеку алгоритмов препроцессинга и предоставлять способ написания пользовательских алгоритмов препроцессинга.

Для формального описания методов трансформирования пространственно-координированных данных, введем следующие понятия.

Источники – это пространственно-координированные объекты, значения или расположение которых не зависят от каких либо факторов (независимые переменные).

Рецепторы – это пространственно-координированные объекты, значения которых зависят от расположения и значения источников (зависимые переменные).

Рассмотрим частный случай решения задачи предварительной обработки данных и применения к ним метода самоорганизации. Целью является анализ зависимости уровня аварийности участка улично-дорожной сети от дислокации технических средств организации дорожного движения. Из всех видов технических средств рассмотрим дорожные знаки, различающиеся по типу, номиналу и количеству на участке УДС. В качестве участка УДС выберем перегон – линейный участок дороги и перекресток нескольких улиц. Уровень аварийности является характеристикой участка и имеет непрерывную величину, измеряемую в интервале [0, 1]. В качестве алгоритма построения модели используем дважды многорядную нейронную сеть.

Для решения поставленной выше задачи строится матрица, описывающая расстояния между каждым источником и рецептором, в дальнейшем будем называть ее *матрица источник-рецептор*. Каждый элемент матрицы высчитывается как расстояние от *i*-го рецептора до *j*-го источника по каждой координате:

$$a_{2\cdot j,i} = x_i^r - x_j^s;$$

$$a_{2\cdot j+1,i} = y_i^r - y_j^s.$$

Матрица источник-рецептор, а также вектор значений рецепторов подаются на вход самоорганизующийся нейронной сети с активными нейронами. Метод делит входные данные на два подмножества: обучающую и тестирующую выборки. В качестве критерия оценки модели на каждом шаге самоорганизации применяется критерий внешнего дополнения.

В ходе самоорганизации искусственной нейронной сети строится оптимальная модель, являющаяся полиномом, связывающим расстояния между источниками и рецепторами и значения рецепторов. Готовую модель можно использовать для расчета значений неизвестных рецепторов. Для случая, когда имеются несколько источников, несколько рецепторов и функция, описывающая их связь, каждое неизвестное значение будет рассчитываться по формуле:

$$c_i = \sum_{j=1}^{N} f(X_{ij}, Y_{ij}) \cdot q_j \qquad (1)$$

где $i = \overline{1, M}$, $j = \overline{1, N}$; N – число источников; M – число рецепторов; X_{ij} и Y_{ij} – расстояние между j-м источником и i-м рецептором по оси абсцисс и ординат, соответственно; $f(X_{ij}, Y_{ij})$ – функция зависимости удаленности i-го рецептора от j-го источника; q_j – значение j-го источника.

Автоматизированная система идентификации зависимостей в пространственно-координированных данных

Основываясь на описанных подходах, построена автоматизированная система идентификации зависимостей в пространственно-координированных данных (рис. 6) [5]. Процесс расчета значений по регрессионной модели начинается с выбора пользователем слоя геоинформационной системы, на котором расположены источники и объек-

ты. Пользователь загружает одну из сохраненных моделей. Выбранные пространственно-координированные объекты преобразуются с помощью адаптера (1) и подаются на вход метода *Вычислить()* экземпляра класса *ПолиномиальнаяМодель*. Вычисленные значения поступают в подсистему обработки пространственно-координированных данных. Они могут добавиться на существующий геослой или на новый слой, в зависимости от выбранных пользователем настроек. Добавленные на карту данные немедленно отображаются в подсистеме управления и визуализации, и пользователи сразу видят результаты расчетов.

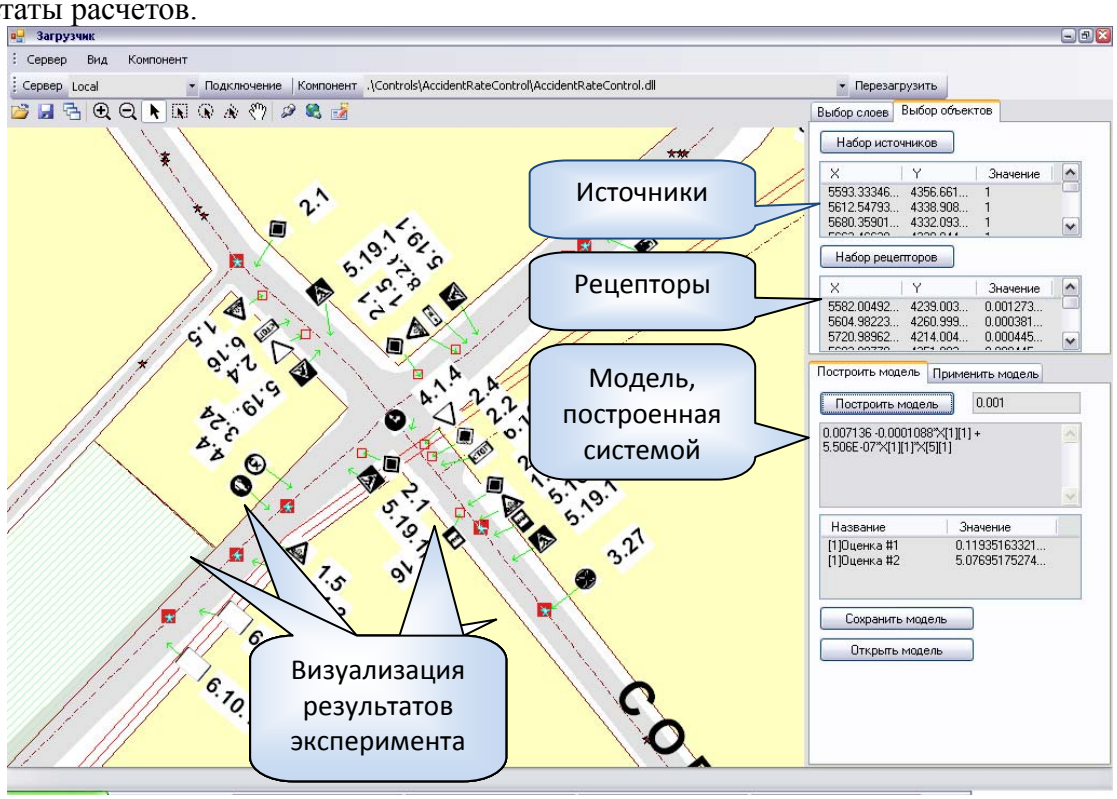


Рис. 6. Экранная форма с отображением результатов эксперимента

Испытания разработанной системы проводились на случайных и реальных данных. В качестве реальных данных использованы слои электронной карты г. Самара, содержащие дислокацию дорожных знаков и значения уровня аварийности участков УДС. Для анализа выбраны перекрестки города, сравнение велось по внешнему критерию перекрестной проверки *PRR(s)* [6]: $PRR(s) = \frac{1}{M_B}\sum_{i=1}^{M_B}(z_i - z_i(B))^2 \to \min$, где M_B – число рецепторов в проверочной выборке, z_i – значение построенной функции в *i*-ой проверочной точке, $z_i(B)$ – значение объекта-рецептора в *i*-ой проверочной точке.

На рис. 6 изображена экранная форма системы идентификации зависимостей в географических данных для исследуемого перекрестка Ново-Садовая ул. - Советской Армии ул.

По методу дважды многорядной нейронной сети (рис. 5) построена функция:

$z=0.003928 + 4.291\cdot10^{-5}\cdot y_1 - 9.397\cdot10^{-7}\cdot y_1\cdot x_3$

где *x, y* – координаты пространственно-координированных объектов-источников на электронной карте города.

Значение критерия $PRR(s) = 1.5904\cdot10^{-9}$. Критерий получил оптимальное значение после первого ряда самоорганизации сети и поэтому построенная функция имеет всего два параметра.

Для сравнения выбраны два наиболее часто используемых алгоритма: метод наименьших квадратов и нейронная сеть обратного распространения ошибки. Алгоритмы реализованы в среде MATLAB, в которую импортированы данные, полученные с каждого перекрестка. По данным построены регрессионные модели с использованием каждого из методов: по методу наименьших квадратов построена линейная полиномиальная модель, по методу нейронной сети - многослойная нейронная сеть с обратным распространением ошибки. Ограничением на построение нейронной сети выбрано число входных нейронов, соответствующее числу переменных, число промежуточных слоев, равное единице, и выходной слой, содержащий один нейрон. Для каждой модели вычислены значения внешнего критерия.

По методу наименьших квадратов получена следующая функция:

$z = 0.0509 \cdot 10^{-4} x_1 - 0.3706 \cdot 10^{-4} y_2 + 0.0521 \cdot 10^{-4} y_9 + 0.1569 \cdot 10^{-4} y_{10}$,

$PRR(s) = 1.3226 \cdot 10^{-5}$.

В многослойной нейронной сети с обратным распространением ошибки взято десять нейронов в скрытом слое, весовые коэффициенты для которого приведены в таблице.

$PRR(s) = 4.4677 \cdot 10^{-7}$.

Таблица. Весовые коэффициенты скрытых нейронов

w_1	w_2	w_3	w_4	w_5
-0.379	*0.2158*	*-1.0041*	*-0.3533*	*1.1399*
w_6	w_7	w_8	w_9	w_{10}
-0.095	*1.3432*	*-0.0006*	*-0.8218*	*0.2087*

Результаты получены и на других наборах данных для более 50 проведенных экспериментов. Метод дважды многорядной нейронной сети позволяет строить модель с лучшими предсказательными характеристиками.

Автоматизированная система идентификации зависимостей в пространственно-координированных данных реализована в среде интеллектуальной транспортной системы г. Самара. Применение системы позволит более эффективно проводить мероприятия, направленные на повышение безопасности дорожно-транспортной инфраструктуры города.

Литература

1. Mueller J.-A., Lemke F. Self-Organizing Data Mining. An Intelligent Approach To Extract Knowledge From Data – Berlin, Dresden – 1999. 225 p.
2. Алгоритм комплексирования аналогов для самоорганизации дважды многорядных нейронных сетей / Ивахненко Г.А. // УСиМ. 2003. № 2. С. 100–106.
3. Ivakhnenko A.G., Ivakhnenko G.A., Mueller J.-A. Self-Organization of Neural Networks with Active Neurons // Pattern Recognition and Image Analysis – 1994. – №4. P. 185-196.
4. Михеева, Т.И. Структурно-параметрический синтез интеллектуальных транспортных систем / Самара: Самар. науч. центр РАН, 2008. 380 с.
5. Mikheev, S.V. Model of zone control of traffic flows / Computer Science and Information Technologies. Mikheeva T.I., Mikheev S.V. //Proceedings of the 10th International Workshop. V. 2. Antalya, Turkey, 2008. P. 121-128.
6. Михеева Т.И., Сапрыкин О.Н. Идентификация зависимостей в пространственно-распределенных данных с использованием нейросетевых технологий // Вестник Самарского гос. техн. ун-та. Серия «Технические науки». Самара: СамГТУ, 2006. С. 93-99.

Исследование качества видеопотоков в беспроводных сетях 3G

А.А. Семенов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Для большинства людей сейчас большой быстрый пакет мобильных возможностей уже далеко не роскошь, а просто повседневность. Специально для особо требовательных и современных людей были разработаны сети третьего поколения – 3G. Эта широкополосная мобильная связь даёт возможность не только пользоваться радиосвязью, но и работать в Интернете с большой скоростью, а также передавать через него видеоматериалы.

Но существует серьезное препятствие для повседневного использования видео сервисов в беспроводных сетях: соответствующее качество связи оставляет желать лучшего. Большой процент потерь пакетов и вариация задержки (сетевой джиттер) делают беспроводные сети малопригодными для передачи данных в режиме реального времени, так как мультимедийный трафик чрезвычайно чувствителен к подобным искажениям. Пакеты потокового видео теряются при передаче по сети, изменяют порядок следования из-за значительной вариации задержек пакетов. На получаемом видеоизображении появляются множественные артефакты, происходит рассинхронизация потока, что приводит к искажениям изображения, а иногда и к полной остановке воспроизведения видео.

Идея субъективного тестирования (MOS) состоит в том, что видео, закодированное одним из кодеков MPEG-2, MPEG-4 или Windows Media Video 9, передается через сеть Интернет и беспроводную сеть стандартов WiFi, 3G или WiMAX. Полученное после передачи видео, демонстрируется группе экспертов, которые выставляют оценки, основываясь на своих впечатлениях от качества.

В настоящей работе рассматривается проблема адаптации современных алгоритмов кодирования и передачи видео для беспроводной сети 3G путем спектрального анализа сетевого трафика, а также сделана попытка найти численную зависимость качества видеоизображения от сетевых параметров.

Математическая модель.

При передаче видео по сети качество связи ухудшается в зависимости от характеристик сетевого соединения. Для того чтобы описать качество передаваемого видео в зависимости от сетевых параметров была рассмотрена универсальная функция $Q_{MOS}(p, f, D, B)$, описывающая качество видео по шкале MOS. Эта функция может быть разложена в степенной ряд по сетевым параметрам, при этом можно ограничиться линейными членами и только одним вкладом второго порядка, описывающим передачу видео с различными скоростями потока.

На конференции TERENA 2005 было показано, что для фиксированной скорости видео потока достаточно рассмотреть только члены разложения, описывающие линейную зависимость от потерь пакетов и сетевого джиттера:

$$Q_{mos} = Q_{ideal} - \alpha p - \beta j \qquad (1)$$

где

Q_{ideal} – максимальное качество видео для данного кодека, баллы от одного до пяти;

p – процент потерь пакетов, %;

j –сетевой джиттер (вариация задержки) в момент ошибки, сек.;

Q_{MOS} – качество видео на приемной стороне, баллы от одного до пяти;

α, β – коэффициенты модели, которые могут быть найдены экспериментально.

Для проведения исследований была разработана единая видео последовательность, которая обрабатывалась кодеками MPEG-4 (DivX), MPEG-2 и Windows Media video 9 с постоянным битрейтом 256 кбит/с.

Отдельно вычисляются коэффициенты α^k и β^k для потока с повреждением ключевых кадров, а также α^w и β^w для последовательностей без повреждения ключевых кадров. Например, α^k_{DivX} будет обозначать усредненный коэффициент, который характеризует ухудшение качества видео, закодированного кодеком MPEG-4 (DivX), если повреждение пришлось на ключевой кадр. Далее в работе будут найдены и сопоставлены α^k_{DivX} и α^w_{DivX}, а также α^k_{Mpeg2} и α^w_{Mpeg2}, чтобы определить в численном виде меру ухудшения качества видеоизображения.

Планирование эксперимента

Для анализа трафика и нахождения коэффициентов уравнения (1) был разработан и проведен ряд экспериментов. Закодированные кодеками MPEG-4 (DivX), MPEG-2 и Windows Media Video 9 видео файлы, пересылались на ноутбук, подключенный к беспроводной сети 3G. На ноутбуке проводилась запись получаемого видео в файл, параллельно записывался сетевой трафик на уровне пакетов при помощи сетевого снифера Wireshark. Таким образом, по полученному видеоизображению можно установить качество видео по шкале MOS, а по сетевым логам – параметры сетевого соединения.

На рис. 1 представлена схема проведения эксперимента.



Рис. 1. Схема проведения эксперимента

В качестве программы для трансляции видео по сети использовался программный продукт VideoLan VLC. Данная программа запускалась с сервера, подключенного к сети Интернет с реальным IP адресом.

На компьютере, который принимал видео, был установлен VLC media player, а также сетевой сниффер Wireshark, записывающий весь получаемый по сети трафик в момент передачи видео. VLC media player был настроен на одновременное отображение принимаемого видео на экран и сохранение его в видеофайл. Таким образом, при проведении каждого опыта записываются два файла для дальнейшего анализа: видеофайл и файл сетевых логов.

Для проведения экспериментов использовались – самарские сегменты сети операторов связи: Мегафон (3G), Билайн (3G), которые бесплатно предоставили оборудование и технические возможности для тестовых испытаний.

Обработка результатов экспериментов

Порядок анализа всех полученных видеофрагментов и сетевых логов одинаков. Далее описан процесс анализа одного из видеофайлов и соответствующего ему файла сетевых логов в моменты ухудшения качества видео:

Основой анализа трафика видеопотоков является построение функции распределения F(t) и плотности распределения P(t) межпакетного интервала (время между получением предыдущего и последующего пакетов) и размера пакетов. К сожалению VideoLan VLC работает лишь со стандартными пакетами (1370 байт) и построение последних зависимостей на основе полученных результатов невозможно.

Для построения функции и плотности распределения межпакетного интервала необходимо воспользоваться данными записанными с помощью сетевого сниффера Wireshark. В сетевой сниффер Wireshark загружаются соответствующие логи сети (http://stream.ip4tv.ru/wireless/WiFi/test1/1divx.pcap), записанные в момент трансляции данного видео. Так как нам известно, точное время получения каждого пакета, то несложно вычислить межпакетный интервал.

Обработка результатов эксперимента для нахождения коэффициентов математической зависимости (1) проводится следующим образом:

Программой VirtualDub открывается видеофайл (http://stream.ip4tv.ru/wireless/WiFi/test1/1divx.mp4), полученный в результате опыта. Ищется кадр, предшествующий искаженному кадру, например это кадр 143, который отображается на 5923 мс видеоряда. Аналогично находится последний искаженный кадр, пусть это 171 кадр, показываемый на 7083 миллисекунде. Таким образом, вычитая из времени начала искажения видео время окончания получится длительность искажения, составляющая 28 кадров или 1160 мс.

В сетевой сниффер Wireshark загружаются соответствующие логи сети (http://stream.ip4tv.ru/wireless/WiFi/test1/1divx.pcap), записанные в момент трансляции данного видео. Во встроенном анализаторе приведен список всех пакетов и красным помечены места, где были потеряны пакеты. Также указаны статистические данные по межпакетному интервалу и джиттеру.

Основная проблема анализа – как соотнести файл сетевых логов с видеозаписью. Для этого в логе пакетов ищется пакет (226), после которого произошла ошибка. По отношению к началу записи логов этот пакет поступил на 10,878914 секунде. В пункте 1 было найдено, что в видеоряде искажение произошло на 143 кадре, который отображался на 5923 мс видеоряда. Таким образом, искомое соотношение определено.

В пункте 1 установлено, что длительность искажений в видеоизображении 1160 мс. Соответственно от 226 кадра, пришедшего на 10,878914 секунде, отсчитывается 1160 мс. Получаем 12,038914 с. По сетевому логу Wireshark видно, что последний пакет, который пришел до 12,038914 секунды это пакет 253.

В Анализаторе RTP пакетов также указан межпакетный интервал и джиттер для каждого пришедшего пакета.

Другая важнейшая задача, возникающая при обработке результатов экспериментов, это оценка качества видео по шкале MOS. Алгоритм оценки следующий:

На первом искаженном кадре видео и последнем устанавливаются метки в программе VirtualDub (см. рис. 2).

Искаженное изображение можно просматривать несколько раз и сравнить его с оригиналом.

Визуально оценивается качество видеоизображения в момент ошибки по шкале MOS от 1 до 5.

В настоящей работе для каждой ошибки свою оценку от 2 до 4 выставляли 4 эксперта. Затем по их оценкам вычислялось среднеарифметическое значение.

Результаты экспериментов

Полученные в результате проведения экспериментов данные были обработаны по описанной в предыдущем разделе методике.

Характерные функции и плотности распределения для кодеков MPEG-2 и MPEG-4 приведены на рисунках 2, 3.

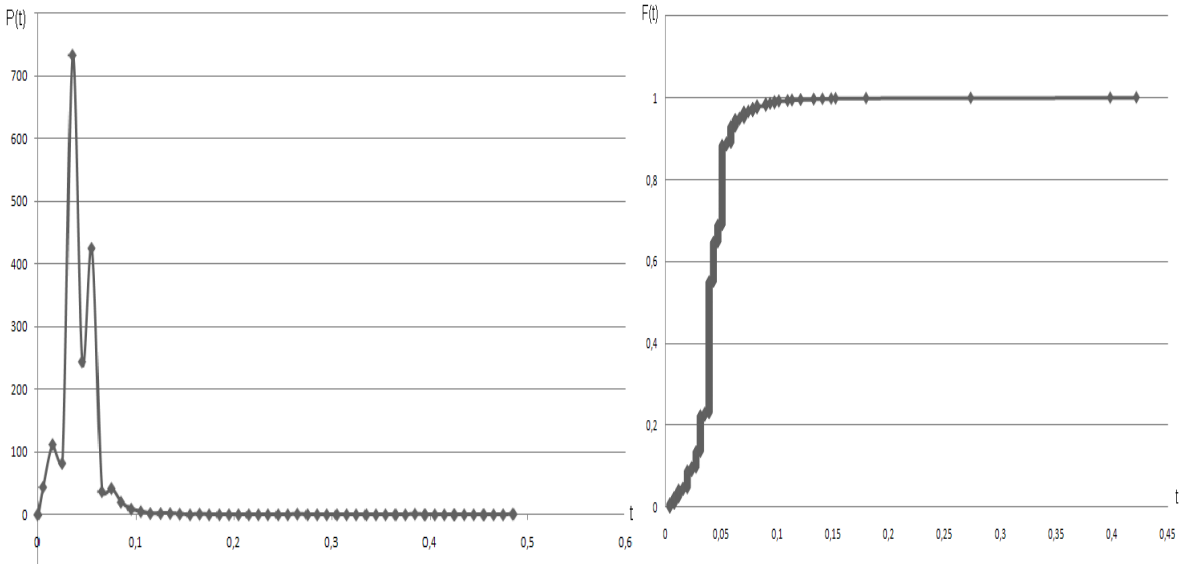


Рис. 2. Функция и плотность распределения межпакетного интервала для кодеков MPEG-4

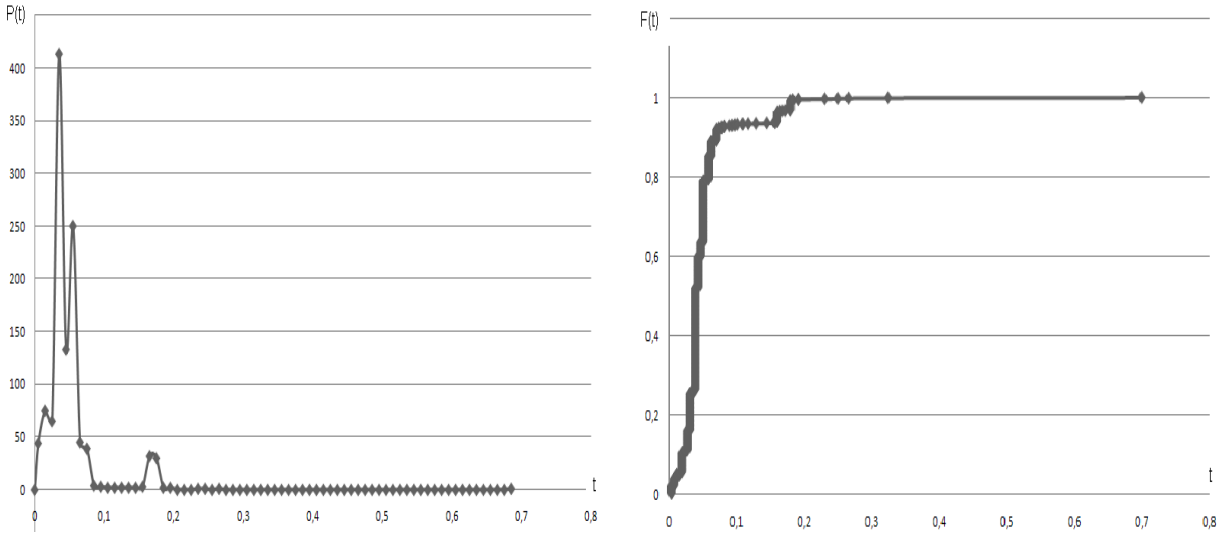


Рис. 3. Функция и плотность распределения межпакетного интервала для кодеков MPEG-2

Найдено субъективное качество видео Q_{mos} в зависимости от процента потерь пакетов p и сетевого джиттера j. Полученные значения коэффициентов сведены в таблицу 1.

Таблица 1. Значения коэффициентов модели для кодеков MPEG2, DivX в сети 3G

№ пп	Кодек	Q_{ideal}	α^k	β^k	α^w	β^w
1	MPEG2	4.2±0.2	0.005±0.002	0.005±0.002	0.004±0.001	0.003±0.001
2	DivX	4.7±0.2	0.01±0.003	0.003±0.001	0.002±0.0005	0.002±0.0008

Было установлено, что качество видеоизображения зависит не только от процента потерь пакетов и сетевого джиттера, но и от типа кадра, на котором произошла ошибка. Поэтому в таблице 1 нами специально выделено два типа коэффициентов – с потерями на ключевых кадрах и без них.

В таблице 1 α^k и β^k - коэффициенты модели с потерями пакетов на ключевых кадрах, α^w и β^w - коэффициенты для неповрежденных ключевых кадров, а Q_{ideal} – оценка по шкале MOS для исходного файла (до пересылки по сети).

Выводы

В данной работе было проанализировано качество передачи видео в беспроводных сетях стандарта 3G на основе построения функции и плотности распределения межпакетного интервала, по которым в дальнейшем планируется исследовать качество беспроводных сетей, и исследовано влияние искажения ключевых кадров на качество видео. Для повышения качества видео в этих сетях нами предложен метод дублирования ключевых кадров сервером потокового видео при использовании кодека MPEG-4 (DivX). Данный метод может повысить качество видео за счет модернизации воспроизводящего плейера функцией отсеивания дублирующихся кадров.

В работе были найдены коэффициенты математической модели, описывающей качество передачи видео для случая беспроводной сети 3G и кодеков MPEG-2 и MPEG-4 (DivX). Для расчета значений коэффициентов была разработана специальная экспериментальная методика, которая позволила собрать и обработать данные.

Мы считаем, что результаты наших исследований должны значительно повысить качество видео при передаче его в ненадежных из-за расстояния и шумов сетях, таких как 3G. Также данное решение может быть применено в беспилотных разведывательных самолетах и спутниках, так как организовать для них надежный беспроводной канал связи для передачи видео очень сложно.

Литература

1. Cisco Systems, “Cisco Visual Networking Index: Forecast and Methodology, 2008–2013”, 9 июня 2009. (http://www.cisco.com/en/US/ solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf)

2. “Traces of video stream in wireless networks (WiMAX, 3G, WiFi)”. (http://www.ip4tv.ru/stati/aaa.html)

3. S. Bradner, J. McQuaid, RFC2544 - Benchmarking Methodology for Network Interconnect Devices, March 1999

4. A. Sukhov, P. Calyam, W. Daly, A. Ilin, Towards an analytical model for characterizing behavior of high-speed VVoIP applications, Computational Methods in Science and Technology, 11(2), 161-167 (2005)

5. International Telecommunication Union, “Methodology for the subjective assessment of the quality of television pictures (ITU-R BT.500-11)” (2002)

Исследование влияния параметров алгоритмов обучения нейронной сети на точность решения задачи семантического анализа текста

О.П. Солдатова, Э.В. Кольбова
Самарский государственный аэрокосмический университет им. академика С.П. Королёва (национальный исследовательский университет), Самара

Введение

Стремительный рост сети Интернет естественным образом породил проблемы поиска и упорядочивания информации. Сегодня в электронных хранилищах по всему миру содержатся терабайты информации. Информационных источников становится все больше, и получить из этого океана нужные знания становится все труднее. В связи с этим встает проблема построения универсального классификатора текстов, имеющего возможность распределения исходного набора текстов по нескольким заранее установленным рубрикам в соответствии с их семантикой, то есть смысловым содержанием. Использование такого классификатора позволило бы сократить трудозатраты на поиск нужной информации, представленной электронными текстами.

Одним из перспективных направлений на сегодняшний день считается использование нейронных сетей в качестве основы подобного рода классификатора. Основным преимуществом нейронных сетей является возможность выявления зависимостей, не поддающихся обнаружению при использовании других методов обработки информации. Нейросетевой подход к анализу текстовой информации обладает достаточным быстродействием и не зависит от языка предметной области, но при этом, в отличие от большинства алгоритмов обработки текстов, реализованных на основе статистического подхода, дает хорошие результаты.

Общая идея применения аппарата нейронных сетей в задачах классификации текстов на естественном языке (ТЕЯ) основывается на представлении описания классов в виде значений синаптических весов искусственных нейронов.

Описание специфики задачи классификации ТЕЯ

Первоочередной задачей для нейросетевого классификатора ТЕЯ является выбор способа преобразования анализируемого текста к виду, приемлемому для подачи на вход нейронной сети. В настоящей работе используется частотное представление. В частотном представлении текст характеризуется вектором частот встретившихся в нем лемм. Лемма – это каноническая форма лексемы языка (проще говоря, лемма – основа слова, лишенная падежных, родовых, числовых и других подобных характеристик).

Общий вид системы классификации текстов представлен на рис. 1. Частотный анализатор оперирует с системным словарем, при помощи которого производится морфологическая обработка текста, и на выходе выдает частотную характеристику текста. Частотная характеристика - это вектор $f=(f_1,...f_n) \in F$, длина которого равна количеству слов в словаре V и каждая компонента f_i которого есть целое неотрицательное число:

$$f_i = \sum_{j=0}^{k} e(v_i, t_j) \;\; ; \;\; e(v_i, t_j) = \begin{cases} 0, & v_i \neq t_j \\ 1, & v_i = t_j \end{cases}$$

Иначе говоря, для каждого слова $v_i \in V$ определяется число его вхождений $f_i \geq 0$ в данный текст $t=t_1t_2t_3...t_k$. Частотную характеристику f можно рассматривать как точку в пространстве признаков F, соответствующую тексту t.

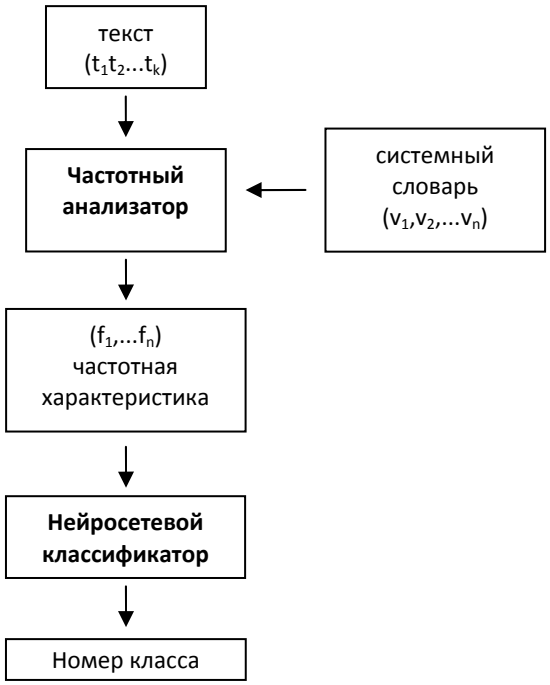


Рис. 1. Общий принцип работы системы классификации ТЕЯ

Нейросетевой классификатор классифицирует вектор частотных характеристик (точку в пространстве признаков), полученный с помощью частотного анализатора, т.е. разделяет все пространство признаков на определенное количество областей. Для решения этой задачи в настоящей работе используется модель многослойного персептрона. Размер входного слоя равен длине вектора частотной характеристики. Размер выходного слоя определяется количеством областей (числом классов), на которые мы будем делить пространство признаков. Сеть такой архитектуры позволяет выделять в пространстве признаков области любой формы и сложности [1]. Итак, на вход нейронной сети подается вектор частотной характеристики, на выходе получаем вектор $(y_0...y_m)$, где m – количество классов. Номер j, для которого выход y_j имеет максимальную активность (т.е. $y_j=max(y_i); i=0...m$), соответствует номеру класса входного образца. Этот номер соответствует одной из заданных пользователем тем (спорт, политика, медицина и т.п.).

Для обучения сети перед началом работы классификатора пользователю необходимо определить классы, с которыми будет работать система, и подобрать множество учебных текстов. Далее из множества учебных текстов, определенным образом выделяются слова - формируется системный словарь. На последнем этапе, используя учебные тексты и полученный системный словарь, обучается нейросетевой классификатор. После процедуры обучения классификатор текстов готов к работе.

Исследование алгоритмов обучения нейронной сети

Для обучения нейронной сети типа «многослойный персептрон» в данной работе использовались четыре алгоритма:

алгоритм последовательного обучения (Incremental Propagation);

быстрого обучения (QickProp);

эластичного обучения (RProp);

пакетного обучения (BatchProp).

Алгоритм Incremental Propagation представляет собой классический алгоритм обратного распространения ошибки - итеративный градиентный алгоритм, который используется с целью минимизации среднеквадратического отклонения (СКО) текущего

выхода многослойного персептрона и требуемого выхода. Другие алгоритмы являются его модификациями. А именно: алгоритм RProp основан на использовании не величины производной ошибки сети, а ее знака; алгоритм BatchProp предполагает осуществлять коррекцию весов и смещений один раз за период обучения - после предъявления всех векторов обучающей последовательности (так называемой «пачки» векторов); алгоритм QickProp построен исходя из допущений о том, что зависимость ошибки сети от каждого весового коэффициента может быть аппроксимирована выпуклой параболой, и что изменение какого-либо веса в процессе настройки не оказывает влияния на другие веса [2].

Как уже было сказано, целью обучения многослойного персептрона является минимизация СКО текущего выхода нейронной сети от требуемого выхода. Рассмотрим динамику уменьшения СКО для каждого алгоритма. Для этого создадим нейронную сеть и будем обучать ее на одних и тех же данных. На рис. 2 приведены графики изменения СКО сети по эпохам для всех четырех исследуемых алгоритмов.

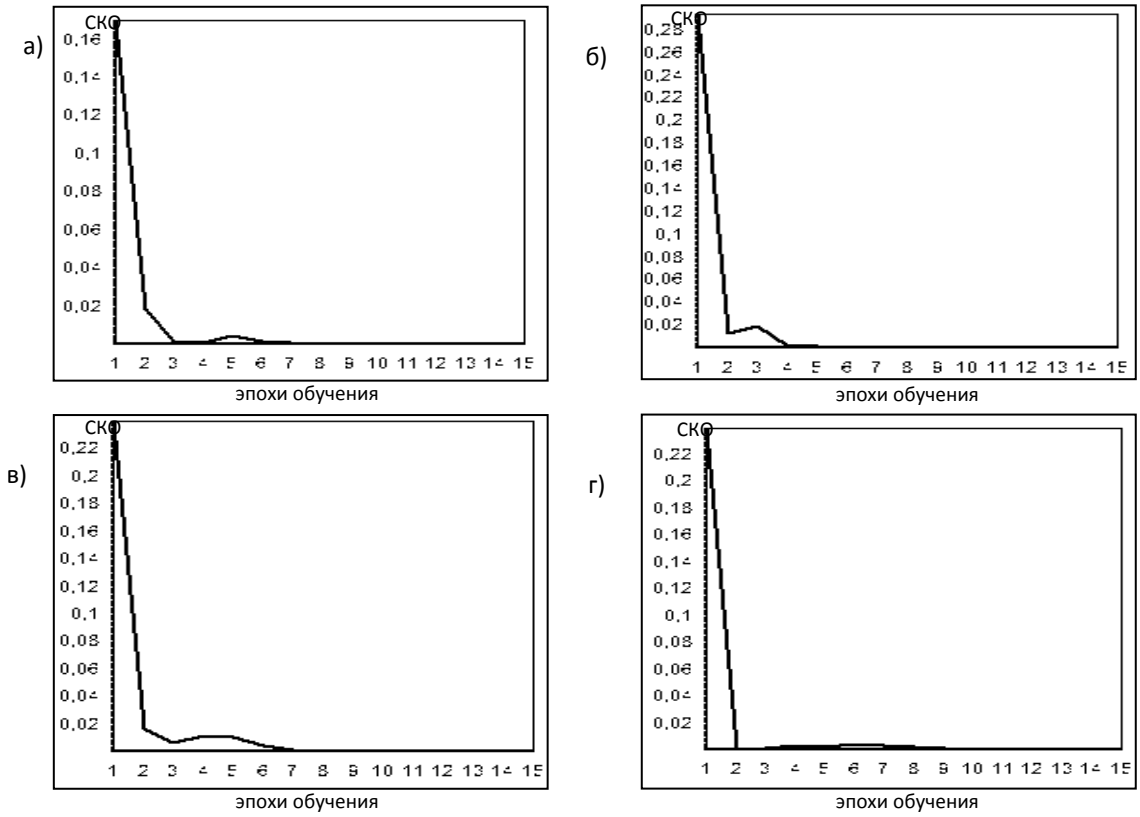


Рис. 2. Изменение СКО сети по эпохам при обучении: а - алгоритмом Incremental Propagation, б - алгоритмом QickProp, в - алгоритмом RProp, г – алгоритмом BatchProp

В ходе исследований было установлено, что на скорость снижения СКО в значительной степени влияют параметры начальной инициализации весовых коэффициентов сети. Начальная инициализация происходит случайными числами в пределах [-r;r]. В табл. 1-3 приведены данные экспериментов по обучению сети с одним скрытым слоем, сигмоидальной функцией активации для скрытого и выходного слоев при различных значениях r. Эксперименты проводились на сети, обученной на 80 примерах для 5 рубрик. Проведены по 10 испытаний для каждого алгоритма обучения, завершение работы алгоритма происходило каждый раз на сотой эпохе. Табл. 4 содержит усредненные значения СКО для каждого алгоритма. Из табл. 4 видно, что наибольшее влияние параметров начальной инициализации на качество обучения многослойного персептрона оказывается при обучении алгоритмами Batch и Rprop.

Таблица 1. Значения СКО для разных алгоритмов обучения нейронной сети при r = 0.5

№	Incremental	QuickProp	Rprop	Batch
1	4,66E-05	6,71E-07	1,15E-05	3,15E-05
2	2,67E-06	3,72E-06	5,99E-05	2,25E-05
3	3,26E-05	5,99E-07	1,98E-05	7,82E-06
4	2,85E-08	7,89E-06	8,53E-06	2,19E-05
5	6,59E-05	6,17E-08	2,43E-05	8,15E-06
6	2,37E-06	2,59E-05	7,37E-05	6,39E-05
7	7,95E-04	4,98E-09	1,64E-05	4,99E-05
8	9,65E-06	4,58E-06	8,53E-06	7,62E-05
9	7,47E-05	6,89E-09	1,33E-05	1,54E-05
10	1,99E-06	4,17E-06	8,97E-06	5,48E-06

Таблица 2. Значения СКО для разных алгоритмов обучения нейронной сети при r = 0.7

№	Incremental	QuickProp	Rprop	Batch
1	2,04E-08	4,35E-07	1,06E-07	5,07E-07
2	1,58E-07	1,57E-05	9,04E-09	1,97E-10
3	4,17E-09	8,06E-06	4,05E-07	1,20E-07
4	1,56E-05	4,35E-07	4,33E-06	1,34E-09
5	1,01E-07	2,13E-07	3,92E-07	6,45E-09
6	6,80E-09	8,17E-06	5,99E-06	4,00E-09
7	9,75E-09	4,47E-06	1,86E-07	2,65E-09
8	6,92E-07	7,97E-07	7,40E-06	9,79E-09
9	3,00E-07	1,57E-05	2,70E-07	7,65E-10
10	3,03E-08	6,43E-07	7,47E-06	3,47E-10

Таблица 3. Значения СКО для разных алгоритмов обучения нейронной сети при r = 0.1

№	Incremental	QuickProp	Rprop	Batch
1	1,44E-08	8,87E-10	2,11E-11	4,45E-11
2	5,00E-08	2,62E-09	2,87E-11	6,33E-12
3	5,01E-09	2,39E-13	5,46E-10	8,75E-12
4	1,83E-11	3,42E-08	3,09E-11	3,98E-11
5	1,72E-08	4,16E-15	4,64E-10	3,46E-12
6	5,78E-09	1,49E-15	1,30E-11	4,75E-12
7	2,10E-08	3,21E-10	5,14E-12	4,87E-11
8	1,03E-10	1,91E-15	4,83E-10	6,93E-12
9	4,22E-09	3,95E-10	2,58E-11	6,26E-12
10	6,84E-09	4,97E-10	5,09E-11	2,98E-11

Таблица 4. Сводная таблица усредненных значений СКО для разных алгоритмов обучения при различных r

	Incremental	QuickProp	Rprop	Batch
r=0,5	1,03E-04	4,76E-06	2,45E-05	3,03E-05
r=0,7	1,69E-06	5,46E-06	2,65E-06	6,53E-08
r=1	1,25E-08	3,89E-09	1,67E-10	1,99E-11

Стоит отметить, что при значениях r>1 СКО снижается еще быстрее, однако на этапе тестирования обнаруживается эффект нулевого выхода (выходной вектор оказывается нулевым), означающий, что сеть не соотносит предъявленный ей текст ни к одной рубрике. Таким образом, за оптимальное значение предела начальной инициализации весовых коэффициентов сети следует взять значение равное 1.

Кроме того, следует отметить, что полученные оказанные результаты означают, что для задачи классификации текстов наиболее эффективными алгоритмами обучения являются Batch и RProp, так как за одно и то же количество шагов они достигают наименьших значений СКО.

Проведены исследования качества классификации при помощи сети, обученной на 80 примерах для 5 рубрик, в зависимости от СКО. Сеть содержит один скрытый слой из 250 нейронов, размер входного слоя 2500 нейронов обучение проводилось алгоритмом Rprop при r=1. Все тексты для тестовых примеров и обучающей выборки были взяты из новостных каналов русскоязычного Интернета. Результаты исследования приведены в табл. 5.

Таблица 5. Зависимость качества работы сети от СКО

СКО сети	Кол-во ошибок сети на 50 тестовых примерах	Время обучения, сек.
2,35E-04	30	124
1,13E-05	24	265
3,12E-06	16	276
1,48E-07	10	324
4,38E-08	8	516
5,34E-09	7	644
1,14E-10	7	698
3,12E-11	6	719
2,67E-12	5	864
2,27E-13	4	986
4,99E-14	4	1327
5,12E-15	3	1815
3,99E-16	3	2431
1,65E-17	3	3903
5,17E-18	3	5673

Как следует из таблицы, для данной обучающей выборки оптимальным можно считать значение СКО порядка -15 степени.

Литература

1. В.А. Головко, под ред. проф. А.И. Галушкина Нейронные сети: обучение, организация и применение [Текст] - М: ИПРЖР, 2001. - 358 с.

2. С. Осовский Нейронные сети для обработки информации [Текст]: [пер. с польск.] / И.Д. Рудинский – М.: Финансы и статистика, 2002. - 344 с.

Экспериментальная проверка модели измерения доступной пропускной способности

Т.Г. Султанов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Доступная пропускная способность канала связи является важной характеристикой его работы и её оценка заслуживает значительного внимания. Измерение доступной пропускной способности необходимо не только для того, чтобы знать состояние сети, но также предоставлять информацию сетевым приложениям с целью управления исходящим трафиком для разделения полосы пропускания сети между ними.

С одной стороны, завышение необходимой пропускной способности ведет к непроизводительным расходам средств. С другой, если канал не сможет обеспечить требуемые скорость и качество связи, то задержка информации управления в сети может быть недопустимо большой. Таким образом, определение оптимальных значений пропускной способности сети передачи данных и времени задержки информации является задачей, имеющей большое практическое значение.

В данной работе сделана попытка представить универсальную и простую модель, позволяющую измерять метрику доступной пропускной способности. Представленный метод базируется на тестировании сети пакетами различного размера. Эта технология известна под названием Переменный Размер Пакета (Variable Packet Size, VPS) и применялась в работе [6]. Технология VPS позволяет оценить полную пропускную способность до *i*-го узла на основе данных о двусторонней задержки пакета (Round-Trip Time, RTT) до *i*-го узла и размере тестового пакета *W*.

Модель и область её применимости

В работе [3] было продемонстрировано, что полная задержка передаваемого пакета есть сумма минимальной составной задержки D_{fixed} и неустранимой переменной части задержки d_{var}, которая распределена экспоненциально. Для того чтобы устранить постоянную часть задержки было предложено тестировать сеть пакетами разного размера, так чтобы размер пакета различался на максимально возможную величину. Тогда для измерения полосы пропускания можно использовать выражение:

$$B_{av} = \frac{W_2 - W_1}{D_2 - D_1} \tag{1}$$

В этой формуле используются следующие переменные:

W_1, W_2 – размеры 1-го и 2-го передаваемого пакета,

D_1, D_2 – полные задержки 1-го и 2-го передаваемого пакета,

B_{av} – доступная пропускная способность канала.

В данной работе была рассмотрена область применимости модели, то есть, какой диапазон скоростей каналов связи можно измерять при помощи данного метода. Доказано, что верхнюю границу для полосы пропускания можно вычислить по формуле:

$$B = \frac{W_2 - W_1}{2\Delta D}\eta \tag{2}$$

,

где η – точность измерений в %,

ΔD – погрешность величины задержки пакета,

B – верхняя граница доступной пропускной способности.

Поэтому, с помощью системы RIPE Test Box, позволяющей находить задержку с погрешностью в 2 микросекунды $\Delta D=2\cdot10^{-6}$ *мкс*, можно измерять полосу пропускания до 300 Mbps с точностью в 10%.

Предложенный метод позволяет найти способ для устранения ограничений, связанных с переменной частью задержки. Переменная часть задержки является причиной довольно большой погрешности измерений других методов.

Экспериментальное сравнение различных методов измерений доступной пропускной способности

Для проверки модели было проведено экспериментальное сравнение различных методов измерений доступной полосы. Эксперимент проводился в два этапа, первый из них - это прецизионное тестирование большим количеством пакетов разного размера с помощью измерительной системы RIPE Test Box [5]. Количество измерительных узлов этой системы достигает 80-ти, эти узлы покрывают основные мировые интернет центры, достигая наибольшей плотности в Европе. Погрешность измерений задержки пакетов не превышает 2 микросекунд. Для реализации эксперимента, три точки RIPE Test Box были установлены в Москве, Самаре и Ростове на Дону в течение 2006-2008 лет по гранту РФФИ 06-07-89074. Нами, для последующего анализа, были собраны несколько наборов данных, содержащих до 3000 измерений на различных направлениях, прежде всего, на участке Самара – Амстердам (tt01.ripe.net – tt143.ripe.net).

На втором этапе было проведено сравнение данных, полученных нашим методом, с результатами уже существующих методик измерения. В качестве инструмента, реализующего альтернативный метод измерений, был выбран программный продукт *pathload* [8]. Данный программный продукт считается одним из известных инструментов для оценки доступной ширины канала.

Помимо этого с помощью утилиты *traceroute* было мы определили «узкое место» маршрута пакетов между Самарой (СГАУ) и Москвой (ИОХ РАН), им оказался внешний маршрутизатор СГАУ, ширина канала на котором была ограничена 30 Мбит/с.

Все данные приведены в Таблице 1 с указанием времени проведения эксперимента.

Таблица 1. Сравнительный анализ результатов измерений

№ пп	Время проведения эксперимента	Направление измерений	Доступная пропускная способность канала по данным RIPE Test Box, Мбит/с	Доступная пропускная способность канала по данным *pathload*, Мбит/с	Данные с маршрутизатора, Мбит/с
1	13.01.2010	tt143->tt146	10.0±2.2 Mbps	21.9±14.2 Mbps	12.1±2.5 Mbps
2	13.01.2010	tt146->tt143	4.4±1.2 Mbps		7.8±3.8 Mbps
3	23.01.2010	tt143->tt146	20.3±5.1 Mbps	41.2±14.0 Mbps	18.7±1.1 Mbps
4	23.01.2010	tt146->tt143	9.3±2.7 Mbps		11.3±2.6 Mbps
5	06.02.2010	tt143->tt146	9.2±1.4 Mbps	67±14 Mbps	12.0±2.0 Mbps
6	06.02.2010	tt146->tt143	3.5±1.2 Mbps		4.5±2.0 Mbps

Из Таблицы 1 видно, что результаты, полученные нашим методом, и данные, собранные утилитой MRTG с маршрутизатора, хорошо согласуются, в то время как результаты *pathload* сильно отличаются. Работы по исследованию типа переменной части задержки [3] дают ответ на вопрос, почему так происходит. Разброс результатов измерений объясняется наличием переменной части задержки d_{var}.

Условие устойчивой работы утилит, измеряющих доступную пропускную способность

Как было уже отмечено в предыдущем разделе, основным недостатком большинства современных утилит является большой разброс значений измеряемой величины – доступной пропускной способности. Это объясняется тем, что механизмы измерений, положение в их основу, не учитывают влияние переменной части задержки. К сожалению, во всех разработанных утилитах нет механизма поправок, компенсирующих случайную компоненту задержки.

Любой метод, который будет давать точные результаты, должен содержать в своей основе механизм, нивелирующий влияние d_{var}. Чтобы понять, как влияет переменная часть на результаты измерений, обратимся к следующему эксперименту. Между точками RIPE Test Box tt01.ripe.net (Амстердам, Голландия) и tt143.ripe.net (СГАУ, Самара) был проведён ряд измерений и получено около 3000 значений задержек пакетов размером 100 и 1024 байта в обоих направлениях. С помощью представленного метода были рассчитаны величины доступной пропускной способности канала, для случаев, когда усреднение ведется по 20, 50 и 100 парам значений задержек пакетов. На Рисунке 1 представлен график, отображающий изменения рассчитанной пропускной способности в зависимости от выбранного для усреднения количества пар пакетов.

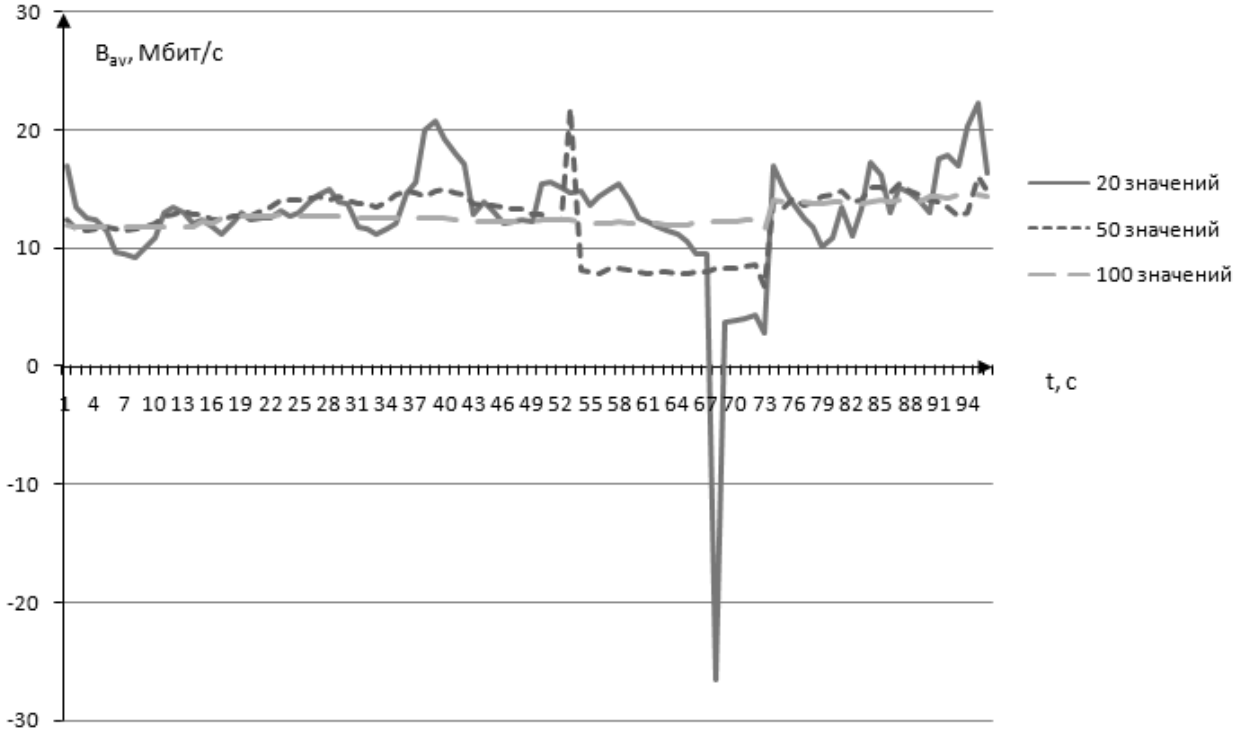


Рис. 1. Зависимость доступной пропускной способности от количества измерений

Как видно из графика, биения рассчитанной пропускной способности являются критичными при 20 значениях задержек, при 50 они менее заметны, а при 100 значениях линия графика практически выравнивается. Прослеживается явная зависимость между количеством измерений и вариацией рассчитанной пропускной способности. Наличие биений обусловлено переменной частью задержки. Отметим, что её роль снижается при увеличении количества измерений.

В данном разделе будет рассчитано необходимое количество измерений при помощи двух методик: на основе анализа экспериментальных данных, полученных от RIPE Test Box и путем симуляции, зная тип распределения для сетевой задержки пакетов.

На основании данных, полученных в ходе эксперимента между точками tt01 и tt143, нами были вычислены среднеквадратические отклонения (СКО) $\sigma_n(B)$ для доступной пропускной способности для пакетов разных размеров (см. Таблицу 2).

Таблица 2. Зависимость СКО от числа измерений

Число измерений, *n*	5	10	20	30	40	50	70	100	200	300
Среднеквадратическое отклонение, $\sigma_n(B)$ (Мбит/с)	22.2	14.9	10.2	8.3	7.3	6.7	5.7	4.9	2.9	2.3
Среднее значение доступной пропускной способности сети, B_{av} (Мбит/с)	13.1									

Как видно из Таблицы 2, необходимо брать как минимум 50 измерений (разность для 50-ти пар пакетов), в этом случае рассчитанное значение пропускной способности в 2 раза превышает СКО, то есть: $B \geq 2\sigma_n(B)$.

Более аккуратный результат можно получить при помощи обобщенной генерирующей функции для описания задержек пакетов. В работе [3] было установлено, что распределение задержки подчиняется экспоненциальному закону, а для длительности задержки можно использовать следующую генерирующую функцию:

$$D = D_{\min} + W/B - (1/\lambda)\ln(1 - F(D,W)), \quad (3)$$

где $\lambda = 1/(D_{av} - D_{\min})$, а функция *F(W,D)* эмулируется стандартным генератором случайных чисел на интервале [0; 1).

Знание формы генерирующей функции дает возможность рассчитать табличные значения функции η_n^T. Для этого при помощи генерирующей функции (3) рассчитаем значения СКО $\sigma_n^T(D_2 - D_1)$ для усредненной разности от задержек пакетов разных размеров [мс], приняв $\lambda^T = 1000$ с$^{-1}$.

Таблица 3. Эмуляция зависимости СКО от количества измерений (λ^T=1000 с$^{-1}$)

Число измерений, *n*	5	10	20	30	50	100	200
Среднеквадратическое отклонение, $\sigma_n^T(D_2 - D_1)$ (мс)	0.661	0.489	0.354	0.284	0.195	0.111	0.075

Для полученных значений СКО из Таблицы 3 произведем расчет табличных значений функции η_n^T. Расчеты будем проводить для следующих величин: $\Delta W^T = W_2 - W_1 = 1000$ байт, $B^T = 10$ Мбит/с, что соответствует $D_2^T - D_1^T = 8 \cdot 10^{-4}$ с.

Таблица 4. Зависимость погрешности от количества измерений

Число измерений, *n*	5	10	20	30	50	100	200
Погрешность измерений, η_n^T (%)	28.6	16.1	14.2	13.5	12.4	11.9	9.4

В ходе реального эксперимента измеряемые значения $\lambda^{\exp}$, $D_2^{\exp} - D_1^{\exp}$ или $B^{\exp}$ принимают произвольные значения, но поправочные коэффициенты позволяют легко рассчитать требуемое количество измерений для оценки экспериментальной погрешности измерений из табличных значений:

$$\eta_n^T = k(D_2 - D_1) \cdot k(\lambda) \cdot \eta^{\exp}, \quad (4)$$

где $k(\lambda) = \lambda^{\exp} / \lambda^{T}$, а $k(D_2 - D_1) = (D_2^{\exp} - D_1^{\exp})/(D_2^{T} - D_1^{T})$.

Подставляя в уравнение (4) значения коэффициентов $k(D_2 - D_1)$, $k(\lambda)$ и требуемую погрешность измерений η_n^T сравним полученные значения с табличными η_n^T и найдем число измерений *n*, требующихся для достижения заданной погрешности.

Выводы

В данной работе проведена экспериментальная проверка нового метода с использованием измерительной инфраструктуры RIPE Test Box. Это позволило обеспечить высокую точность результатов эксперимента. Было произведено сравнение данных, полученных описанным в работе методом, с результатами уже существующих технологий измерений доступной пропускной способности. В качестве сравнительного инструмента была выбрана утилита *pathload*. В результате проверки выяснилось, что разработанный метод обладает преимуществами по сравнению с другими технологиями измерения, так как последние в достаточной мере не учитывают влияние переменной части задержки. Показано, что для точного измерения доступной пропускной способности канала, необходимо нивелировать влияние d_{var}, путем увеличения количества проводимых измерений. Были найдены границы применимости модели на основе точности измерений, проведено компьютерное симулирование и найдена связь между ошибкой измерения доступной пропускной способности и количеством измерений.

Литература

1. Платонов А.П., Сидельников Д.И., Стрижов М.В., Сухов А.М., Измерительная инфраструктура для изучения качества соединений в российском сегменте Интернет, Телекоммуникации, 2009, № 1, с.11-16
2. Choi, B.-Y., Moon, S., Zhang, Z.-L., Papagiannaki, K. and Diot, C.: Analysis of Point-To-Point Packet Delay In an Operational Network. In: Infocom 2004, Hong Kong, pp. 1797-1807 (2004)
3. A.M. Sukhov, N Kuznetsova, What type of distribution for packet delay in a global network should be used in the control theory? 2009; arXiv: 0907.4468
4. Padhye, J., Firoiu, V., Towsley, D., Kurose, J.: Modeling TCP Throughput: A Simple Model and its Empirical Validation. In: Proc. SIGCOMM Symp. Communications Architectures and Protocols, pp. 304-314 (1998)
5. Dovrolis C., Ramanathan P., and Moore D., Packet-Dispersion Techniques and a Capacity-Estimation Methodology, IEEE/ACM Transactions on Networking, v.12, n. 6, December 2004, p. 963-977
6. Downey A.B., Using Pathchar to estimate internet link characteristics, in Proc. ACM SICCOMM, Sept. 1999, pp. 222-223.
7. Ripe Test Box, http://ripe.net/projects/ttm/
8. Jacobson, V. Congestion avoidance and control. In Proceedings of SIGCOMM 88 (Stanford, CA, Aug. 1988), ACM
9. Prasad R.S., Dovrolis C., and B. A. Mah B.A., The effect of layer-2 store-and-forward devices on per-hop capacity estimation, in Proc. IEEE INFOCOM, Mar. 2003, pp. 2090-2100.
10. Jain, M., Dovrolis, K.: End-to-end Estimation of the Available Bandwidth Variation Range. In: SIGMETRICS'05, Ban_, Alberta, Canada (2005)

Исследование производительности компьютерных сетей на основе анализа их трафика

В.Н. Тарасов, Н.Ф. Бахарева, А.Л. Коннов
Поволжский государственный университет телекоммуникаций и информатики, Самара
Оренбургский государственный университет, Оренбург

Введение

Моделирование процессов передачи данных в корпоративных сетях и исследование их производительности не возможно без создания моделей этих сетей. Существует ряд готовых программных систем имитационного (дискретно-событийного) моделирования сетей, таких как: Opnet Modeler [1], NetCracker, COMMNET и др. В работе для исследования компьютерных сетей использована программная система Opnet Modeler IT Guru Academic Edition. Она представляет собой упрощенную коммерческую версию IT Guru 9.1, предоставляемую бесплатно в образовательных целях. Одним из недостатков этой программы дискретно-событийногй имитации, является ограничение по количеству событий произошедших в сети.

Неотъемлемой частью анализа любой компьютерной сети является сбор и анализ циркулирующего в ней сетевого трафика. Необходимость анализа сетевого трафика может возникнуть по нескольким причинам. Кроме показателей производительности, это и аспекты безопасности, поиска узких мест для оптимизации структуры сети, отладки работы сети, контроля входящего/исходящего трафика для оптимизации работы разделяемого подключения к сети Интернет др.

Постановка задачи

Авторами ставилась задача промоделировать компьютерную сеть 14-го и 15-го корпусов, представляющей собой часть корпоративной сети Оренбургского государственного университета и проанализировать полученные результаты на предмет ее возможной модернизации. Данная сеть состоит из 11 подсетей 2-х факультетов, главного коммутатора, сервера и маршрутизатора, показанных на рис. 1.

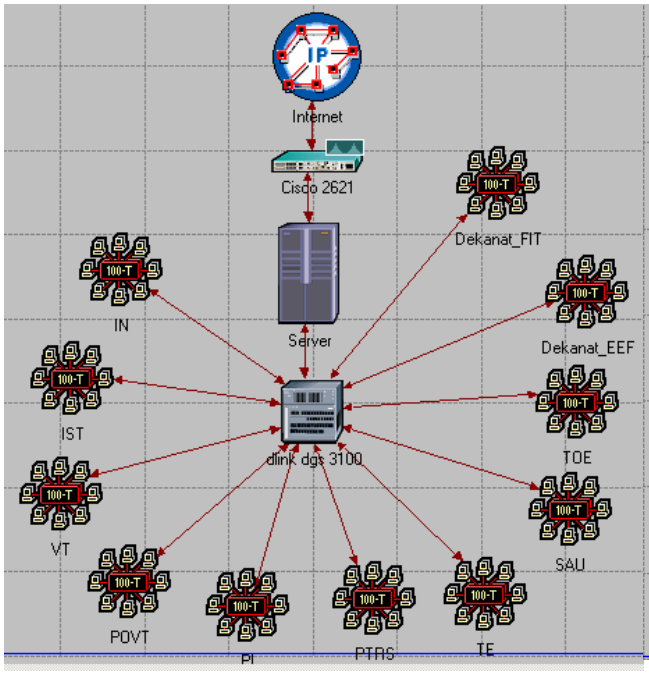


Рис. 1. Модель сети 14-го и 15-го корпусов

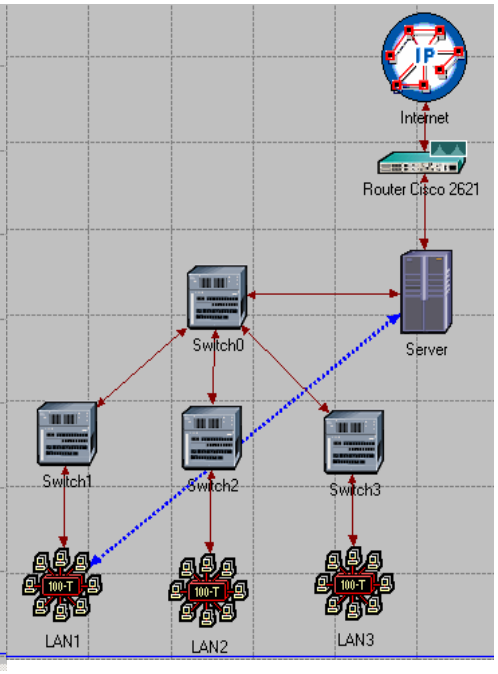


Рис. 2. Модель сети кафедры

Подсети здесь представляют собой отдельные ЛВС кафедр и деканатов, объединенные в общую сеть каналами связи на 100 Мбит/с. В качестве подсети рассмотрим, например, ЛВС кафедры вычислительной техники, хотя все они в основном отличаются друг от друга только количеством компьютеров.

Сеть кафедры ВТ образуют три учебных класса по 10 компьютеров (рис. 2). Коммутаторы учебных классов подключены к главному коммутатору. К этому же коммутатору подключен сервер, предоставляющий такие сетевые сервисы как доступ в Интернет, доступ в локальное файловое хранилище по протоколам FTP и NetBIOS, электронная почта, работа с базами данных.

Решение поставленной задачи

При помощи бесплатной демоверсии программы NetFlow Analyzer был собран трафик в пакетах за периоды - день, неделя, месяц для всех подсетей и всей сети в целом для его анализа. Один из таких трафиков подсети приведен на рис.3.

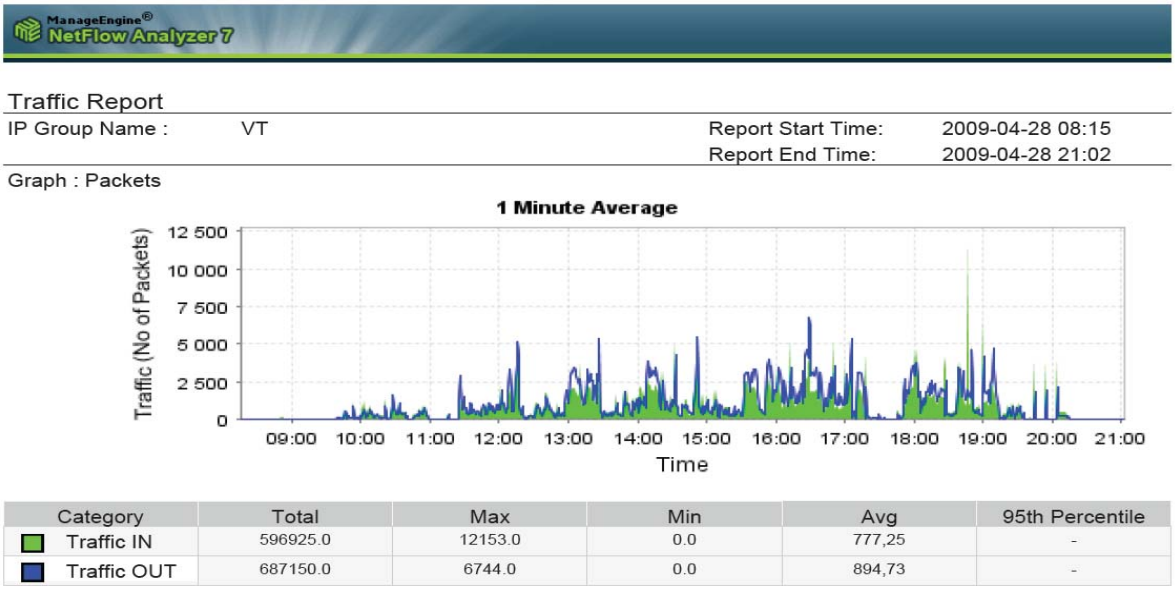


Рис. 3. Трафик сети кафедры ВТ в пакетах за один день

Процесс моделирование начнем с сетей кафедр, т.е. с подсетей. По собранному трафику определяем максимальную интенсивность поступления пакетов от маршрутизатора к серверу, равную 203 пак/с (12153/60), где средняя длина пакета – 763 байта. Наружу, т.е. от сервера к маршрутизатору поступает по максимуму 112 пак/с (6744/60). Поставим **задачу**: какую нагрузку на эту подсеть создает такой трафик?

Задаем в модели эти интенсивности 203 и 112 пак/с трафика в соответствующих направлениях, от маршрутизатора к серверу и наоборот (рис 4). Длительность моделирования – один рабочий день.

Затем весь поступающий трафик делим поровну между 3-мя подсетями (т. к. количество компьютеров в классах одинаково) и получаем интенсивность на входе каждой подсети 67 пак/с, на выходе 37 пак/с. Проделаем это также для второй и третьей подсетей кафедры ВТ.

В результате прогона модели получаем следующие данные по загрузкам каналов, как это показано на рис. 5. Результаты показывают, что загрузки каналов связи на 100 Мбит/с практически малы (максимальная загрузка всего 1,5%). Это позволяет утверждать следующее:

– такие ЛВС имеют большой запас производительности, т.е. входящий трафик может быть увеличен в несколько десятков раз при условии, что эта сеть работает автономно;

– остальные подсети сети факультетов также загружены не более чем на 5%;

– зная входящий и исходящие трафики и пропускную способность всех каналов связи, можно спрогнозировать всю нагрузку на сеть;

– очевидно, что реальные загрузи каналов связи и узлов, будут еще меньше.

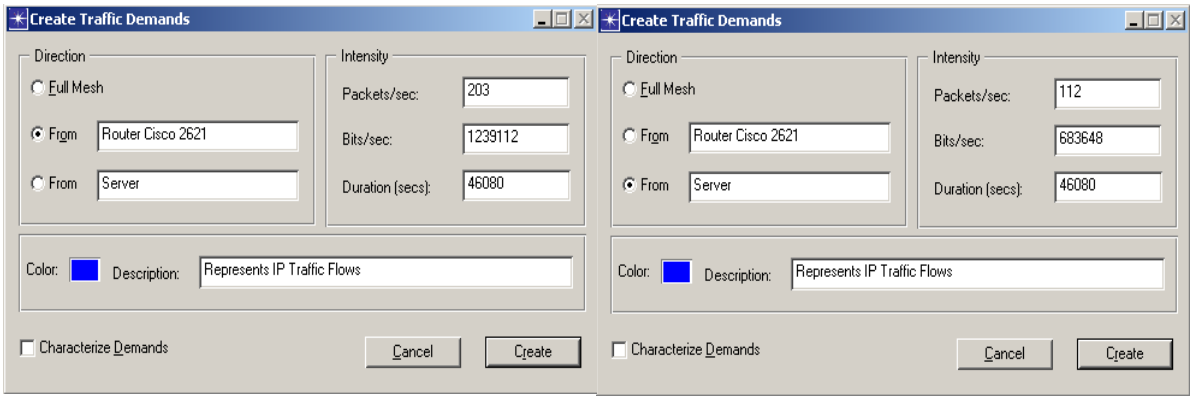


Рис. 4. Задание входящего и исходящего трафика

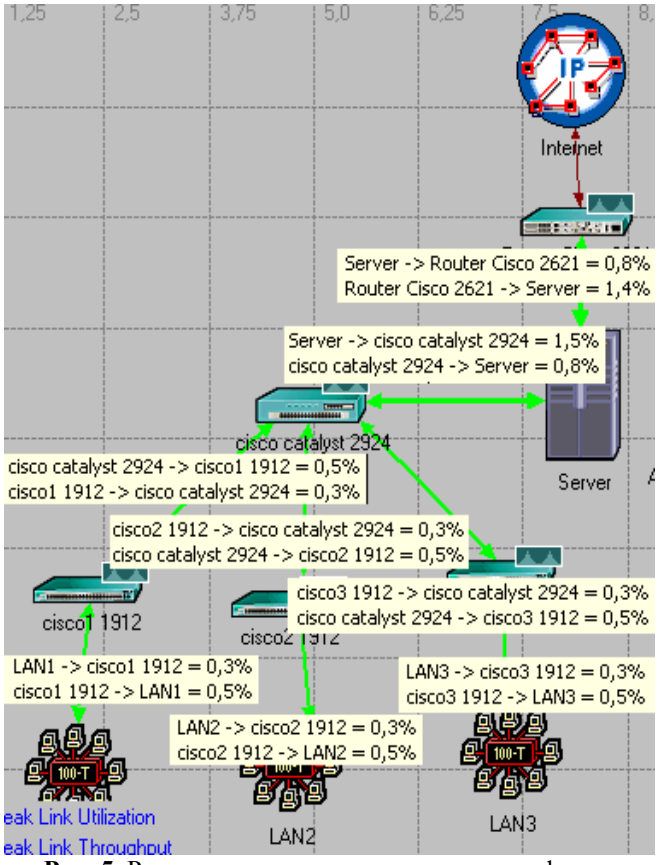


Рис. 5. Результаты эксперимента по сети кафедры

Аналогичным образом исследуем сеть 14-го и 15-го корпусов ГОУ ОГУ (рис.1).

По собранному трафику определяем максимальную интенсивность поступления пакетов от маршрутизатора к серверу, равную 4030 пак/с. Наружу, т.е. от сервера к маршрутизатору поступает по максимуму 3280 пак/с (см. рис. 6).

Зададим эти значения интенсивности трафика в указанных направлениях, а также выше смоделированные интенсивности всех 11 подсетей в обоих направлениях и получим результаты моделирования сети 14-го и 15-го корпусов в таком виде, как это показано на рис.7.

Для большей убедительности сложим полученные загрузки линий всех подсетей в обоих направлениях. Сложение сверху вниз дает: 2,8+3,0+1,4+5,4+1,8+2,4+1,2+1,7+1,7+2,4+0,2=24%. В тоже время результаты эксперимента показывают результат загрузки линии сервер - главный коммутатор в 24,1%. Направление снизу вверх по подсетям (рис.7) дает результат 2,3+2,5+0,8+4,4+1,5+1,9+1,0+1,4+1,4+2,1+4,1=23,4%, а по эксперименту получили 22,7%. Как видим, расхождение составляет всего лишь доли процента.

Разница в загрузках линий маршрутизатор-сервер (26,3/21,6) и сервер-главный коммутатор (24,1/22,7) объясняется тем, что сервер, помимо ответов на запросы пользователей подсетей, еще сам обменивается данными с Интернетом. В реальной сети это может быть при обновлении операционной системы самого сервера.

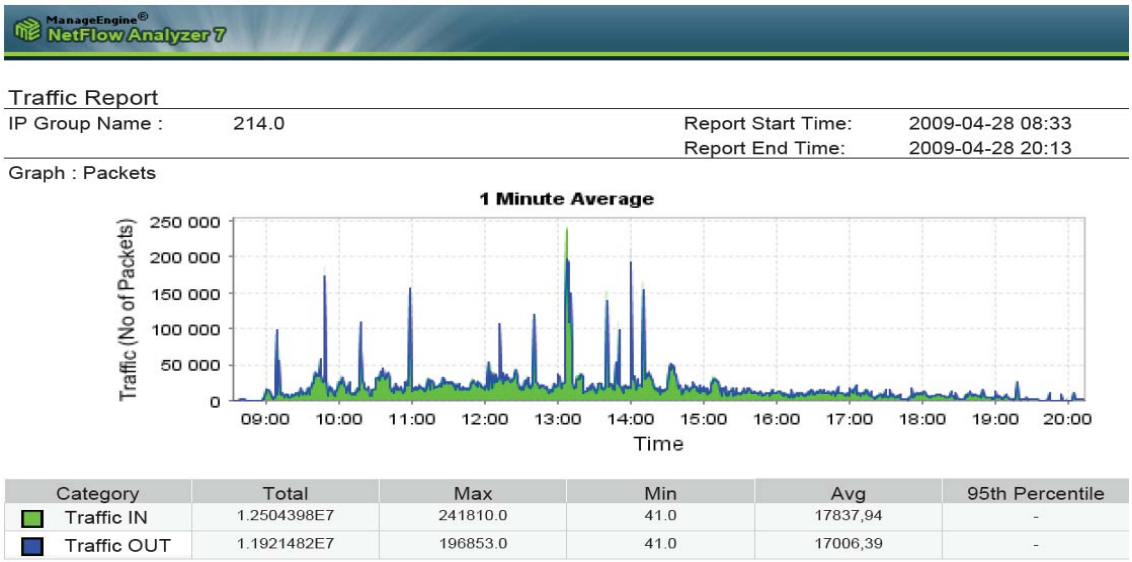


Рис. 6. Трафик сети факультетов в пакетах за один день

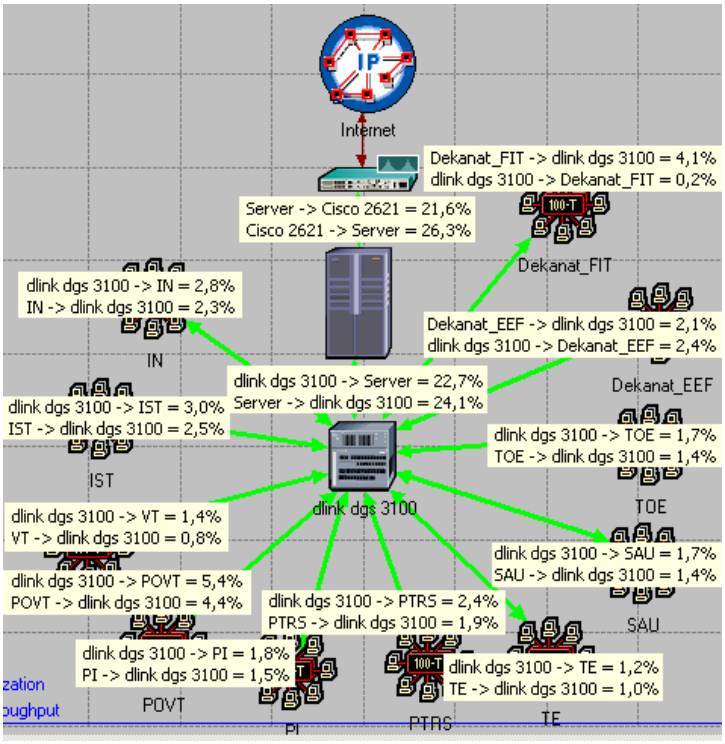


Рис. 7. Результаты эксперимента по сети 2-х факультетов

Как видно из рис. 7, максимальная загрузка каналов связи составляет 26,3%, и это при пиковом значении нагрузки. Следовательно, реальная загрузка каналов и узлов у этой сети намного меньше и имеется также большой запас производительности. Учитывая, что указанная сеть двух факультетов в свою очередь является подсетью корпоративной сети ОГУ, куда входят еще сети остальных 13 корпусов, необходимо дополнительно исследовать «вклад» остальных корпусов в общую нагрузку.

Кроме того, необходимо еще проанализировать другие показатели производительности сети. В их числе основными являются задержка сети и времена отклика приложений. Как показали результаты моделирования при максимальном значении интенсивности входящего трафика, эти показатели находятся на уровне 0,16 мс (см. рис.8) и 0,05 с (см. рис. 9) соответственно. Это для таких сетей очень хорошие показатели.

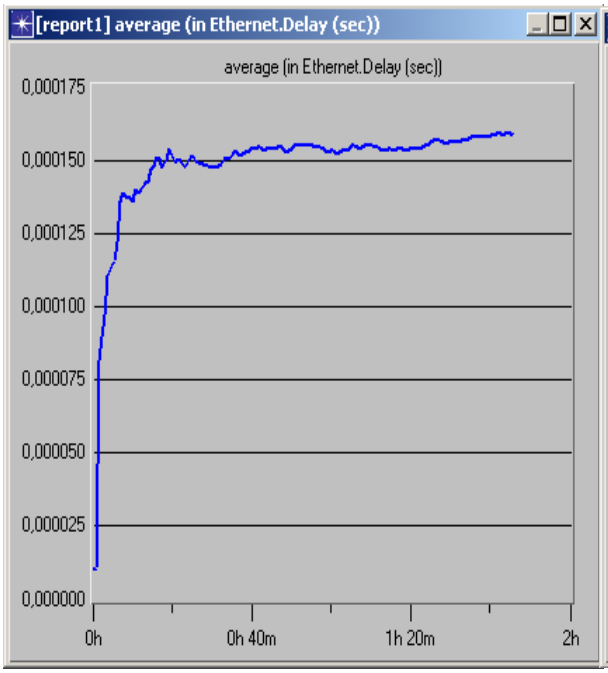


Рис. 8. Задержки Ethernet

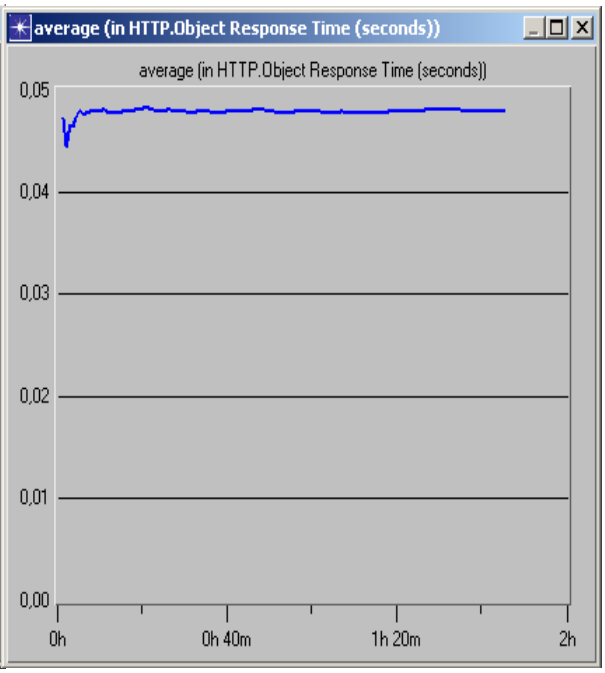


Рис. 9. Время отклика приложений

Выводы

Метод декомпозиции сети на подсети упрощает ее исследование, а учет трафика при этом повышает достоверность моделирования. Таким образом, результаты моделирования вполне адекватно могут отражать процессы функционирования реальных сетей. По результатам эксперимента из рис. 5 видно, что загрузки внешних линий для сети кафедры 0,8%/1,4% полностью совпадают с полученными данными для всей сети – 0,8%/1,4% (рис.7).

Результаты моделирования задержки Ethernet – для сети кафедры 0,08 мс, для сети корпусов – 0,2 мс, а также времен отклика приложений порядка 40-50 мс, говорят о том, что сеть работает устойчиво. Поэтому существует резерв по пропускной способности, позволяющий задействовать в сети дополнительные сетевые приложения как для сетей кафедр, так и для сети факультетов.

Литература

1. Бахарева, Н.Ф. Проектирование и моделирование мультисервисной сети кафедры вуза / Н.Ф. Бахарева, А.Л. Коннов, // Инфокоммуникационные технологии. – 2008. – Т.6, №3. – С. 132 - 138.

2. Тарасов, В.Н. Проектирование и моделирование сетей ЭВМ в системе OPNET Modeler. Лабораторный практикум. / В.Н. Тарасов, Н.Ф. Бахарева, А.Л. Коннов, Ю.А. Ушаков. – Самара, 2008. – 233 с.

3. Тарасов, В.Н. Анализ и оптимизация локальных сетей и сетей связи с помощью программной системы OPNET MODELER /В.Н. Тарасов, А.Л. Коннов, Ю.А. Ушаков // Вестник ОГУ. – 2006. – Т.2 , №6. – С. 197-204.

Подсекция 6
«Информационные технологии в управлении движением»

Фреймовая модель представления знаний в системе планирования и оптимизации маршрутов грузоперевозок

А.И. Белоусов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

При разработке интеллектуальных транспортных систем одной из наиболее часто решаемых задач является задача построения оптимальных маршрутов перевозок различных грузов. Типичная постановка и алгоритмы решения такой задачи для перевозок однородных грузов мелкими партиями изложены в [1]. В качестве критериев оптимальности транспортных маршрутов используются время перевозок, транспортные затраты, а также критерии безопасности маршрута [2]. В связи с возрастающей сложностью решаемых задач возникает необходимость существенного изменения архитектуры соответствующих систем и перехода к системам, основанным на знаниях.

Характерной особенностью таких систем является постоянно изменяемая и расширяемая база знаний. Связано это как с совершенствованием методической и технической базы предприятий и повсеместным внедрением компьютерных сетей, так и с распространением и популяризацией всемирной компьютерной сети Интернет. Представляется актуальной идея разработки систем на базе унифицированного представления знаний, которое будет, с одной стороны, обеспечивать прозрачную интеграцию с объектной и компонентной моделями построения программных систем, с реляционными базами данных, а так же допускать распределение фрагментов знаний по сети с последующим распределенным выводом для решения некоторой прикладной задачи.

Фреймовая модель базы знаний

В качестве модели хранения знаний разрабатываемой интеллектуальной информационной системы (ИИС), была выбрана концепция фреймовых представлений, логически объединяющая системы хранения материальных и абстрактных (информационных) объектов [3, 4]. Единицей хранения и информационным представлением знаний является фрейм. Преимуществом фреймов как модели представления знаний является способность отражать концептуальную основу организации памяти человека, а так же ее гибкость и наглядность. Следуя классическому подходу [5], будем различать фреймы-прототипы и фреймы-экземпляры. Фрейм-экземпляр – конкретная реализация фрейма, описывающая текущее состояние в предметной области. Фрейм-прототип – шаблон для описания объектов или допустимых ситуаций предметной области.

Фреймы системы объединены в единую Базу Знаний (БЗ), которая представлена в виде семантической сети. При организации базы знаний ИИС, фреймы были разделены на системные и пользовательские. Системные фреймы – это фреймы с неизменяемым набором слотов (характеристик), которые описывают внутреннюю организацию инструментальной системы, позволяют вести контроль версий системы,

контроль доступа к разрабатываемым приложениям, а так же обеспечивают хранение знаний о графическом представлении фреймов-данных. Пользовательские фреймы – это фреймы, создаваемые в процессе проектирования приложений. Такое разделение фреймов помогает защитить систему работы с фреймовыми структурами от несанкционированного доступа и уменьшить вероятность повреждения базы знаний, так как доступ к системным фреймам контролируется.

Основной проблемой при работе с большой базой знаний является проблема поиска знаний, релевантных решаемой задаче. В связи с тем, что в обрабатываемых данных может не содержаться явных указаний на значения, требуемые для их обработки, необходим более общий механизм доступа, чем метод прямого доступа (метод явных ссылок). Задача этого механизма состоит в том, чтобы по некоторому описанию сущности, имеющемуся в рабочей памяти, найти в базе знаний объекты, удовлетворяющие этому описанию. Очевидно, что упорядочение и структурирование знаний могут значительно ускорить процесс поиска.

Нахождение желаемых объектов в общем случае уместно рассматривать как двухэтапный процесс. На первом этапе, соответствующем процессу выбора по ассоциативным связям, совершается предварительный выбор в базе знаний потенциальных кандидатов на роль желаемых объектов. На втором этапе путем выполнения операции сопоставления потенциальных кандидатов с описаниями кандидатов осуществляется окончательный выбор искомых объектов.

В инструментальной системе осуществляется нисходящий поиск по дереву. Так как все имена фреймов в базе знаний уникальны, поиск осуществляется по имени.

Для того чтобы ускорить процесс поиска, в подсистеме хранится список ссылок на все прототипы. После того как найден нужный прототип, найти его экземпляр не составляет сложности, так как прототип содержит слот «подкласс». В этом слоте перечислены все наследуемые от него фреймы.

Фрейм, описывающий приложение, в слоте «агрегация» содержит ссылки на все фреймы-экземпляры, созданные в этом приложении. Таким образом, поиск фрейма-экземпляра упрощается, если известно, какому приложению принадлежит этот фрейм-экземпляр.

Система планирования и оптимизации развозочных маршрутов

Для проверки всех функций инструментальной системы был разработан контрольный пример, представляющий собой задачу расчета и оптимизации развозочных маршрутов с баз и складов снабжения и сбыта для перевозки мелкопартийных однородных грузов потребителям.

Для построения модели системы были рассмотрены следующие требования и ограничения: заданы пункты потребления X_i (i = 1, 2, ..., n) и начальный пункт X0 (база). На базе имеется однородный груз в количестве M единиц и транспортные средства грузоподъемности Q единиц. Потребности пунктов потребления в грузе составляют 1, 2 , ..., n единиц. Некоторые пункты X_i , X_j (i, j = 0, 1, 2, ..., n) соединены дорогой, стоимость проезда по которой имеющимся транспортным средством равна C_{ij} (независимо от наличия груза) [1].

При реализации системы расчета развозных маршрутов были разработаны и реализованы следующие пользовательские фреймы-прототипы (в скобках указаны слоты фрейма):

- пункт (название, назначение);
- дорога (расстояние);
- автомобиль (государственный номер автомобиля, грузоподъемность, скорость движения);
- заявка (пункт доставки, пункт отправки, количество груза);

− расписание по маршруту (автомобиль, город отправки, время загрузки, количество груза, маршрут);
− сотрудник (ФИО, должность);
− матрица смежности.

Целью ИИС является составление наборов маршрутов, обеспечивающих потребности всех заказчиков, для которых сумма длин маршрутов, входящих в набор, принимает наименьшее значение и затем составление оптимальных расписаний по маршрутам.

Алгоритм расчета оптимального порядка объезда пунктов состоит из нескольких этапов:

составление схемы размещения пунктов и расстояний между ними в соответствии с заявками на текущий день;

расчет минимального расстояния между пунктами по алгоритму Флойда [6];

расчет кратчайшей связующей сети, которая соединяет все пункты без замкнутых контуров. Расчет ведется по алгоритму составления минимального «остовного» дерева (алгоритм Прима [6]);

группировка пунктов на маршрут с учетом количества ввозимого груза и грузоподъемности единицы движимого состава и учетом времени работы смены сотрудников;

определение порядка объезда пунктов каждого маршрута с помощью матрицы кратчайшего расстояния от каждого до каждого пункта по алгоритму Дейкстры [6] и матрицы для расчета рационального порядка объезда пунктов маршрута;

расчет расписания по каждому маршруту с учетом времени разгрузки и загрузки единицы перевозимого груза и средней скорости движения транспорта.

Критерием оптимального маршрута в данной задаче является минимизация длины маршрута, а также уменьшение времени перевозки с учетом графика работы сотрудников. Результаты работы ИИС и созданной экспертной системы расчета маршрутов представлены в [7].

Выводы и рекомендации

Предложенная модель хранения знаний позволяет реализовать достаточно сложные информационные системы в относительно короткие сроки.

Интеллектуальная информационная среда применяется как для разработки приложений, так и для их исполнения.

Контрольный пример задачи расчета и оптимизации развозочных маршрутов однородного груза показал достаточно хорошие результаты по оптимизации и времени расчета маршрутов.

Таким образом, обобщённая концептуальная модель базы знаний представляет собой сеть фреймов типовой структуры и рекомендуется для виртуализации ресурсов хранения в любой сложной автоматизированной системе хранения.

Литература

1. Неруш, Ю.М. Логистика в схемах и таблицах: учебное пособие. [Текст]/Ю.М. Неруш. – М.: Проспект, 2008. – 192 с.

2. Михеева, Т.И. Использование паттерна проектирования «Decorator» в задачах построения маршрута перевозки опасного груза / Т.И. Михеева, С.В Михеев, И.А. Чугунов, И.А.Рудаков // Математическое моделирование информационных процессов и систем в науке, технике и образовании. – Самара, СГАСУ, 2010. – с. 85-88.

3. Дерябкин, В.П. База знаний системы синтеза и параметрической настройки проблемно-ориентированной информационной компьютерной среды / В.П. Дерябкин // Перспективные информационные технологии в научных исследованиях, проектировании и обучении «ПИТ-2006», Т.1. – Самара, СГАУ, 2006. – с. 65-69.

4. Дерябкин, В.П. Среда визуальной разработки интеллектуальных систем продукционно-фреймового типа / В.П. Дерябкин // Математическое моделирование информационных процессов и систем в науке, технике и образовании. – Самара, СГАСУ, 2010. – с. 48-51.

5. Джарратано, Д. Экспертные системы: принципы разработки и программирования, 4-е издание / Д. Джарратано, Г. Райли. – М.: ООО «И.Д. Вильямс», 2007. -1152 с.

6. Электронная энциклопедия Википедия [электронный ресурс] – http://en.wikipedia.org

7. Белоусов А.И. Подсистема работы с фреймовыми структурами инструментального комплекса разработки и развертывания интеллектуальных информационных систем/ А.И. Белоусов //Дипломный проект – Самара, СГАУ, 2010. – 189 с.

Методы зональной оптимизации координированного управления

С.В. Михеев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Увеличение количества автомобильного транспорта привело в крупных городах к значительным потерям времени в «автомобильных пробках». Для управления транспортными потоками используются светофоры с фиксированными интервалами сигналов, работающие в нескольких режимах: стационарный, пиковый, дежурный. Такие режимы не всегда подходят для управления на перекрестках сложной конфигурации и в условиях широкого диапазона изменения интенсивности движения транспорта. Один из возможных путей решения проблемы – применение новых методов управления транспортными потоками (ТП).

Основные понятия координированного управления

Разновидностью системного управления со сменой программ координации является *координированное управление*, составляющее основу стратегии и тактики управления дорожным движением. За последние годы проводилось широкое исследование проблемы координированного управления ТП [1, 2, 3]. Накопленный материал позволил установить некоторые закономерности изменения физических характеристик ТП. Схема управления должна быть пригодной для различных условий, но достаточно простой, чтобы гарантировать надежность.

Сущность координированного управления сводится к обеспечению безостановочного проезда группы автомобилей с определенной скоростью через все регулируемые перекрестки магистрали (рис. 1). Для этого сдвиг фаз регулирования в направлении координации выбирается, исходя из заданной скорости движения групп и расстояния между перекрестками.

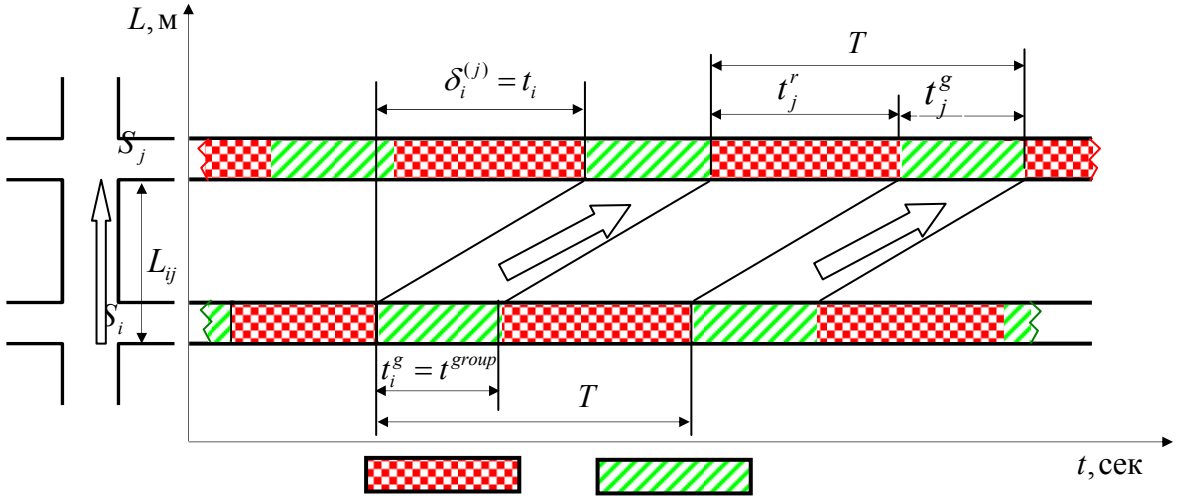


Рис. 1. Диаграмма «Время-путь» при координированном управлении

Сдвиг фаз $\delta_i^{(j)}$ - это интервал времени между включением основного такта на *i*-ом и смежном с ним *j*-ом перекрестках. Управляющими уставками для координированного управления, охватывающего ряд последовательно расположенных перекрестков, является общая для всех перекрестков длительность цикла управления T и совокупность длительностей фаз (t_i^g - зеленый сигнал, t_i^r - красный сигнал) и сдвигов

$\delta_i^{(j)}$, равных времени движения группы транспортных средств по перегону $\delta_i^{(j)} = t_i$. Совокупность управляющих уставок для перекрестков, входящих в район управления, носит название программы координации. Эффективность координированного управления ТП в большей мере определяется качеством расчета программ координации. При полной загрузке перекрестка $t_i^g = t^{group}$, время прохождения группы транспортных средств через перекресток соответствует времени эффективного зеленого сигнала светофора.

Анализ диаграммы (рис. 1) показывает, что транспортные средства пройдут участок магистрали без остановки, если отсутствует влияние потока с конфликтующего направления на перекрестке S_i, и длительность зеленого сигнала перекрестков S_i и S_j находятся в соотношении $t_j^g \geq t_i^g$ при неизменных и одинаковых величинах скорости каждого транспортного средства (ТС) в группе на протяжении участка. Перечисленные условия предполагают, что группа транспортных средств, покидающая перекресток S_i, пройдет перегон L_{ij} и перекресток S_j без изменения параметров. Данные условия выполнимы при $L_{ij} \leq 200$ м, когда распадение групп ТС невелико [2]. Для более сложных реальных условий - средняя и высокая интенсивность движения, $L_{ij} \leq 1000$ м и при появлении особенностей преобразования (распадения, формирования) групп ТС приходится ограничиваться получением локальных (удовлетворительных), но не оптимальных решений.

Методы оптимизации координированного управления

Методы оптимизации координированного управления ТП, т.е. нахождение алгоритма, минимизирующего показатель качества - целевую функцию, делятся на методы изолированного и взаимосвязанного управления на перекрестках. Изолированное управление перекрестками основывается на предположении о пуассоновском характере прибывающего ТП, влияние координации учитывается сдвигом фаз светофорной сигнализации на смежных перекрестках. При взаимосвязанном управлении каждый перегон и перекресток описываются уравнением преобразования ТП с учетом того, что исходящий поток одного перекрестка является входящим для смежных перекрестков. Оба метода могут быть реализованы как «жесткими» алгоритмами, не учитывающими кратковременные флуктуации потоков, так и адаптивными алгоритмами, которые «следят» за прибытием и убытием ТП.

Алгоритмы адаптивного координированного управления основаны на предварительном расчете программы координации контрольных значений параметров потоков, выборе их из «библиотеки» при появлении близкой к контрольной транспортной ситуации и общей коррекции, компенсирующей отклонение реальной ситуации от контрольной. При управлении через фиксированные программы координации определяются характерные периоды времени (утренний и вечерний «пики»), для которых программы рассчитываются заранее и включаются в определенный момент времени. Управление, при котором отсутствует «библиотека», программы координации рассчитываются в реальном масштабе времени за сравнительно небольшое время.

Метод ***зональной оптимизации управления*** (*MZoneTrafficFlowControl*) [1, 3], являясь промежуточным между системным (сетевым) и изолированным управлением, учитывает ситуацию на смежных перекрестках, откуда прибывают транспортные средства, с тем, чтобы обеспечить оптимальное качество управления.

Все известные алгоритмы сетевой оптимизации основаны на последовательной оптимизации управления перекрестками на модели сети. При локальной оптимизации

значительно возрастает гибкость управления при непредусмотренных изменениях условий движения и упрощается алгоритм управления: уравнение преобразования ТП решается по отдельности для каждой зоны.

Механизм формирования уставок $\delta_i^{(j)}$ (сдвигов фаз) координированного управления на регулируемых перекрестках улично-дорожной сети показан на рис. 2, 3, 4. При отсутствии координации параметров светофорного регулирования на смежных светофорных объектах транспортный поток может прерываться по различным сценариям. В каждом случае это приводит к своему типу распределения скорости движения и статистическим параметрам распределения.

При синхронном включении одноименных фаз на смежных перекрестках, т.е. при $\delta_i^{(j)} = 0$ (отсутствии координации) происходит разрыв транспортного потока на регулируемом перекрестке. Часть транспортных средств проходит перекресток без остановки, а часть задерживается. Одним из параметров светофорного регулирования в рассматриваемом методе расчета программ координации является величина уставок $\delta_i^{(j)}$. Выбор соответствующей величины уставки влияет на процесс преобразования групп ТС в зоне действия светофорного объекта. Расчет уставок, начиная с входного перекрестка магистрали [1], регламентирует величину сдвига фаз, которая должна быть меньше длительности цикла $\delta_i^{(j)} \le T$, что требует выполнения ограничения $\delta_i^{(j)} - T \cdot p > 0$ (p – натуральное число).

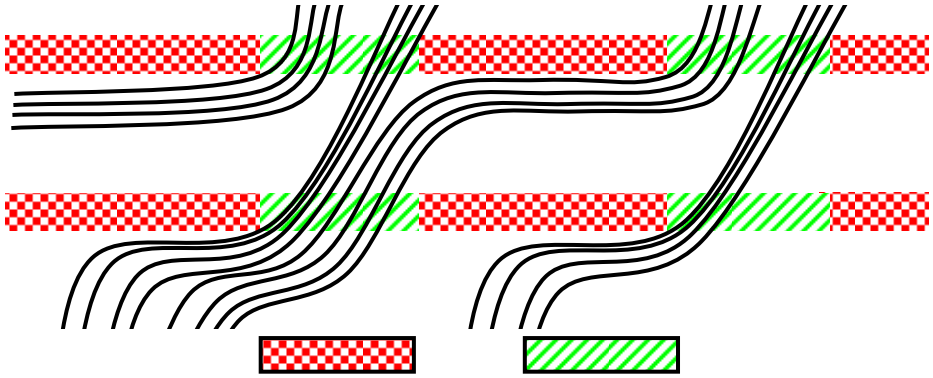


Рис. 2. Траектории движения автомобилей при синхронном включении разрешающего сигнала на смежных перекрестках

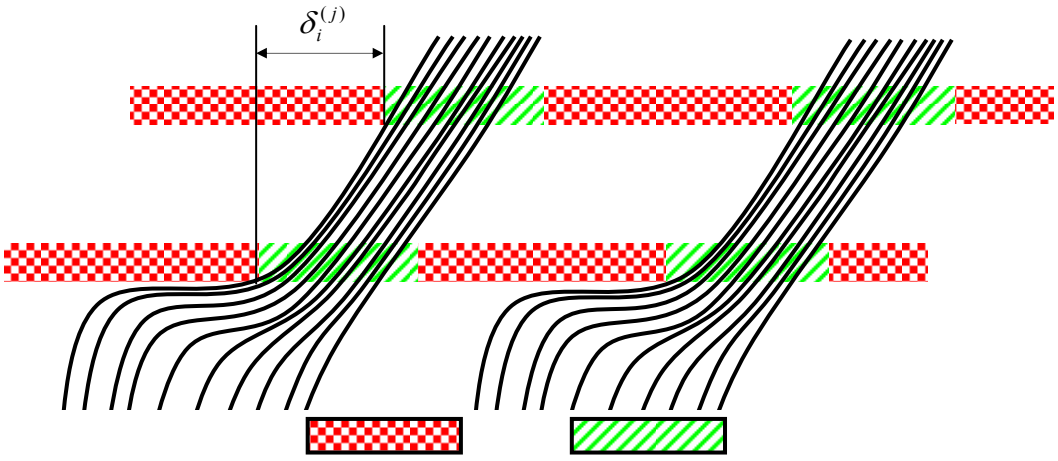


Рис. 3. Траектории движения автомобилей при координированном управлении

Пример. На рис. 4 представлена схема зоны управления в виде участка магистрали, состоящей из трех перекрестков и двух перегонов. Исходные данные: интенсивность движения, время проезда и расстояние между стоп-линиями. Схемы организации движения на всех перекрестках двухфазные, длительности разрешающих фаз (по основному направлению) – 40 с.

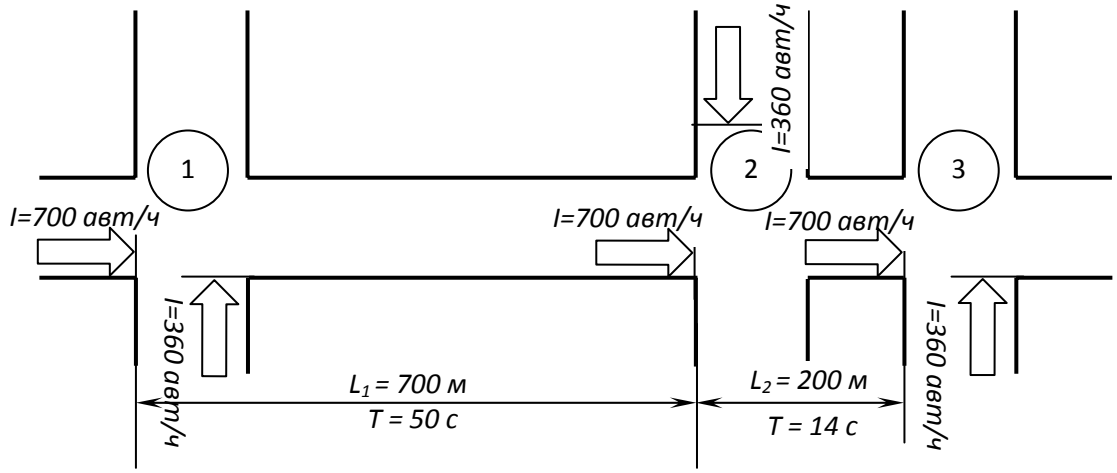


Рис. 4. Схема участка магистрали

Программа координации, рассчитанная с помощью описанного метода, при $T=60\ c$ дала среднюю задержку на участке за цикл D_j^Q=220 с и остановку групп ТС перед вторым перекрестком. Программа координации, рассчитанная с использованием зависимости (1) [1, 3] при $T=60\ c$ дала среднюю задержку за цикл D_j^Q=182 с.

$$D_j^Q = \frac{I_j (t_j^r)^2}{2} + Q_{j-1} t_j^r + \frac{(I_j - C)(t_j^g)^2}{2} + Q_j t_j^g , \qquad (1)$$

где Q_j - число ТС, ожидающих в очереди, I_j - интенсивность движения ТП, C - поток насыщения на перекрестке.

Рис. 5 иллюстрирует программу координации, построенную для участка магистрали, представленного на рис. 4.

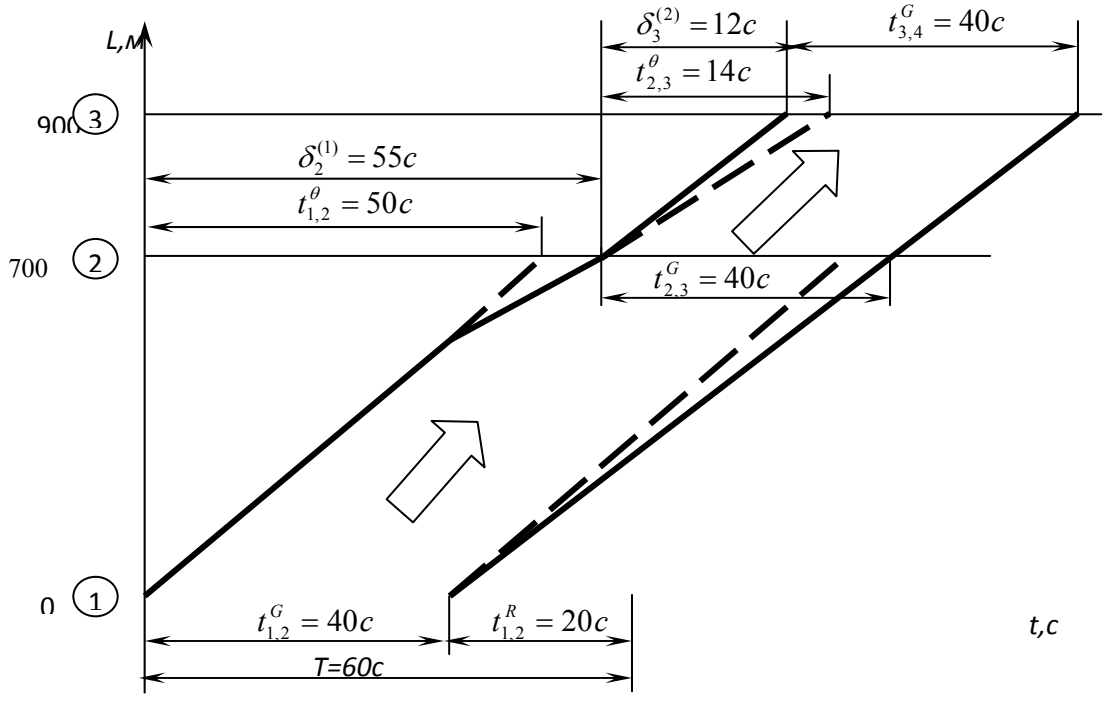


Рис. 5. Программа координации

Выбор уставки $\delta_2^{(1)} = 55$ большей, чем время проезда $t_{1,2}^{\theta}$ перегона между первым и вторым перекрестками на величину $\Delta t = 5$ с обеспечивает формирование группы ТС за счет замедления лидеров. При движении лидера по траектории, показанной пунктиром, и соответствующем выборе уставки группа имела бы задержку с остановкой.

Выбор уставки $\delta_3^{(2)}$ между 2-м и 3-м перекрестками меньшей, чем время проезда перегона между указанными перекрестками, позволяет пропускать группу, проходящую короткий перегон, без дополнительной задержки.

Метод зонального координированного управления *MZoneTrafficFlowControl* поддерживает алгоритмы, обеспечивающие функционал, по меньшей мере, трех режимов [1]:

Режим априорной оптимизации - жесткого регулирования: алгоритм использует информацию, накопленную на основе усреднения данных, изменяемых за большие промежутки времени. Выходные величины (длительность и сдвиги фаз) имеют постоянные для выбранного промежутка времени значения и меняются не чаще чем 3-5 раз в сутки. Малая гибкость данного алгоритма позволяет использовать его на перекрестках с малой интенсивностью движения.

Режим адаптивной оптимизации - алгоритм ориентирован на использование минимального объема предварительных сведений. Алгоритмы зонального управления способны, в принципе, работать без априорных данных о характеристиках ТП. В этом случае необходимо задавать начальное значение уставок светофоров или допустить период обучения, в течение которого определяются параметры ТП, а регулирование ведется по жесткой программе.

Режим on-line оптимизации – алгоритм, в котором исходные уставки определяются в режиме реального времени на основе данных, полученных от детекторов транспорта, в качестве которых могут вы ступать как стационарно установленные на УДС детекторы, так и *GPS*-навигаторы. Такой режим представляется наиболее целесообразным на современной стадии исследований, когда интенсивность ТП возросла, развитые информационные технологии вкупе с современной микропроцессорной техникой позволяют адекватно решать задачи управления в реальном времени.

Литература

1. Михеева, Т.И. Структурно-параметрический синтез интеллектуальных транспортных систем / Самара: Самар. науч. центр РАН, 2008. 380 с.

2. Клинковштейн Г.И. Организация дорожного движения: Учеб. для вузов. – М. : Транспорт, 2001. 247 с.

3. Mikheev, S.V. Model of zone control of traffic flows / Computer Science and Information Technologies. Mikheeva T.I., Mikheev S.V. //Proceedings of the 10th International Workshop. V. 2. Antalya, Turkey, 2008. P. 121-128.

Синтез интегрированных систем мониторинга и управления транспортной инфраструктурой

Т.И. Михеева, О.М. Батищева
Самарский государственный аэрокосмический университет, Самара
Самарский государственный технический университет, Самара

Введение

Неотъемлемый рост объема транспортных услуг в современном обществе, необходимость повышения надежности, безопасности и качества актуализирует потребность в увеличении затрат на развитие транспортной инфраструктуры. Однако с ростом автомобильного парка проявляется ряд существенных проблем: рост количества дорожно-транспортных происшествий (ДТП), токсичные выбросы, шум, низкие скорости движения, заторы в часы «пик», большие потери времени для участников движения, перепробеги, высокий расход топлива, увеличение суммарных эксплуатационных затрат на автомобильные перевозки и др. В первую очередь эти проблемы проявляются в местах концентрации транспортных потоков на участках сети, функционирующих в режимах, близких к пропускной способности. Как правило, это крупные города – мегаполисы с высоким уровнем автомобилизации. Учитывая стохастическую природу дорожного движения, динамичность изменения его характеристик во времени, большую сложность представляет процесс качественной оценки и прогнозирования изменений дорожно-транспортных ситуаций на сети автомобильных дорог.

Качественный скачок в разработке и выпуске мощных информационных компьютерных систем, в развитии современных видов связи, эффективных навигационных систем, технических средств сбора и обработки информации о характеристиках транспортных потоков и дорожной сети диктует необходимость в активном использовании технологий интеллектуальных транспортных систем (ИТС), как систем наблюдения и управления транспортной инфраструктурой.

Термин «интеллектуальные транспортные системы» характеризует комплекс интегрированных средств управления и мониторинга объектов транспортной инфраструктурой, перевозками, применяемых для решения всех видов транспортных задач на основе современных технологий, методов моделирования транспортных процессов, программного обеспечения, организации информационных потоков в реальном масштабе времени [1]. Технологии ИТС имеют в настоящее время около 50 подсистем различных направлений применения, но при реализации своих функций в отдельности не могут быть в полной мере использованы потенциальные возможности каждой подсистемы как части системы [2, 3]. Максимальный эффект от их использования может быть получен только при общей интеграции отдельных подсистем в единую систему.

Основой для создания ИТС являются существующие автоматизированные системы управления дорожным движением, системы управления движением маршрутного транспорта, автоматизированные системы обнаружения дорожно-транспортных происшествий, системы маршрутной навигации, информационные системы управления дорожной сетью и другие подсистемы управления дорожным движением и перевозками [1].

Указанные системы и децентрализованные информационные базы данных объединяются для функционирования в общей ИТС по следующим структурным направлениям:

− объединение по функциональным признакам использования информационных потоков для выполнения различных функций в транспортных

системах (управление движением и перевозками, информационное сопровождение водителей, пассажиров и грузов, выявление аварийных ситуаций, электронная оплата проезда и парковки и т.д.);

− объединение по институциональным признакам использования информации различными организационными структурами (органы государственной власти, органы местного самоуправления, частные фирмы и т.д.);

− интеграция баз данных от множества источников получения и обработки информации о транспортных процессах;

− интеграция во времени для отражения характеристик транспортных процессов, моделирования и анализа ситуации в реальном режиме времени.

Современные тенденции развития ИТС показывают, что одной из основных целей их функционирования является предоставление мультимодальной информации не только для управленческих структур, но и персонально участникам движения.

Системный анализ при проектировании и синтезе интеллектуальной транспортной системы

Анализ и синтез сложной системы, к которым, несомненно, относится транспортная система региона, подразумевает использование системного подхода. Интеллектуальная транспортная система, как сложноорганизованная система, характеризуется следующими особенностями [4]:

- наличие большого числа взаимосвязанных между собой элементов;
- многомерность и иерархичность системы, обусловленная большим числом связей между элементами;
- целостность (эмерджентность) системы;
- многокритериальность, обусловливаемая имманентностью (несовпадением) целей отдельных элементов системы;
- многофункциональность элементов системы;
- управляемость;
- сложность информационных процессов.

С позиций структурно-функционального подхода ИТС можно определить как единство структуры, функций и целостности [1]. Структура характеризует элементы ИТС и их взаимодействие. Функции определяют природу связей между элементами и поведением ИТС. Целостность выражает взаимозависимость структуры и функций ИТС и проявляется в наличии у реальной системы таких свойств, которые не присущи отдельным ее элементам и не выводимы из свойств этих элементов и способов их соединения.

ИТС г. Самара представляет собой сложноорганизованную систему, характеризующуюся интегральным взаимодействием таких факторов как:

− ИТС является сложной динамической иерархической и стохастической системой, состоящей из многочисленных взаимодействующих и взаимосвязанных распределенных компонентов;

− ИТС представляет собой синергию транспортных, материальных, информационных, финансовых и др. потоков и процессов, образующих адаптивную систему, включающую объект и субъект управления;

− при формировании транспортной инфраструктуры региона должна использоваться интегральная парадигма логистики, реализующая общую стратегическую или оперативную цель функционирования компонентов системы при оптимальном использовании в системе транспортных, материальных, финансовых, информационных и трудовых ресурсов и согласовании локальных критериев функционирования компонентов ИТС с глобальной целью оптимизации. Целевая функция оптимизации при этом является, как правило, многокритериальной;

− управление ИТС региона не может быть полностью формализовано (а, следовательно, алгоритмизировано), что вызывает необходимость построения комплекса формализованных моделей и неформальных (эвристических) процедур и представлений;

− информационно-компьютерная поддержка функционирования ИТС должна охватывать как можно большее количество объектов и процессов управления.

Логистический подход требует развития соответствующих методик и математического аппарата, основанного на методологии системного анализа. При этом актуальными становятся постановка и решение задач в области структурного синтеза сложных систем со слабо формализуемыми внутренними связями. Синергетический эффект при проектировании и синтезе ИТС проявляется в форме организационно обусловленного перехода от имманентности к синергии за счет расширенной системной и функциональной интеграции:

− постановка проблем организации движения и перевозок;
− функциональные требования пользователей и государства;
− разработка концепции функционирования ИТС;
− определение функциональных возможностей ИТС;
− транспортная, экономическая, информационная логистика;
− разработка управленческих решений по интеграции ИТС;
− функциональная, институциональная, временная интеграция, интеграция баз данных;
− развитие подсистем в каждой функциональной группе;
− интеграция информационных потоков между подсистемами;
− реализация конкретных функций ИТС.

ИТС является синтезом субъектов и объектов логистического управления, экономически и функционально обособленных, со своими организационно-функциональными структурами и локальными критериями оптимизации функционирования, которые в общем случае могут не совпадать с глобальной целью функционирования ИТС.

Интеллектуальная транспортная система «ITSGIS»

Функциональная схема ИТС представлена на рисунке. В оперативно-дежурную часть города поступает информация от различных юридических и физических лиц, органов исполнительной власти, дорожно-патрульной службы, оперативных служб города: «01», «02», «03», «ГО и ЧС» и др. Данные записываются в соответствующие базы данных и обрабатываются по мере поступления запросов к ним.

Комплекс программных средств включает в себя следующие компоненты:

− супервизор – сервер приложений;
− оцифрованная карта города на базе геоинформационной системы;
− единое хранилище семантических данных;
− система обработки данных;
− система генерации отчетов.

Сервер приложений – это многокомпонентная система, реализованная как набор COM-объектов. Основные COM-объекты бизнес-уровня системы:

Административный модуль, отвечающий за настройку конфигурации системы. Осуществляет управление системой авторизации (добавление, удаление «пользователя», изменения уровня доступа), обеспечивает разграничение прав доступа \ изменения критических данных, служебно-технической информации.

Компоненты, отвечающие за функционирование всей системы. Приложения-клиенты имеют доступ только к интерфейсу супервизора, т.е. прямой доступ к базам данных и геоинформационной системе (ГИС) невозможен. Сам супервизор также не

имеет прямого доступа к уровню хранения данных, взаимодействуя с СУБД\ГИС посредством промежуточных интерфейсов (IDB\IMAP, соответственно). Такая организация системы позволяет полноценно использовать преимущество многоуровневой архитектуры ИТС.

Компонент, предоставляющий доступ к СУБД через MS ADO.

Компонент, предоставляющий доступ к ГИС. Доступ от этого компонента к ГИС осуществляется через DCOM\COM интерфейс.



Рис. Функциональная схема ИТС

Географическая информационная система – одна из подсистем ИТС, обеспечивающая сбор, хранение, обработку, доступ, отображение и распространение пространственно-координированных данных.

Основу информационного обеспечения подсистем ИТС составляет совокупность данных, необходимых для выполнения процесса проектирования. Реляционная модель системы затрагивает те объекты, состояние которых требуется фиксировать в *базе данных*. Это картографические данные о геообъектах, их семантические атрибуты, справочники и некоторые дополнительные данные, необходимые для функционирования системы [5].

Система обработки данных включает в себя ряд подсистем, решающих задачи обработки данных, экспертизы и моделирования.

Подсистема сбора, хранения, редактирования и обработки информации, касающейся улично-дорожной сети (УДС) и объектов на ней, содержит программные модули: «Паспортизация дороги», «Реестр дорожных знаков», «Реестр светофорных объектов» [5], «Реестр железнодорожных переездов» и др.

Подсистема сбора, хранения, редактирования и обработки оперативной информации: «Мониторинг оперативной информации», «Контроль за состоянием УДС», «Учет и анализ ДТП» и др.

Экспертные системы дислокации технических средств организации дорожного движения на УДС города, дислокации постов дорожно-патрульной службы, проектирования структуры светофорного цикла, прогнозирования интенсивности движения транспортных потоков.

Системы моделирования зонального координированного управления транспортным потоком; оптимального распределения транспортных потоков на улично-дорожной сети города и др.

Комплекс технических средств включает в себя специализированные дорожные лаборатории, средства организации дорожного движения. Передвижные дорожные лаборатории осуществляют сбор и обработку информации о состоянии УДС, характеристиках транспортных потоков, параметрах ДТП и др. Технические средства организации дорожного движения включают в себя дорожные знаки и светофорные объекты, расположенные на УДС города [5]. Через комплекс технических средств осуществляется управление транспортными потоками, их оптимальное распределение на УДС, обеспеченное соответствующими модулями комплекса программных средств ИТС.

На основе анализа информации отделом ГИБДД УВД города выдаются указания, предписания и рекомендации соответствующим службам: водоканалу, муниципальным предприятиям «Самарагорсвет», «Самарские тепловые сети», «Трамвайно-троллейбусное управление», «Управление зеленого хозяйства» и др., т.е. службам, отвечающим за содержание, ремонт и реконструкцию УДС. После анализа отдаются распоряжения штабу дорожно-патрульной службы на расстановку постов в местах концентрации ДТП, другим службам по проведению мероприятий для предотвращения аварийных ситуаций в будущем, проведению исследований. По запросам формируются отчеты для вышестоящих организаций и средств массовой информации.

Выводы

ИТС является открытой системой, расширяемость ее достигнута за счет использования компонентно-ориентированного подхода к проектированию систем. Данный подход предполагает разбиение приложения на ряд независимых модулей-компонентов, а также добавление компонентов, написанных другими разработчиками без их перекомпиляции. В настоящий момент отдельные модули ИТС внедрены и используются в работе различных организаций г. Самара и Самарской области.

Литература

1. Михеева, Т.И. Структурно-параметрический синтез интеллектуальных транспортных систем / Самара: Самар. науч. центр РАН, 2008. 380 с.

2. Буслаев, А.П. Вероятностные и имитационные подходы к оптимизации автодорожного движения / Буслаев А.П., Новиков А.В., Приходько В.М. [и др.] // Под ред. В.М. Приходько — М. : Мир, 2003. 368 с.

3. Гвоздев, В.Е., Проблемы информационного обеспечения систем анализа состояния природно-технических объектов на основе математико-геоинформационного моделирования. Гвоздев В.Е., Ильясов Б.Г., Павлов С.В., Ямалов И.У. / Проблемы управления и моделирования в сложных системах.– Самара: ИПУСС РАН, 2000. С. 323-326.

4. Михеева, Т.И. Модели для синтеза интеллектуальных транспортных систем. Михеева Т.И., Михеев С.В. / Труды Всероссийского семинара по вопросам организации дорожного движения. М.: ФГУ «Дирекция Программы повышения безопасности дорожного движения». 2008. С. 69-77.

5. Михеева, Т.И. Геоинформационная система дислокации светофорных объектов Михеева Т.И., Михеев С.В. Рудаков И.А., Чугунов И.А. //Математическое моделирование информационных процессов и систем в науке, технике и образовании: Межвузовский сборник научных трудов. Самарск. гос. арх.-строит. ун-т. – Самара:, 2010. С. 89-92.

ГИС в составе информационной системы паспортизации объектов транспортной инфраструктуры

Т.И. Михеева, Н.В. Гаманова
Самарский государственный аэрокосмический университет, Самара
Самарский государственный технический университет, Самара

Введение

Для улучшения ситуации на улично-дорожной сети города существует два основных решения. Первое решение заключается в изменении структуры улично-дорожной сети: создание новых участков сети, расширение уже существующих дорог, введение дополнительных многоуровневых транспортных развязок. Хотя данное решение приводит к глобальному улучшению ситуации, оно применяется крайне редко потому, что требует значительных материальных затрат и занимает длительное время, в течение которого изменяемый участок улично-дорожной сети (УДС) будет перекрыт для движения. Второй путь решения проблемы – оптимизация дорожного движения. Данное решение проблемы не приводит к глобальному улучшению ситуации на УДС города, но является более доступным и применяется значительно чаще. Оптимизация дорожного движения может производиться по различным критериям: безопасность движения транспортных средств, увеличение пропускной способности улично-дорожной сети, уменьшение временной задержки транспортных средств на перекрестке и др.

При реализации мероприятий, направленных на улучшение транспортной ситуации на УДС, особую роль отводят корректной и оптимальной (необходимой и достаточной) дислокации управляющих объектов на улично-дорожной сети. К таким объектам относятся технические средства организации дорожного движения (ТСОДД): дорожные знаки, средства светофорного регулирования, дорожная разметка, дорожные ограждения. Правила установки каждого типа объекта регламентируются ГОСТом 52289-2004 «Технические средства организации дорожного движения. Правила применения».

Любой объект ТСОДД, устанавливаемый на УДС, оказывает управляющее воздействие на транспортный поток, которое должно быть скоординировано с воздействием, оказываемым уже установленными объектами. Анализ взаимного влияния объектов ТСОДД и проверка корректности дислокации с точки зрения ГОСТ является сложной многокритериальной задачей. В большинстве случаев человеку, проводящему данный анализ трудно выявить все факторы влияния объектов друг на друга. В этой связи становится актуальной задача разработки методов и программных средств, позволяющих автоматизировать комплексный анализ дислокации ТСОДД на УДС города с учетом различных критериев оптимальности организации дорожного движения и ГОСТов по установке и эксплуатации данных объектов.

Информационная система паспортизации объектов транспортной инфраструктуры

Интеллектуальные транспортные системы представляют собой комплекс интегрированных средств управления транспортной инфраструктурой (улично-дорожной сетью, техническими средствами организации дорожного движения, транспортными потоками), применяемых для решения задач организации дорожного движения, на основе современных информационных технологий, организации информационных потоков о функционировании транспортной инфраструктуры в реальном режиме времени. Многоуровневая, сложноорганизованная интеллектуальная транспортная система (ИТС) представляет собой гибридную систему, состоящую из

множества разнородных систем, сложным образом взаимодействующих друг с другом – управляющих, классифицирующих, прогнозирующих, экспертных, принимающих решения или поддерживающих эти процессы, объединенных для достижения единой цели. Одной из таких систем, предназначенных для учета и мониторинга разнообразных объектов, является информационная система паспортизации объектов транспортной инфраструктуры. Кроме того, в состав ИТС входит геоинформационная система (ГИС) ITSGIS.

Геоинформационная система ITSGIS

Визуализация процесса установки ТСОДД на карту в ГИС позволяет пользователю системы наглядно оценивать результаты работ.

ГИС ITSGIS позволяет отображать различные данные, имеющие пространственную привязку. В ITSGIS заложены следующие возможности:

- создание, отображение точечных, линейных, полигональных объектов;
- связь географических объектов с атрибутивной информацией;
- поиск географических объектов;
- методы тематической закраски карт;
- методы создания и редактирования легенд;
- поддержка наборов форматов данных;
- доступ к удаленным базам данных (БД) и распределенная обработка данных.

ITSGIS позволяет получать информацию о местоположении по адресу или имени, находить пересечения улиц, границ, производить автоматическое и интерактивное географическое кодирование, наносить на карту объекты из базы данных. Форма представления информации в ГИС может иметь вид таблиц карт, диаграмм, текстовых справок. ITSGIS дает возможность проводить специальный географический анализ и графическое редактирование. В дополнение к традиционным для систем управления базами данных (СУБД) функциям ITSGIS позволяет собирать, хранить, отображать, редактировать и обрабатывать картографические данные, хранящиеся в базе данных, с учетом пространственных отношений объектов.

Встроенный язык запросов SQL, благодаря географическому расширению, позволяет организовывать выборки с учетом пространственных отношений объектов, таких как удаленность, вложенность, перекрытия, пересечения, площади объектов и т.п. Запросы к базе данных можно сохранять в виде шаблонов для дальнейшего использования.

Дислокация дорожных знаков в ITSGIS

Одним из ключевых моментов в ИТС является отображение установленных объектов на электронной карте.

При реализации системы для каждого вида объектов в геоинформационной системе создается собственный слой, который содержит только эти виды объектов или связанные с ними. Например, в системе реализован слой «Дорожный знак». Пример этого слоя приведен на рисунке 1.

Следующей задачей при реализации ИТС является отображение объектов системы (дорожных знаков, светофоров, разметки и др.) на созданных слоях. Геоинформационная система ITSGIS содержит векторную масштабируемую карту, поэтому при отображении объектов системы, необходимо, чтобы они тоже могли подвергаться масштабированию. При отображении простых объектов, например столбов, групп дорожных знаков, опор выбраны простые геометрические фигуры: окружности, многоугольники, линии. Сложные объекты, такие как дорожный знак или светофор, невозможно обозначить таким образом, поскольку они имеют более

сложную структуру. Для их обозначения реализованы специальные шрифты, для каждого вида объектов реализован свой шрифт. Сложные объекты устанавливаются на электронную карту города в виде символов специального шрифта.



Рис. 1. Слой с дислокацией дорожных знаков

Автоматизированная система дислокации дорожных объектов может функционировать в двух режимах:

• режим просмотра информации;

• режим полного доступа к системе.

Режим работы пользователя зависит от типа пользователя, который задается в момент «создания» нового пользователя в системе и определяется в момент авторизации, после ввода имени и пароля. Данная информация сохраняется в сессии пользователя и впоследствии используется при работе пользователя в системе.

В режиме просмотра информации пользователь может только получать информацию об уже установленных объектах, но не имеет возможность редактировать или вносить новую информацию в систему.

В режиме полного доступа пользователь может просматривать, изменять и вносить новую информацию.

В случае если необходимо установить новый дорожный объект на участок УДС, то в системе производится анализ допустимости дислокации нового объекта на выбранный пользователем участок УДС. Если анализ выдал положительный результат, то производится сохранение информации о новом объекте и его отображение на электронной карте города. Если система приняла решение о недопустимости дислокации объектов, то пользователю выдается отчет о причине принятия такого решения.

Рассмотрим алгоритм добавления в систему нового объекта на примере добавления нового дорожного знака на УДС города.

Шаг 1. Пользователь, используя соответствующие инструменты перемещения карты, определяет место дислокации нового дорожного знака — участок улично-дорожной сети города. Пример карты с проезжими частями улично-дорожной сети, жилых кварталов, домов приведен на рисунке 2.

Шаг 2. При помощи соответствующего пункта меню системы пользователь выбирает вид объекта, который он хочет установить. В системе предусмотрена возможность дислокации различных объектов транспортной инфраструктуры города: точечных, линейных, полигональных. Объекты типа «Дорожный знак» или «Светофор» относятся к типу точечных объектов. На рисунке 3 приведен пример дислокации дорожного знака.

Шаг 3. При помощи курсора мышки пользователь определяет точное место дислокации дорожного знака на карте и место привязки данного дорожного знака к дуге графа УДС города.

Шаг 4. Из справочника дорожных знаков пользователь выбирает дорожный знак, который он хочет установить (рис. 4).

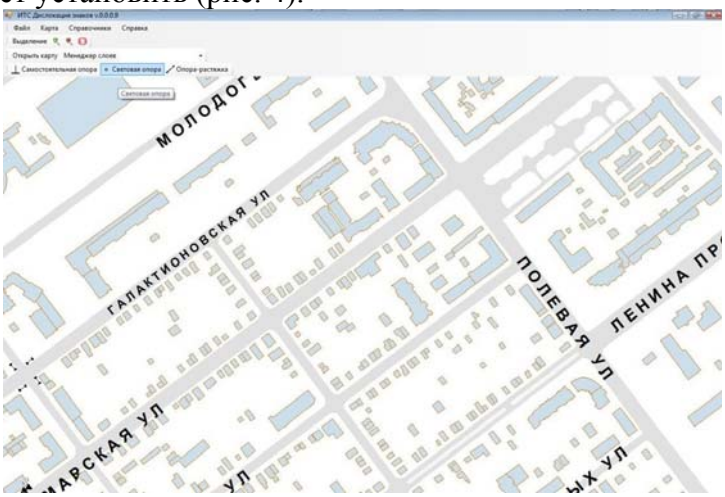


Рис. 2. Выбор участка УДС

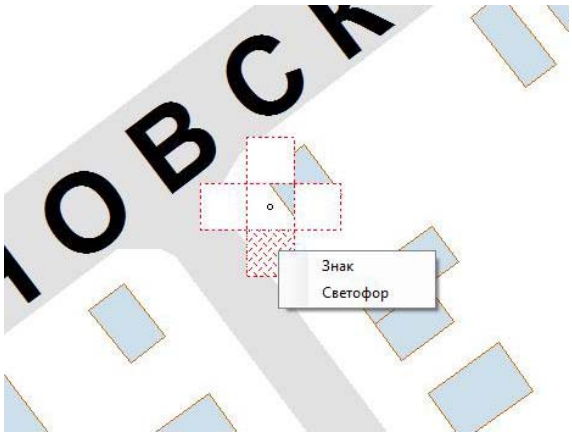


Рис. 3. Дислокация нового дорожного знака

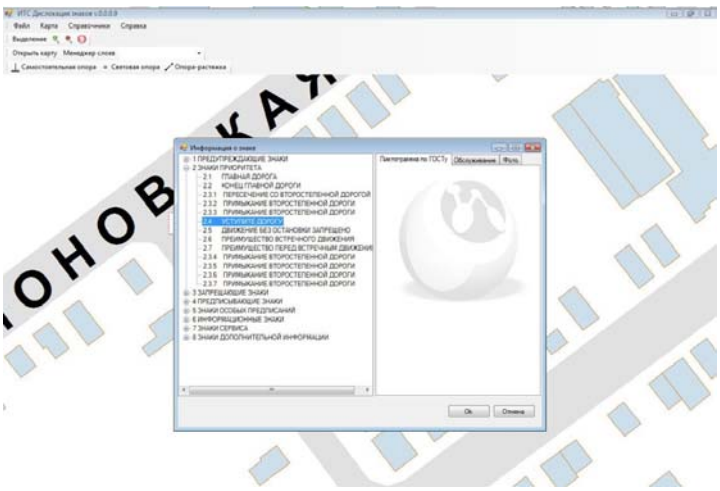


Рис. 4. Выбор дорожного знака

Шаг 5. Пользователь заполняет информацию о дорожном знаке: район и улицу, где установлен знак; сторона улицы; ориентир; качество знака. Таким образом, на

каждый дорожный объект (в примере – дорожный знак) создается «Карточка объекта», которую в дальнейшем при необходимости можно отредактировать, добавить новую информацию и т.д. Пример заполнения информации о знаке приведен на рисунке 5.

Шаг 6. На каждый объект транспортной инфраструктуры в карточку объекта могут быть добавлены фотографии. В окне информации о знаке пользователь загружает фотографию существующего знака (рис 6).

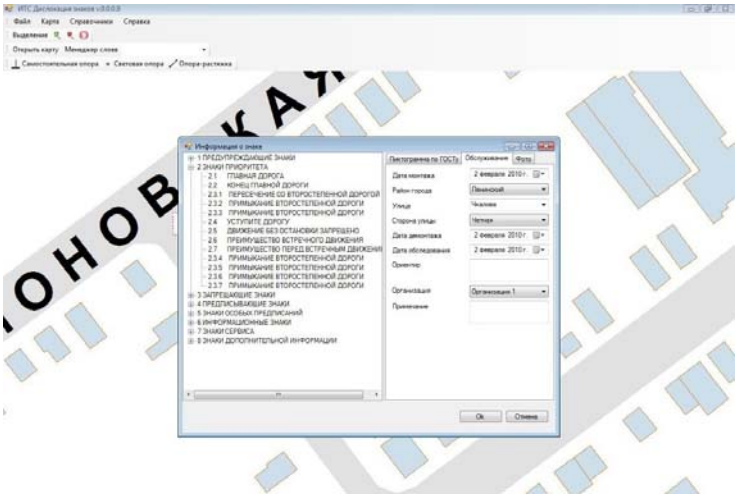


Рис. 5. Заполнение информации о дорожном знаке

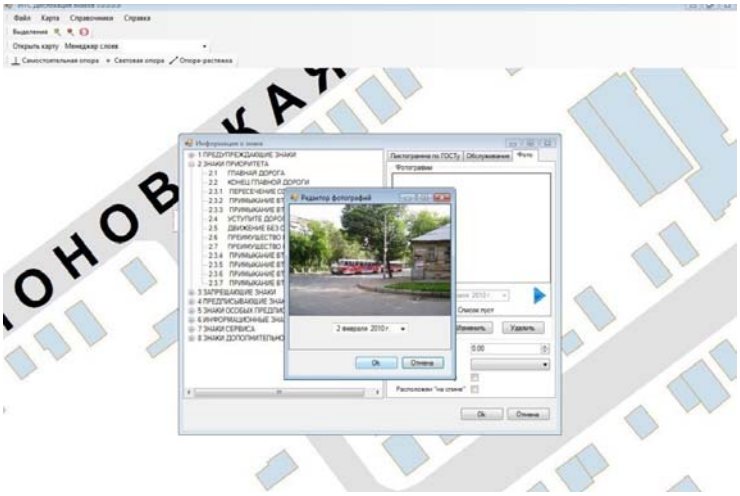


Рис. 6. Добавление фотографии знака

Шаг 7. Информация о дорожном знаке, участке на которой происходит установка знака и координат на дуге передается в сервер приложения.

Шаг 8. Запускается алгоритм дислокации объекта, который сохраняет введенные данные в системе хранения информации.

В настоящее время в ИТС осуществляется дислокация дорожных знаков города Самара.

Литература

1. Михеева Т.И. Построение математических моделей объектов улично-дорожной сети города с использованием геоинформационных технологий // Информационные технологии. 2006. №1. С. 69–75.

2. Михеева Т.И. Интеллектуальная транспортная система. Дислокация дорожных знаков // Вестник Самарского гос. техн. ун-та. Серия «Технические науки» №32. - Самара:СамГТУ, 2005. С.53-63.

Система автоматизированного управления интерактивным контентом

Михеева Т.И., Головнин О.К.
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

«Информационный киоск» – автоматизированный программно-аппаратный комплекс, предназначенный для предоставления различного вида справочной информации. Информационные киоски собирают на базе персонального компьютера, оснащенного сенсорным монитором.

Одним из перспективных направлений использования информационных киосков является отображение картографической информации, например, схем крупных бизнес-центров (развлекательных, торговых, офисных). Такие схемы облегчают поиск требуемого помещения, маршрута движения к нему.

С целью управления информацией, представляемой в информационных киосках, создана система автоматизированного управления интерактивным контентом. Эта система позволяет создавать интерактивные схемы бизнес-центров, управлять отображением схем на экране информационных киосков с использованием шаблонов оформления.

Система содержит в своем составе геоинформационную систему ITSGIS, позволяющую создавать, редактировать, удалять картографические объекты различных видов: точки, линии, полигоны. Картографические объекты располагаются на разных тематических слоях. Так, например, для крупных бизнес-центров офисные помещения (полигональные объекты) хранятся в специальном слое «Офисы», топология электрических и компьютерных сетей (полилинии), дислокация пожарных гидрантов (точки) располагаются в соответствующих слоях электронной карты-схемы. В системе предусмотрена возможность управлять слоями электронной карты: создавать новые слои, удалять, редактировать, перемещать, делать видимыми и т.д.

Кроме того, система дополнена редактором рекламы, позволяющим управлять отображением рекламы.

Для представления информации в информационном киоске создана подсистема отображения, в функции которой входит не только представление картографической информации, но и обеспечение взаимодействия с пользователем. Система предоставляет пользователю удобные средства поиска требуемых помещений, средства масштабирования и перемещения по схеме бизнес-центра, просмотра информации о выбранных помещениях, просмотра рекламных акций и многое другое. В настоящее время ведется разработка алгоритмов, отображающих маршруты к требуемым помещениям.

Подсистема отображения информации использует технологию WPF, которая обеспечивает широкие возможности по работе с графикой. Кроме того, эта технология позволяет строить интерфейсы в графической среде Expression Blend, разделяя работы программиста и дизайнера.

Система написана на языке C# для платформы .NET, которая является на данный момент наиболее развивающейся и перспективной.

Первая версия АСУ интерактивным контентом введена в эксплуатацию в гостинице «Моя». За время эксплуатации получены положительные отзывы и конструктивные предложения по расширению функциональных возможностей.

Методы оценки уровня безопасности транспортной инфраструктуры

Т.И. Михеева, А.В. Золотовицкий
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

На современном этапе развития транспортной инфраструктуры большое внимание уделяется решению частных задач оптимизации транспортных процессов. Однако создание транспортно-логистической системы на уровне региона требует комплексного рассмотрения всех аспектов указанных процессов. Приоритетной становится задача построения оптимальных и безопасных маршрутов перевозки на территории России для обработки транзитных и внешнеторговых грузопотоков [1]. Как показывает практика, оптимальность маршрутов может обеспечиваться не только сокращением сроков перевозок или транспортных затрат, но и безопасностью маршрута [2].

Транспортную систему можно рассматривать в виде совокупности двух множеств элементов: «узлы» системы и «потоки» между узлами. Связями в данной трактовке системы будут отношения инциденции узлов и потоков. Такое представление позволяет применить концепции и методы теории графов, как к построению моделей транспортной системы и составляющих ее элементов (транспортная сеть, поток), так и к решению задач маршрутизации [3].

Наличие большого количества ограничений, сложность системы и, как следствие, большая размерность решаемых задач создают значительные трудности при построении оптимальных маршрутов традиционными методами. Применение интеллектуальных технологий, разработка специализированных структур и алгоритмов, позволяющих наилучшим образом решать задачи оптимизации транспортных потоков и обеспечения безопасности, значительно облегчит этот процесс и обеспечит переход на новый, качественно более высокий уровень проектирования и управления транспортными системами.

Метод конфликтных точек

Метод оценки уровня безопасности движения на пересечениях, примыканиях и транспортных узлах связан с оценкой транспортных потоков (ТП) по взаимному контакту – конфликтным точкам (рис. 1). При помощи таких контактов оценивается сложность транспортного узла, перекрестка: $M = \sum_{i=1}^{n} k_{di} \cdot m_i$, где M -показатель сложности транспортного узла; i - вид конфликта; k_{di} - коэффициент сложности i-го конфликта; m_i - число конфликтных точек i-го вида.

Определено три вида конфликтов: пересечение, слияние, ответвление. Транспортный узел, у которого численное значение M больше, считается более опасным. Метод конфликтных точек, учитывающий контакты между транспортным средством (ТС) и пешеходом, угол пересечения между конфликтующими направлениями, заключается в расчете показателя сложности транспортного узла в виде суммы произведений суммарных интенсивностей движения I_i в каждой конфликтной точке на соответствующие коэффициенты k_{di} их тяжести, причем конфликтная точка определяется как точка слияния, ответвления, пересечения траекторий ТС $M = \frac{\sum_{i=1}^{n} I_i \cdot k_{di}}{10^7}$, где n - число конфликтных точек.

Для расчета степени опасности пересечения применяется метод вероятностного определения количества ДТП на конфликтных точках. Каждая из конфликтных точек на пересечении представляет тем большую опасность, чем больше интенсивности движения пересекающихся на этой точке ТП. Вероятное количество ДТП на *i*-й конфликтной точке на 10 млн. проехавших автомобилей z_i определяется следующим образом: $z_i = \frac{25 \cdot k_i \cdot I_i^{'} \cdot I_i}{k_y \cdot 10^7}$, где k_i - относительная аварийность данной конфликтной точки**.** ДТП/10 млн. авт.*;* $I_i^{'}$, I_i - интенсивности пересекающихся в данной конфликтной точке транспортных потоков, авт/сут; k_y- коэффициент годовой неравномерности движения по месяцам. Коэффициент 25 вводится в формулу для учета среднего количества рабочих дней в месяц, в течение которых загрузка дорог резко превышает загрузку в выходные дни [2].

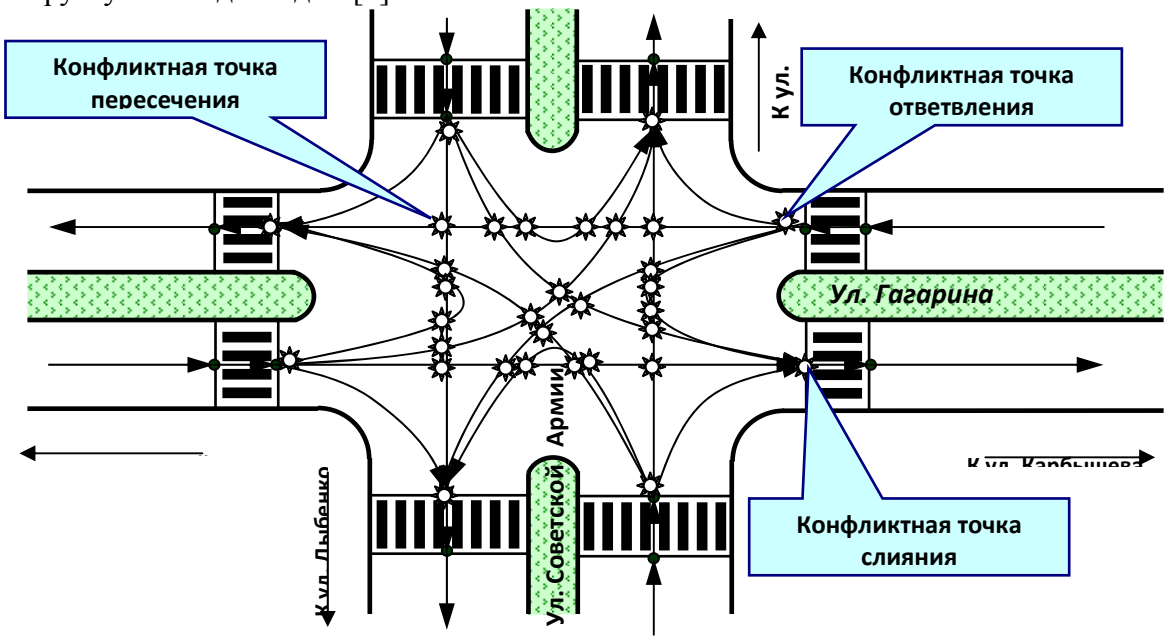


Рис. 1. Конфликтные точки на перекрестке

Вероятное число ДТП на пересечении в целом $Z_g = \sum_{i=1}^{n} z_i$, где n - число конфликтных точек. Уровень обеспечения безопасности движения на пересечениях оценивается коэффициентом безопасности пересечения: $k_{cd} = \frac{Z_g \cdot k_y \cdot 10^7}{25 \cdot \left(I^{'} + I\right)}$, где $I^{'}$, I - интенсивности движения на пересекающихся дорогах, авт/сут.

Метод оценки безопасности маршрута

Еще одним методом оценки безопасности маршрутов является метод экспертных оценок. Метод формирования групповой оценки экспертов представляет собой ряд последовательно осуществляемых процедур, направленных на формирование группового мнения по вопросам, по которым ощущается недостаток информации [1, 4]. Характерные черты метода: анонимность, достигаемая применением специальных вопросников; регулируемая обратная связь, осуществляемая за счет проведения нескольких туров опроса, результаты каждого из которых обрабатываются с помощью статистических методов и сообщаются экспертам. Применение статистических методов позволяет уменьшить разброс индивидуальных оценок и получить обобщенное мнение экспертов.

Среди показателей транспортной сети, которые оцениваются данным методом, выделим следующие:

Свойства, влияющие на безопасность движения на участке УДС.

Наличие светофорной сигнализации на перекрестке/перегоне.

Коэффициент качества дорожного покрытия (среднее).

Погодные условия по времени года (мокрый асфальт, гололед, туман).

а) Свойства, приводящие к повышению опасности.

Наличие ДТП на участке УДС (0 – 2 ДТП/год).

Наличие очага аварийности на заданном участке УДС (более 3-х ДТП/год).

Пересечение с ж/д путями со шлагбаумом на перегоне.

Пересечение с ж/д путями на перекрестке.

Наличие ж/д переезда без шлагбаума.

Пересечение с трамвайными путями на перегоне.

Пересечение с трамвайными путями на перекрестке.

Наличие пассажирского транспорта (троллейбус отдельно).

Приоритетность движения на данном участке (второстепенная).

Наличие левоповоротного движения на перекрестках при наличии встречного движения.

Наличие левого поворота на нерегулируемых перекрестках.

Наличие пешеходного перехода.

Наличие движения грузового транспорта.

Наличие детских учебных заведений (наличие знака 1.21).

Количество полос на дороге 2-3 при наличии встречного движения.

Приоритетность движения на данном участке (равнозначная).

Расстояние видимости на участке (до 50 м).

Расстояние видимости на участке (от 50 до 100 м).

Расстояние видимости на участке (от 100 до 150 м).

Наличие дорожных знаков, предупреждающих об опасности.

Коэффициент качества дорожного покрытия (плохое).

Сложные погодные условия по времени года (зима: гололед, туман).

Сложные погодные условия по времени года (весна: мокрый асфальт, гололед, туман).

б) Свойства, приводящие к повышению безопасности.

Отсутствует ДТП на участке.

Участок УДС имеет одностороннее движение.

Наличие дорожной разметки.

Приоритетность движения на данном участке (главная).

Количество полос на дороге 4 и больше при наличии встречного движения.

Наличие освещения на данном участке (наличие исправной системы освещения на данном участке).

Коэффициент качества дорожного покрытия (хорошее).

Расстояние видимости на участке (от 150 до 300 м).

Расстояние видимости на участке (более 300 м).

Сформированный перечень показателей использован для оценки уровня безопасности k_s различных видов участков. Для расчета уровня безопасности введен дополнительный параметр C_j для показателя:

+1 – для свойства, приводящего к повышению безопасности.

-1 – для свойства, приводящего к повышению опасности.

Значение показателя рассчитывается как среднее арифметическое оценок экспертов: $k_{pj} = \frac{\sum_{i=1}^{m} R_{ij}}{m}$, где m – количество экспертов.

Эмпирически выбран диапазон значений каждого показателя: 0.7-1.3, характеризующего уровень безопасности: высокое значение коэффициента соответствует высокому значению уровня безопасности. Для приведения показателей к требуемому диапазону используется выражение:

$$k'_{pj} = 1 + \lambda \cdot c_j \cdot k_{pj}, \quad (1)$$

где k'_{pj} - приведенное значение коэффициента; c_j - коэффициент, определяющий вид показателя; λ - коэффициент приведения. Для приведения к диапазону 0.7-1.3 значение λ=0.03. Коэффициент безопасности участка рассчитывается $k_s = (\sum_{j=1}^{n} k'_{pj}) / n$, где n - количество показателей безопасности. Расчет безопасности сложноорганизованного маршрута обеспечивается следующим выражением:

$$S_r = \frac{\sum_{j=1}^{n} (l_j \cdot k_{s_j})}{\sum_{j=1}^{n} l_j}, \quad \left(j = \overline{1,n}\right) \quad (2)$$

где l_j – длина j-го участка; n – количество участков в маршруте.

Интеллектуальная транспортная система перевозки опасных грузов

Методы учета степени опасности участков транспортной сети применяются в интеллектуальной транспортной системе перевозки опасных грузов. В подсистеме учета маршрутов перевозки опасных грузов построение транспортных маршрутов возможно по различным критериям, таким как: безопасность маршрута, расстояние и др. В процессе исследования транспортной системы г. Самара построены матрицы смежности графа улично-дорожной сети города участки, захватывающие маршрут от Кряжской нефтебазы до пересечения ул. Арцыбушевская и ул. Буянова. В таблицах 1 и 2 приведены построенные маршруты по различным критериям.

Таблица 1. Маршрут оптимальный по расстоянию

	Расстояние	Безопасность
Эльтонская/Грозненская ул.		
Грозненская ул./Пугачевский тракт	3000	0,8
Пугачевский тракт/Новокуйбышевское ш.	3500	0,95
Новокуйбышевское ш./Кряжская дамба	4500	0,7
Кряжская дамба/Шоссейная ул.	3000	0,75
Шоссейная ул./Главная ул.	800	0,74
Водников ул./Главная ул.	300	0,8
Водников ул./Венцека ул.	1300	0,9
Венцека ул./Бр. Коростелевых ул.	1500	1,3
Бр. Коростелевых ул./Л.Толстого ул.	900	1,2
Л. Толстого ул./Арцыбушевская ул.	200	0,8
Результирующее значение	19000	0,894

При построении маршрута по критерию безопасности он полностью совпал с эталонным маршрутом.

Результаты построения маршрута в подсистеме учета маршрутов перевозки опасных грузов показаны на рис. 2.

Результаты работы интеллектуальной транспортной системы (ИТС) используются при построении маршрутов перевозки опасных грузов в различных городах России. На подсистему ИТС, предназначенную для построения безопасных маршрутов и оценки уровня безопасности транспортной инфраструктуры получено свидетельство об официальной регистрации программы для ЭВМ.

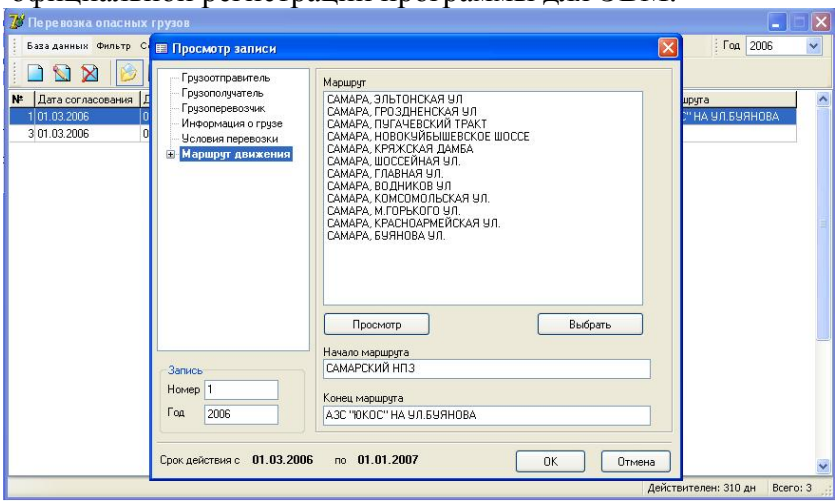


Рис. 2. Маршрут перевозки опасного груза

Таблица 2. Маршрут оптимальный по безопасности

	Расстояние	Безопасность
Эльтонская/Грозненская ул.		
Грозненская ул./Пугачевский тракт	3000	0,8
Пугачевский тракт/Новокуйбышевское ш.	3500	0,95
Новокуйбышевское ш./Кряжская дамба	4500	0,7
Кряжская дамба/Шоссейная ул.	3000	0,75
Шоссейная ул./Главная ул.	800	0,74
Водников ул./Главная ул.	300	0,8
Водников ул./Комсомольская ул.	400	0,75
Комсомольская ул./М. Горького ул.	1500	1
М. Горького ул./Л. Толстого ул.	1100	0,8
Л. Толстого ул./Арцыбушевская ул.	2200	0,8
Результирующее значение	20300	0,809

Литература

1. Михеева, Т.И. Структурно-параметрический синтез интеллектуальных транспортных систем / Самара: Самар. науч. центр РАН, 2008. 380 с.
2. Вероятностные и имитационные подходы к оптимизации автодорожного движения / Буслаев А.П., Новиков А.В., Приходько В.М. [и др.] // Под ред. В.М. Приходько — М. : Мир, 2003. 368 с.
3. Михеева, Т.И. Методы решения задач управления транспортными потоками в автоматизированной системе. Михеева Т.И., Золотовицкий А.В. / Интеллектуальные системы: Труды Шестого международного симпозиума (INTELS'2004) / Под ред. К.А. Пупкова. - М.: РУСАКИ, 2004. С. 413-416.
4. Золотовицкий, А.В. Расширенная графовая модель решения задач маршрутизации. Золотовицкий А.В., Михеева Т.И., Михеев С.В. / Проблемы автоматизации и управления в технических системах // Труды международной научно-технической конференции. Пенза: 2008. С. 298-301.
5. Золотовицкий, А.В. АИС составления маршрутов перевозок опасных грузов. Золотовицкий А.В., Абросимова Д.В. / Свидетельство об официальной регистрации программы для ЭВМ № 2007613779. – М.: Федеральная служба по интеллектуальной собственности, патентам и товарным знакам, 2007.

Алгоритмы координации управления транспортными потоками на связанных перекрестках

Т.И. Михеева, С.В. Михеев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Объектом управления в интеллектуальных транспортных системах являются транспортные потоки (ТП). При их описании приходится принимать во внимание ряд особенностей, характерных для сложного социального объекта [1]:

- необходимость математических моделей при управлении;
- стохастичность поведения ТП, связанная с его сложностью и наличием второстепенных по отношению к цели управления процессов;
- сопротивляемость ТП управлению, связанная со спецификой ТП и объясняемая возможностью возникновения отличий цели управления объектом от целей его элементов:
- нестационарность объекта управления, проявляющаяся во временном изменении характеристик и параметров, описывающих поведение ТП;
- неповторяемость экспериментов, что проявляется в отличии их результатов при одних и тех же воздействиях на объект и параметрах, которые используются при описании его поведения.

Все изложенное приводит к тому, что цель управления в полной мере не достигается, поскольку для синтеза и реализации управления требуется некоторое время, в течение которого объект изменяется. Это приводит к необходимости использования прогноза при определении характеристик ТП и расчета воздействий с упреждением и минимальными затратами времени.

Экспериментальные исследования процесса движения транспортных средств в городах показывают, что характеристики ТП по времени суток испытывают сильные колебания, но эти же самые колебания носят устойчивый вероятностный характер с постоянными параметрами в различные дни недели, что связано с наличием детерминированных тенденций в движении транспортных средств (работа общественного транспорта, поездки, связанные с трудовой деятельностью и т.д.).

Свойством объекта управления, определяющим принципы управления, является групповая (пачкообразная) структура ТП [2].

В силу постоянного изменения микрохарактеристик ТП [3] может наблюдаться бесконечное множество сочетаний параметров, характеризующих процесс движения. В идеальной системе управления каждому возможному состоянию, в котором система может пребывать время Δt, должна соответствовать вполне определенная совокупность управляющих воздействий. Такое управление может быть реализовано только в том случае, если указанная совокупность может быть рассчитана и обработана за время, меньшее времени пребывания системы в данном состоянии. Так как технические средства управления дорожным движением не позволяют применить такое управление даже для небольшого числа перекрестков, входящих в состав системы, то задача управления сводится к многошаговому процессу принятия решений со следующей последовательностью операций:

- измерение характеристик потока, усредненных за Δt;
- прогнозирование состояния, в котором будет находиться система управления;
- выработка управляющих воздействий, соответствующих упомянутому состоянию.

Модели транспортных потоков

Движение транспортных средств по улично-дорожной сети определяется поведением, как одного, так и коллектива водителей [2, 3, 4]. Отдельный водитель, пытаясь достичь собственного оптимального решения, вступает в конфликт с другими, которые взаимодействуют с ним посредством обгонов, перестроения, смены полосы

движения и т.д. Такая модель рассматривается в рамках ***микроскопического подхода*** [3, 5]. Маневры каждого автомобиля могут быть расценены как вероятностные события. Однако, в случаях, когда много автомобилей движется в группе, транспортный поток может быть рассмотрен как детерминированный и непрерывный. Применение микроскопических моделей (как и любое увеличение степени детализации описания) влечет за собой увеличение точности описания и числа параметров. Таким образом, с одной стороны, при увеличении степени детализации описания объекта растет точность модели, а с другой - рост параметров ведет к уменьшению ее точности. При решении многомерных оптимизационных задач управления возрастают ресурсные затраты (время и память), затрудняющие получение приемлемого решения.

Взаимодействие нескольких транспортных средств может привести к некоторой коллективной стратегии, которая моделируется в рамках *макроскопического подхода* [2, 5]. В этом случае транспортный поток отображается как стационарное явление, представляемое общей средней скоростью, плотностью потока и интенсивностью движения в определенный момент времени в определенной точке сети. Кроме того, в современных системах управления сигналы светофоров адресуются не какому-либо определенному участнику движения, а их совокупности.

Для описания ТП используем следующие обозначения:

S_i – перекресток со светофором и номером *i*;

t_i^R, t_i^r – длительность красного и эффективного красного сигнала на S_i, *c*;

t_i^G, t_i^g – длительность зеленого и эффективного зеленого сигнала на S_i, *c*;

t_i^Y – длительность желтого сигнала на S_i, *c*;

T – длительность цикла, *c*;

C_i – поток насыщения на S_i для (*i, j*), *авт/с, авт/час*;

$\delta_i^{(j)}$ – сдвиг начала красного сигнала на S_j перекрестке относительно S_i, ($0 \le \delta_i^{(j)} \le T$), *c*.

Связи приведенных показателей иллюстрирует рис. 1.

Пусть $Q(t,x)$ - количество автомобилей, которые в момент *t* приближаются к сечению $x \in H$ по одной полосе, тогда

$$I(t,x) = \frac{\partial Q}{\partial t} \text{ - интенсивность ТП,} \quad (1)$$

$$k(t,x) = \frac{\partial Q}{\partial x} \text{ - плотность ТП.} \quad (2)$$

Предполагаются следующие свойства транспортного потока [1].

Временная зависимость интенсивности $I(t,x)$ транспортного потока, прибывающего к перекрестку, работающему в координированном режиме, - является периодической функцией с периодом, равным циклу T на перекрестке, откуда поток прибывает:

$$I(t,x) = I(t + n \cdot T, x), \quad (3)$$

где *n* – произвольное целое число. Для изолированного перекрестка $I(t,x) \equiv const$.

На S_i отсутствуют заторы (случай заторов оговаривается особо), что означает:

$$\int_0^T I(t)dt < C_i t_i^g \quad (4)$$

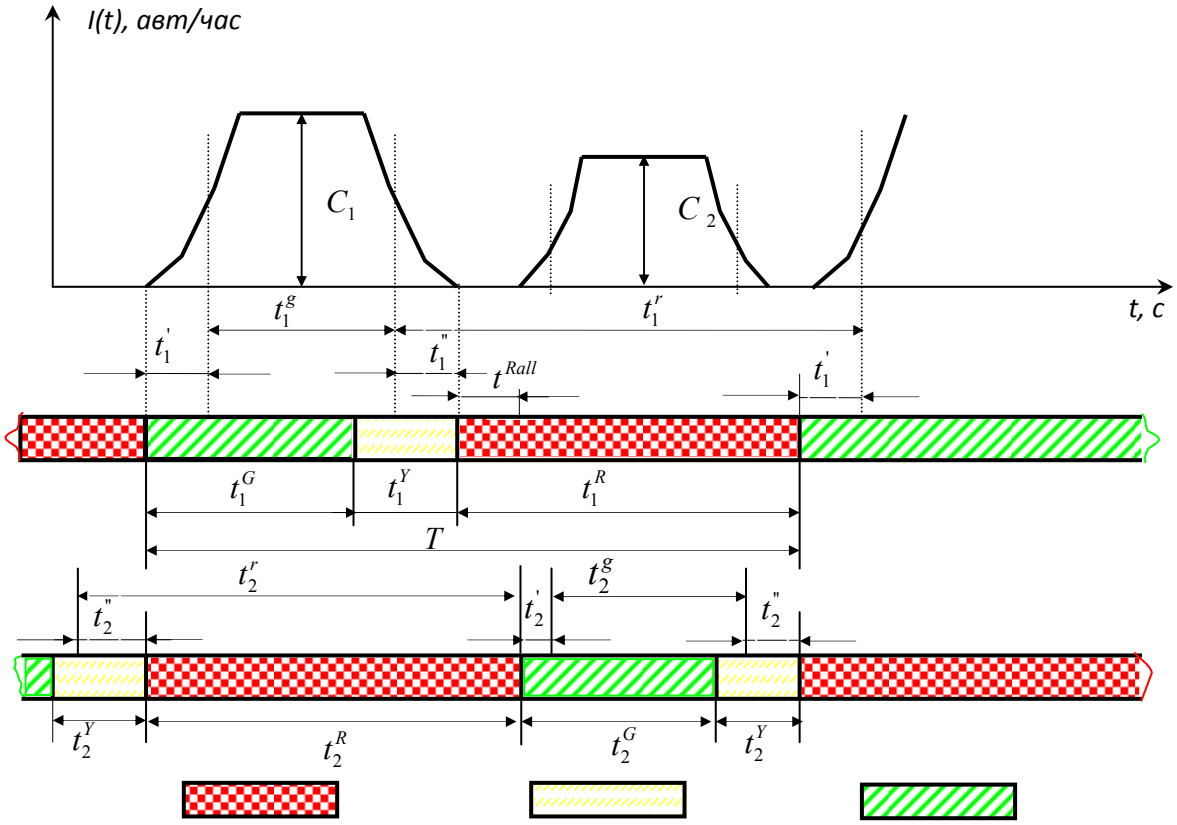


Рис. 1. Связь управляющих сигналов на перекрестке

Значение интенсивности транспортного потока не может превышать значения потока насыщения, т.е. $\forall t : I(t) \le C_i$. На рис. 1 приведен график временной зависимости интенсивности транспортного потока на выходе перекрестка при большой степени его загрузки $\int_0^T I(t)dt \approx C_i t_i^g$.

Транспортный поток характеризуется плотностью потока $k(t,x)$ и средней скоростью потока $v(t,x)$. Параметры транспортного потока связаны следующим основным уравнением:

$$v(t,x) = \frac{I(t,x)}{k(t,x)}. \quad (5)$$

Различают два вида средней скорости транспортного потока: *среднюю пространственную скорость* v_S и *среднюю временную скорость* v_t, которые связаны следующим соотношением, выведенным для случая движения по дороге без пересечений [2]:

$$v_t = v_S\left(1+\left(\frac{\sigma_S}{v_S}\right)^2\right), \quad (6)$$

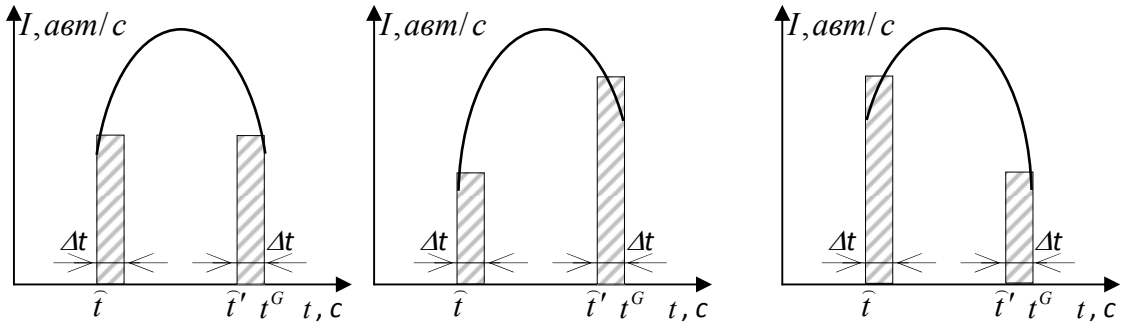
где σ_S^2 – дисперсия средней пространственной скорости; v_S – средняя пространственная скорость, т.е. средняя скорость n автомобилей, находящихся на заданном участке дороги в определенный момент времени; v_t – средняя временная скорость, т.е. средняя скорость n автомобилей, проходящих через заданное сечение дороги за определенный промежуток времени.

Алгоритмы координации

Алгоритмы координации управления транспортными потоками на связанных перекрестках используют условие минимизации числа остановленных ТС либо суммарного времени задержек [1].

Алгоритм минимизации числа остановленных ТС [6] основан на том утверждении, что минимум достигается при условии равенства мгновенных интенсивностей движения в моменты непосредственно после разъезда очереди $\hat{t}$ и перед переключением сигнала с разрешающего на запрещающий $\left(t^G, t^G + T\right)$: $\tilde{I}\left(\hat{t}\right) = \tilde{I}\left(\hat{t}\,'\right) = const$.

Оптимальная форма интенсивности на выходе перекрестка показана на рис 2 а. В качестве интенсивностей будем использовать число транспортных средств $\hat{Q}$ и $\hat{Q}'$, проезжающих через перекресток в течение промежутков времени Δt после $\hat{t}$ и после $\hat{t}'$ ($\hat{t}' = t^G - \Delta t$), на рис. 2 а им соответствуют заштрихованные площади.



а) б) в)

Рис. 2. Различные ситуации использования алгоритма:
а) оптимальные значения; б) сдвиг фаз следует увеличить; в) сдвиг фаз следует уменьшить

Значения Δt должно быть по возможности малым, чтобы обеспечить достаточную точность сравнения интенсивностей. Вместе с тем Δt должно быть достаточно большим, чтобы уменьшить случайные колебания числа проехавших ТС. На основе экспериментальных данных выбрано Δt=3-5 с, при низких интенсивностях движения Δt=8-10 с для усредненных $\hat{Q}$ и $\hat{Q}'$ за несколько периодов.

Правило изменения сдвигов фаз состоит в следующем: если $\hat{Q} < \hat{Q}'$ (рис. 2 б), переключение фаз происходит слишком рано, так что для приведения интенсивности к виду рис. 2 а значения t^R и t^G надо увеличить на Δt. Если $\hat{Q} > \hat{Q}'$ (рис. 2 в), то сдвиг фаз, определяемый значениями t^R и t^G надо уменьшить на Δt. После коррекции t^R и t^G проверка их оптимальности повторяется, и процесс оптимизации продолжается.

В случае управления двухсторонним движением производится определение желательных изменений t^R и t^G для обоих направлений. Если рекомендации о направлении изменений фаз совпадают, производится соответствующий сдвиг этих фаз; при противоположных рекомендациях, а такие при выполнении условий оптимальности фазы не изменяются.

Алгоритм минимизации задержки состоит в равенстве средней интенсивности прибытия ТС в период существования очереди с интенсивностью в момент начала образования очереди $\left(t^G, t^G + T\right)$. Суммарные и средние задержки ТС могут принимать минимальное значение только в том случае, когда мгновенная интенсивность прибытия ТС в момент включения запрещающего сигнала равна средней интенсивности поступления ТС в очередь, образовавшуюся после этого момента:

$$\tilde{I}\left(t^G\right) = \frac{1}{\hat{T}}\left(Q\left(t^G + \hat{T}\right) - Q\left(t^G\right)\right), \qquad (7)$$

где $Q(t)$ – длина очереди в момент t, $\hat{T}$ – время накапливания очереди.

Для проверки этого условия достаточно располагать значениями величины $\widehat{Q}'$ и длительности существования очереди в период разрешающей фазы $\widehat{t}$, по которой рассчитывается параметр:

$$\widehat{Q} = \frac{C\widehat{t}}{t^R + \widehat{t}}\Delta t . \qquad (8)$$

Преимуществом рассмотренных алгоритмов является небольшое количество используемой информации, сравнительная простота и универсальность. Оптимум достигается путем применения процесса последовательных приближений. Подобные алгоритмы являются, в основном, средством коррекции расчетного времени прибытия ТП к перекрестку.

В результате реализации алгоритма формируется базовая программа координации, рассчитанная по критерию оптимизации выбранного показателя эффективности функционирования транспортного потока. Базовая программа воздействует на ТП. В результате измерения параметров ТП и анализа их взаимодействия с управляющими воздействиями программы координации через заранее заданный интервал времени $\Delta\widetilde{t}$ определяется действительный показатель длительности задержек за интервал $\Delta\widetilde{t}_i$.

В течение первого интервала включается алгоритм коррекции программы координации управления транспортными потоками на связанных перекрестках, воздействующей на объект. В дальнейшем алгоритм коррекции включается через $\Delta\widetilde{t}$ при условии, что предыдущая коррекция привела к уменьшению длительности задержки транспортных средств. Методы оптимизации обеспечивают за относительно малое число шагов нахождение локального экстремума. В принципе рассчитанный базовый набор управляющих воздействий соответствует глобальному оптимуму.

Литература

1. Михеева, Т.И. Структурно-параметрический синтез интеллектуальных транспортных систем / Самара: Самар. науч. центр РАН, 2008. 380 с.
2. Клинковштейн, Г.И. Организация дорожного движения: Учеб. для вузов. – М. : Транспорт, 2001. 247 с.
3. Семенов, В.В. Математическое моделирование динамики транспортных потоков мегаполиса. - М.: Институт прикладной математики РАН, 2004. 44 с.
4. Математическое моделирование динамики автотранспортных потоков и вызываемого ими загрязнения атмосферного воздуха в автомобильных тоннелях / Н. Н. Смирнов, А. Б. Киселев, В. Ф. Никитин [и др.].— Наукоемкие техн., 2003. Т.4, № 9. С. 29-43.
5. Mikheev, S.V. Model of zone control of traffic flows / Computer Science and Information Technologies. Mikheeva T.I., Mikheev S.V. //Proceedings of the 10th International Workshop. V. 2. Antalya, Turkey, 2008. P. 121-128.
6. Mikheeva, T.I. Models for Synthesis of Intellectual Transport Systems. Mikheeva T.I., Mikheev S.V. / Computer Science and Information Technologies // Proceedings of the 11th International Workshop on Computer Science and Information Technologies (CSIT'2009), Vol. 2. Crete, Greece, 2009. Pp. 81-85.

Информационная система дислокации светофорных объектов с использованием цифровых моделей местности

Михеева Т.И., Япрынцева О.А.
Самарский государственный аэрокосмический университет, Самара
Самарский государственный технический университет, Самара

Введение

Вследствие растущей загруженности автомагистралей, возрастающей интенсивности автомобильных перевозок в условиях современных мегаполисов остро ощущается необходимость регулирования транспортных потоков. Для обеспечения комфорта участникам дорожного движения и их информирования о текущем состоянии транспортной сети, для снижения материальных и финансовых издержек были созданы системы управления дорожным движением, получившие название Intelligent Transportation Systems (Интеллектуальные транспортные системы – ИТС), которые уже долгое время успешно функционируют во всем мире. Увеличение количества автотранспорта, усложнение инфраструктуры современного города приводит к постоянному ухудшению показателей безопасности дорожного движения, уменьшается пропускная способность транспортной сети города, средняя скорость транспортных средств, увеличивается время простоя транспортных средств в «пробках».

Задача регулирования движения транспортных потоков на улично-дорожной сети (УДС) решается путем расстановки на УДС средств регулирования - дорожных знаков, светофоров, разметки - технических средств организации дорожного движения. Помимо технических средств организации дорожного движения на улично-дорожной сети могут располагаться и другие объекты, непосредственно или опосредованно влияющие на организацию дорожного движения: световые опоры, рекламные сооружения, объекты дорожного сервиса и др. Правила установки каждого типа объекта регламентируются ГОСТом 52289-2004 «Технические средства организации дорожного движения. Правила применения» [1].

Задача оптимальной и корректной, с точки зрения обеспечения безопасности движения, дислокации светофорных объектов на улично-дорожной сети требует использования современных интеллектуальных информационных технологий с привлечением геоинформационных систем (ГИС). Оптимальные параметры управления светофорной сигнализацией и дислокации технических средств организации дорожного движения обеспечивают равномерное распределение транспортных потоков, позволяют снизить задержки транспортных средств на локальном перекрестке и в целом на улично-дорожной сети.

Дислокация светофорных объектов в интеллектуальной транспортной системе

В настоящее время светофоры являются самыми распространенными приборами для регулирования дорожного движения. Они используются на перекрестках, примыканиях дорог, пешеходных переходах, железнодорожных переездах и во многих других случаях. Светофор сегодня это полностью автоматический, запрограммированный световой прибор, который переключает сигналы автоматически через равный промежуток времени, независимо от ситуации, сложившейся на перекрестке. Поскольку именно они наиболее часто встречаются на дорогах города, необходимо знать какие виды светофоров существуют, для чего они предназначены и что обозначают те или иные сигналы [1]. Эта информация нужна при работе с информационной системой дислокации светофорных объектов, учета их работы и сохранения в специализированной базе данных.

Интеллектуальная транспортная система содержит среди своих модулей – модуль дислокации на электронную карту объектов транспортной инфраструктуры с наглядной визуализацией их географической и атрибутивной информации. Одним из таких объектов является светофорный объект.

При реализации системы для каждого вида объектов в геоинформационной системе создается собственный слой, который содержит только эти виды объектов или связанные с ними. Например, в системе реализованы слой «Светофоры». Пример этого слоя приведен на рисунке 1.

При отображении простых объектов, например столбов, групп дорожных знаков, опор выбраны простые геометрические фигуры: окружности, многоугольники, линии. Сложные объекты, такие как дорожный знак или светофор, невозможно обозначить таким образом, поскольку они имеют более сложную структуру. Для их обозначения реализованы специальные шрифты, для каждого вида объектов реализован свой шрифт. Сложные объекты устанавливаются на электронную карту города в виде символов специального шрифта.

Интеллектуальная транспортная система позволяет решать следующие задачи, связанные со светофорными объектами [2]:

- установка опор различных типов для светофоров и знаков;
- установка светофоров на электронной карте города с выбором место-положения и угла поворота на опоре;
- удаление светофора, установленного на электронной карте города;
- добавление и редактирование детальной информации по светофорному объекту;
- мониторинг установленных на улично-дорожной сети светофоров;
- добавление, удаление и редактирование программ работы светофорных объектов;
- идентификация светофоров по типам, указанных в ГОСТ Р 52282-2004 с выбором типа по изображению.

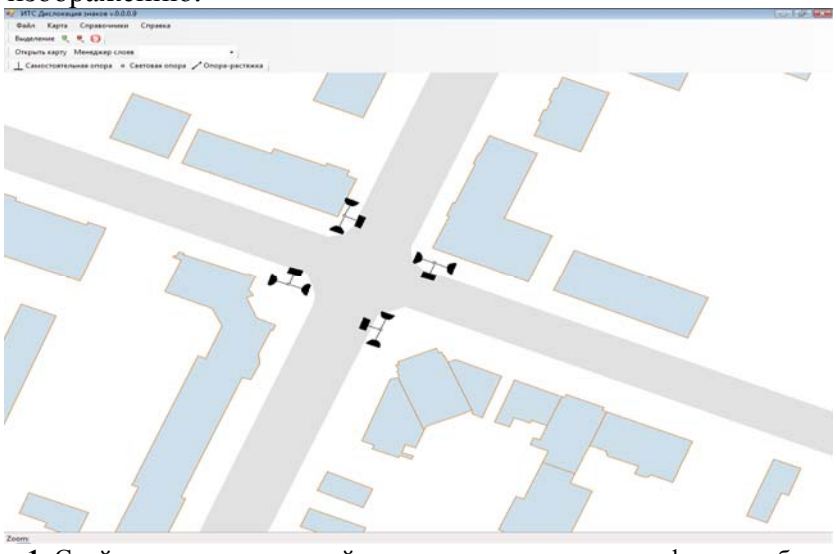


Рис. 1. Слой автоматизированной системы дислокации светофорных объектов

Рассмотрим алгоритм добавления в систему нового объекта на примере добавления нового светофорного объекта на УДС города.

Шаг 1. Пользователь, используя соответствующие инструменты перемещения карты, определяет место дислокации нового светофорного объекта - участок УДС города, являющийся одной из цифровых моделей местности (рис. 2).

Шаг 2. При помощи соответствующего пункта меню системы пользователь выбирает вид объекта, который хочет установить (рис 3).

Шаг 3. При помощи курсора мышки пользователь определяет точное место дислокации светофора на карте и его место привязки к дуге графа УДС города.

Шаг 4. Из справочника светофоров пользователь выбирает тип светофорного объекта по ГОСТу (рис. 4).

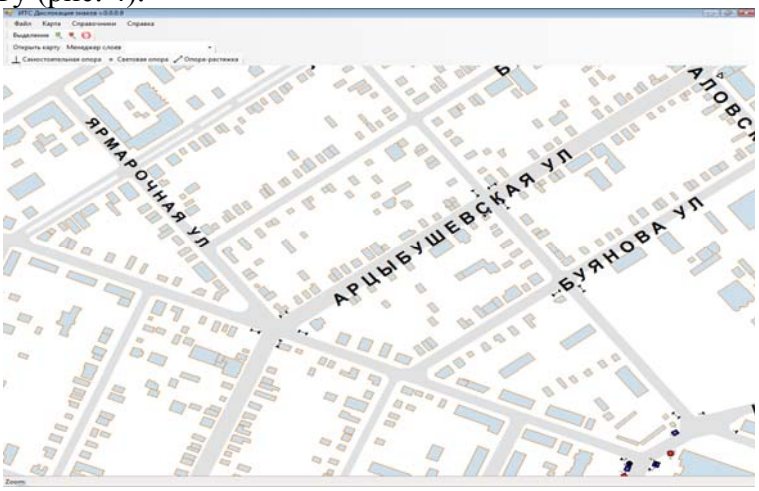


Рис. 2. Выбор участка УДС

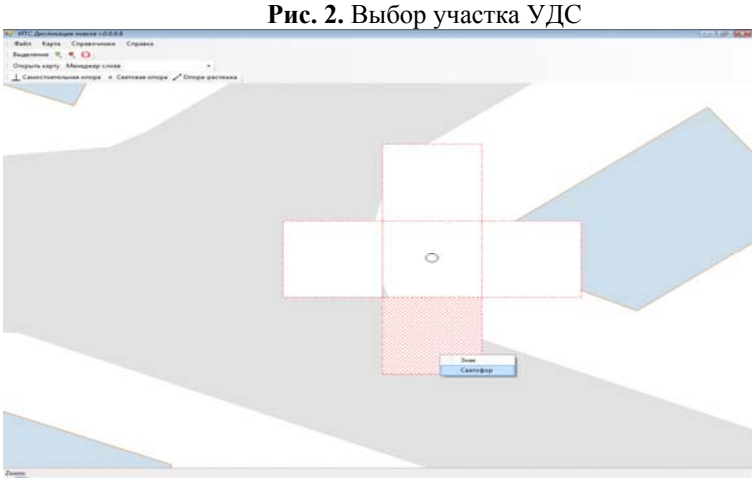


Рис. 3. Дислокация нового светофорного объекта

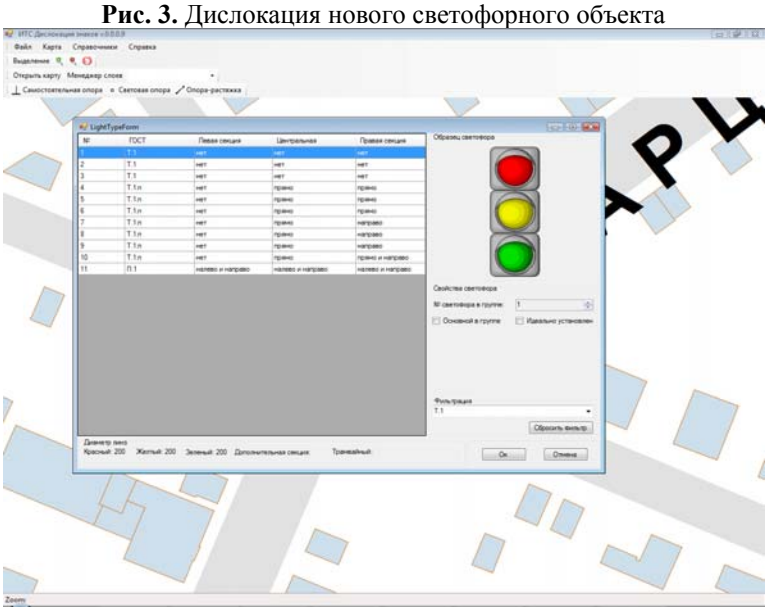


Рис. 4. Выбор светофорного объекта

Шаг 5. Пользователь заполняет информацию о светофорном объекте: свойства светофора (номер светофора в группе, способ установки).

Шаг 6. Информация о светофорном объекте, участке, на который происходит установка светофора, и координаты на дуге передаются в сервер приложения.

Шаг 8. Запускается алгоритм дислокации объекта, который сохраняет введенные данные в системе хранения информации.

В настоящее время дислокация светофорных объектов г. Самара ведется в среде интеллектуальной транспортной системы с использованием специализированной ГИС «ITSGIS» и автоматизированной информационной системы дислокации объектов транспортной инфраструктуры [3].

Литература

1. ГОСТ Р 52289-2004 Технические средства организации дорожного движения. Правила применения дорожных знаков, разметки, светофоров, дорожных ограждений и направляющих устройств. — М. : Стандартинформ, 2006. 94 с.

2. Михеева. Т.И. Построение математических моделей объектов улично-дорожной сети города с использованием геоинформационных технологий // Информационные технологии. 2006. №1. С. 69–75.

3. Михеева, Т.И. Геоинформационная система дислокации светофорных объектов Михеева Т.И., Михеев С.В. Рудаков И.А., Чугунов И.А. //Математическое моделирование информационных процессов и систем в науке, технике и образовании: Межвузовский сборник научных трудов. Самарск. гос. арх.-строит. ун-т. – Самара:, 2010. С. 89-92.

Бортовая система учета пассажиров общественного транспорта

Б.В. Скворцов, Е.С. Перцович
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Применение бортовой системы учета пассажиров общественного транспорта позволяет решить социально значимую задачу оптимального использования единиц общественного транспорта, оперативного регулирования городскими маршрутами, финансового и организационного контроля транспортных единиц, что приведет к значительной экономии бюджета государственных и частных транспортных компаний, улучшению обслуживания населения.

Назначение бортовой системы учета пассажиров общественного транспорта - учет общего количества пассажиров, перевезенных транспортным средством за определенный период времени. Фиксация времени прихода транспортного средства на каждую остановку и количества пассажиров вошедших и вышедших на каждой остановке. Передача данных диспетчеру.

В основу бортовой системы учета пассажиров общественного транспорта положено использование ультразвуковых датчиков, фиксирующих факт входа или выхода пассажира из салона транспортного средства. Информация датчиков обрабатывается универсальным контроллером, в котором сохраняются все вышеуказанные данные до конца смены или до прихода транспортного средства на диспетчерский пункт, где она может быть прочитана диспетчером на индикаторе контроллера или передана по последовательному порту или радиоканалу в центральный компьютер автотранспортного предприятия.

Возможны две схемы расположения датчиков, они показаны на рис. 1 и 2.

1.

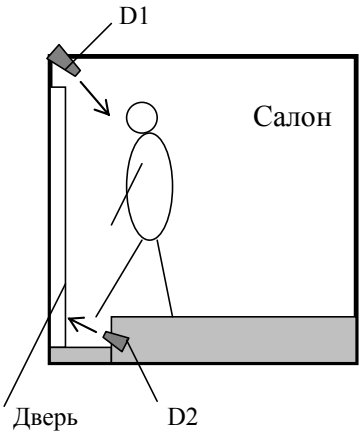


Рис. 1

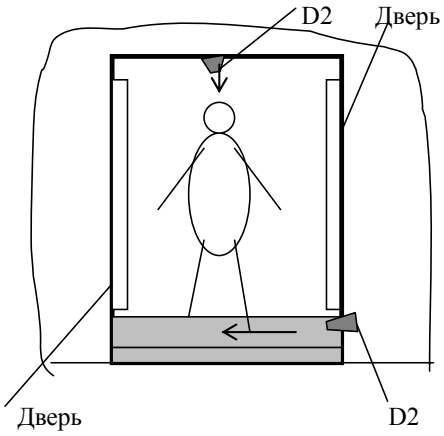


Рис. 2

По схеме на рис. 1 первый ультразвуковой датчик D1 крепится в нижней ступеньке с диаграммой направленности излучения из двери (на улицу), второй датчик D2 над дверью с диаграммой направленности в салон. Зона действия датчиков от 6 до 130 см (настраивается отверткой через отверстие на корпусе датчика) Датчики посылают ультразвуковые импульсы, которые отражаются от человека Полный угол диаграммы направленности 10 град. Отраженный сигнал приводит к формированию импульса на выходе датчика. Верхний датчик срабатывает от головы или плеч пассажира. Нижний датчик от – от ног, вставших на ступеньку.

В схеме на рис. 2 нижний датчик D2 крепится сбоку ступеньки и посылает сигнал вдоль нее. Появление ноги в зоне ступеньки приведет к срабатыванию датчика.

Основной научно-технической проблемой при разработке системы является повышение надежности распознавания объекта учёта в моменты прохода через дверь, когда идет непрерывный поток пассажиров. Здесь возможны конструкторские и алгоритмические методы решения, а также их сочетания.

Сигналы датчиков передаются на контроллер, куда также посылаются сигналы об открытии и закрытии двери, а также сигнал от ключа зажигания (если необходим контроль времени работы двигателя). Контроль времени производится внутренним таймером контроллера. Все данные записываются в память контроллера по сигналу закрытия двери. В зависимости от последовательности срабатывания датчиков пассажир идентифицируется как вошедший или вышедший. В конце рейса или рабочего дня диспетчер может списать все показания на экране контроллера, вызвав их на индикацию с помощью клавиатуры на передней панели.

Возможны другие схемы установки в зависимости от конкретных условий и технических правил эксплуатации транспортного средства и крепления датчиков.

Структурная схема системы показана на рис. 3.

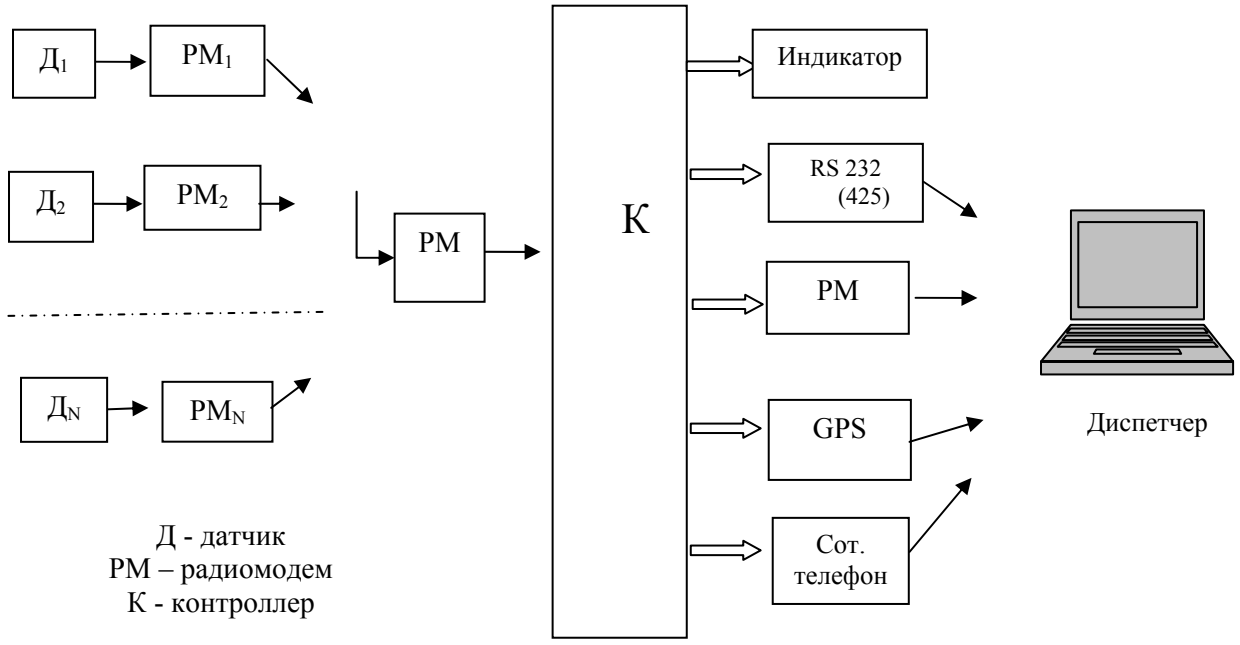


Рис. 3.

Информация датчиков обрабатывается универсальным контроллером, в котором сохраняются все вышеуказанные данные до конца смены или до прихода транспортного средства на диспетчерский пункт, где она может быть прочитана диспетчером на индикаторе контроллера или передана по последовательному порту или по радиоканалу в центральный компьютер автотранспортного предприятия. Для оперативной передачи данных предполагается использование автоматизированной системы мониторинга «Навигатор-С» (в дальнейшем будет использоваться термин Диспетчерский Пункт «Навигатор-С» или ДП «Навигатор-С») представляет программно – технический комплекс, предназначенный для мониторинга подвижных (мобильных) и стационарных объектов.

Под подвижным объектом мониторинга понимается транспортное средство, оснащённое системой контроля МТ (мобильный терминал) позволяющей

контролировать состояние (до 8 контактных датчиков и до 4 параметрических датчиков) и местоположение (а также скорость движения, пройденное расстояние и др.) транспортного средства и передавать информацию по каналу GSM (SMS, Data, др.).

Ядро системы – Monitor.exe – программный модуль «Navigator Monitor Service», отвечающий за взаимодействие с клиентами, базой данных, приёмо-передающими модулями (базовыми станциями).

Connection.exe – программный модуль «Navigator Connection Service», осуществляющий взаимодействие между приёмопередающей станцией (базовой станцией) и модулем Monitor . Используется для удалённого подключения (по локальной сети, а также через Интернет) базовых станций к серверу.

Модуль GPRSService - служит для приёма-передачи данных от терминалов по каналам Интернет (GPRS). Модуль устанавливается на сервер с выходом в Интернет и с постоянно выделенным IP адресом. Терминалы по указанному им распорядку подключаются к серверу и передают модулю данные. Модуль держит постоянную связь с сервером Navigator Monitor Service

Database \ Monitoring.gdb – база данных формата Interbase для хранения информации об объектах мониторинга.

Navigator.lic – файл лицензии, используется для защиты ПО. Каждая лицензия связана с определённым HASP ключом, поэтому программное обеспечение (серверная часть) не будет корректно работать без наличия файла лицензии и соответствующего ключа HASP. Файл лицензии ограничивает максимальное количество контролируемых объектов и количество подключений к серверу.

WWW сервис – web-модуль доступа к серверу. Используя любой компьютер с выходом в Интернет, пользователь может обращаться к серверу через браузер (например, Internet Explorer) для наблюдения за местоположением и состоянием мобильных объектов.

Преимущества и новизна предложенной разработки состоит в том, что за счет применения ультразвуковых датчиков присутствия удается повысить точность контроля пассажиропотока, особенно в часы «пик», когда идет плотный поток пассажиров, регистрация которого с помощью применяемых в настоящее время оптических датчиков невозможно. За счет применения интеллектуального датчика с передачей информации по радиоканалу повышается мобильность системы, уменьшается время монтажа, система становится более надежной и удобной в обслуживании. В России таких систем в настоящее время не производится. Монопольным и единственным производителем систем контроля пассажиропотока в мире является фирма IRIS GMBA, г. Берлин, которая продает их по всему миру по цене от 4500 до 5500 Евро. Однако такая система для России не пригодна так как использует оптические датчики, что не обеспечит требуемой точности.

Городские пассажирские автотранспортные предприятия, трамвайно-тролейбусные тресты, метро, автозаводы, производящие городской общественный транспорт, где предлагаемая система может устанавливаться стационарно, непосредственно при изготовлении транспортной единицы. Потребность в системах в г.Самаре составляет не менее 300 штук, в Москве по данным Минтранспорта 3000 штук в стране 200000 штук, Красноярское пассажирское автотранспортное предприятие готово приобрести 150 штук таких систем для автобусов ПАЗ 3205. Себестоимость изделия для однодверного транспортного средства составляет 1000 Евро. Себестоимость увеличивается с ростом числа дверей на 20 % на каждую дверь. Рыночная цена устройства для однодверного автобуса 1400 Евро. Себестоимость изделия для микроавтобуса типа «Газель» составит 2500 руб. Продажная цена – 3000 – 4000 руб.

При условии комплектования такой системой каждой единицы общественного транспорта рынок практически неисчерпаем. При этом себестоимость и цена могут быть значительно снижены.

В рамках проекта бортовой системы учета пассажиров общественного транспорта решается социально значимая задача оптимального использования единиц общественного транспорта, оперативного регулирования городскими маршрутами, финансового и организационного контроля транспортных единиц, что приведет к значительной экономии бюджета государственных и частных транспортных компаний, улучшению обслуживания населения.

Подсекция 7
«CALS/ИПИ-технологии»

Использование 3D-моделей переменной плотности в задачах оптимального проектирования авиационных конструкций

А.В. Болдырев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Рассматриваются следующие взаимосвязанные задачи, решаемые на ранних стадиях проектирования летательных аппаратов:

– определяются нагрузки, действующие на конструкцию;

– выбираются наиболее эффективные по массе силовые схемы;

– определяются (прогнозируются) значения масс агрегатов.

Эти задачи приходится решать в условиях неопределенности, когда отсутствует необходимая информация об объекте проектирования, в частности, еще не известны деформации создаваемой конструкции. При этом обычно используются специфические для каждой задачи математические модели и данные о прототипах проектируемого изделия.

В сложившемся порядке проектирования самолетов на ранних этапах *задача определения нагрузок*, как правило, решается в предположении абсолютной жесткости конструкций. В то же время известно, что изгибные деформации, например, стреловидного крыла вызывают отрицательные изменения углов атаки в концевых сечениях крыла. Учет этих изменений может дать другое, более выгодное с точки зрения прочности, распределение нагрузки и соответствующее снижение массы конструкции. Для этого необходим достоверный прогноз переменных состояния разрабатываемых упругих систем.

Силовая схема конструкции определяется типом силовых элементов, их числом, расположением в пространстве и способами соединения между собой. Решения, связанные с выбором силовых схем, на практике зачастую принимаются эвристическими методами, на основании опыта и интуиции, а также с использованием результатов расчетов на упрощенных математических моделях. Более достоверное исследование свойств объекта проектирования с применением метода конечных элементов (МКЭ) при этом производится лишь в конце процесса его разработки на стадии проведения поверочных расчетов. С целью повышения качества проектов в работе [1] предложена технология формирования силовой схемы по условиям прочности с привлечением МКЭ еще до начала полномасштабной разработки конструкции на этапе рабочего проектирования.

Весовое проектирование конструкций часто основывается на использовании так называемых "весовых формул", получаемых из статистического анализа построенных самолетов и использующих упрощенную балочную теорию. В работе [2] предложен подход к весовому проектированию авиационных конструкций на основе дискретных конечно-элементных моделей и специфического критерия – силового фактора.

Современные авиационные конструкции традиционных форм близки к исчерпанию своих аэродинамических и весовых характеристик, поэтому во всем мире ведется интенсивный поиск новых технических решений. Характерными примерами результатов такого поиска могут служить телескопические крылья, самолеты интегральной компоновки, крылья малого удлинения необычной формы, исследования,

направленные на поиск рациональных подкреплений тонкостенных конструкций в зонах больших вырезов и другие инновационные работы для которых еще не накоплены статистические данные, достаточные для использования традиционных методик проектирования.

Все это свидетельствует, во-первых, о *противоречии* между существующими способами проектирования и потребностью внедрения в конструкцию летательных аппаратов новых технических решений и, во-вторых, об *актуальности* вопроса разработки методики, использующей высокоточное математическое моделирование уже на этапе эскизного проектирования авиационных конструкций.

В настоящем исследовании предлагается развитие технологии проектирования [1, 2], позволяющее решать указанные задачи на основе специальной модели МКЭ.

Моделирование авиационных конструкций

На стадии эскизного проектирования конструкции еще не известны основные параметры сечений силовых элементов, которые определяются в процессе дальнейшей проработки объекта. В этих условиях для обоснованного выбора силовой схемы конструкции с учетом ограничений на эквивалентные напряжения, обобщенные перемещения и значения критических нагрузок, приводящих к выпучиванию упругой системы, может быть использована гипотетическая упругая среда переменной плотности, вписанная в геометрические ограничения разрабатываемого объекта.

Идея применения в процессе проектирования силовых конструкций материала с переменной по объему плотностью и жесткостью имеет четкий отечественный приоритет [3, 1]. Эта идея использована в исследованиях зарубежных ученых в виде пористого специфического материала с переменными прочностными и упругими свойствами несколько позже [4, 5].

В настоящей работе используется материал [3, 1], для которого модуль упругости и прочностные характеристики пропорциональны плотности ρ:

$$E = \rho \, \overline{\overline{E}}, \tag{1}$$

$$\overline{\sigma} = \rho \, \overline{\overline{\sigma}}, \tag{2}$$

где $\overline{\sigma}$ – допускаемое напряжение материала; $\overline{\overline{E}}$ и $\overline{\overline{\sigma}}$ – модуль упругости и допускаемое напряжение при единичной плотности.

Допустимая область объемом V, внутри которой может размещаться проектируемая конструкция, разбивается на n достаточно малых трехмерных конечных элементов. Непрерывная упругая среда со свойствами (1) – (2) (континуальная модель) потенциально содержит внутри заданных геометрических ограничений все мыслимые силовые схемы конструкции, которые могут быть образованы сгустками материала переменной плотности. Оптимизация распределения материала в континуальной модели позволяет определить проект, которому соответствует рациональная с точки зрения передачи усилий силовая схема конструкции.

Для анализа достоверности континуальной модели на ранних стадиях проектирования летательных аппаратов решаются две задачи. Во-первых, исследуется адекватность моделирования переменных состояния объемного деформируемого твердого тела на основе сопоставления численных результатов, получаемых на континуальной модели, и точных аналитических решений. Во-вторых, в условиях, приближенных к реальному проектированию, анализируются характеристики проекта, получаемые при переходе от континуальной модели к реальной тонкостенной конструкции. Сущность исследования заключается в следующем. Берется существующая конструкция, для которой имеется адекватная конечно-элементная модель и выполняется расчет ее напряженно-деформированного состояния. Затем в геометрические ограничения конструкции вписывается трехмерная модель переменной плотности и производится оптимизация распределения плотностей ее элементов на те же внешние нагрузки, что и уже спроектированная конструкция. Далее сравниваются переменные состояния объекта и континуальной модели в идентичных случаях нагружения.

В качестве объектов исследования взяты крылья легкого маневренного и военно-транспортного самолетов, а также отсек фюзеляжа среднего магистрального самолета. Анализируются перемещения и напряжения, возникающие при нагружении конструкций. Данные, полученные на континуальных моделях, удовлетворительно согласуются с соответствующими параметрами тонкостенных моделей реальных объектов. При соблюдении рекомендаций, сформулированных в [6], погрешность результатов, полученных на континуальных моделях, в настоящей работе составила 3-5%, что свидетельствует о применимости модели переменной плотности на ранних стадиях проектирования авиационных конструкций.

Оптимизация распределения материала в континуальной модели

Рассматривается следующая оптимизационная задача. За проектные переменные принимается плотность материала в трехмерных конечных элементах ρ_i при фиксированных величинах $\overline{\overline{E}}$ и $\overline{\overline{\sigma}}$, где i - номер элемента.

Минимизируется масса упругой среды

$$m(\rho) = \sum_{i=1}^{n} \rho_i V_i \to \min \tag{3}$$

при функциональных ограничениях g_j

$$g_j(\rho) = C_j(\rho) - \overline{C}_j \le 0, \quad (j=1, 2, ..., p), \tag{4}$$

где C_j и $\overline{C}_j$ – j-я переменная состояния конструкции и ее допускаемое значение; p – количество функциональных ограничений. Переменными состояния конструкции C_j являются обобщенные перемещения и критические усилия потери устойчивости.

Область поиска задается так:

$$\rho_i \ge \rho_i^{\min}, \quad (i = 1,2,...,n), \tag{5}$$

где $\rho_i^{\min}$ – ограничение снизу для i-ой проектной переменной, которое учитывает ограничение на напряжение в соответствующем элементе на основе концепции *полнонапряженности* [3,1] в следующем виде:

$$\rho_i^{\min} = \rho_i \max_j(\sigma_{ij} / \overline{\sigma}), \quad (j=1,2,...,s), \tag{6}$$

где s – количество случаев нагружения; σ_{ij} – эквивалентные напряжения по принятой теории прочности в i-ом элементе при j-м случае нагружения.

В работе [7] предложен алгоритм решения задачи оптимизации (3) – (6) с учетом ограничений по прочности и жесткости. Движение в области поиска осуществляется на основе последовательного чередования "пробных" и "рабочих" шагов. На этапе "пробных" шагов определяются значения проектных переменных на основе линейных аппроксимаций функциональных ограничений. "Рабочий" шаг характерен тем, что для него выполняется анализ напряженно-деформированного состояния и устойчивости конструкции в полном объеме. В работе [8] этот подход дополнен учетом ограничений на значения критических нагрузок, приводящих к выпучиванию упругой системы.

Алгоритм [7, 8] реализован на языке Visual Basic for Applications в среде прикладного программного интерфейса МКЭ-системы NASTRAN. Для испытания работоспособности алгоритма и программного обеспечения проведено исследование на специально подобранных тестовых задачах, показавшее удовлетворительную сходимость алгоритма. Для практических целей оказалось достаточно 10 – 30 "рабочих" шагов алгоритма.

Проектирование авиационных конструкций

Анализ распределения материала и основных путей передачи сил в теоретически оптимальной континуальной конструкции, полученной по алгоритму [7, 8], позволяет, во-первых, разработать варианты рациональных силовых схем с использованием стратегии [1], во-вторых, спрогнозировать массу конструкции на основе подхода [2] с учетом статической аэроупругости [9, 10], а также широкого спектра функциональных ограничений и конструктивно-технологических требований.

В настоящей работе рассматриваются задачи проектирования стреловидных крыльев, несущих поверхностей малого удлинения [11], в том числе круглого в плане крыла, фюзеляжей в зоне больших вырезов.

При проектировании стреловидного крыла удалось выявить дополнительные элементы и места их расположения [12], позволяющие уменьшить угол закручивания концевого сечения крыла на 32% при увеличении массы полнонапряженной конструкции на 2%.

Исследования на континуальной модели отсека фюзеляжа с прямоугольным вырезом позволили выявить новое техническое решение, для которого теоретически необходимая масса материала для компенсации выреза на 14% меньше, чем для традиционного варианта силовой схемы.

Таким образом, идея модели переменной плотности использована в развитии технологии проектирования авиационных конструкций. Для обоснованного выбора рациональной силовой схемы и достоверной оценки массы разрабатываемой конструкции технология учитывает функциональные ограничения по прочности, жесткости и устойчивости, а также зависимость распределения аэродинамических нагрузок от деформаций упругой системы. Новая технология проектирования позволяет на ранних этапах разработки выявить и реализовать существенные резервы снижения массы конструкции.

Литература

1. Комаров В.А. Проектирование силовых схем авиационных конструкций // Актуальные проблемы авиационной науки и техники. М.: Машиностроение, 1984. С.114-129.
2. Комаров В.А. Весовой анализ авиационных конструкций: теоретические основы // Полет. 2000. №1. С. 31-39.
3. Комаров А.А. Основы проектирования силовых конструкций. Куйбышев: Куйбышевск. книж. изд-во, 1965. 82 с.
4. Bendsoe M.P., Kikuchi N. Generating Optimal Topologies in Structural Design Using a Homogenization Method // Computer Methods in Applied Mechanics and Engineering. 1988. V. 71. P. 197–224.
5. Eschenauer H.A., Olhoff N. Topology optimization of continuum structures: A review // Appl. Mech. Rev. 2001. V. 54. № 4. P. 331-389.
6. Болдырев А.В., Комаров В.А., Лаптева М.Ю. Об оценке точности прогнозирования деформаций крыла на основе модели переменной плотности // Вестник КГТУ. 2009. № 3. С. 13-15.
7. Болдырев А.В. Структурная оптимизация крыльев с учетом требований прочности и жесткости // Вестник МАИ. 2009, Т16, № 3, С. 15-21.
8. Болдырев А.В. Развитие технологии проектирования авиационных конструкций на основе модели переменной плотности // Полет**.** 2009. № 11. С. 23-28.
9. Болдырев А.В., Комаров В.А., Лаптева М.Ю., Попович К.Ф. Учет статической аэроупругости на ранних стадиях проектирования // Полет. 2008. № 1. С. 34–39.
10. Болдырев А.В., Комаров В.А. Структурная оптимизация несущих поверхностей с учетом статической аэроупругости // Известия вузов. Авиационная техника. 2008. № 2. С. 3-6.
11. Болдырев А.В. Весовой анализ крыльев нетрадиционной конфигурации // Полет**.** 2009. № 10. С. 57-60.
12. Болдырев А.В. Структурная модификация тонкостенных конструкций с учетом требований жесткости // Проблемы прочности и пластичности. 2008. Вып. 70. С. 175-183.

Программное обеспечение онтологии концептуального проектирования самолета

Н.М. Боргест, А.В. Глушков, А.В. Петровский, Д.В. Шустова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Онтология проектирования

Онтология в информатике рассматривается как попытка всеобъемлющей и детальной формализации некоторой области знаний. Такая формализация состоит из структуры данных, содержащей все релевантные классы объектов, их связи и правила, принятые в этой области [2]. Этот изначально философский термин в информатике принял самостоятельное значение. Существенные отличия заключаются в необходимости иметь формат, который сможет обрабатывать компьютер, и конкретные цели решения проектных и конструкторских задач.

Проектирование как процесс и область знаний об артефакте является предметом онтологического анализа ученых-исследователей и узких специалистов. Онтология проектирования как научное направление включает в себя:

- исследование понятийного аппарата и разработку на его основе тезауруса,
- анализ критериев и моделей проектируемого объекта, методов и сценариев проектирования,
- сбор и обработку информации об объекте как системы и составляющих его элементах.

Онтология проектирования - это структурированное знание об объекте и методах его проектирования, это совокупность тезауруса, баз данных и процедур, алгоритмов оптимизации и учета проектной неопределенности.

Концептуальное проектирование самолета

Создание любого технического объекта, особенно такого сложного как самолет, начинается с выработки его концептуальных решений. Формирование технического задания, критериев оценки будущего самолета, требований, ограничений и параметров, выбор схемных компоновочных решений (внешних аэродинамических форм и внутренних - кабины пилотов и пассажирского салона), силовой установки, состава систем, оборудования и авионики – все это служит основой для принятия решения о дальнейшей проработке проекта [4].

Онтология концептуального проектирования самолета содержит весь предыдущий опыт проектирования подобного класса самолетов: подробные базы данных по летно-техническим характеристикам, геометрии, аэродинамике, весовому совершенству и оборудованию самолетов, по параметрам и характеристикам силовых установок. Формализуемые этапы, методы и алгоритмы проектирования составляют базу процедур, а сценарии и выработанные правила базу правил. Общая совокупность баз и тезаурусов предметной области может рассматриваться как база знаний [1].

«Работающая» онтология предполагает общение специалиста и потенциального заказчика с компьютером, в результате которого можно будет получить первый образ «желаемого» самолета.

Программное обеспечение онтологии проектирования

Программное обеспечение должно позволять разрабатываемой онтологии концептуального проектирования самолета обладать:

- обширной базой знаний,
- удобным пользовательским интерфейсом,
- умением генерировать на основе заложенной информации новые знания,

– возможностью получить на основе запроса пользователя первоначальный облик самолета, а также оценки полученных решений по аванпроекту.

Подбор инструментов в виде программного обеспечения является важным этапом в создании интеллектуальных систем. Исследования проводились по следующим направлениям:

–сравнительный анализ и выбор СУБД,

–сравнительный анализ и выбор онтологической системы,

–обзор инженерных систем,

–обзор языков высокого уровня.

Выбор СУБД. На основе анализа публикаций и форумов, с привлечением экспертов проводилось сравнение таких СУБД как, Microsoft Access, Interbase, Firebird, MySQL, Microsoft SQL Server, Oracle по нескольким критериям:

- скорость работы при размерах таблицы в 10000 записей,
- удобство работы,
- возможность самообучения,
- возможность интеграции в другие СУБД,
- стоимость лицензии,
- требования к аппаратному обеспечению.

Сравниваемые системы ранжировались по общему количеству набранных ими баллов. Для того чтобы найти оценку системы по i-му критерию, необходимо каждый балл полученный системой по i-му критерию умножить на вес этого критерия [6]. Результаты метода экспертных оценок представлены в таблице 1.

Таблица 1. Оценка СУБД методом экспертных оценок

	Вес критерия	Оценка по i-му критерию [Экспертная оценка]					
		Access	SQL Server	***Interbase***	MySQL	Oracle	Firebird
Высокая скорость работы	0.1	5 [0.5]	5 [0.5]	5 [0.5]	5 [0.5]	5 [0.5]	5 [0.5]
Удобство работы (интуитивность)	0.15	5 [0.75]	3 [0.45]	2 [0.3]	1 [0.15]	4 [0.6]	1 [0.15]
Возможность самообучения	0.15	5 [0.75]	3 [0.45]	4 [0.6]	3 [0.45]	2 [0.3]	1 [0.15]
возможность интеграции в другие СУБД	0.2	1 [0.2]	5 [1]	1 [0.2]	5 [1]	5 [1]	5 [1]
Стоимость лицензии	0.3	3 [0.9]	1 [0.3]	3 [0.9]	5 [1.5]	2 [0.6]	5 [1.5]
Требования к аппаратному обеспечению	0.1	5 [0.5]	4 [0.4]	5 [0.5]	5 [0.5]	3 [0.3]	5 [0.5]
ΣI интегральная экспертная оценка систем		3.6	3.1	3	4.1	3.3	3.8

Выбор Онтологической Системы. При создании онтологий целесообразно пользоваться подходящими инструментами. Инструментальные программные средства, созданные специально для проектирования, редактирования и анализа онтологий, называются *редакторами онтологий* или *онтологическими системами*.

В настоящее время на рынке программных средств достаточно активно продвигаются более 50 редакторов онтологий и систем поддержки их жизненного

цикла [3]. Для выбора онтологической системы были использованы результаты «Сравнения популярности редакторов онтологий», приведенных на сайте [9]. Из наиболее популярных и развивающихся онтологических систем были выбраны Protégé, OpenCyc, OntoStudio. Так же включили в сравнительный анализ систему Magenta, использующуюся для обучения студентов СГАУ по дисциплине «Онтология производственной сферы» [2]. В качестве критериев для сравнительного анализа онтологических систем были рассмотрены:

- открытость системы,
- удобство работы (интуитивность),
- наличие доступного материала для самообучения,
- возможность импортирования данных из СУБД,
- поддержка формата OWL.

Результат анализа онтологических систем представлены в таблице 2.

Таблица 2. Оценка онтологических систем методом экспертных оценок

	Вес критерия	Оценка по i-му критерию [Экспертная оценка]			
		Protégé	**OpenCyc**	**OntoStudio**	**Magenta**
Открытость системы	0.3	5 [1.5]	5 [1.5]	4 [1.2]	2
Удобство работы (интуитивность)	0.15	4 [0.6]	2 [0.3]	3 [0.45]	2 [0.3]
Наличие доступной помощи для возможности самообучения	0.2	3 [0.6]	2 [0.4]	2 [0.4]	2 [0.4]
Возможность импортирования данных из выбранной СУБД - MySQL	0.15	5 [0.75]	1 [0.15]	5 [0.75]	1 [0.15]
Возможность использования языка онтологий OWL	0.1	5 [0.5]	1 [0.1]	5 [0.5]	5 [0.5]
Мультиагентность	0.1	5 [0.5]	1 [0.1]	5 [0.5]	5 [0.5]
ΣI — интегральная экспертная оценка систем		3.95	2.15	3.3	1.35

Примеры создания онтологий приведены на Рисунках 1 (А,Б). На Рисунке 1(А) рассмотрен пример решения задачи по выбору самолета под заданное техническое задание в системе Magenta. На Рисунке 1 (Б) – Аэроонтология, созданная в Protégé.

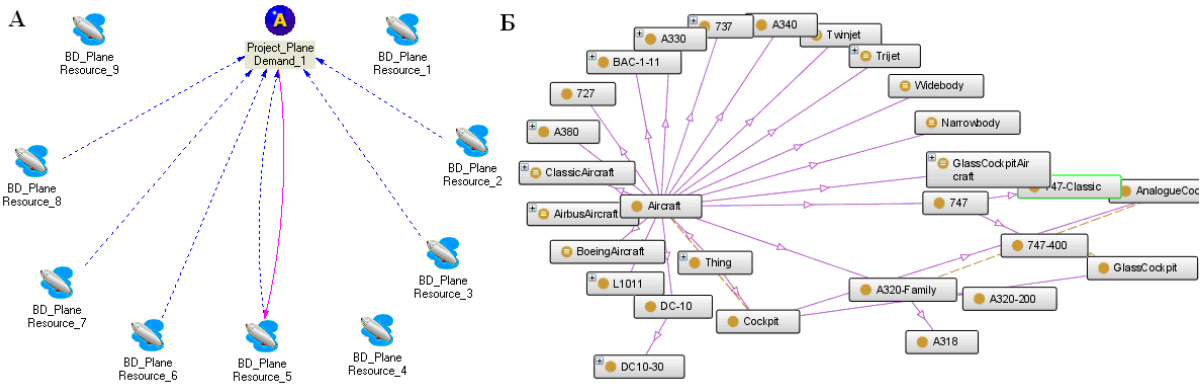


Рис. 1. (А) Выбор самолета по ТЗ (Magenta), (Б) Аэроонтология (Protégé)

В результате исследований онтологических систем Protégé, OpenCyc, OntoStudio, Magenta были сделаны следующие выводы:

– системы Protégé и OntoStudio являются актуальными, новые версии были выпущены в этом году,

– установочные файлы систем Protégé, OpenCyc и OntoStudio можно скачать с официальных сайтов программ. Это позволяет свести к нулю затраты на приобретение и владение продуктом.

– компания Magenta Technology предоставляет бесплатные лицензии университетам, научно-исследовательским и академическим учреждениям, но это лишь базовые версии, в основном компания работает с клиентами в индивидуальном порядке [10]. С одной стороны это позволяет учитывать потребности заказчика, а не предлагать найти в стандартной версии то, что было бы необходимо именно ему. С другой – такой подход не является бесплатным, и поэтому нельзя назвать онтологическую систему Magenta общедоступной.

– онтологические системы отличаются спецификой работы и для детального анализа исследователю нужно обладать начальной подготовкой, т.к. встроенной помощи системы не имеют, а материалы и статьи [2,5,8], позволяют лишь немного облегчить освоение незнакомых программ,

– было отмечено, что все исследуемые системы имеют минимальные системные требования и просты в установке,

– все исследуемые системы, кроме OpenCyc поддерживают язык онтологий OWL (Web Ontology Language), который необходим для построения онтологий,

– Protégé, OntoStudio, Magenta являются мультиагентными системами. Они представляют собой самоорганизующуюся сеть программных объектов, постоянно и параллельно работающих над установлением, пересмотром и адаптацией связей.

– для оценки удобства работы была предпринята попытка решить задачу по выбору двигателя на этапе предварительного проектирования самолета [4] в системах Protégé, OpenCyc, OntoStudio и Magenta. Было отмечено, что Protégé обладает более интуитивным интерфейсом по сравнению с другими системами.

В результате проведенных исследований была выбрана онтологическая система Protégé. Помимо выбранных критериев Protégé обладает рядом преимуществ:

– является наиболее распространенной онтологической системой для создания и сопровождения онтологий [9],

– отличается всеобъемлющими функциями интуитивного моделирования онтологии,

– имеет возможность импорта различных структур, схем и моделей,

– поддерживает форматы файлов OWL, RDF, РИФ, ObjectLogic,

– имеет возможность импортирования в UML 2.0, базы данных (Oracle, MS-SQL, DB2, MySQL), Excel, Outlook.

Инженерные системы. Для получения первоначального облика самолета могут быть использованы CAD-системы, такие как AutoCad, Компас, SolidWorks, и CAD/CAM/CAE-системы, такие как CATIA, ProIngineer, UNIGRAPHICS и др. Эти системы поддерживают большое количество расширений, что дает возможность импорта и экспорта изображений и чертежей.

Языки программирования. В прикладных онтологиях фиксируются знания и договоренности специалистов предметных областей. Поэтому язык представления онтологий должен быть одновременно удобен для специалистов и открыт для пользователей. При этом внутреннее представление онтологий для компьютерного использования и межмашинного обмена должно быть стандартизованным. В работах

[1,7] рассмотрено применение языков программирования Prolog и CLIPS для создания экспертных систем.

На основе изложенной концепции [1] и выбранного программного обеспечения в настоящее время в СГАУ разрабатывается онтология концептуального проектирования самолета.

Онтологический подход позволяет выявлять, накапливать и передавать извлеченные и структурированные знания при подготовке авиационных специалистов, получать новые знания на основе анализа структуры предметной области, а так же повысить качество проектирования.

Литература

1. Боргест Н.М. Автоматизация предварительного проектирования самолета. Учеб. пособие. Самара: САУ, 1992 – 92 с.
2. Боргест Н.М., Симонова Е.В. Основы построения мультиагентных систем, использующих онтологию. Учеб. пособие. Самара: СГАУ, 2009 – 80 с.
3. Гаврилова Т.А. Формирование прикладных онтологий, Санкт-Петербург, СПбГПУ – 7с.
4. Комаров В.А., Боргест Н.М. и др. Концептуальное проектирование самолета. Учеб. пособие. Самара: СГАУ, 2007 – 92 с.
5. Муромцев Д.И. Онтологический инжиниринг знаний в системе Protege: Методическое пособие. - СПб: СПбГУ ИТМО, 2007. - 62 с.
6. Отбор проектов для финансирования: экспертные методы, Методический журнал [Электронный ресурс] / А.В. Корчагин// Режим доступа: http://www.reglament.net/bank/credit/2005_2_article_1.htm
7. Частиков А.П., Гаврилова Т.А., Белов Д.Л. РАЗРАБОТКА ЭКСПЕРТНЫХ СИСТЕМ. СРЕДА CLIPS. - СПб: «БХВ-Петербург», 2003 – 393 с
8. Natalya F. Noy and Deborah L. McGuinness. ``Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
9. http://ais-portal.ru/?p=161 «Сравнения популярности редакторов онтологий»
10. http://www.magenta-technology.ru/ru/technology/overview/different/

Тезаурус как основа онтологии концептуального проектирования самолета

Н.М. Боргест, Д.В. Шустова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Одним из перспективных направлений искусственного интеллекта является изучение онтологий конкретных предметных областей и построение их моделей. Проектирование летательных аппаратов - сложный наукоемкий процесс, на каждом этапе которого число данных и число участников непрерывно увеличивается. Взаимопонимание между участниками различных организаций в большей степени зависит от имеющейся у них информации и способности ее эффективно использовать. Поэтому существует необходимость управлять созданием, распространением, обработкой и использованием информации [7]. Совокупность этих процессов есть *управление знаниями,* которое является ядром концепции *онтологического инжиниринга* [3].

Онтология определяет термины предметной области, дает их толкование, содержит утверждения, которые ограничивают смысл этих терминов. Совокупность терминов предметной области образуют ее терминологическую систему. Онтология предметной области есть явное представление концептуализации предметной области, совокупность соглашений [4]. Терминологическая система онтологии предметной области может быть представлена в виде тезауруса - особой разновидности словаря, в котором указаны семантические отношения между лексическими единицами [9]. Структура тезауруса может быть выстроена на основе семантических сетей, которые отражают семантику предметной области в виде понятий и отношений и являются одним из наиболее удобных способов представления знаний. Семантические сети могут быть записаны практически на любом языке программирования. Самые популярные в этом отношении языки Lisp и Prolog, так как они содержат встроенный механизм, в котором логическая операция вывода является основной. Например, язык Prolog был успешно применен для построения экспертной системы, предназначенной для выбора схемы самолета [2].

Кроме языков программирования для создания тезауруса могут быть использованы инструментальные программные средства, созданные специально для проектирования, редактирования и анализа онтологий, называемые *редакторами онтологий* или *онтологическими системами*. Они позволяют применять онтологический подход для решения широкого круга задач прикладного характера. Например, специализированная система «OntoEditor» была применена при исследовании характеристик АвиаОнтологии. На Рисунке 1 представлен визуальный образ Авиаонтологии [5]. При разработке онтологии, для того чтобы не начинать работу «с нуля», могут быть использованы терминологические и лингвистические ресурсы, например тезаурус русского языка РуТез, разработанный АНО Центр информационных исследований. Это позволит существенно сократить время, которое было бы потрачено на разработку нового ресурса. Так же для создания тезауруса могут быть использованы книги и информационно-справочные порталы для профессионалов и любителей авиации, на основе которых уже разрабатывалась интеллектуальная справочная система по самолетам [8]. Визуальные методы проектирования онтологий способствуют более быстрому и полному пониманию структуры знаний предметной области, что особенно ценно для исследователей, работающих в новой предметной области.

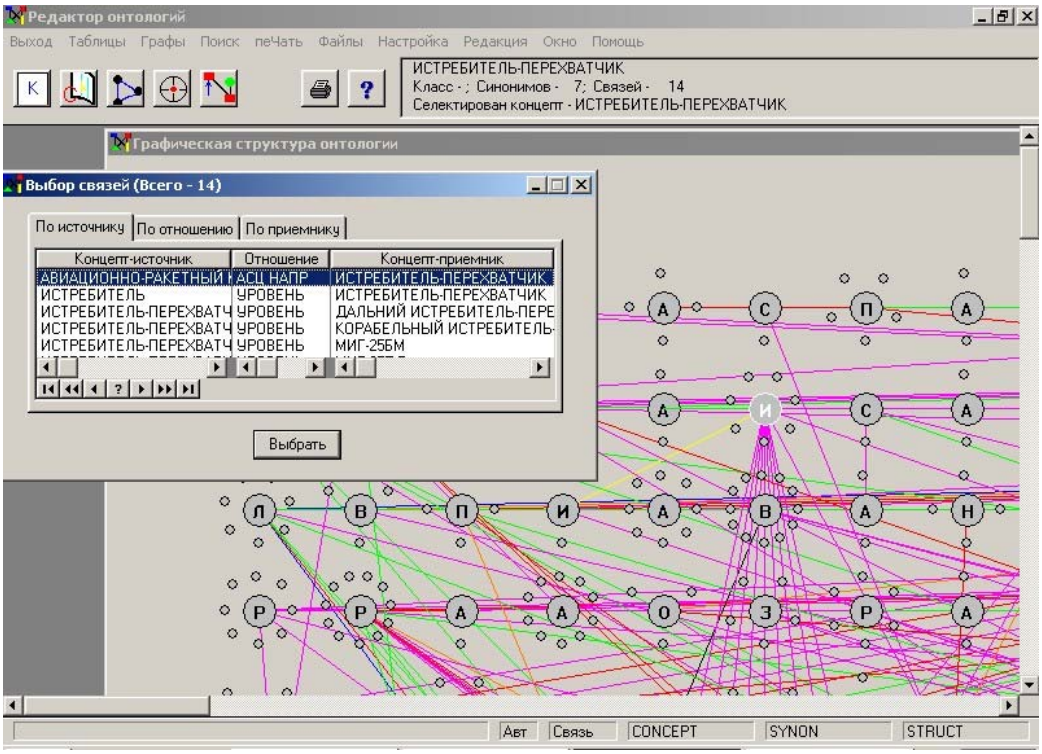


Рис. 1. Визуальный образ АвиаОнтологии

Во многих дисциплинах сейчас разрабатываются стандартные онтологии, которые могут использоваться экспертами по предметным областям для совместного использования и аннотирования информации в своей области. Например, в области медицины созданы большие стандартные, структурированные словари, такие как SNOMED и семантическая сеть Системы Унифицированного Медицинского Языка (the Unified Medical Language System). Также появляются обширные общецелевые онтологии. Например, Программа ООН по развитию (the United Nations Development Program) и компания Dun&Bradstreet объединили усилия для разработки онтологии UNSPSC, которая предоставляет терминологию товаров и услуг [11]. Кроме того, при создании большой онтологии, можно интегрировать несколько существующих онтологий, описывающих части большой предметной области или использовать повторно основную онтологию, такую как UNSPSC, и расширить ее для описания интересующей предметной области.

Protégé – одна из наиболее популярных систем работы с онтологиями, созданная в Стэндфордском университете (США). В системе Protégé все понятия предметной области делятся на классы, подклассы, экземпляры. Экземпляры могут быть как у класса, так и подкласса и описываются они фреймом [6]. Разработка онтологий для Protégé состоит из 5 шагов [10]:

1. выделение области, масштаба, границ онтологии,
2. определение классов,
3. организация иерархии классов,
4. формирование фреймов для описания классов, подклассов, экземпляров, через определение слотов, т.е. свойств,
5. определение значений.

В Protégé существует несколько стратегий проектирования онтологий:

– нисходящая (сверху вниз),
– восходящая,
– комбинирование стратегий.

Узлы в онтологии должны находиться на одном уровне. Их должно быть не меньше двух и не больше двенадцати, а иначе выделяется еще один подкласс.

Слот относится к самому высокому классу в иерархии. Слоты делятся на:

– внутренние (определяются внутренним свойством объекта),
– внешние (определяются внешним свойством объекта),
– части,
– другие.

PProtege состоит из закладок (tabs), которые отображают различные аспекты модели знаний. Наиболее важной закладкой, является закладка классов (Classes). Обычно классы соответствуют объектам или типам объектов, в некой предметной области [6]. В примере с выбором самолета по заданным параметрам, классы включают в себя классы самолетов, а именно, Гражданские и Военные, а также их виды, такие как Магистральные, Местных авиалиний, Грузовые и т.д.

Классы в Protege отображаются в виде иерархии наследования (inheritance hierarchy), которая располагается в области просмотра, называемой Class Browser (или навигатор классов) в левой части закладки классов (Рисунок 2 – Создание класса «Гражданские самолеты»).

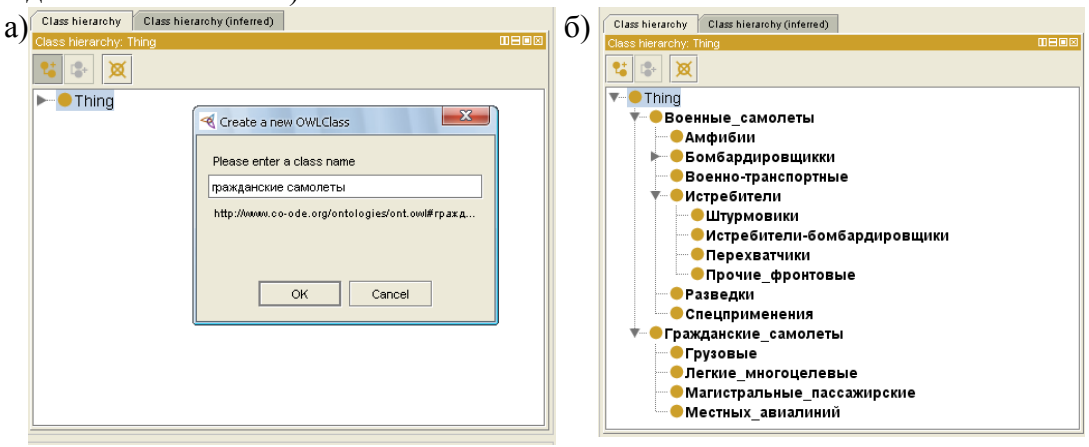


Рис. 2. (а) Создание класса «Гражданские самолеты», (б) Иерархия наследования.

Экземпляры классов – это данные базы знаний. Перед вводом конечных данных необходимо проверить структуру проекта, так как необходимость ее изменения может повлечь за собой потерю уже введенной информации. На любом этапе создания тезауруса можно посмотреть семантические связи между его элементами (Рисунок 3).

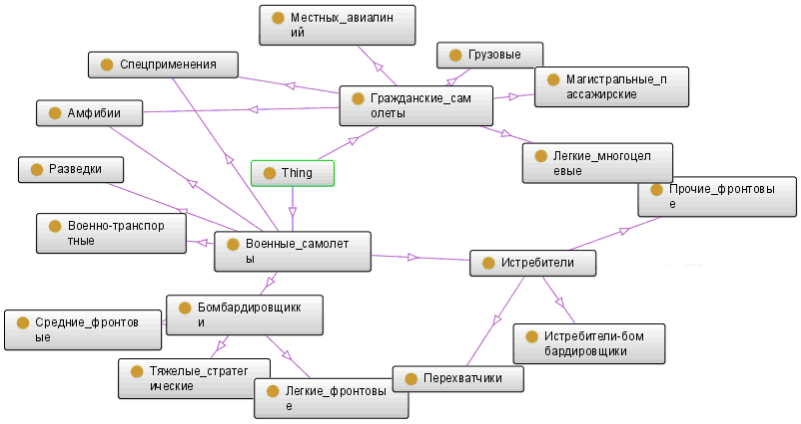


Рис. 3. Семантическая сеть

Онтология всегда отражает взгляд аналитика, т.е. всегда субъективна, а поэтому не существует единственно правильного способа ее создания. Несмотря на это можно выделить некоторые основные принципы, на которых базируется создание онтологии:

Ясность — онтология должна эффективно передавать смысл введенных терминов, ее определения должны быть объективны, а для их объективизации должен использоваться четко фиксированный формализм.

Согласованность — все определения должны быть логически непротиворечивы.

Расширяемость — необходимо проектировать онтологию так, чтобы ее словари терминов можно было расширять, используя уже существующие понятия.

Минимум влияния кодирования — концептуализация онтологии должна быть специфицирована на уровне представления, а не символьного кодирования.

Минимум онтологических обязательств — онтология должна содержать только наиболее существенные предположения о моделируемой предметной области, чтобы оставлять свободу расширения и специализации.

Для создания тезауруса существует стандарт [1], который устанавливает правила разработки, структуру, состав и форму представления информационно-поисковых тезаурусов, разрабатываемых в рамках автоматизированных информационных систем и сетей научно-технической информации.

Отношения в тезаурусе могут быть разделены на два класса: иерархические и ассоциативные.

Основным преимуществом использования онтологии в качестве инструмента познания является системный анализ предметной области. Прикладное значение онтологического подхода позволяет:

– выявлять, накапливать и передавать извлеченные и структурированные знания при подготовке авиационных специалистов;

– разработать дружественный интерфейс интеллектуальных систем автоматизированного проектирования;

– получить новые знания на основе анализа структуры предметной области;

– повысить качество проектирования.

Литература

1. ГОСТ 7.25-2001. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления
2. Боргест Н.М. Автоматизация предварительного проектирования самолета. Учеб. пособие. Самара: САУ, 1992 – 92 с.
3. Гаврилова Т.А., Хорошевский В.Ф., 2000. Базы знаний интеллектуальных систем / Учебник для вузов. – СПб, Изд-во “Питер”, 2000
4. Клещев А.С., Артемьева И.Л. Математические модели онтологий предметных областей. Часть 1. Существующие подходы к определению понятия "онтология" // Научно–техническая информация, Серия. 2, 2001, № 2. С. 20–27.
5. Н.В. Лукашевич, О.А. Невзорова. АвиаОнтология-2004: анализ современного состояния ресурса//Компьютерная лингвистика и интеллектуальные технологии: Тр. Междунар. Конференция Диалог'2004.– М.: Наука, 2004. – С.424-430.
6. Муромцев Д.И. Онтологический инжиниринг знаний в системе Protege: Методическое пособие. - СПб: СПбГУ ИТМО, 2007. - 62 с.
7. Шустова Д.В. Разработка тезауруса прикладной онтологии. «Аспирантский вестник Поволжья», № 1-2, 2010. – с.108-110.
8. Щемелев А. Описание ИИС по самолетам.
9. http://www.slovopedia.com. Большой энциклопедический словарь (БЭС)
10. www.protege.org
11. http://www.unspsc.org

Математическое моделирование весовой и аэродинамической эффективности телескопических крыльев

А.А. Вырыпаев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Обсуждается потребность в поиске новых аэродинамических компоновок ЛА для повышения их эффективности и расширения функциональных возможностей. Анализируются и классифицируются различные направления в поиске крыльев с изменяемой конфигурацией, называемые в англоязычной технической литературе «Morphing wing».

Отмечается, что большинство публикации, в основном в Интернет, касаются предложения новых идей и обсуждения их достоинств, цифровые оценки и методы проектирования практически отсутствуют. В тоже время сложность и стоимость проектов современных летательных аппаратов требует развития и применения проектных технологий «точного попадания в цель» - «Concurrent design». Эти технологии основываются на использовании высокоточного математического моделирования на ранних стадиях проектирования.

Применение новых технологий проектирования рассматривается на примере телескопического крыла.

Для оценки весовых (массовых) характеристик крыла используется безразмерный коэффициент силового фактора, предложенный в работе: Комаров В.А. Весовой анализ авиационных конструкций: теоретические основы. «Полёт», 2000, №1. Коэффициент силового фактора вычисляется через оптимизацию распределения материала в 3D модели крыла переменной плотности.

Для расчёта распределения циркуляции скорости по размаху крыла и определения нагрузок используется метод дискретных вихрей.

В докладе приводятся результаты исследования влияния ряда геометрических параметров на вес и аэродинамическое качество телескопического крыла. В качестве проектных переменных рассмотрены: соотношения длины подвижной и неподвижной части, отношение хорд, положение подвижной части крыла по хорде неподвижной и некоторые другие. Численные исследования выполнены в среде Nastran, Tornado и оригинальной реализации МДВ СГАУ.

Сформулирована задача об оптимизации параметров телескопического крыла. В качестве целевой функции используется взлётная масса ЛА.

Обсуждаются области рационального применения телескопических крыльев.

Информационная модель технологических данных предприятия ракетно-космической техники

А.А. Греднева
Государственный научно-производственный ракетно-космический центр "ЦСКБ-Прогресс",
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Государственный научно-производственный ракетно-космический центр "ЦСКБ-Прогресс" – ведущее российское предприятие по разработке, производству и эксплуатации ракет-носителей среднего класса и автоматических космических аппаратов для дистанционного зондирования Земли и научного назначения.

«ЦСКБ-Прогресс» выполняет крупномасштабные и наукоёмкие проекты, характеризующиеся многономенклатурностью изделий и длительным циклом разработки, производства и эксплуатации. Для предприятия наиболее актуальна проблема обеспечения интеграции и совместного использования информации, порождаемой на всех этапах жизненного цикла изделия, в рамках единого информационного пространства.

В настоящее время на «ЦСКБ-Прогресс» осуществляется программа по внедрению информационных технологий и совершенствованию методов проектирования и хранения данных об изделиях. Специалисты конструкторских и технологических подразделений осваивают CAD/CAM-систему высокого уровня «Pro/Engineer» (компания «PTC») и PDM-систему с широкими функциональными возможностями «Windchill» («PTC»). Совместное использование систем «Pro/Engineer» и «Windchill» позволило реализовать технологию сквозного нисходящего проектирования изделий. Также осуществляется внедрение САПР ТП «Вертикаль» и её интеграция с системой «Windchill».

Выбор PDM-системы

Для предприятия «ЦСКБ-Прогресс» характерно проведение большого количества изменений в документации. Возникает необходимость не только отслеживать все происходящие изменения, но и управлять ими. Система «Лоцман PLM» (на предприятии с 2006 года) подходит для хранения данных как электронный архив, но не позволяет эффективно управлять изменениями. Специалистами отделов САПР была проведена большая работа по анализу функциональных возможностей PDM-систем разного уровня. В результате остановили выбор на системе «Windchill», которая отвечает всем необходимым требованиям. В настоящее время «Windchill» - единая среда разработки изделий на предприятии. Система позволяет пользователям получать актуальную информацию о состоянии конструкторско-технологической подготовки производства по изделию.

Нисходящее проектирование в САПР «Pro/Engineer»

В качестве базовой САПР для предприятия была выбрана «Pro/Engineer» как система с оптимальным соотношением цены и функциональных возможностей. Использование нисходящего проектирования предусматривает создание многоуровневой управляющей структуры, содержащей всю основную геометрию и базовые параметры проектируемого изделия. В основе управляющей структуры лежат модели мастер-геометрии (скелетоны). Данные из мастер-геометрии верхнего уровня передаются на нижестоящий уровень и дополняются уточняющей геометрией, позволяя, таким образом, сформировать концептуальную схему проектируемого изделия. Разветвленная схема управляющей структуры дает возможность организовать

параллельную работу разных подразделений проектантов и конструкторов. Заключительным этапом является создание реальных конструкторских моделей деталей и узлов со ссылками на мастер-геометрию и выпуск конструкторской документации (КД).

Электронная модель изделия

В настоящее время электронная модель изделия (ЭМИ) в системе «Pro/Engineer» является электронным конструкторским документом, содержащим всю необходимую информацию об изделии; служит основой для создания чертежей, проведения различных видов анализов, разработки технологической оснастки и управляющих программ для станков с ЧПУ, подготовки интерактивных электронных документов. ЭМИ состоит из геометрической модели изделия, определенного количества атрибутов модели и включает технические требования. При разработке КД электронные модели являются первичными по отношению к КД на бумажном носителе. На всех стадиях разработки изделия обеспечивается соответствие (ассоциативная связь) электронных моделей и выпускаемых по ней чертежей.

Внедрение PDM-системы «Windchill»

PDM-система «Windchill» является универсальной системой с широкими возможностями. Приспособить такую систему под нужды конкретного предприятия и его подразделений – важная задача, которая стоит перед специалистами отделов САПР предприятия и консалтинговой фирмы.

Проект по внедрению информационных технологий в каждом подразделении содержит следующие основные этапы:

1. Организационный этап
2. Исследования существующих рабочих процессов. Создание модели процессов «как есть»
3. Подготовка предложений по совершенствованию процессов. Создание модели «как должно быть»
4. Техническое обеспечение проекта
5. Обучение специалистов отделов САПР
6. Настройка программного обеспечения
7. Подготовка к опытной эксплуатации
8. Опытная эксплуатация
9. Промышленная эксплуатация.

Внедрение «Windchill» началось с конструкторских подразделений. Были созданы контексты – «изделия» и «библиотеки». В контекстах «изделия» подразумевается расположение структуры разрабатываемых изделий (в виде конечного изделия и частей) и связанных с ними объектов: CAD-документов и документов (приказов на запуск в производство, листов запуска в производство, извещений об изменении деталей и сборочных единиц и др.). «Библиотеки» содержат части, которые используются более чем в одном изделии, а также стандартные элементы.

Информационная модель технологических данных в «Windchill»

Конструкторская спецификация изделия в «Windchill» носит название eBOM (engineering Bill of Material) и представляет собой структуру изделия «как спроектировано» (as-designed). Технологическая спецификация называется mBOM (manufacturing Bill of Material) – структура изделия «как будет собираться» (as-planned). Группы пользователей конструкторских и технологических подразделений отнесли к разным «организациям» (доменам) с целью разграничения прав доступа. Поэтому структуры изделий – конструкторская и технологическая – располагаются каждая в своей «организации» и разворачивается соответственно конструкторами и технологами.

Рассмотрим работу технологов. После того, как создан объект «изделие», развернута его структура и конструкторской документации присвоено состояние «выпущено», на входящие в него детали и сборочные единицы создаются расцеховки, в которых указывают маршрут их изготовления – номера цехов. Расцеховку согласовывают ведущие технологические отделы, прорабатывая КД на технологичность. Технологи каждого цеха-изготовителя создают объект «внутрицеховой техпроцесс», прикрепляя его к учетной карточке компонента (части) и связывая его с расцеховкой. Для управления расцеховками и техпроцессами используется модуль «Windchill MPMLink».

Для написания техпроцессов на «ЦСКБ-Прогресс» используется САПР ТП «Вертикаль». Специалисты предприятия совместно с консалтинговой фирмой «PTS» создали модуль интеграции «Вертикали» и «Windchill». Системы взаимодействуют следующим образом. Технолог запускает «пустой» техпроцесс в САПР ТП «Вертикаль» и вводит необходимые данные (наполняет техпроцесс). После выполнения действия «сдать на хранение», в Windchill передаются следующие файлы:

- файл с информацией о ТП во внутреннем формате VTP программы – среды разработки «Вертикали»;
- *.xml-файл, который содержит атрибуты ТП и ассоциируется с объектом ТП;
- *.pdf-файл, предназначенный для визуализации ТП.

Техпроцесс становится доступным для других пользователей.

При открытии техпроцесса в «Вертикали» технолог выбирает основной материал из справочника «Материалы и сортаменты», указывает номер цеха-изготовителя. Затем добавляет операции и переходы, необходимые для изготовления детали (или сборки узла). На операцию назначаются оборудование, инструмент, указывается исполнитель.

Если для выполнения операции техпроцесса необходима специальная технологическая оснастка (СТО) или управляющая программа (УП), то на операцию назначается заявка на СТО (заявка на УП). Заявка заполняется в «Вертикали». На основе заявок при выполнении соответствующей команды технолог генерирует документ «ведомость оснащения». В качестве первичного содержимого документа выступает *.pdf-файл, в качестве вторичного - *.xml-файл.

Также технолог создает документы «техническое задание (ТЗ) на проектирование СТО», «ТЗ на УП». Затем генерирует операции, которые отражаются в окне структуры техпроцесса, и ресурсы: станки, квалификации, СТО, комплекты УП.

Если после проведения изменения на деталь возникла необходимость изменить техпроцесс, технолог создает новую версию объекта «техпроцесс», запускает его в САПР ТП «Вертикаль» и вносит необходимые изменения. После сдачи на хранение новой версии техпроцесса, технолог генерирует документ «Изменение ведомости оснащения».

Ресурс «Комплект УП» ссылается на документ «заявка на УП», связан с CAD- и CAM-документами, описывается документами «ТЗ на УП», «УП», «карта наладки», «расчетно-технологическая карта (РТК)», «акт внедрения УП». Документ «РТК» выпускается на деталь, поэтому может относится и к другим операциям, если для их выполнения также необходим комплект УП.

Ресурс «СТО» ссылается на «заявку на СТО», эквивалентен CAD-документу и описывается документом «ТЗ на проектирование СТО». Для ресурса «СТО» создается расцеховка, в которой указываются цех-проектировщик, цех-изготовитель и отдел по подготовке производства, ответственный за сроки. Технолог цеха-изготовителя создает цеховой техпроцесс, генерирует ВО 2-го порядка и ресурсы.

Освоение систем «Вертикаль» и «Windchill» осуществляется при поддержке консалтинговой фирмы ООО «ПТС». В настоящее время проект по информационной модели технологических данных находится в стадии опытной эксплуатации для ряда производств: приборостроительного, механического и сборочного. Для подразделений, занимающихся проектированием и изготовлением специальной технологической оснастки и цехов, работающих на станках с ЧПУ, — на стадии настройки программного обеспечения. Результатом введения информационной модели технологических данных в промышленную эксплуатацию станет автоматизация выпуска технологической документации по изделиям.

Главное преимущество, которое предприятие «ЦСКБ-Прогресс» получило благодаря внедрению современных информационных технологий, - рост конкурентоспособности предприятия на международном рынке. На предприятии повысилась эффективность управления: руководители получили возможность отслеживать процесс и получать необходимую информацию о степени готовности изделия.

Литература

1. Зильбербург, Л.И. Информационные технологии в проектировании и производстве [Текст]/Л.И. Зильбербург, В.И. Молочник, Е.И. Яблочников. – Спб: Политехника, 2008. – 304с.

2. Методика нисходящего проектирования при разработке изделий в системе Pro/ENGINEER [Текст]: регламент предприятия «ФГУП ГНП РКЦ «ЦСКБ-Прогресс». - 53с.

3. Никонова, Н. Внедряются современные информационные технологии. Интервью Н.Г.Петрова, начальника отдела автоматизации разработки технологической документации ФГУП ГНП РКЦ «ЦСКБ-Прогресс» [Текст]/ Вестник ракетно-космического Центра. - 2010. - 4 августа.

4. Стандарт предприятия ФГУП ГНП РКЦ «ЦСКБ-Прогресс» СТП 0162-2010. Система менеджмента качества. Электронная модель изделия. Общие положения. – 19с.

5. Суханова, А.Ю. «Мы владеем уникальной методологией нисходящего проектирования». Интервью А.Н.Филатова, директора по ИТ ФГУП ГНП РКЦ «ЦСКБ-Прогресс» [Текст]/ CAD/CAM/CAE Observer. - 2010. - №5(57). - С.10-23.

Методы и средства интеграции информационных систем и оценки эффективности инновационных проектов предприятий наукоемкого машиностроения

В.Ф. Денисов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Современные концепции развития информационных систем (ИС) предприятий основаны на понятии Архитектура предприятия и применении методологии открытых систем, обеспечивающих организационную, семантическую и техническую интероперабельность двух и более взаимодействующих систем [1-4]. В основе интеграции разнообразных ИС предприятий лежат стандарты описания продукции, процессов и ресурсов предприятия, применяемых информационно - коммуникационные технологий (ИКТ), а также регламенты взаимодействия, интерфейсы и стандарты обмена данными для принятия решений на разныхх уровнях управления. Основные проблемы создания эффективных ИС связаны с вопросами выбора моделей систем управления, неопределенностью оценок состояния внутренней и внешней среды, вероятностными характеристиками процессов, а также различной интерпретацией терминов, мотивами поведения, групповых и индивидуальных предпочтений персонала и корпоративных групп инфраструктуры поддержки инновациоого разного общества, необходимостью использования наследуемых ИС, накопленной информационной базы и эксплуатируемых средств ИКТ.

Анализ опыта ведущих предприятий наукоемкого машиностроения показывает, что внедрение отдельных автоматизированных систем управления отдельными процессами предприятия и, поддерживающих их ИКТ, без учета архитектуры и стратегий развития предприятия часто не оправдывают ожиданий по следующим причинам:

- недостаточный учет отраслевых особенностей продукции процессов предприятия;
- несоответствие функций типовых ИКТ и реализуемых в них методов, свойствам реальных объектов управления;
- недостаточный учет поведенческо-психологических факторов и личностных качеств персонала в реальных слабо формализуемых ситуациях;
- ограниченность открытого взаимодействия отдельных компонент ИС между собой и внешней средой;
- большая стоимость типовых ИКТ и сложность настройки систем на параметры реального предприятия.

Разработки ИС предприятия должны выполняться по собственной, возможно заимствованной, но принятой на предприятии методологии, доступной для понимания руководством и ведущим персоналом предприятия. Ориентация на применение "фирменных" прикладных программ может приводить к рискам потери управляемости предприятия в силу закрепления зависимости предприятия от технических решений поставщика, возможных изменений его стратегий продвижения своих продуктов и услуг на рынке, отказом от сопровождения и другим негативным последствиям. Для получения значимых эффектов и минимизации затрат на инновационные проекты предприятий необходима системная оценка предлагаемых на рынке средств ИКТ на соответствие архитектуре предприятия, стандартам и базовым профилям открытых систем [2, 3], а разработки и поставки методического, программного и технического

обеспечения ИС должны выполняться под проектные спецификации предприятия. При выборе методологии и стратегий развития ИС предприятия учитываются [4-6]:

- базовые функционально полные модели деятельности предприятия и сферы взаимодействия с внешней средой;
- экспертно-статистические методы оценки сложности объектов, процессов и ресурсов;
- методы упорядочения ключевых показателей деятельности и их ранжирования по различным критериям оценки потенциала развития предприятия;
- оценки привлекательности инновационных проектов в "точки роста" экономики предприятия и региона (стратегические продукты и услуги, востребованные заказчиками);
- состояние рынка современных средств ИКТ (базы данных, инструментальные средства, моделирующие программные системы, сети ЭВМ и электронных коммуникаций);
- методы согласования интересов, координации действий и узаконивания решений, основанные на знаниях объективных законов развития предприятий и установленных обществом "правилах" и стандартах деятельности;
- процедуры и технологии проблемно-ситуационного моделирования совместной деятельности специалистов в сложных слабо-формализуемых ситуациях.

Оценка результативности и эффективности ряда инновационных проектов ИС отечественных предприятий показывает ряд существенных недостатков и проблем системного обоснования целесообразности создания и интеграции таких систем, недостатков в проектировании, комплектации и техническом обслуживании, обеспечении организационной, семантической и технической интероперабельности. Требуется ряд существенных мер по правовому и организационно-методическому обеспечению, координации и консолидированному ресурсообеспечению инновационных проектов предприятий, оценки рисков неадекватного межведомственного взаимодействия при экспертизе и принятии решений о реализации проектов. Особая роль при разработке таких мер отводится вопросам унификации и стандартизации системной архитектуры и компонент технического и программного обеспечения ИС, применении методов математического моделирования, рационального управления и сертификации в области системной и программной инженерии [4].

Актуальность решения задач стандартизации ИКТ-систем предприятий наукоемкого машиностроения и, в частности, аэрокосмической отрасли проявляется в сферах:

- стандартизации жизненного цикла систем и определении роли предприятия в реализации инновационных процессов развития общества;
- определения и унификации описаний инновационных проектов для анализа их привлекательности, согласования интересов участников, принятия решений по поддержке проектов и определения корреляций с другими проектами;
- унификации нормативно-справочной информации, гармонизации регламентов и стандартов в смежных областях деятельности;
- определении функций целевых автоматизированных систем предприятия и применения в них различных ИКТ- систем и платформ;

- оценки совместимости ИКТ- платформ, взаимодействующих предприятий;
- унификации оборудования офисов подразделений предприятия и минимизации стоимости владения средствами ИКТ;
- выбора моделей ИТ – служб и интерфейсов обмена данными с внутренними и внешними пользователями информационных ресурсов; планирования, координации и консолидированного ресурсообеспечения сложных проектов.

Для конструктивного описания задач и выбора объектов типизации и стандартизации ИКТ используется системная модель предприятия в виде частично упорядоченных множеств вида <X, F, N, W, U, P, S, R, D>, где:

X - описание объектов и субъектов деятельности предприятия (входные данные);

F - описание функций структурных элементов предприятия (модель поведения);

N - заданные нормы и показатели деятельности на этапах жизненного цикла продукции и услуг;

W - характеристики внешних событий и возмущений на объекты и процессы;

U - решения и соглашения органов управления предприятием (проектом);

P - запросы от других активных элементов (структурных элементов и процессов) на услуги и "приобретение" ресурсов для реализации процессов;

S - характеристики доступных ресурсов (персонал, оборудование, программные средства, финансы и др.);

У - выходные результаты деятельности предприятия (продукты и услуги);

R - показатели качества продукции и услуг, эффективности использования ресурсов;

D- альтернативные решения по применению продукции и услуг предприятия.

Такое представление предприятия и его окружения подлежит детализации на разных уровнях управления и хорошо согласуется с известными принципами функционального моделирования организационно - технических систем на основе стандартов IDEF0, UML, архитектурными моделями Захмана, а также 3-D моделями предприятия и перспективными направлениями развития сервис-ориентированной архитектуры прикладных информационных систем (SOA), социальных сетей, "облачных вычислений" в корпоративных и глобальных сетях ЭВМ и реализации концепций "Программный продукт как услуга".

По результатам анализа применения ИКТ на ряде предприятий [5-7] сформулированы общие требования к функционально-полному комплексу моделей ИС обеспечения деятельности предприятий (профиль функцй), в который включаются:

- системная модель архитектуры предприятия и его окружения;
- модели ключевых процессов деятельности подразделений предприятия;
- типовые модели описания продукции, процессов и ресурсов предприятия;
- модели потоков данных, внутреннего и внешнего документооборота;
- информационно-логические модели баз данных и знаний;
- модели актуализации данных и защиты информационных ресурсов;
- соглашения и интерфейсы по обмену данными со смежными системами;
- расчетно-аналитические модели оценки показателей качества,

оценки рисков нарушения целостности и безопасности функционирования объектов предприятия;

- модели оценки потребностей в ресурсах и управления проектами реализации текущих и перспективных проектов предприятия;
- модели оценки деятельности должностных лиц и эксплуатационного персонала;
- модели технического обслуживания и сопровождения средств ИКТ.

Принципиальным является то, что структура этих моделей является открытой. Они могут быть детализированы и реализованы на разных ИКТ - платформах с использованием программных средств различных производителей или собственной разработки (программной реализации типовых проектных решений, адаптированных для условий конкретного предприятия). Системное применение этих моделей обеспечивает организационную, семантическую и техническую интероперабельность систем, информационную базу для проектирования, реализации процедур оценки и экспертиз инновационных проектов предприятий. Унификация входных - выходных данных этих моделей обеспечивает решение задач интеграции и сборки "индивидуальных" ИС из "готовых" типовых ИКТ - компонент, их тестирование на совместимость и соответствие международным и национальным стандартам, включая требования к безопасности продукции, персонала, процессов и технологий, а также средствам защиты информационных ресурсов и интеллектуальной собственности.

Следует отметить, что подготовка инновационных проектов ИС предприятий требует учета междисциплинарных связей при оценке и принятии решений по их поддержке различными структурами инновационного развития общества [8]. Основная проблема заключается в сложности оценки привлекательности инновационной продукции и встраивании проектов в реально действующие организационно-технические системы предприятия. Разнообразие продукции и услуг, моделей деятельности инновационных предприятий, отсутствие единого понятийного аппарата и четкой системы критериев оценки продуктов, процессов и ресурсов затрудняет решение вопросов экспертизы и согласования решений по поддержке проектов потребителями, инвесторами, исполнителями и другими участниками проектов.



Рис. Структура комплекса средств информационных технологий анализа инновационных проектов предприятий

Рассмотренные подходы положены в основу создания программно - методического комплекса средств ИТ анализа продукции, процессов и ресурсов инновационных проектов предприятий (КСИТ АИП), разработанного по результатам ряда НИОКР и консультационных проектов для предприятий машиностроения, органов административно-территориального управления и ВУЗов. на кафедре информационных систем и технологий Самарского государственного аэрокосмического университета Функциональная структура КСИТ АИП приведена на рисунке.

Для описания инновационных проектов в КСИТ АИП используется функционально-полная модель деятельности предприятия, дополненная специальной нотацией (на основе обобщения стандартов DFD, IDEF0, UML) для понимания специалистами в конкретной предметной области деятельности. Функциональные модули КСИТ АИП обеспечивают ввод данных о характеристиках инновационных проектов, оценку их сложности и потребностей в ресурсах, проектирование составляющих компонентов, моделирование и оценку показателей качества, расчет статистических оценок случайных процессов функционирования, контроль состояния процессов и управления проектами. Работа комплекса поддерживается специализированными базами общего и специального назначения. Предлагаемые технологии позволяют в доступной форме представлять инновационные проекты предприятий, отображать и анализировать продукты, процессы и ресурсы проектов, осуществлять информационную поддержку, поиск аналогов и прототипов организационных и технических решений, вести подготовку и презентацию проектов для согласования и принятия решений о целесообразности их реализации.

Полномасштабная реализация КСИТ АИП планируется в рамках проекта создания ИТ-парка «Жигулевская долина» и предусматривает:

- - подготовку организационно-методического обеспечения КСИТ АИП с учетом международных и национальных стандартов в сфере технического регулирования, информационных технологий, обеспечения качества и безопасности продукции и процессов ее производства;
- формирование баз данных о состоянии проектов, ресурсов и интеллектуальной собственности инновационных предприятий;
- подготовку конструкторско-технологической документации для производства модулей КСИТ АИП, их тестирования и организации поставок потребителям;
- организацию специализированных учебных курсов и семинаров по применению, техническому обслуживанию и сопровождению КСИТ АИП;
- подготовку услуг по применению КСИТ АИП в режиме «Программное обеспечение как услуга», в т.ч с использованием технологий «облачных вычислений».

Расчетные модули КСИТ АИП обеспечивают оценку влияния инновационных проектов на показатели деятельности предприятия, развитие региона и степень влияния рисков на их реализацию, включая социально-политические, законодательные, природные (экологические); технические, экономические , человеческий фактор.

Основными показателями эффективности инновационных проектов являются: показатели привлекательности и результативности; показатели качества продукции и услуг предприятий, применяемых ИКТ ,показатели использования ресурсов; обобщенные (интегральные) статистические показатели эффективности деятельности и уровня зрелости предприятия; технико-экономические показатели проекта.

Применение в КСИТ АИП специализированных модулей оценки этих показателей позволяет существенно повысить достоверность и точность расчетов эффективности использования ресурсов, получить дополнительный синергетический

эффект, способствовать привлечению внутренних и внешних инвестиций в проекты предприятия и позволяет:

- обеспечить применение стандартов и типовых решений в ИС предприятий, легитимных процедур принятия решений по поддержке инновационных проектов;
- сократить сроки проектирования, согласования и принятия решений по реализации инновационных проектов;
- повысить уровень обоснованности и качества проектных решений;
- существенно уменьшить совокупную стоимость владения средствами ИКТ в ИС предприятий, затраты на их техническое обслуживание и сопровождение.

Для решения задач управления проектами и интеграции компонент ИС, предлагаемых различными поставщиками ИКТ-решений, следует учитывать важность подготовки персонала организации по эксплуатации, техническому обслуживанию и сопровождению ИС, включая специальную подготовку эксплуатационного персонала предприятий. Однако в реальных программах обучения в ВУЗах этим вопросам уделяется явно недостаточное внимание. Имеется ряд серьезных проблем в использовании промышленных и образовательных стандартов, процедур управления реальными проектами, их "встраивания" в существующие структуры подготовки персонала и реализации программ высшего и дополнительного профессионального образования. Промышленные стандарты "де-юре" и "де-факто" весьма часто разнятся, причем последние с развитием рынка ИКТ- индустрии изменяются более динамично. Требуется согласование базовых понятий, применяемых в различных "промышленных" и образовательных стандартах а также их гармонизации со стандартами в сфере строительства, обеспечения безопасности зданий и сооружений, жизнедеятельности персонала, образования и других и междисциплинарных социально значимых проектах .Для упорядочения понятийного аппарата в конкретных проектах ИС предприятий, унификации и стандартизации терминов и понятий предметной области (онотологий) в рамках конкретных междисциплинарных проектов и определения связей между понятиями различных дисциплин используются формальные процедуры определения отношений: системности, информативности, координируемости и технологичности на множестве локальных описаний различных задач управления предприятием (проектом).[7].

Весьма актуальными являются вопросы гармонизации и применения в учебных дисциплинах международных и национальных стандартов на компоненты ИКТ, интерфейсы, технологические процессы проектирования и управления проектами на основе стандартов ИСО 15288, ИСО 12207, ADEF0, UML и др. При этом в рамках развития новых Государственных Образовательных Стандартов необходима разработка (Вузами и наукоемкими предприятиями) системы формирования требований к компетенции специалистов, обучения и дополнительной подготовки персонала в сфере создания и эксплуатации ИКТ - систем, которая может быть модифицирована и тиражирована для применения в различных сферах экономики и быть предметом поставки для организаций инфраструктуры инновационного развития регионов и отдельных предприятий.

Литература

1. Батоврин В.К.,Зиндер Е.З. Результаты и перспективы "тихой революции" архитектуры предприятия и сервисного подхода //Материалы IV конф. “ Стандарты в проектах современных информационных систем.- М. : ФОСТАС, 2007 г.

2. Денисов В.Ф. Методы и средства проектирования функциональных профилей систем управления в организационно-технических системах./. Сб. научн.трудов «Перспективные

информационные технологии в научных исследованиях, проектировании и обучении», том 3 - Самара, СГАУ, 2006 г.,с.32- 39.

3. Денисов В.Ф. Методы и средства упорядочения функций и формирования профилей распределенных автоматизированных систем Предприятий //Первая международной конф. «Стандартизация информационных технологий и интероперабельность» (СИТОП-2007) - М.: ОИТ и ВС РАН, ФАИТ и др., с 50-56.

4. Костогрызов А.И., Нистратов Г.А. Стандартизация, математическое моделирование, рациональное управление и сертификация в области системной и программной инженерии. М.: Изд. «Вооружение, политика, конверсия», 2004 – 395с.

5. Денисов В.Ф., Дерябкин В.П. Корнев Е.П., Кузьминов С.А., Кулаков Г.А., Лычев В.Ф. Стратегия развития и методология разработки комплекса средств управления предприятием.//сборн. трудов IV Всерос. практ. конф. "Стандарты в проектах современных информационных систем", М.: ФОСТАС, изд-во «Открытые системы» 2004г.

6. Прохоров С.А., Федосеев А.А., Денисов В.Ф., Иващенко А.В. Методы и средства проектирования профилей интегрированных систем обеспечения комплексной безопасности предприятий наукоемкого машиностроения. // Самара, СНЦ РАН, 2009- 199с., илл.

7. Денисов В.Ф., Чекин В.И. Опыт использования промышленных и государственных образовательных стандартов при разработке базовых профилей информационных систем/ сборн. трудов III Всерос. практ. конф. «Стандарты в проектах современных информационных систем» - М.: Фостас, изд-во «Открытые системы», 2003 г.

8. Денисов В.Ф. Инфраструктура инновационного развития общества и информационно – коммуникационные технологии.//Материалы XI междунар.научн. конф. "Проблемы развития экономики и общества"- М. Высшая школа экономиики, 6-8 апреля 2010 г.

Проблемы обеспечения согласованного взаимодействия в интегрированной информационной среде предприятия

А.В. Иващенко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В настоящее время большую популярность приобретают подходы к построению автоматизированных систем управления процессами и поддержки принятия решений по планированию загрузки ресурсов, основанные на принципах децентрализации. Это связано с необходимостью функционирования в гетерогенной информационной среде и сложностью решения оптимизационных задач, сформулированных относительно ключевых показателей эффективности, на которые влияет большое количество параметров.

В связи с этим все чаще используются технологии, основанные на имитации природных механизмов самоорганизации [1], когда для решения задач эффективного использования ресурсов строится некоторый виртуальный мир, взаимодействие элементов которого подчиняется законам, справедливым для реальных живых систем. Однако непосредственное применение этих технологий при построении программного и информационного обеспечения в задачах проектирования, производства, транспортной логистики и при управлении в распределенных системах, связано с некоторыми проблемами.

Например, тяжело доказать преимущества получаемых в результате работы таких алгоритмов результатов в сравнении с другими подходами, сложно оценивать устойчивость получаемых решений, отсутствует повторяемость получаемых результатов в одинаковых условиях. Применение таких подходов при построении компонентов интегрированной информационной среды предприятия оправдано, например, для обеспечения интеллектуальной поддержки принятия решений. Однако при этом нужно использовать новые принципы построения такой среды, управления ею и оценивания показателей эффективности ее использования. Одной из наиболее актуальных задач в данном случае является обеспечение согласованного взаимодействия в интегрированной информационной среде предприятия.

Особенности задачи

Применение моделей, основанных на имитации природных механизмов самоорганизации, оправдывает себя в случае, когда в ходе решения задачи требуется учесть большое количество параметров, сложно сформулировать цели в связи с их неоднозначностью, имеются субъективные факторы, влияющие на принятие решений, а также присутствует высокая степень неопределенности. Весьма многие задачи управления на современных предприятиях имеют эти особенности: в транспортной логистике, на производстве, при управлении распределенными системами в режиме реального времени задача поиска оптимального управления видоизменяется при поступлении новой информации.

Следует отметить, что эволюционный характер развития такого рода систем существенно влияет на выбор подхода к их анализу и оптимизации. Эволюционный подход к построению решения [2] приводит к тому, что действия, произведенные с целью локальной оптимизации в некоторый момент времени могут придать системе ряд неустранимых свойств, которые потребуется учитывать в дальнейшем. В этом смысле, сложность решаемой задачи постоянно возрастает по мере развития системы, как

возрастает количество параметров, исторических факторов, наследуемых свойств и причин, которые нужно учесть для принятия правильного решения.

Накопления этих знаний в информационном пространстве предприятия с одной стороны, позволяет их учитывать комплексно на каждом шаге, но с другой стороны, требует дополнительных усилий по выборке полезной информации для ее предоставления всем заинтересованным лицам. В таком контексте решать одну и ту же задачу с течением времени становится все сложнее.

Применение аналогии с живыми системами позволяет отойти от задачи выработки оптимального решения на каждом шаге и использовать лишь косвенное управление, определяющее возможности пользователей интегрированной информационной среды. Несмотря на то, что более правильным является глубокое исследование решаемой проблемы, поиск закономерностей, формализация задачи и ее решение, практика показывает, что решения, найденные в результате отслеживания развития модели исследуемой системы как живого организма, среди множества прочих решений часто оказываются близки к оптимальному. Этот аспект с учетом постоянной изменчивости критериев и особенностей постановки задачи, позволяет успешно применять такие модели в практических приложениях.

Таким образом, полезным является рассмотрение принципов построения единого информационного пространства предприятия, как среды, интегрирующей все информационные и программные ресурсы его компонентов, на основе теории живых систем [3]. Такой подход возможен, благодаря современным достижениям в области организации параллельных вычислений, построению онтологий и интеллектуальных систем поддержки принятия решений.

Способы исследования процессов взаимодействия в интегрированной информационной среде

Согласно описанному подходу, существенно меняется постановка задачи проектирования единого информационного пространства предприятия. Создавая модель интегрированной информационной среды, проектанты должны предусмотреть пути ее развития. В ходе ее непрерывного использования необходимо обеспечить проведение мониторинга процессов взаимодействия пользователей в этой среде и постоянно повышать эффективность своевременного их обеспечения актуальной информацией.

В этой области можно определить новый класс задач анализа динамического развития интегрированной информационной среды. Такого рода проблемы достаточно давно решались разработчиками программного обеспечения, в основном при тестировании и анализе качества функционирования сложных комплексов, для отслеживания «плавающих» ошибок и выяснения узких мест, влияющих на производительность. В частности, в дополнение к применению средств автоматического тестирования, при контроле качества программного обеспечения широко применяются методы имитационного моделирования, позволяющие провести нагрузочные или ситуационные тесты в сложных условиях.

Применив этот опыт для решения задач мониторинга состояния единого информационного пространства с учетом эволюционного характера его развития, можно выделить новые способы повышения эффективности взаимодействия пользователей интегрированной информационной среды, направленные на обеспечение согласованного принятия решений в режиме реального времени.

В качестве исходных данных для такого анализа можно использовать журналы, или «логии», принятия решений в интеллектуальных информационных системах. Кроме этого, отслеживание сообщений, которыми обмениваются компоненты распределенных систем, позволяет оценивать особенности их взаимодействия между

собой и с пользователями единого информационного пространства предприятия. В этой области можно выделить два направления анализа: семантическое исследование сообщений с целью выявления закономерностей в процессах обмена информацией [4] и исследование формальных особенностей взаимодействия [5], влияющих на своевременность принятия решений.

В первом случае хорошие результаты показывает выявление ключевых слов – тегов, что позволяет достаточно просто и универсально определять, насколько разные сообщения в рамках независимых информационных потоков близки по содержанию. Такой подход позволяет оценивать близость пользователей единого информационного пространства предприятия и объединять их в группы при необходимости обеспечения специальных стратегий взаимодействия для адекватной реакции на поступающие события.

Во втором случае можно эффективно применять алгоритмы корреляционно-спектрального анализа процессов и потоков событий, характеризующих взаимодействие пользователей и компонентов единого информационного пространства. В частности, исследование ритмичности взаимодействия пользователей корпоративных информационных систем позволяет своевременно определять разнообразные проблемы согласования решений.

Анализ процессов взаимодействия в интегрированной информационной среде может производиться как на этапе ее проектирования (на основе использования результатов имитационного моделирования), так и в процессе эксплуатации, в ручном и автоматическом режиме. В последнем случае могут применяться встраиваемые инструменты анализа процессов взаимодействия, результатом работы которых является изменение структуры или наполнения базы знаний, используемой при принятии решений и реализованной, например, в виде онтологии.

Решение практических проблем

Описываемый подход может быть применен в различных проектах по построению интегрированной информационной среды предприятия. В транспортной логистике [6], исследование взаимодействия водителей транспортных средств, диспетчеров и операторов по приему заказов позволяет выявлять недостатки в обеспечении нужной информацией лиц, принимающих решения, а также проблемы, связанные с нехваткой времени на согласование решений.

При планировании распределения производственных ресурсов [7] обеспечение удаленного взаимодействия руководства, планово-диспетчерских подразделений, сотрудников складов и кладовых, руководства цехов, участков, мастерских и смен, а также рабочих средствами новых информационно-коммуникационных технологий является существенным и необходимым дополнением к известным алгоритмам и программным средствам производственного планирования.

В задачах управления безопасностью предприятия [8] согласование взаимодействия экспертов по оценке рисков позволяет построить эффективную систему противодействия угрозам.

Организация виртуального круглого стола проектантов различной специализации по работе с единым планом позволяет повысить эффективность их взаимодействия, сократить время на выработку решений, а также снизить вероятность ошибок. В качестве примера можно привести решение задачи интеллектуальной поддержки планирования грузопотока и расчета ресурсов Международной космической станции, которой посвящены две статьи в данном сборнике.

Отдельно стоит отметить задачи обеспечения взаимодействия компонентов сложных технических систем с высокой степенью автономности.

Обеспечение новых принципов управления распределенными системами, широко внедряемый сейчас при решении такого рода задач, невозможно без применения программных средств исследования особенностей взаимодействия пользователей и интеллектуальных компонентов интегрированной информационной среды. В связи с чем, применение методов и алгоритмов оценивания ритмичности взаимодействия компонентов на основе интервально-корреляционного анализа и алгоритмов выделения виртуальных сообществ пользователей ЕИП, моделей единого информационного пространства и общих принципов его проектирования имеет весьма широкие перспективы.

В случае, когда разные сегменты интегрированной информационной среды удалены друг от друга, а протоколы информационного обмена предусматривают множество возможностей по организации обмена данными, задача обеспечения ритмичного взаимодействия весьма важна для обеспечения работы системы управления в режиме реального времени.

Выводы

Применение алгоритмов исследования процессов взаимодействия в интегрированных информационных системах может быть весьма полезно при построении единого информационного пространства предприятия, в случае, когда в его состав включаются компоненты, функциональность которых основана на имитации природных механизмов самоорганизации. Комбинирование средств моделирования виртуальных сообществ и интервально-корреляционного анализа временных рядов дает возможность оценивать характеристики процессов использования интегрированной информационной среды и позволяет вырабатывать мероприятия по ее совершенствованию в случаях, когда это необходимо.

Литература

1. Leitão, P. Holonic rationale and bio-inspiration on design of complex emergent and evolvable systems, Transactions on large-scale data- and knowledge-centered systems [Text]. – Lecture notes on Artificial Intelligence 5740. – Germany: Springer. – p. 243 – 266.

2. Эбелинг, В. Физика процессов эволюции. Синергетический подход / В. Эбелинг, А. Энгель, Р. Файстель. – М-Иж.: Эдиториал УРСС. – 2001. – 328 с.

3. Иващенко, А.В. Технология автореинжиниринга единого информационного пространства предприятия. – Вестник Самарского государственного аэрокосмического университета. – 2009. – № 2. – с. 124 – 134

4. Иващенко, А.В. Моделирование виртуальных сообществ пользователей интегрированной информационной среды / А.В. Иващенко, Е.С. Пугачева, С.С. Погодина. – Управление большими системами. Выпуск 29. М.: ИПУ РАН. – 2010. – с. 68 – 87

5. Иващенко, А.В. Интервально-корреляционный анализ ритмичности взаимодействия в интегрированной информационной среде предприятия. – Системы управления и информационные технологии. – 2010. – № 1(39). – с. 32 – 36

6. Glaschenko, A. Multi-agent real time scheduling system for taxi companies [Text]/ A. Glaschenko, A. Ivaschenko, G. Rzevski, P. Skobelev. – 8th International conference on autonomous agents and multiagent systems (AAMAS), Budapest, Hungary. – 2009. – pp. 29 – 36

7. Andreev, M. A Multi-agent platform design for adaptive networks of intelligent production schedulers [Text]/ M. Andreev, A. Ivaschenko, P. Skobelev, A. Tsarev. – 10th IFAC Workshop on Intelligent Manufacturing Systems (IMS'10), Lisbon, Portugal. – 2010. – p. 87 – 92

8. Федосеев, А.А. Комплексное управление безопасностью в едином информационном пространстве предприятия [Текст]/ А.А. Федосеев, С.А. Прохоров, А.В. Иващенко. – Программные продукты и системы. – 2008. – № 4. – с. 132 – 135

Мультиагентные технологии управления распределением производственных ресурсов

М.В. Андреев, И.О. Бабанин, А.В. Иващенко, П.О. Скобелев
Научно-производственная компания «Разумные решения», Самара

Введение

Управление современным производственным предприятием является сложной задачей, связанной с необходимостью обеспечения требуемого уровня качества, как выпускаемой продукции, так и производственных процессов, постоянного совершенствования производства, применения современных методов оперативного планирования, основанных на использовании актуальной информации. В настоящее время существует большое количество разнообразных подходов к планированию производственных ресурсов [1, 2, 3]. Однако, современные требования по обеспечению адаптивного распределения ресурсов, сложность выявления единых критериев планирования для разных заказов, обуславливают необходимость поиска новых способов решения данной задачи.

В частности, следует специально отметить тенденцию по построению систем автоматизированного управления производственными ресурсами с помощью технологий, основанных на имитации природных механизмов самоорганизации [4]. В основе этих современных разработок лежат принципы построения децентрализованных систем, в которых задается логика функционирования отдельных элементов и обеспечивается взаимодействие между этими элементами.

Такой подход оправдывает себя в случаях, когда логика взаимодействия участников процесса планирования и распределения ресурсов сложна для описания, а поступающие события постоянно приводят к небольшим изменениям условий принятия решений, что не позволяет использовать классические оптимизационные методы в связи с большими временными затратами на точную постановку задачи оптимизации. В этом смысле, применение новых технологий, базирующихся на знаниях и аналогиях с природными организмами эволюции, не противопоставляется классическим алгоритмам решения задач поддержки принятия решений, но дополняет их, позволяя строить адаптивные системы управления ресурсами, способные функционировать в условиях неопределенности и высокой динамики изменения условий принятия решений.

Одним из таких новых способов, получивших, однако, в последнее время достаточно широкое распространение, является применение мультиагентных технологий [5], позволяющих автоматизировать распределение ресурсов и оперативное построение расписаний и создать основу для последующей оптимизации, контроля и развития производства. В данной работе дано краткое описание особенностей мультиагентного алгоритма управления распределением производственных ресурсов, а также затронуты вопросы применения мультиагентных технологий в задачах оптимизации распределения ресурсов по выбранному критерию их загрузки.

1. Роль и место мультиагентной системы управления распределением ресурсов в едином информационном пространстве предприятия

С учетом современных требований по автоматизации всех этапов жизненного цикла изделия включающего и процесс производства, в управлении современными предприятиями широко используются автоматизированные системы управления и поддержки принятия решений. Эти системы обычно классифицируют по выполняемым функциям: это системы инженерной подготовки производства и управления жизненным циклом изделия (PDM/PLM), системы управления ресурсами (ERP), системы автоматизированного проектирования, финансового и бухгалтерского учета и другие. В частности, весьма востребованными являются производственные исполнительные (MES) системы, основным назначением которых является автоматизация планирования и управления производством.

Системы планирования и управления производством имеют достаточно уникальное положение, что определяет их высокую важность в едином информационном пространстве предприятия. С одной стороны, они имеют доступ к знаниям о производственных ресурсах, обработка которых ведется в ERP системе. С другой стороны, они обладают актуальными данными о текущем статусе производственного процесса. Именно благодаря такому положению этим системам на многих предприятиях отводится главная роль, а, следовательно, в условиях высокой динамики развития предприятий, к ним предъявляются особые требования.

В частности, для обеспечения эффективной автоматизации управления производством необходимо автоматизировать планирование производственных ресурсов в режиме реального времени. Для этого требуется применение новых алгоритмов, которые позволяют за ограниченное время найти оптимальное решение. При этом в ходе планирования и собственно управления производством могут меняться критерии и ограничения, а каждый из многочисленных объектов и субъектов планирования может иметь индивидуальную логику принятия решений. Такая картина наиболее точно отражает реальный мир, в котором план производства строится в процессе непрерывного согласования, разрешения конфликтов и поиска компромиссов между лицами, принимающими решения, которые отвечают за разные аспекты деятельности предприятия: технологов, мастеров, экономистов и др.

Наиболее эффективными технологиями, которые к тому же позволяют проиллюстрировать важные особенности современного управления производством с активным использованием единого информационного пространства, на наш взгляд, являются мультиагентные технологии, которые базируются на новом подходе к описанию и моделированию сложных систем. В отличие от классических систем, в мультиагентной системе внутрицехового планирования каждое предприятие моделируется как динамическая сеть программных агентов потребностей и возможностей [6]. В такой сети могут быть представлены различные подразделения, конкретные производственные заказы (на готовое изделие или его компоненты, отдельную операцию станка и т.д.) и конкретные ресурсы (например, рабочие, детали или станки).

Главной задачей такой системы является построение и поддержание баланса интересов всех участников производственного процесса. Для такого рода систем становится характерным переход от централизованных решений к распределенным; замена иерархий на сетевую организацию, команд-инструкций «сверху-вниз» – на переговоры равноправных сторон, жестких планов – на гибкие планы, фиксированных цен – на договорные и т.д. В условиях современной экономики эти принципы обеспечивают более высокую гибкость и эффективность управленческих решений.

Интеграция мультиагентной системы распределения производственных ресурсов в единое информационное пространство предприятия целесообразно выполнить на основе веб-сервисов, что позволяет обеспечить гибкость в настройке системы и активно использовать имеющийся задел на предприятиях в области построения единого информационного пространства.

2. Алгоритм мультиагентного планирования ресурсов

Мультиагентный алгоритм распределения производственных ресурсов предприятия может быть представлен следующим образом. При поступлении нового заказа создается событие на его планирование. Во время обработки события новый заказ получает агента, который обращается к агенту процесса своего выполнения. Получив описание этапов выполнения графика для планируемого заказа, агент заказа создает агентов для каждого этапа (операции). По стратегии Just in Time (JIT) агент заказа начинает планировать операции от последней к первой. Отметим, что планирование можно осуществлять и по стратегии As soon as possible (ASAP) – от первой к последней. Для этого он отправляет сообщение агенту последней операции о начале планировании. Агент этапа после получения сообщения высылает сообщение агенту указанного ресурса с просьбой его запланировать.

Агент ресурса получает сообщение, проверяет свои предпочтения (по недоступности, сменам и т.п.) и, если размещение возможно, возвращает ответ агенту

этапа, что все успешно. В случае если агент ресурса не может разместить этап в предпочитаемое время, он в ответном сообщении возвращает контрпредложение.

В случае успеха агент этапа возвращает агенту заказа сообщение об успешном планировании. Агент заказа определяет по стратегии JIT следующий этап и посылает его агенту сообщение о необходимости планирования. Агент этого этапа должен учесть, что в расписании уже есть запланированные этапы и планироваться с учетом этого. Остальные этапы планируются аналогично. После размещения нового заказа агенты этапов получают возможность проактивно улучшить свое состояние согласно своим предпочтениям. В качестве критериев для этапов могут использоваться размещение на как можно более раннее время, выполнение в срок, обеспечения резерва времени на выполнение и т.п.

В ходе процесса переговоров агентов производится построение квазиоптимального, сбалансированного по многим критериям плана производства с учетом индивидуальных ограничений и предпочтений, а также целей предприятия в целом. В случае возникновения непредвиденных событий (поломка станка, опоздание рабочего), агенты могут динамически, в режиме реального времени, перераспределить задания на другие доступные ресурсы, без пересмотра всего плана производства.

Важным свойством, являющимся следствием распределенного планирования, является адаптивность. Построение локальных изменений производится не по жесткому централизованному алгоритму, а является результатом совместной работы отдельных агентов, учитывающих свои состояния и действующих по обстоятельствам.

3. Особенности применения мультиагентной системы распределения производственных ресурсов (на примере предприятия тяжелого машиностроения)

Управление производственным подразделением включает планирование и мониторинг исполнения производственных заказов. При этом необходимо добиваться наиболее точных результатов планирования и гибко реагировать на возникающие события, изменяя планы и адаптируясь к происходящим изменениям в режиме реального времени. Планирование на предприятии тяжелого машиностроение имеет ряд особенностей представления графиков загрузки производства, которые нужно учитывать при автоматизированном распределении производственных ресурсов.

Обычно область деятельности такого предприятия связана с производством различного оборудования, изготавливаемого согласно технологическим процессам (графикам), которые имеют сложную структуру, то есть включают в себя связанные между собой технологические операции (этапы). Каждый этап выполняется в рабочем центре, выступающем в качестве ресурса. На вход системы планирования поступает цикловой график по основным ресурсам, а также графики производства. Результатом работы планировщика является производственное расписание, описывающее распределение операций по рабочим центрам.

В данной задаче в качестве критериев обычно выступают: требование максимальной загруженности ресурсов и минимизация срока выполнения заказов. Агенты заказов и ресурсов могут вступать в непосредственные связи между собой и инициировать процесс взаимного пересмотра и согласования планов по мере возникновения ожидаемых или заранее непредвиденных событий (например, недоступности ресурса).

Проблема применения мультагентных систем управления распределением производственных ресурсов для решения такой задачи заключается в том, что в результате переговоров происходит построение квазиоптимального плана. Это означает, что в результате полного перебора всех вариантов в большинстве случаев можно найти решение, лучшее в смысле оптимизации выбранных показателей эффективности, чем то, которое было получено в ходе мультиагентных переговоров.

Однако, на практике, в случае высокой интенсивности потока поступающих событий, данная проблема не является критической, так как каждое новое событие приводит к частичному пересмотру плана, а, следовательно, изменению ключевых показателей эффективности. При этом важнее локализовать изменения в расписании, и не производить полного перепланирования в ответ на каждое поступающее событие.

Таким образом, критерием достаточности решения, полученного мультиагентной системой, является его близость к глобальному оптимуму.

Для исследования данного свойства был проведен эксперимент по планированию одного заданного графика двумя способами: полный перебор возможных вариантов результирующего расписания и построение варианта с помощью мультиагентной системы планирования. В качестве критерия оптимальности расписания был задан минимум времени выполнения (интервал времени между началом самого первого этапа и окончанием самого последнего). В результате сравнения решения, полученного с помощью мультиагентной системы (без специальной оптимизации) и путем «полного перебора» возможных вариантов было выявлено, что лишь менее 4 % решений лучше чем то, которое находит мультиагентная система.

Для более сложного примера (с группой графиков) «полный перебор» невозможен, ввиду большого количества вариантов. Однако возможно задание некоторого количества индивидуальных для конкретного случая эвристик: изменение последовательности планирования графиков и этапов в зависимости от их длительности, изменение стратегии мультиагентных переговоров и т.п. Для оценки целесообразности такой настройки, был рассмотрен типовой случай: планирование 7 графиков разной длины – от 3 до 49 этапов в каждом. Сравнительно с решением, полученным за первые 10 запусков, в течение последующих 90 запусков было найдено лишь одно решение, лучшее по выбранному показателю эффективности, что составляет около 1%, причем улучшение по значению оценки составило 0.51%.

Заключение

Использование новых технологий управления распределением производственных ресурсов на основе имитации природных механизмов самоорганизации является незаменимым на практике для управления производством сложных изделий в режиме реального времени, требующего учета множества индивидуальных особенностей производства каждого элемента в условиях заранее непредсказуемых изменений спроса и предложения. Результаты внедрения реализованной согласно этим принципам мультиагентной системы управления производством демонстрирует возможность использования мультиагентных технологий для решения задачи оптимизации использования производственных ресурсов в смысле повышения эффективности их загрузки.

Литература

1. Леньшин, В.Н., Производственные исполнительные системы (MES) – путь к эффективному предприятию [Текст]/ В.Н. Леньшин, В.В. Куминов, Е.Б. Фролов, Р.А. Будник. – М.: САПР и графика. – 2003. – № 6 (сайт www.sapr.ru)

2. Van Brussel, H., Reference architecture for holonic manufacturing systems: PROSA / H. Van Brussel, J. Wyns, P. Valckenaers, L. Bongaerts, P. Peeters. – Computers in Industry, 1998, 37 (3). – p. 255 – 274

3. Загидуллин, Р.Р., Вопросы синтеза математических моделей оперативно-календарного планирования [Текст]. – М.: Технология Машиностроения. – 2006. – № 1. – с.76 – 78

4. Leitão, P., Implementation of a holonic control system in a flexible manufacturing system [Text]/ P. Leitão, F. Restivo. – IEEE Transactions on Systems, Man and Cybernetics. Part C: Applications and Reviews. – 2008. – vol. 38. – № 5. – p. 699 – 709

5. Andreev, M., A Multi-agent platform design for adaptive networks of intelligent production schedulers [Text]/ M. Andreev, A. Ivaschenko, P. Skobelev, A. Tsarev. – 10th IFAC Workshop on Intelligent Manufacturing Systems (IMS'10), Lisbon, Portugal. – 2010. – p. 87 – 92

6. Виттих, В.А., Метод сопряженных взаимодействий для управления распределением ресурсов в реальном масштабе времени [Текст]/ В.А. Виттих, П.О. Скобелев. – М.: Автометрия. – 2009. – № 2. – с. 78 – 87

Разработка РКТ с использованием прототипов в системе нисходящего проектирования

Л.А Комарова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Методология нисходящего проектирования используется для вновь создаваемых изделий, требующих проектно-конструкторских работ, с целью эффективной организации управления данными и конструкторскими изменениями.

Использование нисходящего проектирования предусматривает создание многоуровневой управляющей структуры, содержащей всю основную геометрию и базовые параметры проектируемого изделия. В основе Управляющей структуры лежат модели мастер-геометрии. Данные из мастер-геометрии верхнего уровня передаются на нижестоящий и дополняются уточняющей геометрией, позволяя таким образом сформировать концептуальную схему проектируемого изделия. Разветвленная схема Управляющей структуры дает возможность организовать параллельную работу разных подразделений проектантов и конструкторов. Заключительным этапом является создание реальных конструкторских моделей деталей и узлов со ссылками на мастер-геометрию и выпуск КД.

Нисходящее проектирование основано на использовании сборки управляющих файлов, называемой Управляющей структурой. Рабочий проект изделия формируется в отдельной сборке, именуемой Конструкторской Сборкой изделия. На одну Управляющую структуру может ссылаться множество Сборок изделия, таким образом, реализуются различные варианты разрабатываемого изделия. Управляющая структура создается независимо от Сборки изделия и служит для организации проекта и разработки структуры изделия. Управляющая структура состоит из верхней Управляющей сборки и входящих в неё нижестоящих Управляющих подсборок.

Нисходящее проектирование сложного изделия с большой группой разработчиков невозможно организовать без системы Windchill. Система Windchill позволяет обеспечить постоянный доступ к моделям Pro/ENGINEER в рамках проекта.

Система Windchill заносит электронные данные в специальное защищенное хранилище, которое обеспечивает целостность информации и контроль всех изменений.

В докладе приводятся иллюстрации процесса создания Управляющих структур изделий РКТ:

- согласование модели границ для разрабатываемого изделия (модуля);
- разработка мастер-геометрии верхнего уровня и создание публикуемых наборов геометрии;
- размещение упрощенных моделей приборов и агрегатов и создание модели распределения пространства;
- разработка мастер-геометрии отсеков и узлов;
- разработка эскизных моделей деталей и узлов на основе файлов мастер-геометрии и модели распределения пространства.

В результате проведенной работы были показаны основные процессы проектирования перспективных изделий РКТ.

Особенности электронного макетирования космических аппаратов дистанционного зондирования

Е.В. Космодемьянский
Научно – производственный ракетно-космический центр «ЦСКБ – Прогресс», Самара

Работа посвящена вопросам применения информационных технологий при создании космических аппаратов дистанционного зондирования.

Обостряющаяся борьба на рынке космических аппаратов дистанционного зондирования Земли (КА ДЗЗ) требует постоянного уменьшения сроков разработки изделий и повышения качества, а это невозможно без применения современных информационных технологий.

На сегодняшний день трехмерное моделирование становится одним из основных инструментов для разработки изделий космической и другой техники. Соответственно, с увеличивающейся сложностью и объемами создаваемых моделей, формируются правила моделирования, требования к трехмерным моделям, методики, способствующие оптимальной работе с ними.

В работе рассматривается процесс создания электронного макета космического аппарата дистанционного зондирования Земли. Акцентируется внимание на создании и ведении трехмерной модели с помощью системы управления данными об изделии. Рассматривается проблема создания и отображения различных состояний конструкции такого сложного объекта, как космический аппарат на разных этапах жизненного цикла.

Продемонстрированы основы технологии принятой в ГНП РКЦ «ЦСКБ-Прогресс» при создании электронного макета КА ДЗЗ, технические и организационные сложности создания такого макета. Проанализированы особенности КА ДЗЗ, как объекта для электронного макетирования и сделаны выводы об эффективности таких работ.

Методы и средства программной инженерии при проектировании интегрированной информационной системы предприятия аэрокосмической отрасли

М.Е. Кременецкая, П.В. Сафин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В настоящее время большое количество отечественных предприятий осуществляют переход на цифровые технологии и проводят комплексную автоматизацию производства. Аэрокосмическая отрасль не является исключением. Актуальным является вопрос создания единой для всего предприятия информационной системы, объединяющей новые внедряемые программные продукты и уже используемые средства автоматизации. Один из вариантов формирования такой системы – это использование готовых программных средств, очень широко представленных на рынке. Важным при таком подходе является один аспект – возможность адаптации под сферу деятельности предприятия-заказчика. Другое решение – создание системы «с нуля», исходя из требований заказчика. Порою второй вариант получается более затратным, но выигрышным, так как зачастую оправдывает ожидания заказчиков и удовлетворяет всем заявленным требованиям при согласованном взаимодействии команды предприятия и команды разработчиков программного обеспечения. Причём неважно, будут ли использоваться типовые программные решения, либо решения, созданные специально под заказ, необходимо использовать некоторую совокупность визуальных моделей как для однозначного ознакомления с деятельностью предприятия, так и для интерпретации структуры и поведения интегрированной информационной системы (ИИС) в целом и дополнительно разрабатываемых модулей, автоматизирующих бизнес-процессы.

В данной статье автор предлагает использовать современные методы и средства программной инженерии для однозначной интерпретации и визуализации как бизнес-процессов предприятия, так и структуры и поведения ИИС, автоматизирующей эти процессы. В дальнейшем данные модели могут помочь в развитии и расширении ИИС за счёт создания и программирования новых модулей, что, в свою очередь, повысит уровень информационных технологий на предприятии.

Прежде чем приступить к проектированию ИИС, необходимо ознакомиться с предметной областью – сферой деятельности предприятия, его организационной структурой, изучить и задокументировать все процессы (состояния «как есть» и «как должно быть»), а также ответственных за их выполнение. Наилучшим образом с поставленной задачей помогает справиться процессный подход на основе методологии ARIS. В зависимости от типа процессов, лучше использовать определённую нотацию моделирования. Например, если существует потребность написания регламентов, то ARIS EPC, не смотря сложность восприятия – один из лучших вариантов. Однако в настоящее время из-за простоты понимания и освоения сотрудниками организации логично и применение нотаций вида «Процедура» приложения Business Studio, либо обычных блок-схем. Преимущество ARIS EPC в том, что она поддерживается большинством средств визуального моделирования, в том числе Microsoft Visio, ARIS Toolset, Business Studio, Бизнес Инженер. Что касается процессов для автоматизации в системе класса BPM, то, безусловно, в данном случае необходимо применять только нотацию BPMN.

Параллельно с обследованием деятельности и формализацией процессов, протекающих на предприятии, логично составить план разработки ИИС, в который

необходимо включить все этапы: от анализа предметной области до внедрения системы. Наилучшим образом справиться с поставленной задачей помогает методология структурного анализа и моделирования SADT и стандарт IDEF, суть которых заключается в последовательной декомпозиция задач от общей к частным. Таким образом, составив полную декомпозицию и применяя соответствующие CASE-средства, поддерживающие данную нотацию моделирования (такие как CA ERwin Data Modeler, Microsoft Visio, Business Studio), мы получаем план работ по формированию ИИС предприятия. Ключевой задачей данного этапа наряду с формированием диаграмм является имитационное моделирование и функционально-стоимостной анализ (ФСА) работ. ФСА проводится для определения и расчёта денежных затрат, времени выполнения каждой из работ и работы верхнего уровня на контекстной диаграмме. Имитационное моделирование необходимо для того, чтобы «проиграть» реальную ситуацию и по результатам моделирования принять соответствующие решения по оптимизации выполнения работ.

Выполнив работы по анализу предметной области и формализации процессов, а также планированию, имитации, расчёту стоимости и оптимизации выполнения работ, необходимо приступить непосредственно к проектированию ИИС. В настоящее время базой к проектированию программного обеспечения является объектно-ориентированный подход (ООП), который связан с применением объектно-ориентированных методологий (например, RUP, MSF). К преимуществам ООП можно отнести:

- уменьшение сложности программного обеспечения;
- повышение надежности программного обеспечения;
- обеспечение возможности модификации отдельных компонентов программного обеспечения без изменения остальных его компонентов;
- обеспечение возможности повторного использования отдельных компонентов программного обеспечения;
- стандартизированный язык моделирования - UML.

Проектирование информационных систем выполняется на основе объектно-ориентированного анализа предметной области и делится на три больших этапа: концепция системы, логический уровень и физическая реализация. Очень важно использовать все необходимые совокупности диаграмм UML для представления структуры и поведения системы. Весь полученный и взаимосвязанный набор диаграмм впоследствии может быть использовать для доработки требуемых дополнительных модулей или для изменения существующих.

Начальным этапом жизненного цикла информационной системы является документирование требований к разрабатываемому продукту. Диаграммы вариантов использования, описывая поведение системы, помогают лучшим образом справиться с этой задачей. Причём, важно не только выделить те функции, которые должна выполнять система (системные требования), но и учесть взаимодействие внешних по отношению к ней пользователей и получение ожидаемого эффекта (бизнес-требования). При необходимости каждый вариант использования может быть раскрыт с помощью диаграмм взаимодействия (то есть диаграммами последовательности или кооперации), либо декомпозирован повторно, если представляет собой сложную структуру. Полученные диаграммы легки в понимании и помогают представить концепцию системы. Соответственно, они могут быть продемонстрированы заказчику с целью их согласования и выявления противоречий, а также использоваться в дальнейшем при непосредственном проектировании и кодировании модулей и компонентов ИИС. Пример формализации бизнес-требований можно наблюдать на рис. 1.

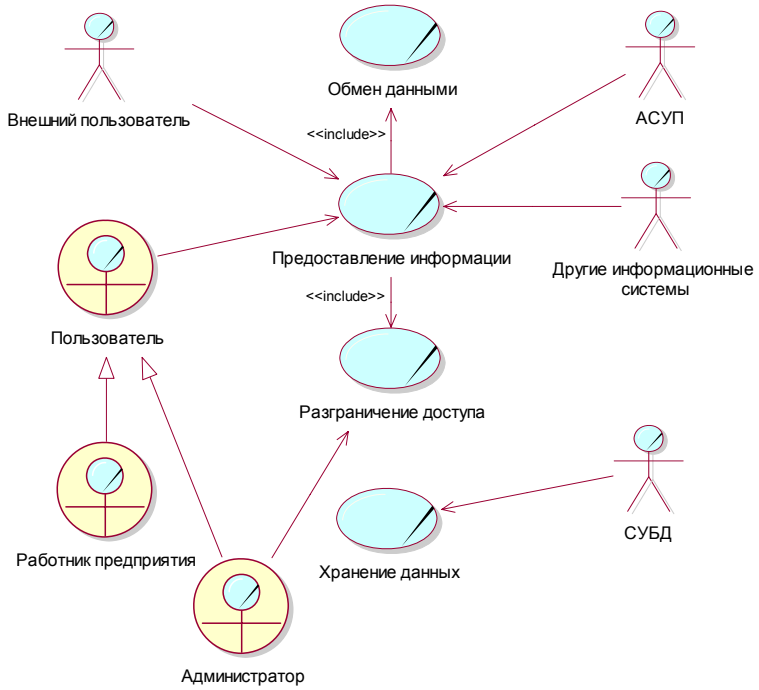


Рис. 1. Бизнес-прецеденты использования ИИС

Проведя объектно-ориентированный анализ предметной области, уже на этапе логической реализации по существующим требованиям можно приступить к моделированию структуры ИИС с помощью диаграмм классов (см. рис. 2), на которой обычно изображаются объекты, имеющие атрибуты и методы (операции), и связи между объектами. Стоит отметить, что каждая диаграмма классов должна показывать систему с определённой точки зрения, а в совокупности данные диаграммы должны составлять единую семантическую сеть, описывающую структуру разрабатываемой системы, то есть, по сути, являться онтологией. При необходимости, используя диаграммы состояний, также можно формализовать состояние каждого из объектов проектируемой системы.

Завершив основной этап структурного моделирования, можно приступать непосредственно к кодированию объектов. По окончании процесса кодирования программист получает компоненты или отдельные модули будущей информационной системы, которые позже будут реализованы на определённой вычислительной платформе. Таким образом, происходит переход на заключительный – физический уровень проектирования информационной системы, на котором определяется её архитектура, платформа реализации и прочие нюансы. Как правило, современные информационные системы реализуются средствами локальной вычислительной сети, используя в качестве элементов совокупность автоматизированных рабочих мест (АРМ). Недалеки и те времена, когда место проводной локальной сети займёт беспроводная вычислительная сети, которая строится по технологиям Wi-Fi.

Физический уровень реализации предполагает использование диаграмм компонентов, на которых обычно располагаются артефакты ИИС в виде файлов, модулей, компонентов базы данных и диаграмм развёртывания, показывающих физическую реализацию системы. Иногда данные диаграммы объединяются воедино, чтобы показать те артефакты, которые будут реализованы непосредственно на определённой вычислительной машине. Имея весь комплект моделей, хранящихся в репозитории одного из специализированных средств моделирования (например, IBM Rational Rose), можно не беспокоиться за базу знаний об информационной системе. В случае возникновения потребности в доработке модулей, либо отдельных компонентов,

а также восполнения недостающих знаний о принципах работы системы, её структуре и поведении, что особенно важно для предприятия, из которого ушёл человек, ответственный за проектирование информационной системы, всегда можно воспользоваться всей совокупностью моделей.

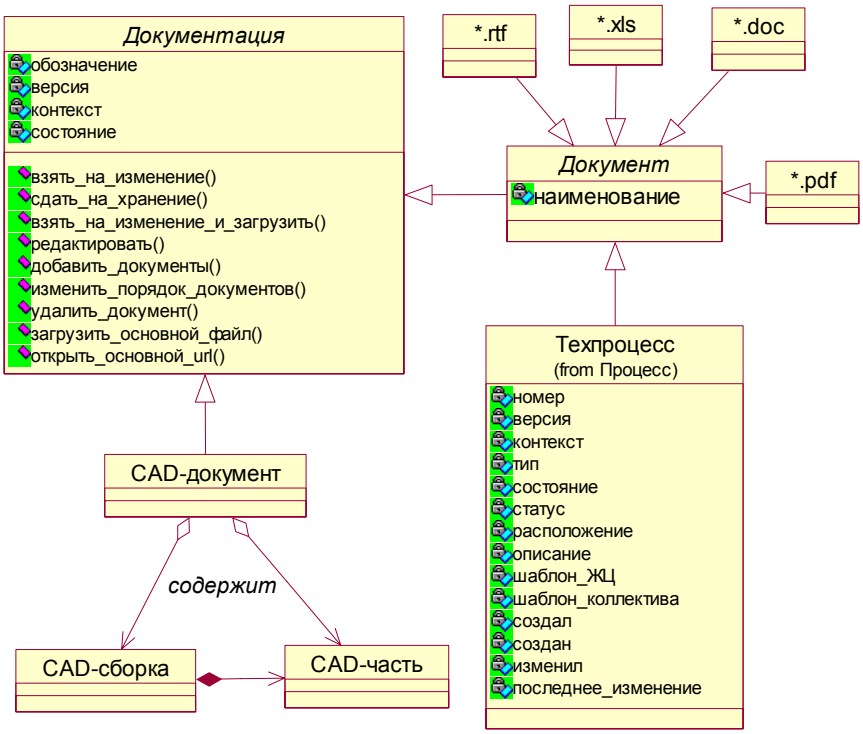


Рис. 2. Диаграмма классов ИИС «Документы и связи между ними»

Таким образом, в данной работе описан весь процесс проектирования интегрированной информационной системы. Большое внимание уделено именно визуальному моделированию. Визуальные модели просты в понимании, позволяют построить взаимосвязь между заказчиками и непосредственно создателями системы, они являются своего рода «чертежами» программного обеспечения, подобно чертежам в строительстве или машиностроении.

В настоящее время успешными являются те организации, которые идут в ногу со временем, используя современные технологии в области проектирования заказного программного обеспечения. Одним из аспектов такого успеха является использование визуального моделирования – как деятельности предприятия, так и создаваемой ИИС в частности.

Формирование облика самолёта на основе высокоточного математического моделирования

А.С. Кузнецов
Самарский государственный аэрокосмический университет имени академика С.П.Королёва (национальный исследовательский университет), Самара

В сложившемся порядке проектирования формирование облика летательного аппарата (ЛА) ведётся с определённым приоритетом требований аэродинамики. Требования весовой эффективности учитываются на этапе аэродинамического проектирования косвенно, зачастую на основе экспертных оценок без использования количественных показателей весовой эффективности. Существенную роль в процессе принятия решений играет квалификация и опыт разработчиков.

Современные требования к экономической эффективности требуют исключения ошибок в определении облика ЛА и, соответственно, повышения точности расчётов характеристик. Поиск путей повышения экономической эффективности приводит к исследованию новых нетрадиционных компоновок, что делает проблему повышения точности ещё более острой. Основным путём для решения этой проблемы является использование высокоточного математического моделирования на ранних стадиях проектирования. Это, в свою очередь, требует формализации процесса проектирования и замены качественных показателей эффективности на количественные.

На этапе предварительного проектирования, когда неизвестны абсолютные размеры конструкции и действующие нагрузки, удобно оперировать безразмерными параметрами и характеристиками. Так, в аэродинамических расчётах использование безразмерных коэффициентов аэродинамических сил и моментов и безразмерного коэффициента аэродинамического совершенства – аэродинамического качества, позволяет производить сравнение различных схем ЛА без привязки к конкретным размерам и условиям полёта. Критерием весовой эффективности обычно выступает относительная масса конструкции. Однако расчёт относительных масс агрегатов ЛА традиционно ведётся с использованием весовых формул, в основе которых лежат упрощенные математические модели, статистические данные, и в состав которых входят абсолютные размеры конструкции.

Работа направлена на развитие методов формирования облика ЛА основанных на высокоточном математическом моделировании. Метод направлен на повышение точности расчётов характеристик ЛА нетрадиционной формы на ранних стадиях проектирования. Предлагается алгоритм выбора рациональных геометрических параметров крыла, в котором в качестве критерия аэродинамической эффективности используется аэродинамическое качество, а в качестве критерия весовой эффективности – величина *коэффициента силового фактора*. Объединение критериев аэродинамической и весовой эффективности в комплексную целевую функцию происходит через взлётную массу самолёта. Для определения аэродинамических характеристик ЛА нетрадиционной формы используется панельный метод высокого порядка, а для вычисления величины коэффициента силового фактора – специальная *конечно-элементная модель переменной плотности*. Алгоритм реализован в виде единой программы выбора параметров крыла ЛА нетрадиционной формы.

Работа алгоритма протестирована на хорошо изученной проблеме выбора удлинения и сужения прямого трапециевидного крыла. Продемонстрирована работоспособность алгоритма и адекватность используемых математических моделей. Указано, что использование внутри единого алгоритма численных методов аэродинамики и весового анализа позволяет повысить точность расчёта нагрузок - учесть массовую разгрузку крыла и его деформации.

Применение генетических алгоритмов для проектирования подкрепленных панелей из композиционных материалов

А.В. Черняев
Самарский государственный аэрокосмический университет им. академика С.П. Королева (Национальный исследовательский университет), Самара

Рассматриваются вопросы разработки программного обеспечения для проектирования подкрепленных панелей из волокнистых композиционных материалов, работающих на сжатие. Отмечаются потенциальные сложности, возникающие при проектировании силовых конструкций из композитов.

Формулируется задача проектирования композитных панелей. Панели должны отвечать требованиям прочности и устойчивости (общей и местной) и обладать минимальной массой при действии заданной сжимающей силы. Параметры панелей, подлежащие выбору в ходе проектирования – количество и направление укладки монослоев композиционного материала, а также последовательность их чередования.

Отмечается дискретный характер сформулированной задачи проектирования и обосновывается целесообразность использования генетических алгоритмов (ГА) для отыскания проектов, обладающих минимальной массой. Рассматриваются особенности описанных в литературе подходов к решению данной задачи с использованием перестановочных генетических алгоритмов (Z.Gurdal, R. Haftka). Отмечаются их достоинства и недостатки. Предлагается оригинальный способ представления данных о параметрах и порядке чередования слоев в обшивке и ребрах панелей, позволяющий существенно снизить размерность задачи проектирования при ее решении с использованием двоичного генетического алгоритма.

Описывается разработанное программное обеспечение. Первый модуль разработанного ПО является модулем оптимизации, реализующим генетический алгоритм на языке технических расчетов Matlab. Его задача – генерация новых проектных решений путем выполнения операторов генетического алгоритма. Набор проектных параметров, соответствующий новому решению, передается во второй модуль, написанный на языке ANSYS Parametric Design Language и отвечающий за автоматическую генерацию геометрической модели подкрепленной панели с заданным набором параметров, ее дискретизацию конечными элементами, задание граничных условий и запуск на расчет решателя ANSYS для выполнения статического расчета и расчета на устойчивость. Данные о массе проекта панели, максимальном значении критерия прочности и собственное значение, характеризующее критическую силу потери устойчивости, передаются в первый модуль и служат для оценки "пригодности" (fitness) данного решения генетическим алгоритмом.

В качестве примера использования разработанного программного обеспечения рассматривается задача выбора рациональных параметров свободно опертой по четырем сторонам углепластиковой панели, подкрепленной П-образными стрингерами. Описываются особенности, связанные с дискретизацией панели конечными элементами и конструктивные ограничения, накладываемые на проектные параметры (максимальное и минимальное количество монослоев, укладка ориентированная на снижение межслойных напряжений). Даются оценки времени работы разработанного ПО. Демонстрируется выполнение наложенных ограничений (прочность, устойчивость) для полученного после оптимизации проекта. Определяется дальнейшее направление работы по разработке инструмента рационального проектирования подкрепленных композитных панелей.

Подсекция 8
«Информационные технологии в задачах проектирования ракетоносителей и космических аппаратов»

Усовершенствование системы управления угловой скоростью космического аппарата

Д.А. Азарных
Московский физико-технический институт, Долгопрудный
Ракетно-космическая корпорация «Энергия» имени С.П. Королёва, Королёв

Введение

Управление угловой скоростью КА осуществляется по средствам Исполнительных Органов (ИО). Классификация ИО представлена на рис. 1.

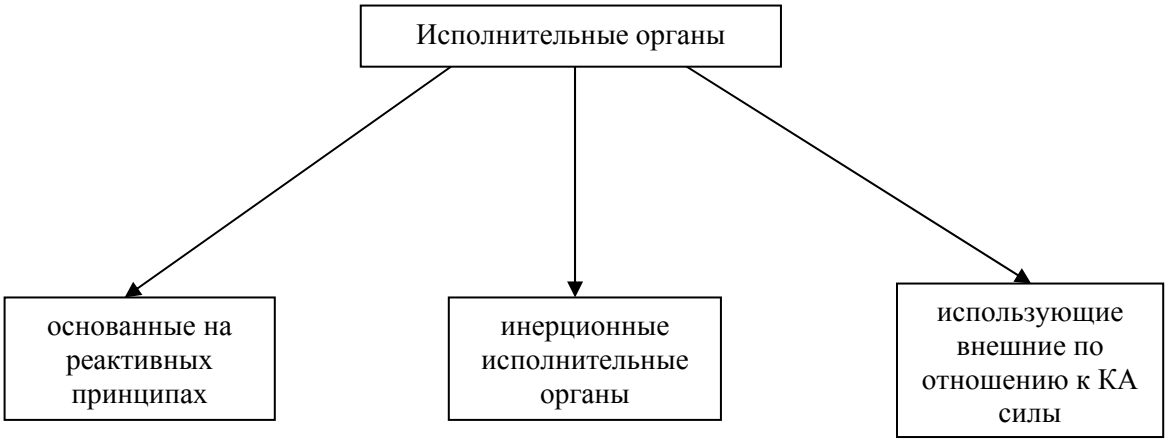


Рис. 1. Классификация исполнительных органов КА

Инерциальные исполнительные органы (ИИО) в КА представляют собой гироскопы. В зависимости от модификации ИИО используются одностепенные или двухстепенные гироскопы. В данной работе рассматривается система управления угловой скоростью КА по средствам одностепенных гироскопов (маховиков).

Формулировка задачи

Момент маховика в ИИО $\overrightarrow{M_i}$. Управляющий момент ИИО представляет собой сумму моментов маховиков: $\overrightarrow{M} = \sum_{i=1}^{i=n} M_i$, $\overrightarrow{M} \in M$.

Ограничение управляющего момента обусловлено существованием предельной угловой скорости маховиков. В современных ИИО используется $n = 3$ маховика. Использование $n > 3$ маховиков дает расширение области ограничения $\mathbf{M}$. Задача оптимального управления $n > 3$ маховиков:

$$\overrightarrow{W_B} = J^{-1}S\vec{u} - J^{-1}\left[W_B\left(JW_B + S\vec{h}\right)\right].$$

С вектором управления $\vec{u}$, вектором состояния $\vec{x} = \left(\overrightarrow{W_B}, \vec{h}\right)$. Здесь S – матрица, характеризующая расположение маховиков. Вектор управления выбирается так, что $\overrightarrow{M} = S\vec{u}$. На задачу накладываются ограничения: $\vec{h} \in \mathbf{H}$, $\vec{u} \in \mathbf{U}$, $\overrightarrow{W_B} \in \mathbf{\Omega}$.

Требуется найти такое управление $\vec{u}$, что в момент времени t_k:

1) $\overrightarrow{W_B}(t_k) = \overrightarrow{W_T}(t_k)$;

2) $T = t_k - t_0 \to min$.

Метод решения задачи

Данная задача является нелинейной и неразрешима в аналитическом виде. Вычислительные методы требуют больших ресурсов БЦВМ. Однако можно найти приближенное решение, считая $\left|\vec{u}\right| \to \infty$, и разбив задачу на 2 части:

1. Расчет вектора потребного момента.

Считаются известными $\overrightarrow{W_B^*}$ – начальная угловая скорость КА, $\overrightarrow{W_T}$ – потребная угловая скорость и $\overrightarrow{G_B}$ – общий кинетический момент. Требуется найти оптимальный управляющий момент $\overrightarrow{M} = \overrightarrow{M_T}$, который удовлетворял бы условиям задачи. Данная задача разрешима с помощью принципа максимума Понтрягина, который, однако, не дает аналитическое решение в нашем случае. Предлагается выбрать линейное управление, дающее асимптотическую устойчивость:

$\overrightarrow{M_T} = F(\vec{x}) + \overrightarrow{W_B}\overrightarrow{G_B} + J\overrightarrow{W_T}$,

где

$$F(\vec{x}) = \begin{cases} k\Delta K, & при\ \Delta K < \Delta K_0; \\ \left\|\overrightarrow{\Delta K}\right\| M_{max}, & при\ \Delta K > \Delta K_0, \end{cases}$$

$$M_{max} = \min_{i \in (1..n)} \left[\frac{a_i - m_0 sign(b_i)}{b_i} \right].$$

2. Расчет оптимальных управляющих сигналов.

Известными считаются потребный момент $\overrightarrow{M_T}$ и необходимо найти оптимальное управление по средствам n маховиков. То есть решить задачу оптимального управления: $\overrightarrow{M} = S\vec{u}$, $\overrightarrow{H} = S\vec{h}$, $\vec{u} = \vec{h}$ с ограничениями: $\vec{u} \in U$, $\vec{h} \in \mathbf{H}$.

Требуется найти такое управление $\vec{u}$, для которого верно:

1) $\overrightarrow{M}(t_k) = \overrightarrow{M_T}(t_k)$;

2) $T = t_k - t_0 \to min$.

Данную задачу предлагается решить методом наименьших квадратов, минимизируя квадратичный функционал. Полученное оптимальное управление:

$\vec{u} = S^T (SS^T)^{-1} \overrightarrow{M_T} - K_h \left(E_n - S^T (SS^T)^{-1} S\right) J_M \vec{p}$.

Выводы

Разработан общий алгоритм управления угловой скоростью КА для ИИО, содержащих $n > 3$ активных одностепенных гироскопов.

Литература

1. Раушенбах Б.В., Токарь Е.Н. “Управление ориентацией космических аппаратов” М: 1974.
2. Бранец В.Н., Шмыглевский И.П. “Применение кватернионов в задачах ориентации твердого тела”. М: “НАУКА”, 1973.
3. Амелькин Н.И. “Кинематика и динамика твердого тела: учебное пособие”. М: МФТИ, 2000.
4. Брайсон А., Хо Ю-Ши. “Прикладная теория оптимального управления”.
5. Понтрягин Л.С. “Принцип максимума в оптимальном управлении”.
6. Бранец В.Н., Черток М.Б., Казначеев Ю.В. “Оптимальный разворот твердого тела с одной осью симметрии”, журнал Космические исследования, т.22, №3 – 1984.
7. Ким Д. П. “Теория автоматического управления”, т.1, М.: “ФИЗМАТЛИТ”, 2003.
8. Егоров А. И. “Основы теории управления”. М.: “ФИЗМАТЛИТ”, 2004.
9. Понтрягин Л.С., Болотянский В. Г., Гамкрелидзе Р. В., Мищенко Е. Ф. “Математическая теория оптимальных процессов”. М: “НАУКА” ,1969.

Моделирование динамики сложной системы «экипаж – космический аппарат с постоянной закруткой»

В.А. Акулов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Закрутка космического аппарата (КА) как средство противодействия невесомости с использованием центробежной силы инерции была предложена К.Э. Циолковским более 100 лет назад. Анализируя эту кинематическую схему с позиций моделирования, отметим, что речь идет о переходе от системы отсчета, близкой к инерциальной (ИСО, Земля), к явно выраженной неинерцианной системе (НСО, КА) [2]. Такой переход неизбежно сопряжен с появлением различий на физическом уровне. Их необходимо оценить (адекватность моделей), наметить пути управления такими различиями и, что особенно важно, их минимизировать. Пребывание человека на Земле (объект моделирования) характеризуется следующими основными факторами: действием силы тяжести $\vec{P} = m\vec{g}$, крайне незначительной величиной кориолисовой силы ($1{,}43 \cdot 10^{-2}$ % от P) и однородностью [2, 3]. Тогда условие адекватности НСО (КА), как рукотворной системы радиуса R, совершающей вращение с угловой скоростью ω, и ИСО (Земля), принимает вид:

$$\begin{cases} \omega^2 R = g, \\ \left[\vec{V}_{отн}, \vec{\omega}\right] \approx 0. \end{cases} \qquad (1)$$

Здесь $\vec{V}_{отн}$ – скорость тела в относительном движении.

Как следует из (1), возможны два варианта построения и управления закруткой КА:

вариант 1: в качестве аргумента выбирается ω. Тогда зависимой переменной становится R:

$$R = g / \omega^2; \qquad (2)$$

вариант 2: задается R (т. е. габариты), и зависимой переменной становится ω:

$$\omega = \pm\sqrt{g / R}\,. \qquad (3)$$

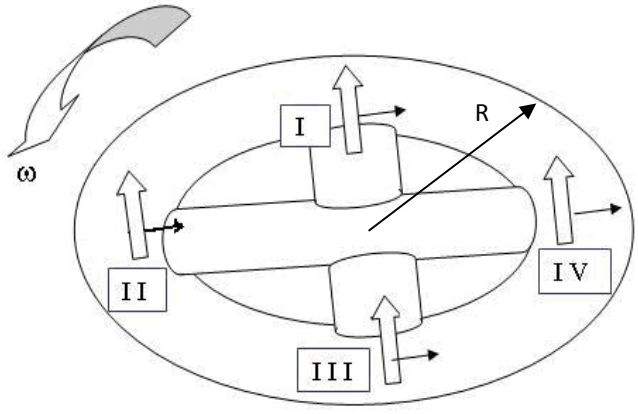


Рис. 1. Схема действия $F_{кор}$ на КА типа «Колесо»
⟹ Направление $V_{отн}$, ⟶ направление $F_{кор}$

Результаты моделирования

Моделирование кинематики и динамики было выполнено применительно к закрученному КА типа «Колесо» (рис. 1). К достоинствам схемы относится значительная свобода перемещений членов экипажа (см. поз I – IV) с грузом и без него. Если следовать рекомендациям космической медицины (n ~ 1 ÷ (2) об/мин) [1], то получим, что R ~ 890 ÷ (224) м. Совершенно очевидно, что на современном этапе подобные габариты КА не приемлемы. Результаты оценок кинематических параметров, выполненных по варианту 2 для 4-х величин R, приведены в таблице 1. Сравнивая

полученные данные (строка №1) с условиями Земли ($\omega = 0{,}73 \cdot 10^{-4}\, c^{-1}$), заключаем: а) отличия значительны и достигают по величине 4-х порядков; б) вращающееся «жилище» не моделирует земные условия по данному параметру. По данным КМ при вращении с частотой $n > 3$, а фактически $n > 4$ (строка №2), возникает проблема длительной (6 – 8 сут) адаптации экипажа к весьма некомфортным условиям, что существенно снижает научно - практическую ценность полета [1].

Таблица 1. Зависимость кинематических и динамических параметров КА от R при моделировании земной силы тяжести (g)

	R, м	20	30	40	50
№ №	*1*	*2*	*3*	*4*	*5*
1	ω, c^{-1}	0,7	0,6	0,5	0,4
2	n, об / мин	6,7	5,5	4,7	4,2
3	$W_{кор}/g$, %	17	14	12	11

Оценивая данные табл. 1 с позиций аналитической механики отметим, что относительная величина кориолисова ускорения ($W_{кор}/g$) значительна и составляет от 17 до 11 % (строка № 3), в то время, как для Земли этот показатель не превышает 0,02%. В связи с этим, были оценены величины кориолисовой силы по формуле

$$\vec{F}_{кор} = 2(m + M)\left[\vec{V}_{отн}, \vec{\omega}\right],$$

учитывающей как массу космонавта (m), так и массу переносимого груза (M). Особый интерес представляет случай, когда M – масса реальных скафандров (M = 10 – 136 кг, см. табл. 2 и примечание к ней).

В обследованном диапазоне масс и относительных скоростей кориолисова сила изменялась в значительных пределах. В частности, при $V_{отн}$ = 0,8 м/с величина $\vec{F}_{кор}$ достигает 77 ÷ 207 н, или в относительных единицах 10 ÷ 26%. С ростом Vотн это рассогласование только увеличивается, достигая 59% (колонка 7, скафандр EMU). Таким образом, не выполняется условие динамического подобия, представленное уравнением системы (1).

Таблица 1. Оценка $F_{кор}$ при перемещениях космонавтов в КА типа «Колесо» (R = 30 м, ω = 0,6 c^{-1}, n = 5,5 об/мин)

	Кориолисова сила [Н] / [$F_{кор}$ / Р, %%]					
1	*2*	*3*	*4*	*5*	*6*	*7*
Масса, кг	80	80 + **10**	80 + 40	80 + 50	80+**110**	80+**136**
V= 0,8 м/с	77/10	**86/11**	115/15	125/16	**182/23**	**207/26**
V= 1,2 м/с	115/15	**130/17**	173/22	187/24	**274/35**	**311/40**
V= 1,8 м/с	173/22	**194/25**	259/33	281/36	**410/52**	**467/59**

Примечание: числитель – $F_{кор}$ в Н, знаменатель – относительная величина в % к собственному весу на Земле Р. Колонки 3, 6, 7 – данные со скафандрами соответственно «Сокол», «Орлан» (Россия) и «EMU» (США).

Отметим, что отличия вращающегося «жилища» от земных условий столь значительны, что необходимо учитывать как величину, так и направление действия $\vec{F}_{кор}$. В частности, при движении от оси вращения (рис. 1, поз. I) $\vec{F}_{кор}$ будет смещать космонавта вправо со значительными усилиями, определенными в табл. 2. При движении к оси (поз. III) следует учитывать два фактора: $\vec{F}_{кор}$ направлена в сторону вращения, а $\vec{F}_{цб}$ препятствует радиальному перемещению. Очевидно, что потребуются устройства типа трапов, а от экипажа приобретение навыков перемещения с участием всех 4-х конечностей (аналог вертикальная лестница на Земле) со сменой поз: «голова вперед», «разворот», «голова назад». В ситуациях (II) и (IV) будет происходить

изменение величины искусственного веса (рис. 1). Причем, в первом случае она будет существенно убывать (аналог: движение под гору), а во втором – возрастать (аналог: подъем в гору, увеличение физической нагрузки). Это означает необходимость в приобретении специфических навыков «оптимизации» внутрикорабельного маршрута.

Выводы

Физическое моделирование естественной гравитации путем постоянной закрутки КА оказывается весьма проблематичным. Если следовать медицинским рекомендациям (n = 1 – 2 об/мин), то внешний радиус аппарата достигает 220 – 890 м, что не приемлемо по техническим соображениям. Если исходить из реальных габаритов (десятки метров), то возникают проблемы самочувствия и работоспособности членов экипажа.

Увеличение центробежной силы за счет интенсификации вращения (ω) с целью имитации силы тяжести сопровождается существенным ростом кориолисовой силы. Ее величина достигает десятков процентов от веса на Земле, что приводит к значительным затруднениям с перемещениями экипажа, подготовкой и осуществлением внекорабельной деятельности, затягивает погрузочно-разгрузочные работы.

Апробация авторских разработок (модели, компьютерные программы) с привлечением реальных данных показала их эффективность и целесообразность применения при выборе проектных параметров и назначении режимов закрутки пилотируемых космических аппаратов.

Литература

1. Котовская А. Р. и др. Медико-биологические аспекты проблемы создания искусственной силы тяжести / Котовская А. Р., Шипов А. А., Виль - Вильямс И. Ф. – М.: Слово, 1996. – 201 с.: ил.

2. Рогачев Н. М. Курс физики. Книга 1. Механика. Молекулярная физика: Учеб. пособие. – Самар. гос. аэрокосм. ун-т. Самара, 2006. - 192 с.

3. Тарг С. М. Краткий курс теоретической механики: Учеб. для ВТУЗов/ С. М. Тарг. – 12 изд., стер. – М.: Высш. шк., 2002. – 416 с.: ил.

Показано, что практическая реализация идеи К. Э. Циолковского о космическом «жилище», совершающим вращение, сопряжена с решением ряда междисциплинарных проблем, требующих всестороннего анализа со стороны многочисленных специалистов, в числе которых, проектировщики, космонавты, испытатели, баллистики, медики.

Расчет продолжительности оптимального по времени разворота космического аппарата дистанционного зондирования Земли с использованием инерционных исполнительных органов-маховиков, при смене режимов ориентации

Р.С. Алиев
Московский физико-технический институт, Долгопрудный
Ракетно-космическая корпорация «Энергия» имени С.П. Королёва, Королёв

В данной работе рассматривается угловое движение автоматического космического аппарата дистанционного зондирования Земли. Виток состоит из двух участков (см. рис. 1): рабочего (в окрестности апогея), на котором аппарат решает целевую задачу, и дежурного.

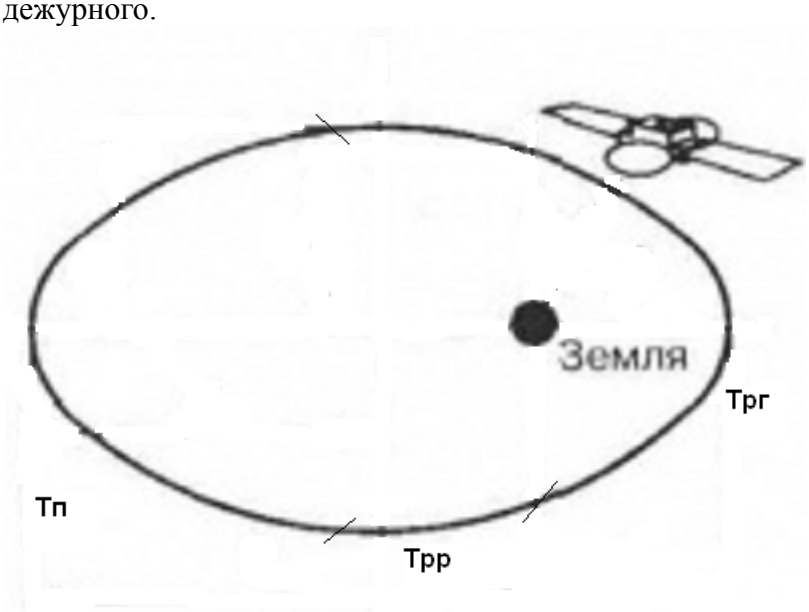


Рис. 1. Движение спутника по высокоэллиптической орбите: Трр – искомое минимальное время разворота; Трг – суммарное время разгрузки маховиков; Тп – время полезной работы

На дежурном участке орбиты решаются задачи выполнения коррекций орбиты с использованием электрореактивных двигателей и гравитационной разгрузки накопленного кинетического момента инерционных исполнительных органов - маховиков. При этом для перехода в режимы коррекции орбиты или для перехода в ориентацию, требуемую на рабочем участке, совершаются программные развороты космического аппарата.

Формулировка задачи

Необходимо определить время начала разворота в ориентацию рабочего участка или в ориентацию, требуемую для проведения коррекции орбиты и разработать бортовой алгоритм расчета минимальной продолжительности разворота космического аппарата дистанционного зондирования Земли вокруг вектора конечного поворота.

Метод решения задачи

Рассмотрим разворот космического аппарата относительно инерциальной системы координат [1, 2]. Полагаем, что область вариации кинетического момента инерционных исполнительных органов имеет форму сферы с радиусом R. Обозначим суммарный вектор накопленного кинетического момента космического аппарата + инерционные исполнительные органы через G_0. Считаем, что за время разворота приращением вектора G_0 за счет действия моментов внешних сил можно пренебречь.

Тогда вектор G_0 остается неизменным (по величине и направлению) в системе координат $X_A Y_A Z_A$

Из кватерниона рассогласования положения космического аппарата найдём мгновенную ось разворота, относительно которой для оптимальности времени будем совершать плоский разворот в пространстве $\vec{e} = \vec{e}(e_1, e_2, e_3)$ [2] (см. рис.2).

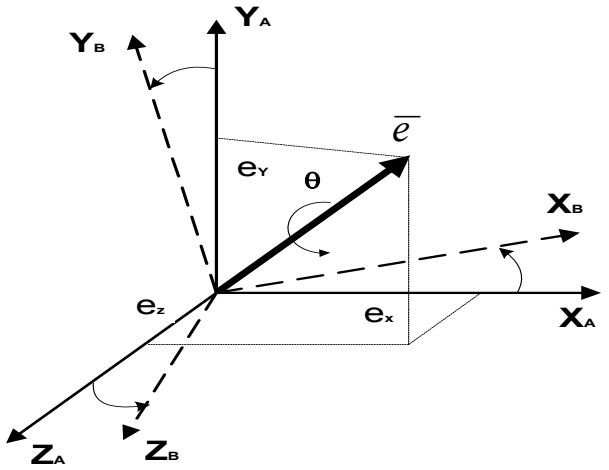


Рис. 2. Эйлерова ось вращения, определяющая поворот системы координат

Зная максимальный управляющий момент управляющего двигателя-маховика $M_{max} = 0{,}15 Нм$ (оценка) – управляющий момент системы маховиков в произвольном направлении, тензор инерции космического аппарата $I = \begin{pmatrix} I_{11} & 0 & 0 \\ 0 & I_{22} & 0 \\ 0 & 0 & I_{33} \end{pmatrix}$, и мгновенную ось разворота, найдём максимальное угловое ускорение, создаваемое системой маховиков вокруг вектора конечного поворота

$$\varepsilon = \frac{M_{max}}{\sqrt{(J_1 e_1)^2 + (J_2 e_2)^2 + (J_3 e_3)^2}}.$$

Разделим угол разворота вокруг оси мгновенного разворота: $\theta = 2\arcsin(\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2})$ на 20 равных частей – компромисс между точностью расчёта и вычислительных возможностей на борту космического аппарата, для последовательного итерационного решения задачи.

Обозначим угловую скорость вокруг вектора конечного поворота через ω. Тогда $\mathbf{G_0} = \mathbf{H} + \omega \mathbf{Je}$ [1], где через H обозначен суммарный кинетический момент инерционных исполнительных органов, а через J тензор моментов инерции космического аппарата. При развороте с максимальной угловой скоростью модуль вектора H равен R. В этом случае получаем уравнение $R^2 = (\mathbf{G_0} - \omega \mathbf{Je})^2$. Решая его, найдем

$$\omega = \frac{(\vec{G_0} J\vec{e}) \pm \sqrt{(\vec{G_0} J\vec{e})^2 - (J\vec{e})^2 (G_0^{\,2} - R^2)}}{(J\vec{e})^2}.$$

В начале разворота осуществляется набор угловой скорости разворота с ускорением ε. Угловая скорость при этом равна $\omega_н = \sqrt{2\varepsilon\theta_T}$, где θ_T – текущий угол разворота.

В конце разворота (при торможении) осуществляется гашении угловой скорости разворота с ускорением ε угловая скорость $\omega_к = \sqrt{2\varepsilon\theta_т}$.

Суммируя на каждом шаге итерации времена, в итоге, получим искомое оптимальное время разворота. Так решается поставленная задача.

Результаты расчетов продолжительности разворота

С использованием разработанной программы на языке Си, был проведен расчет продолжительности разворотов.

Использовались следующие варианты начального накопленного суммарного кинетического момента:

1. $\mathbf{G_0} = (0,\ \ 0,\ \ 0)$;
2. $\mathbf{G_0} = (5\,Нмс,\ \ 5\,Нмс,\ \ 5\,Нмс)$;
3. $\mathbf{G_0} = (10\,Нмс,\ \ 10\,Нмс,\ \ 10\,Нмс)$.

Разворот проводился по 26 различным направлениям, с углом разворота 180°. Ниже приведены результаты 6 разворотов по направлениям осей системы координат.

Направления вектора конечного поворота	1)	2)	3)
+X	1036.3 / 0.00375	963.5 / 0.00427	931.0 / 0.00456
−X	1036.3 / 0.00375	1195.3 / 0.00302	1527.1 / 0.00206
+Y	1231.9 / 0.00300	1145.2 / 0.00342	1098.7 / 0.00365
−Y	1231.9 / 0.00300	1160.0 / 0.00242	1908.9 / 0.00165
+Z	761.5 / 0.005	761.5 / 0.005	761.5 / 0.005
−Z	761.5 / 0.005	761.5 / 0.005	870.3 / 0.0411

Приводятся следующие значения: длительность разворота в секундах/максимальная угловая скорость в течение разворота в рад/с.

Как показывают результаты моделирования, при разбиении угла разворота на 20 интервалов ошибка в расчете продолжительности разворота не превышает 5%.

Выводы

Разработан бортовой алгоритм расчета минимальной продолжительноси разворота космического аппарата дистанционного зондирования Земли с использованием маховиков при смене режимов ориентации.

Перспективы

Сейчас в разработке находится алгоритм оптимального разворота КА с максимальной проекцией вектора угловой скорости на заданное направление с целью его дальнейшего сравнения с исходным алгоритмом для выявления наиболее оптимального расчёта.

Литература

1. Б.В. Раушенбах, Е.Н. Токарь. Управление ориентацией космических аппаратов. – М.: Наука, 1974.

2. Бранец В.Н., Казначеев Ю.В., Черток М.Б. Оптимальный разворот твердого тела с одной осью симметрии. Космические исследования. Том XXII вып.3. – М.: Наука, 1984.

Повышение эффективности авиационных работ по распределению веществ в сельском хозяйстве на основе оптимизации траектории движения воздушного судна

Р.Б. Алтынбаев
Оренбургский государственный университет, Оренбург

Введение

Современное стабильное и экономически развитое государство не может существовать без надежного агропромышленного комплекса. При соблюдении данного условия государство будет поддерживать отечественного поставщика сельскохозяйственных продуктов и не будет в зависимости от импорта иностранных товаров, а это уже соответствует интересам национальной безопасности.

Большинство развитых стран мира выбрало направление «зелёной революции», основанной на выращивании высокоурожайных культур нового поколения, полученных с использованием достижений селекции и генетики, но низкая сопротивляемость этих культур против болезней и вредителей требует использования высокоэффективных средств защиты в виде специально вносимых веществ, которые наиболее оптимально с экономической, экологической и технологических точек зрения вносить авиационным методом. Следует также отметить, что один из самых перспективных и эффективных способов опрыскиваний – малообъёмное и ультрамалообъёмное, производится только с помощью средств авиации [1].

Особо остро стоит проблема повышения эффективности технологической подготовки проведения авиационного распределения веществ и биологических объектов (АРВ и БО) в виду их значимости в планировании и оценке предстоящих работ. В настоящее время определение параметров технологического процесса АРВ и БО проводится по устаревшим методикам с применением старых справочников, использующих приближённые расчёты: учитываются только стандартные способы обработки, а именно челночный или загонный, в то время когда существуют более рациональные способы (например, комбинированный); по существующим методикам расчёта предполагается, что сельскохозяйственный участок имеет строго прямоугольную форму, что не всегда соответствует действительности; некорректно учитываются перелёты от аэродрома до участка, а именно не рассматривается их взаимное расположение и так далее.

Вышеуказанные проблемы обусловили насущную необходимость проведения научно-исследовательских работ в данной области.

Этапы исследований

С целью повышения эффективности проведения авиационного распределения веществ и биологических объектов необходимо выполнить следующие этапы научно-исследовательских работ:

– моделирование технологического процесса АРВ и БО;

– выбор и обоснование критерия эффективности технологического процесса АРВ и БО;

– разработка алгоритма определения оптимальной траектории движения воздушного судна при проведении АРВ и БО;

– разработка программного обеспечения для реализации предложенного алгоритма на языке высокого уровня с возможностью визуализации технологического процесса АРВ и БО;

– подтверждение достоверности разработанной модели и её программной реализации реальным объектам и процессам;

– разработка методики определения оптимальных параметров технологического процесса АРВ и БО, повышающих эффективность его функционирования.

Моделирование технологического процесса

В качестве исходных данных для моделирования технологического процесса используется традиционный набор лётно-технических характеристик воздушного судна (ВС) для каждого варианта сельскохозяйственного оборудования и подробные данные о технологии различных видов АРВ и БО и их способах. Общая схема технологического процесса представлена на рис.1.

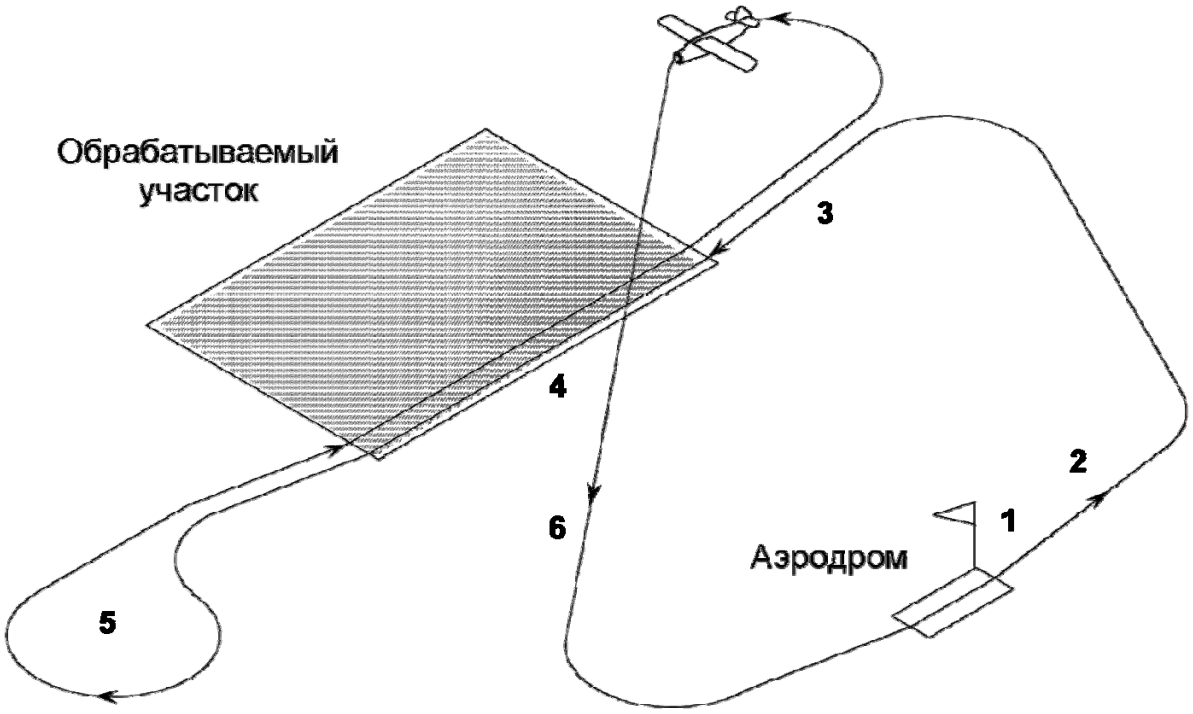


Рис. 1. Общая схема технологического процесса АРВ и БО

После загрузки ВС веществом пилот взлетает с рабочего аэродрома (1), набирает высоту (2) и кратчайшим путем направляется к участку, предназначенному под обработку.

Определив с воздуха местоположение массива и установив направление полёта и порядок обработки, пилот не менее чем за 100 – 200 м до границы обрабатываемого участка (в зависимости от типа ВС) переводит ВС в режим «бреющего полёта» (к «бреющим» относятся полёты, осуществляемые на высоте от 5 до 50 м над рельефом местности), т. е. снижается до установленной рабочей высоты (3). На этой высоте он распыливает вещество (4), включая с этой целью в работу опыливатель (или опрыскиватель) в начале массива и выключая его в конце участка.

Пролетев на рабочей высоте до выходного сигнала, пилот набирает 50 м высоты и разворачивает ВС на 180° для захода на следующий гон (5). Выполнение на бреющем полёте каких-либо доворотов и разворотов с целью изменить направление полёта для следования по изгибам обрабатываемого участка категорически запрещается [2].

Выполнив разворот, пилот вновь направляет ВС к обрабатываемому участку и не менее чем за 100 – 200 метров до него снижается до рабочей высоты и описанным выше способом распыливает вещество над следующей полосой участка (гоном) и так далее. Израсходовав таким образом всё содержимое бака опыливателя (или опрыскивателя), пилот набирает высоту и по кратчайшему направлению возвращается

на аэродром для очередной загрузки (6).

В зависимости от характера и условий выполняемых работ количество отдельных полётов доходит до $30-50$, а иногда и более за рабочий день. Таким образом, технологический процесс, складываясь из отдельных полётов на протяжении рабочего дня, носит ярко выраженный циклический характер.

В свою очередь, каждый производственный цикл складывается из повторяющихся операций (элементов) цикла:

– взлёт ВС (разбег, набор высоты и поворот на курс);

– полёт к обрабатываемому участку;

– работа ВС над участком (распределение веществ и биологических объектов, развороты для захода на очередной гон);

– возвращение ВС на аэродром;

– посадка ВС (поворот с курса, планирование и пробег);

– загрузка ВС (подруливание к месту загрузки, загрузка опыливателя (или опрыскивателя) веществом, выруливание на старт).

Все эти операции, как правило, обязательно встречаются в цикле. Распыливание вещества над обрабатываемой поверхностью или растительностью занимает относительно небольшой удельный вес в общем балансе рабочего времени. Более подробно моделирование технологического процесса АРВ и БО изложено в [3].

Выбор критерия эффективности

В основе повышения эффективности работ по АРВ и БО положена идея оптимизации траектории движения воздушного судна. При нахождении оптимальной траектории движения можно было использовать такие частные критерии, как максимум дальности полёта, минимум расхода топлива, минимум (максимум) времени полёта. Однако решение, оптимальное по частному критерию, может оказаться в целом неэкономичным, то есть здесь имеется задача многокритериальной оптимизации технологического процесса. В общем виде она имеет следующий вид:

$$\max\{f_1(x) = F_1\},$$

$$\max\{f_2(x) = F_2\},$$

…

$$\max\{f_k(x) = F_k\}, \text{ при } x \in X,$$

где X – множество допустимых значений переменных x;

k – число целевых функций (критериев);

F_i – значение i-го критерия (целевой функции);

max – означает, что данный критерий нужно максимизировать.

Заметим, что по существу многокритериальная задача отличается от обычной задачи оптимизации только наличием нескольких целевых функций вместо одной.

Выбранные критерии эффективности представлены в соответствии с рис. 2. Принятые обозначения следующие:

P – производительность, определяется как площадь обработанного участка в единицу времени.

M – экономичность, определяется как стоимость обработки единицы площади в затратах авиационного топлива.

$S^{знв}$ – площадь зоны нецелевого воздействия. Такие факторы, как ветер, ошибки сигнальщиков, техническое состояние воздушного судна, погрешность при включении и выключении опрыскивающей аппаратуры, а также опыт и стаж работы пилотов приводят к тому, что не всегда удается выдерживать строго параллельные заходы и вследствие этого получаются зоны нецелевого воздействия: необработанные зоны, зоны с двойной обработкой, зоны загрязнения.

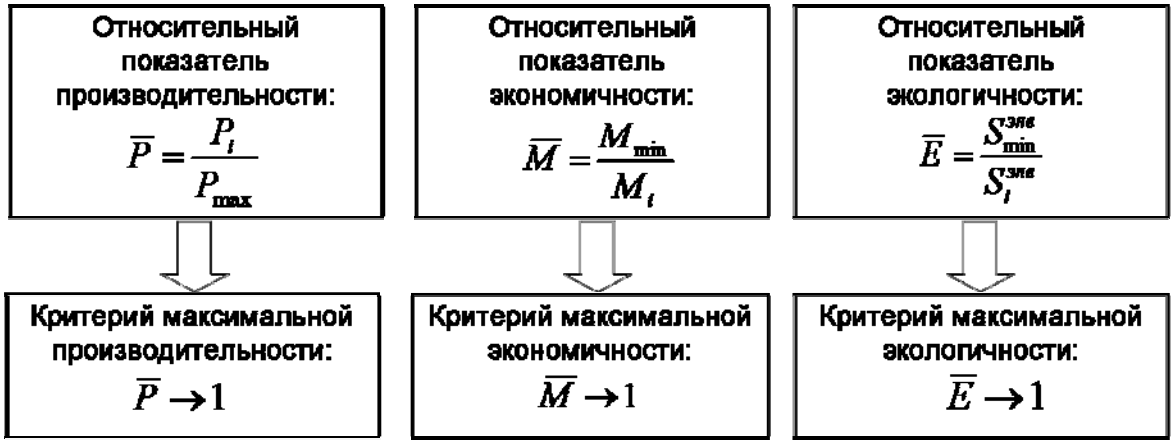


Рис. 2. Показатели и критерии эффективности

Существует множество вариантов траектории движения ВС при проведении АРВ и БО, для каждого из них можно найти показатели Pi, Mi, $S_i^{знв}$, а также минимальные и максимальные значения соответствующих показателей: P_{max}, M_{min}, $S_{\min}^{знв}$. Переход к относительным показателям осуществляется по следующим формулам:

$$\overline{P} = \frac{P_i}{P_{\max}},\ \overline{M} = \frac{M_{\min}}{M_i},\ \overline{E} = \frac{S_{\min}^{знв}}{S_i^{знв}},$$

где $\overline{P}$ – относительный показатель производительности, $\overline{M}$ – относительный показатель экономичности, $\overline{E}$ – относительный показатель экологичности.

Целевая функция и критерий эффективности в данном случае примут следующий вид:

$$F = \overline{P} \cdot kp + \overline{M} \cdot km + \overline{E} \cdot ke,$$

$$F \to \max,$$

где kp, km, ke – весовые коэффициенты соответствующих показателей.

Алгоритм оптимизации

Сущность алгоритма сводится к выполнению следующих операций:

1. Задаются координаты углов многоугольника соответствующего участка, предназначенного для обработки.
2. Задаётся вектор текущего направления движения ВС.
3. Строится траектория движения ВС согласно заданной ширине гона и выбранного способа движения.
4. В процессе вращения вектора линии гона от 0 до 359° с приращением на 1° по часовой стрелке строится траектория движения ВС.
5. Относительно рассчитанной траектории определяются технологические параметры АРВ и БО.

По окончании работы алгоритма мы получим оптимальный вариант траектории движения ВС. Схема алгоритма представлена на рис. 3.

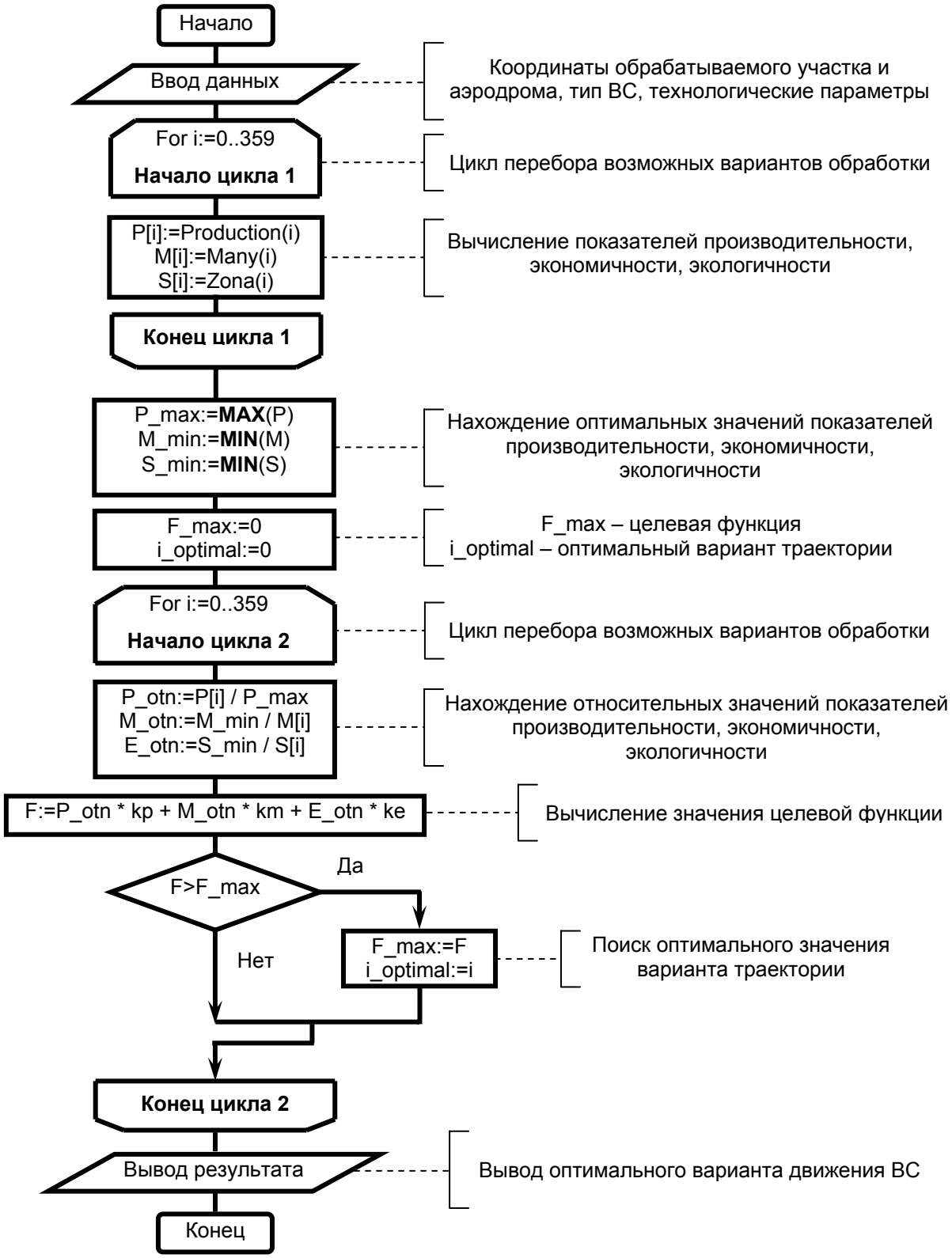


Рис. 3. Схема алгоритма определения оптимальной траектории

Программная реализация

Разработанный алгоритм позволяет создать программное средство для расчёта оптимальной траектории движения ВС с возможностью визуализации технологического процесса АРВ и БО. Внешний вид программного средства показан на рис. 4.

Основные возможности программного средства:

– загрузка цифровой версии географической карты реального района с расположенными на ней участками сельскохозяйственных угодий;

– выбор критериев оптимизации траектории;

– получение выходных параметров на каждом шаге расчёта с возможностью вывода на печать.

– построение графиков по результатам вычислений.

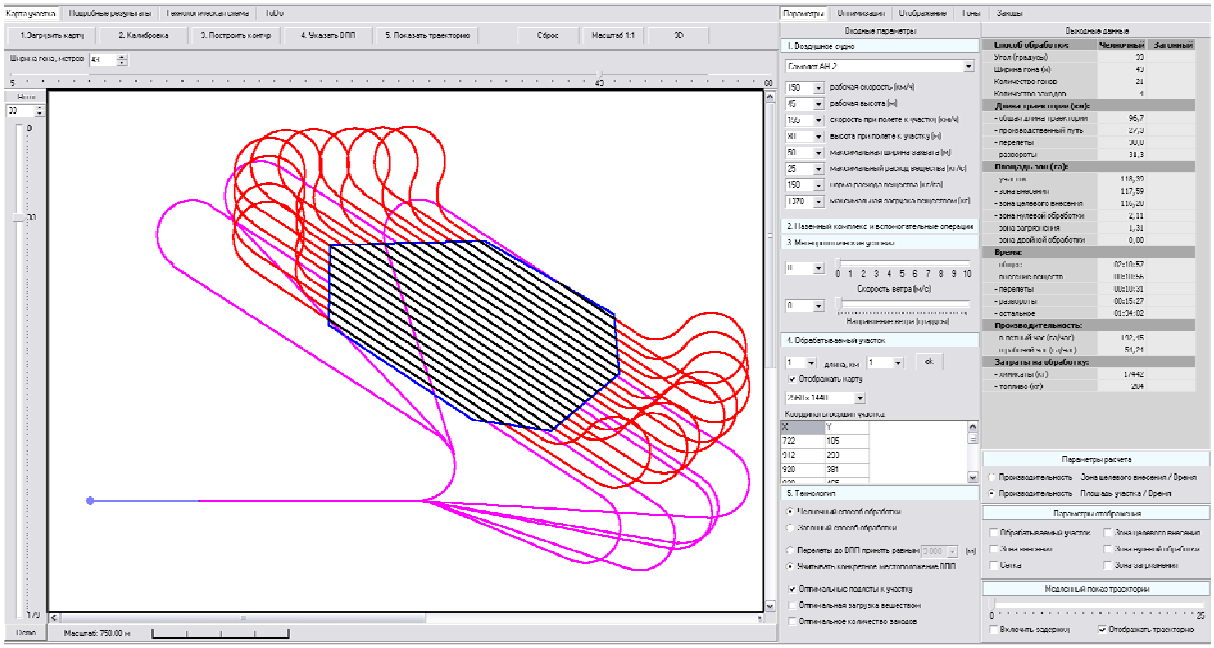


Рис. 4. Внешний вид программного средства по расчёту оптимальной траектории

Заключение

В настоящее время исследования находятся на стадии подтверждения достоверности разработанной модели и её программной реализации реальным объектам и процессам. При этом уже разработанное программное средство позволяет проводить различные вычислительные эксперименты и получать предварительные выводы.

Литература

1. Бондаренко В.А. [и др.] Инновационные процессы в авиационно-химических работах – экологический аспект. – Оренбург: ОГУ, 1998. – 200 с.

2. Назаров В.А. [и др.] Применение авиации в сельском и лесном хозяйстве. – 2-е изд., перераб. и доп. – М.: Транспорт, 1975. – 312 с.

3. Алтынбаев Р.Б., Хибатуллин С. Г. Повышение точности расчёта времени выполнения авиационных работ по распределению веществ и биологических объектов / Научно-технические ведомости СПбГПУ. Серия «Информатика. Телекоммуникации. Управление». 2010. – №5

Управление колебаниями спутника на стационарной орбите

Л.К. Бабаджанянц, Ю.Ю. Пупышева
Санкт-Петербургский государственный университет,
факультет прикладной математики – процессов управления, Санкт-Петербург

Введение

В работах [1-3, 6] были предложены различные алгоритмы оптимального управления по расходу, как для конкретных задач, так и в общей постановке для линейных систем. При этом управление предполагалось кусочно-постоянным, а оптимизация проводилась по моментам переключения. В настоящей работе предлагается аналогичный алгоритм в задаче управления вращательным движением спутника в окрестности любого из четырех его устойчивых положений относительного равновесия на круговой орбите, а оптимизация проводится не только по моментам переключения, но и по высотам ступеней управления. Алгоритм допускает автоматический (компьютеризируемый) режим вычисления оптимального управления.

Рассмотрим механическую систему, описываемую уравнением

$$d\,x/d\,t = Ax + U(t) \tag{1}$$

относительно вектор-функции $x(t) = (x_1,...,x_n) \in R^n$ аргумента t при условии

$$x(0) = x_0, \tag{2}$$

где $x_0 = (x_{10},...,x_{n0}) \in R^n$, $U(t) = (u_1,...,u_n) \in R^n$, A – постоянная $n \times n$ - матрица.

Компоненты u_i управления $U(t)$ предполагаются кусочно-постоянными функциями времени с конечным числом точек переключения, последняя из которых обозначается символом T, т.е. ступенчатой функцией с конечным числом ступеней. Предполагается также, что выполнены неравенства

$$-\infty \le u_{*,i} \le u_i \le u_i^* \le +\infty\,. \tag{3}$$

При кусочно-постоянном управлении решение задачи (1), (2) будет суммой нескольких слагаемых, отвечающих тем или иным собственным значениям матрицы A. Слагаемое, отвечающее собственному значению λ, будем называть частотной составляющей решения, отвечающей этому λ.

В качестве минимизируемого функционала рассматривается величина

$$J = \sum_{k=1}^{n} \int_0^T |u_k(\theta)|\,d\theta\,, \tag{4}$$

которую естественно назвать функционалом типа «расход топлива». Допустимым будем считать удовлетворяющее неравенствам (3) управление U, которое в момент T обращает в нуль одну или несколько избранных частотных составляющих решения. Обозначая сумму избранных частотных составляющих $\tilde{x}(t)$, запишем это условие:

$$\tilde{x}(T) = 0\,. \tag{5}$$

О выполнении условия (5) мы будем говорить также как о гашении избранных частотных составляющих решения или, ради краткости, просто как о гашении избранных частот. Допустимые управления мы рассматриваем далее в следующем представлении:

$$u_k(t) = \sum_{i=1}^{2r_k} h_i^k (-1)^{i+1} H(t - t_i^k) + \sum_{i=1}^{2q_k} \tilde{h}_i^k (-1)^i H(t - \tilde{t}_i^{\,k})\,. \tag{6}$$

Здесь управление разбито на положительные и отрицательные ступени (r_k – число положительных, q_k – число отрицательных ступеней компоненты u_k). Величины

$t_i^k, \widetilde{t}_i^k \in [0,T]$ – моменты времени, соответствующие переключениям этих ступеней, $H(t)$ – функция Хевисайда ($H(t)=1$ при $t \ge 0$ и $H(t)=0$ при $t<0$). Относительно величин $t_i^k, \widetilde{t}_i^k$ предполагается, что они удовлетворяют неравенствам

$$0 \le t_1^k < t_2^k < \ldots < t_{2r_k}^k \le T,\ 0 \le \widetilde{t}_1^k < \widetilde{t}_2^k < \ldots < \widetilde{t}_{2q_k}^k \le T. \qquad (7)$$

Коэффициенты $h_i^k, \widetilde{h}_i^k$ удовлетворяют условиям $h_i^k = h_{i+1}^k, \widetilde{h}_i^k = \widetilde{h}_{i+1}^k$ при $i = 1,3,5,\ldots$. Кроме того, $h_i^k, \widetilde{h}_i^k$ удовлетворяют неравенствам

$$h_{k,*} \le h_i^k \le h_k^*,\ \widetilde{h}_{k,*} \le \widetilde{h}_i^k \le \widetilde{h}_k^*, \qquad (8)$$

где величины $h_{k,*}, \widetilde{h}_{k,*} \in [0,+\infty), h_k^*, \widetilde{h}_k^* \in [0,+\infty]$ считаются фиксированными.

Общая постановка задачи следующая: при заданном числе ступеней (импульсов) допустимого управления найти точки переключения и высоты ступеней этого управления (включая и точку T), удовлетворяющие необходимым условиям экстремума функционала расхода (4).

1. Уравнения движения спутника около центра масс. Введем в рассмотрение системы координат, связанные с движением спутника относительно своего центра масс O, в гравитационном поле Земли (рис. 1).

Начало орбитальной системы координат $O\xi\eta\zeta$ совпадает с центром масс O спутника. Ось $O\zeta$ направлена вдоль радиус-вектора, соединяющего центры масс Земли и спутника. Ось $O\xi$ направлена по трансверсали орбиты, а ось $O\eta$ – по нормали к плоскости орбиты. Система $Oxyz$ жестко скреплена со спутником и ее оси x, y, z направлены по его главным центральным осям инерции, соответствующим моментам инерции I_1, I_2, I_3, причем $I_2 > I_1 > I_3$. На рис. 1 отмечены углы: ψ – угол между проекцией оси Oz на плоскость $O\eta\zeta$ и осью $O\zeta$; θ – угол между осью Oz и плоскостью $O\eta\zeta$; φ – угол, образуемый осью Ox с плоскостью $Oz\xi$. Системы $Oxyz$

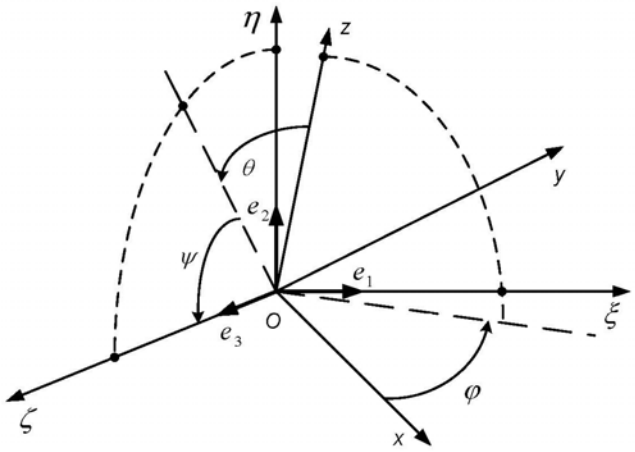


Рис. 1.

и $O\xi\eta\zeta$ совмещаются тремя последовательными поворотами на углы ψ, θ, φ вокруг осей $O\xi, O\eta, Oz$. Взаимное положение этих систем определяют таблицей направляющих косинусов $\alpha, \ldots, \gamma''$. Направляющие косинусы обычно задают как функции углов ψ, θ, φ при $\varphi, \psi \in [0, 2\pi), \theta \in (\pi/2, 3\pi/2)$ [4], или $\theta \in [0, 2\pi]$ [7], или в каких-то других областях изменения этих углов. Мы используем вариант представленный в табл. 1.

Таблица 1. Таблица направляющих косинусов

$$\varphi,\psi \in [0,2\pi),\ \theta \in (0,\pi),$$

$$\alpha = \sin\theta\cos\varphi;\quad \alpha' = -\sin\theta\sin\varphi;\quad \alpha'' = -\cos\theta;$$

$$\beta = \cos\psi\sin\varphi - \sin\psi\cos\theta\cos\varphi;\quad \beta' = \cos\psi\cos\varphi + \sin\psi\cos\theta\sin\varphi;\quad \beta'' = -\sin\psi\sin\theta;$$

$$\gamma = \sin\psi\sin\varphi + \cos\psi\cos\theta\cos\varphi;\quad \gamma' = \sin\psi\cos\varphi - \cos\psi\cos\theta\sin\varphi;\quad \gamma'' = \cos\psi\sin\theta.$$

Если размеры спутника малы по сравнению с расстоянием R между центрами масс Земли и спутника, то, используя соответствующие аппроксимации для силовой функции и моментов, определяющих действие на спутник гравитационного поля Земли, можно получить следующие уравнения управляемого движения ИСЗ относительно центра масс, которые состоят из динамических уравнений Эйлера и кинематических соотношений Пуассона [4]:

$$\begin{gathered}
I_1\,dp/dt + (I_3 - I_2)qr = 3\mu R^{-3}(I_3 - I_2)\gamma'\gamma'' + u_p,\\
I_2\,dq/dt + (I_1 - I_3)pr = 3\mu R^{-3}(I_1 - I_3)\gamma\gamma'' + u_q,\\
I_3\,dr/dt + (I_2 - I_1)pq = 3\mu R^{-3}(I_2 - I_1)\gamma\gamma' + u_r,\\
d\gamma/dt = \gamma' r - \gamma'' q + \omega\alpha,\quad d\gamma'/dt = \gamma'' p - \gamma r + \omega\alpha',\quad d\gamma''/dt = \gamma q - \gamma' p + \omega\alpha'',\\
d\alpha/dt = \alpha' r - \alpha'' q - \omega\gamma,\quad d\alpha'/dt = \alpha'' p - \alpha r - \omega\gamma',\quad d\alpha''/dt = \alpha q - \alpha' p - \omega\gamma'',\\
d\beta/dt = \beta' r - \beta'' q,\quad d\beta'/dt = \beta'' p - \beta r,\quad d\beta''/dt = \beta q - \beta' p,
\end{gathered} \tag{9}$$

где p,q,r – проекции абсолютной угловой скорости вращения спутника соответственно на оси Ox,Oy,Oz; ω – величина угловой скорости движения центра масс по орбите. Далее будем считать, что орбита спутника круговая.

2. Положения относительного равновесия спутника на круговой орбите. В случае круговой орбиты ($\mu/R^3 = \omega^2 = const$) существует относительное равновесие, при котором спутник все время одной стороной "смотрит" на Землю (например, случай Луны, которая приблизительно всегда обращена одной стороной к Земле). Всего существует 24 геометрически различных положений относительного равновесия спутника на орбите [4, 7], четыре из которых являются устойчивыми. При рассмотрении движения спутника в окрестности устойчивого положения относительного равновесия удобно ввести относительные угловые скорости:

$$\overline{p} = p - \omega\beta,\quad \overline{q} = q - \omega\beta',\quad \overline{r} = r - \omega\beta''. \tag{10}$$

Четыре устойчивых положения относительного равновесия спутника на круговой орбите представлены здесь при помощи табл. 2, 3 и рис. 2.

Таблица 2. Положения равновесия спутника на круговой орбите

n	$\overline{p},\overline{q},\overline{r}$	$\alpha=\delta_7$	$\alpha'=\delta_8$	$\alpha''=\delta_9$	$\beta=\delta_{10}$	$\beta'=\delta_{11}$	$\beta''=\delta_{12}$	$\gamma=\delta_4$	$\gamma'=\delta_5$	$\gamma''=\delta_6$
2	0, 0, 0	1	0	0	0	1	0	0	0	1
3	0, 0, 0	1	0	0	0	−1	0	0	0	−1
4	0, 0, 0	−1	0	0	0	1	0	0	0	−1
1	0, 0, 0	−1	0	0	0	−1	0	0	0	1

Пояснение: положению равновесия типа «Луна-Земля» соответствует первая строка этой таблицы ($n=2$), в которой $\alpha=\delta_7=1$, $\alpha'=\delta_8=0$, $\alpha''=\delta_9=0$, $\beta=\delta_{10}=0$, $\beta'=\delta_{11}=1$, $\beta''=\delta_{12}=0$, $\gamma=\delta_4=0$, $\gamma'=\delta_5=0$, $\gamma''=\delta_6=1$, остальные строки расшифровываются аналогично.

3. Линеаризованные уравнения вращательного движения на круговой орбите. Введем новые переменные [3, 5]:

$$x_1=\bar{p},\quad x_2=\bar{q},\quad x_3=\bar{r},\quad x_4=\gamma,\quad x_5=\gamma',\quad x_6=\gamma''-\delta_6,$$
$$x_7=\alpha-\delta_7,\ x_8=\alpha',\ x_9=\alpha'',\ x_{10}=\beta,\ x_{11}=\beta'-\delta_{11},\ x_{12}=\beta'', \tag{11}$$

где величины $\delta_6,\delta_7,\delta_{11}$ обозначают значение ненулевых направляющих косинусов γ'',α,β' в положении соответствующего равновесия (табл. 2).

Таблица 3. Положения равновесия спутника на круговой орбите

n	ψ	φ	θ	промежуток изменения θ
2	0	0	$\pi/2$	$(0,\pi)$
3	π	0	$\pi/2$	$(0,\pi)$
4	π	π	$\pi/2$	$(0,\pi)$
1	0	π	$\pi/2$	$(0,\pi)$

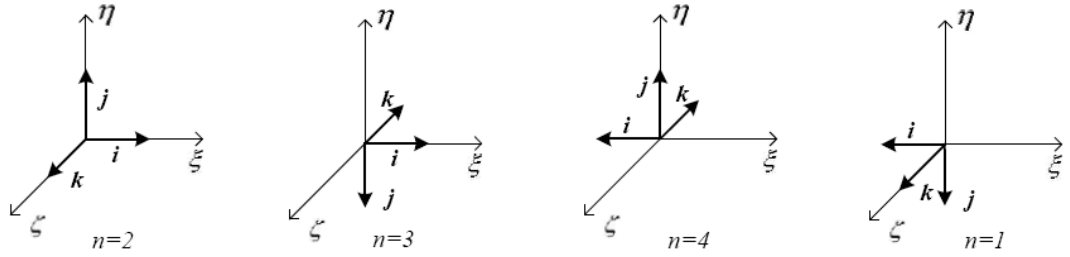


Рис. 2. Четыре устойчивых положения относительного равновесия спутника на круговой орбите

Используя (10), можно записать уравнения (9) для введенных переменных (11) и отбросив нелинейные слагаемые, получаем линейную систему (при $\dot{\omega}=0$), причем решение $x_i=0\ \left(i=\overline{1,12}\right)$ будет ее положением равновесия.

С учетом свойств матрицы направляющих косинусов

$$\alpha^2+\alpha'^2+\alpha''^2=1,\quad \beta^2+\beta'^2+\beta''^2=1,\quad \gamma^2+\gamma'^2+\gamma''^2=1, \tag{12}$$
$$\alpha\beta+\alpha'\beta'+\alpha''\beta''=0,\quad \alpha\gamma+\alpha'\gamma'+\alpha''\gamma''=0,\quad \beta\gamma+\beta'\gamma'+\beta''\gamma''=0.$$

перейдем к автономной системе линеаризованных уравнений шестого порядка относительно величин x_1,x_2,x_3 и трех из девяти направляющих косинусов в зависимости от $\delta_6,\delta_7,\delta_{11}$ (табл. 2). Получим для $y=\left(x_1,x_2,x_3,x_4,x_5,x_8\right)$ систему вида

$$\dot{y}=Ay+V \tag{13}$$

с постоянной матрицей A, имеющей структуру:

$$A=\begin{pmatrix} 0 & 0 & (-1-N_1)\delta_{11} & 0 & 4\omega N_1\delta_6 & 0 \\ 0 & 0 & 0 & 3\omega N_2\delta_6 & 0 & 0 \\ (1-N_3)\delta_{11} & 0 & 0 & 0 & 0 & \omega N_3\delta_7 \\ 0 & -\delta_6\omega^{-1} & 0 & 0 & 0 & 0 \\ \delta_6\omega^{-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\delta_7\omega^{-1} & 0 & 0 & 0 \end{pmatrix} \tag{14}$$

при $N_1=\dfrac{I_3-I_2}{I_1}, N_2=\dfrac{I_1-I_3}{I_2}, N_3=\dfrac{I_2-I_1}{I_3}, \tilde{u}_p=\dfrac{u_p}{I_1\omega}, \tilde{u}_q=\dfrac{u_q}{I_2\omega}, \tilde{u}_r=\dfrac{u_r}{I_3\omega}$. Здесь вектор управления и начальные условия имеют вид:

$$V=\left(\tilde{u}_p,\tilde{u}_q,\tilde{u}_r,0,0,0\right), \tag{15}$$
$$y_0=\left(x_1^0,x_2^0,x_3^0,x_4^0,x_5^0,x_8^0\right)=y(0),\quad \dot{y}(0)=\dot{y}_0.$$

Заметим, что y и, аналогично, y_0 содержат шесть компонент $x_1, x_2, x_3, x_4, x_5, x_8$, а остальные выражаются через них следующим образом:

$$x_9 = -x_4\delta_6\delta_7, \quad x_{12} = -x_5\delta_6\delta_{11}, \quad x_{10} = -x_8\delta_7\delta_{11}, \quad x_6 = x_7 = x_{11} = 0.$$

Величины $\delta_6, \delta_7, \delta_{11}$ берутся из строки табл. 2 соответствующей рассматриваемому положению равновесия (т.е. соответствующему *n*).

В докладе (который мы не можем здесь привести полностью из-за недостатка места) будет представлен алгоритм, который позволяет находить оптимальное по расходу топлива управление для случая, когда матрица A имеет несколько различных пар чисто мнимых собственных значений. Соответствующие этим парам частотные составляющие гасятся последовательно.

Задачу минимизации функционала (4) при допустимых управлениях мы будем рассматривать как две последовательно решаемые задачи: вначале мы найдем точки переключения как функции высот ступеней управлений, а затем, подставив эти функции в функционал (4) найдем оптимальные значения высот ступеней.

Литература

1. Бабаджанянц Л.К., Голубева Н.И., Новоселов В.С. Оптимальное демпфирование быстрых линейных колебаний стационарного ИСЗ с маховиком, “Проблемы механики управляемого движения”, Вып. 3, Пермь, 1973. С. 18–25.
2. Бабаджанянц Л.К., Потоцкая И.Ю. Управление по критерию расхода в механических системах, 2003. СПбГУ, СПб.
3. Бабаджанянц Л.К., Потоцкая И.Ю., Пупышева Ю.Ю. «Управление вращением спутника по критерию расхода». Сборник трудов международной конференции «Устойчивость и процессы управления», СПб, июнь 2005. С. 1052−1059
4. Белецкий В.В. Движение спутника относительно центра масс, 1965, Наука, М.
5. Мгоян П.Б. Оценки в теории возмущенного движения, дисс. к.ф.-м н., 1987, ЛГУ, Л.
6. Пупышева Ю.Ю. Алгоритмы нахождения точек переключения кусочно-полиномиального управления в линейных механических системах, дисс. к.ф.-м н., 2003, СПбГУ, СПб.
7. Сарычев В.А., Сазонов В.В. Оценка границы колебательных движений спутника, “Препринты Ин–та прикл. мат. АН СССР”, 1974, №130.

Программный комплекс расчета параметров противометеорной защиты

А.В. Бабенко, Н.Т. Каргин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

На сегодняшний день одной из важных областей исследований является проблема безопасности космических полетов при воздействии факторов внешней среды. К одному из этих факторов можно отнести воздействие механических частиц естественного и искусственного происхождения на обшивку космического аппарата. Для расчета и выбора, необходимых для безопасности полета, параметров защитных экранов при попадании в них метеорных частиц был написан специальный программный комплекс, включающий в себя графический модуль, наглядно демонстрирующий процесс и результаты соударения.

Исходными данными для проведения расчета являются параметры ударника, такие как масса, плотность материала, скорость соударения, угол соударения и данные преграды – толщина, плотность материала, временное сопротивление разрыву материала, признак пористости и расстояние до следующей преграды. Количество преград не ограничено, для каждой из них отводится специальная вкладка, в которой можно задать все необходимые параметры.

После ввода исходных данных, появляется окно, содержащее графическую интерпретацию процесса соударения частицы с преградой при введенных параметрах. В этом окне последовательно представлены преграды, с соответствующей толщиной и расстоянием между ними, воздействие частицы на них, а после соударения и лидирующий элемент (наибольший по массе и скорости элемент, среди осколков частиц). В качестве результата взаимодействия, выдается сообщение о пробое или не пробое преграды. В случае непробоя пишется, какая преграда оказалась не пробитой, а в случае пробоя появляется сообщение об аварийной ситуации – означающее, что необходимо изменить параметры преград для обеспечения безопасности космического аппарата, например, увеличить их толщину, взять другой материал или принять иные меры. После просмотра графических результатов, их можно закрыть и проанализировать появившиеся результаты расчета, в качестве которых программный комплекс выводит данные о диаметре частицы, предельно-пробиваемой толщине, скорости дробления, диаметре пробоя на свету, массе лидирующего элемента, его диаметре, скорости, также о диаметре зоны повреждения преграды и угле разброса осколков и обломков.

На основании работы данного программного комплекса, позволяющего моделировать процесс соударения метеорных частиц с обшивкой космического аппарата, можно сделать вывод о целесообразности использования конкретного типа защиты, его усовершенствовании или установке новой защитной конструкции.

Обработка измерений спутниковых навигационных систем

С.Ю. Гиндин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Спутниковая радионавигация – это, несомненно, самый распространенный метод позиционирования в наше время. Бытовые навигационные приемники легкодоступны и хорошо работают в городских условиях. Но, когда речь идет об объектах с высокими динамическими характеристиками (самолет, космический аппарат), то даже специализированные навигационные приемники не дают пригодного для исследований результата. Также при запуске КА с навигационной аппаратурой на борту, существует ограничение в выборе производителя приемников (производитель должен иметь разрешение на использование выпускаемого оборудования в космосе). Выход из этой ситуации – самостоятельное решение навигационной задачи с применением собственного алгоритма. Для осуществления этого используем информацию, получаемую приемником от навигационных спутников.

В докладе обсуждаются возможные методы решения указанной выше проблемы. Предлагается не ограничиваться решением частной задачи, привязанной к конкретному приемнику и конкретному алгоритму, а обобщить её для различных приемников и различных алгоритмов. Этапы реализации проекта:

1. создание средства разработки для решения навигационной задачи, распространяемой в виде библиотеки динамической компоновки (DLL - Dynamic-link library) содержащей функции, необходимые для реализации всевозможных алгоритмов решения навигационных задач
2. реализация с помощью библиотеки стандартного алгоритма получения позиционного и скоростного решения в соответствии с ИКД ГЛОНАСС-2007 и IS-GPS-200
3. отработка алгоритмов и полученных результатов на блоке имитации СН – 3803М

Техническое задание на разработку библиотеки:

1. Поддержка входных данных в формате RINEX (The Receiver Independent Exchange Format)
2. Поддержка выходной информации по протоколу обмена NMEA 0183 (National Marine Electronics Association)
3. Работа с навигационными системами GPS, ГЛОНАСС
4. Работа в различных системах координат
5. Наличие руководства разработчика с примерами
6. Совместимость с различными языками программирования

Реализация данного проекта производится на платформе MS Framework .NET на языке программирования C#.

Выбор этой платформы основывается на следующих преимуществах:

\+ Совместимость с приложениями, написанными на различных языках программирования.

\+ Использование управляемого кода (managed code), что предотвращает трудно отлавливаемые утечки динамической памяти, и тем самым существенно сокращает время на отладку программ.

Стандартный алгоритм обработки спутниковых измерений состоит из следующих этапов:

✓ Определение гринвичских координат и относительной скорости навигационных спутников для заданного момента времени

✓ Определение погрешности часов навигационных спутников

✓ Нахождение поправок координат и относительной скорости навигационного спутника, вызванные вращением Земли

✓ Нахождение углов возвышения и азимута навигационного спутника

- Модель ионосферной задержки для корректировки измерений
- Модель тропосферной задержки для корректировки измерений
- Алгоритм определения координат приемника при помощи кодовых псевдодальностей
- Алгоритм определения скорости при помощи доплеровских измерений

Структурная схема данного алгоритма приведена на рисунке 1.



Рис.1. Структурная схема алгоритма определения координат объекта

Для реализации библиотеки DLL используются следующие методы: теория матриц, численные методы решения дифференциальных уравнений, методы объектно-ориентированного программирования, методы математического моделирования, методы исследования динамических систем.

С помощью разработанной библиотеки был реализован стандартный алгоритм определения местоположения объекта. Для проверки результатов использовались данные эксперимента, проводимого на базе лаборатории управления и навигации механико-математического факультета МГУ им. М. В. Ломоносова. Результаты решений совпали.

Разработка данной библиотеки поможет быстро и автоматизировано работать с эфемеридной информацией и переводить формат RINEX в удобную для дальнейшего использования форму. С ее использованием облегчается процесс расчета координат объекта.

Последовательно – оптимальные программы навигационных определений КАДЗЗ

А.П. Долгинцев, И.В. Елисеев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Одним из путей совершенствования систем обработки и управления космическим аппаратом дистанционного зондирования (КАДЗЗ) является совершенствование алгоритмов управления проведением сеансов навигационных определений (СНО) по спутниковой радионавигационной системе ГЛОНАСС. Выбор моментов определений в процессе полета производится путем использования результатов решения задачи программного управления, эквивалентной исходной задаче оптимального планирования [1]. Задача оптимального размещения сеансов навигационных определений (СНО) рассматривается на интервале оценивания параметров движения КАДЗЗ (t_1, t').

Данная задача имеет следующие специфические черты:

Специфика решаемой задачи заключается в вспомогательном характере задачи навигационных определений. Для ее решения отводится время, свободное от решения целевой задачи КАДЗЗ.

Структура сеанса навигационных определений не исследуется: сеанс представляется точками на интервале времени выполнения полетного задания $t_i, i = 1,2,\ldots$. СНО может рассматриваться как мгновенное событие, включающее в себя проведение минимально необходимой для оценивания параметров движения выборки измерений разностей псевдодальностей и разностей радиальных псевдоскоростей $\overline{w}(t_i) = (\Delta D_j(t_i), \Delta \dot{D}_j(t_i), j = \overline{1,3})$. Данная задача успешно решается при помощи анализа полей навигационной точности [1].

Минимальная выборка измерений обрабатывается при помощи конечного метода решения навигационной задачи, формирующего оценку вектора состояния КАДЗЗ $\overline{q}(t_i) = (q_j(t_i), j = \overline{1,6})$ с погрешностями, оцениваемыми при помощи матрицы вторых моментов $K_{\Delta\overline{q}}(t_i) = (F^T(t_i) W^{-1}(t_i) F(t_i) + K^{-1}_{\Delta q0}(t_i))^{-1}$, где $F(t_i) = \left[\frac{\partial \overline{w}(t_i)}{\partial \overline{q}(t_i)}\right]$ – матрица коэффициентов чувствительности разностных измерений $\overline{w(t_i)}$, $W(t_i) = cov(\overline{w}^T(t_i)\overline{w}(t_i))$ – ковариационная матрица случайной составляющей погрешностей навигационных измерений $\overline{w}(t_i)$, $K_{\Delta\overline{q}0}(t_i)$ – ковариационная матрица априорной оценки вектора состояния КАДЗЗ, определяемая показаниями сопряженной инерциальной системы навигации (ИНС).

Число измерений на протяжении времени полета КАДЗЗ не является фиксированным, оно имеет тенденцию к увеличению (после обработки 10 – 20 СНО полет КАДЗЗ продолжается, а, следовательно, продолжаются и сеансы определений – до 100 и более). Навигационные определения, проводимые КАДЗЗ, проводятся многократно на технологической части витка и могут при появлении окна в программе полета возобновляться. Моменты навигационных определений заранее неизвестны и определяются оперативно по мере выполнения полетного задания. Момент времени t' процесса окончания решения навигационной задачи КАДЗЗ является дополнительным

резервом оптимизации, выбираемым с учетом ограничения $t_i < t' < t_{max}$, где t_{max} удовлетворяет целевому назначению КАДЗЗ.

Исходными данными при планировании следующего навигационного определения является текущая на данный момент времени картина поля навигационных определений и показания ИНС. Оба данных фактора обычно приводятся в виде сферической погрешности навигационных определений (оценка среднеквадратических отклонений по положению и скорости σ^n и σ^c). При рассмотрении проекций погрешностей навигационных определений на оси орбитальной системы координат $OXYZ$ исследуются: X – радиальное, Y – трансверсальное (продольное), Z – бинормальное отклонения.

Результаты планирования оцениваются при помощи векторного интервального критерия оптимальности $\bar{I}(t_i) = \{I_k(t_i), k = \overline{1,5}\}$:

- первый критерий $I_1(t_i) = \sigma_X^2(t_i)$ связан с погрешностью знания координат КАДЗЗ по высоте полета;
- второй критерий $I_2(t_i) = K_1^2\left(\sigma_{VX}^2(t_i) + K_2^2\sigma_Y^2(t_i)\right)$ связан со смазом изображения (функция от продольной составляющей скорости и радиальной дисперсии), K_1, K_2 – эмпирически выбираемые нормирующие коэффициенты;
- третий критерий $I_3(t_i) = \sigma_Y^2(t_i)$ связан с информационной производительностью (интервалом между СНО, и в свою очередь, является функцией временного интервала $(t_{i+1} - t_i)$;
- четвертый критерий $I_4(t_i) = \sigma_X^2(t_i) + \sigma_Y^2(t_i) + \sigma_Z^2(t_i)$ является сферической погрешностью навигационных определений по положению;
- пятый критерий $I_5(t_i) = \sigma_{VX}^2(t_i) + \sigma_{VY}^2(t_i) + \sigma_{VZ}^2(t_i)$ является сферической погрешностью навигационных определений по скорости.

Таким образом, критериальный базис $\bar{I}(t_i)$ формируется с использованием следующих соображений.

Критерии $I_4(t_i)$ и $I_5(t_i)$, рассматриваемые как интервальный критерий A – оптимальности формируют показатели $I_4^{mid}(t_i)$, $I_5^{mid}(t_i)$, минимизирующие среднюю дисперсию оцениваемых параметров положения и скорости, а при рассмотрении минимаксной задачи формируют показатели $I_4^{max}(t_i)$, $I_5^{max}(t_i)$, оценивающие параметры положения и скорости в минимаксной постановке.

Критерий смаза изображения $I_2(t_i)$ (интервальный $I_2^{mid}(t_i)$ или минимаксный $I_2^{\max}(t_i)$) является критерием F – оптимальности, минимизирующим дисперсию нескольких оцениваемых параметров, с нормирующей матрицей вида:

$$C = \begin{bmatrix} 0 & K_1^2K_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & K_1^2 & 0 & 0 \end{bmatrix},$$

определяющей Ψ – преобразование матрицы вторых моментов $K_{\Delta\bar{q}}$.

Планирование результатов эксперимента сводится к формированию структуры $\varepsilon(t_1, t_N)$ вида [2]:

$$\varepsilon(t_1, t_N) = \begin{Bmatrix} t_1 & t_2 & \dots & t_N \\ \bar{I}(t_1) & \bar{I}(t_2) & \dots & \bar{I}(t_n) \end{Bmatrix},$$

где $\left(t_i,\ i=\overline{1,N}\right)$ – моменты времени проведения СНО, $\bar{I}(t_i)=\left\{I_k(t_i),k=\overline{1,5}\right\}$ – показатели эффективности, образующие критериальный базис для решения навигационной задачи КАДЗЗ. При этом конечная точка с номером сеанса N не фиксирована. Сеансы могут в любой момент прекратиться или возобновиться.

Методология линейного последовательного оптимального планирования сеансов определений [2] базируется на последовательном выборе на каждом шаге плана $\varepsilon(t_i)$:

$$\varepsilon(t_i)=\begin{Bmatrix} t_i \\ \bar{I}(t_i) \end{Bmatrix}, \varepsilon=\bigcup_{t_i\in(t_i,t')}\varepsilon(t_i),$$

где ε – искомый план навигационных определений на интервале времени $[t_i,t']$, что сводится к выбору квазиоптимального значения свертки дисперсии поверхности отклика $\varphi(t_i,\varepsilon(t_i))$ на каждом шаге t_i:

$$\max_{t_i}\varphi(t_i,\varepsilon(t_i))=\max_{t_i}\lambda(t_i)\ \bar{I}\left[K_{\overline{\Delta q}}(\varepsilon(t_i))\quad F(t_i)\quad \lambda(t_i)\quad F^T(t_i)\quad K_{\overline{\Delta q}}(\varepsilon(t_i))\right],$$

где $\lambda(t_i)=K^{-1}_{\overline{\Delta q}}(t_i)$ – функция эффективности проводимых навигационных определений, связанная с расчетом и анализм полей навигационной точности [3]. Поле точности зависит от времени $\left(t_i,\ i=\overline{1,N}\right)$, погрешностей навигационных измерений $\bar{I}(t_i)$ и геометрии наблюдаемой конфигурации НИСЗ $\Xi=\left\{S_1^{N_1}\quad S_2^{N_2}\quad ...\quad S_n^{N_n}\right\}=\bigcup_{t_i\in[t_i,t']}S_{t_i}^{N_i}$ [3].

Целью исследований является анализ возможности использования методологии линейного последовательного оптимального планирования для построения оптимальных планов измерений в дискретные моменты времени $\left(t_i,i=\overline{1,N}\right)$, а также определение границ применимости данной методологии, в том числе границ линеаризации зависимости свертки дисперсии поверхности отклика:

$$\varphi(t_i,\varepsilon(t_i))=\Psi(t_i),$$

где Ψ – оператор невырожденного линейного преобразования, связанного с видом оптимизируемого показателя эффективности [2]: например, $I_3^{max}(t_i)=\Psi(t_i,\varepsilon(t_i))=\min M\left[\Delta Y\Delta Y^T\right]=\min\sigma^2_{\Delta Y}$. Если используются минимаксные планы измерений с дисперсией, то получим

$$I_3(t_i,\varepsilon(t_i))=\min_{t_i}\max\ \Psi(t_i,\varepsilon(t_{i-1}))=\min_{t_i}\max\ \sigma^2_{\Delta Y}.$$

Все виды рассматриваемых критериев A-,LG, LM,-,F – оптимальности являются частными случаями критерия F – оптимальности. Главная особенность процесса последовательной оптимизации заключена в том, оптимизируются виды сверток, вид которых определяется видом дисперсии поверхности отклика. Поверхностью отклика определена векторным пространством, образованным векторами критериального базиса $\bar{I}\left(t_i,\varepsilon(t_i)\right)$ в последовательные моменты времени t_i, принадлежащие интервалу (t_1,t'). Например, если используются оценка математического ожидания отклонений продольной составляющей на отрезке (t_i,t'), критерий оптимальности примет вид

$$\bar{I}(t_i,\varepsilon(t_i)) = min \int_{(t_i,t')} I_3(t_i,\varepsilon(t_i))dt = min\ \sigma^2_{\Delta Y}.$$

$$(t_1,t')$$

Тогда для оценивания отклонения дисперсии следует знать среднеквадратичное отклонение продольной составляющей на интервале (t_i,t').

В табл. 1 представлены виды интервальных критериев оптимальности и соответствующие им дисперсии поверхностей отклика.

Таблица 1. Виды дисперсий поверхностей отклика

Критерий	Вид линейного преобразования $\Psi\left[K_{\Delta\bar{q}}\right]$	Пошагово оптимизируемое значение $\varphi(t_i,\varepsilon(t_i))$
Интервальный $I_1^{LG}(t_1,t')$	$I_1^{LG}(t_1,t') = \min \int_{t_1}^{t'} \sigma^2_{Xi}(t)dt$ $t_i, i=\overline{1,N}$	$\min \int_{t\in(t_i,t')} \left\{\sum_{\beta=1}^{6}\left[\frac{\partial\overline{w(t)}}{\partial\overline{q(t)}}\right]_\beta K_{\Delta\bar{q}}(\beta,X(t))\right\}^2 dt$
$I_3^{LG}(t_1,t')$	$I_3^{LG}(t_1,t') = \min \int_{t_1}^{t'} \sigma^2_{Yi}(t)dt$ $t_i, i=\overline{1,N}$	$\min \int_{t\in(t_i,t')} \left\{\sum_{\beta=1}^{6}\left[\frac{\partial\overline{w(t)}}{\partial\overline{q(t)}}\right]_\beta K_{\Delta\bar{q}}(\beta,Y(t))\right\}^2 dt$
Интервальный минимаксный $I_1^{LM}(t_1,t')$	$I_1^{LM}(t_1,t') = \min \max\sigma^2_X(t_i)$ $t_i\in(t_i,t')$	$\min \left\{\sum_{\beta=1}^{6}\left[\frac{\partial\overline{w(t_i)}}{\partial\overline{q(t_i)}}\right]_\beta K_{\Delta\bar{q}}(\beta,X(t_i))\right\}$ $t_i\in(t_{i-1},t)$
$I_3^{LM}(t_1,t')$	$I_3^{LM}(t_1,t') = \min \max\sigma^2_Y(t_i)$ $t_i\in(t_i,t')$	$\min \left\{\sum_{\beta=1}^{6}\left[\frac{\partial\overline{w(t_i)}}{\partial\overline{q(t_i)}}\right]_\beta K_{\Delta\bar{q}}(\beta,Y(t_i))\right\}^2$ $t_i\in(t_{i-1},t)$
Интервальный по положению $I_4^A(t_1,t')$ и скорости $I_5^A(t_1,t')$	$I_4^A(t_1,t') = \min \int_{t_1}^{t'} tr\ K^{\Pi}_{\Delta\bar{q}}(t)dt$ (t_1,t') $I_4^A(t_1,t') = \min \int_{t_1}^{t'} tr\ K^{C}_{\Delta\bar{q}}(t)dt$ (t_1,t')	$tr\ \lambda(t_i)F^T(t_i)K^{\Pi}_{\Delta\bar{q}}(t_i)\overline{M}K^{\Pi}_{\Delta\bar{u}}(t_i)F(t_i),$ *где* $\overline{M} = \int_{t\in(t_i,t')} F(t)F^T(t)dt$ $tr\ \lambda(t_i)F^T(t_i)K^{C}_{\Delta\bar{q}}(t_i)\overline{M}K^{C}_{\Delta\bar{q}}(t_i)F(t_i),$ *где* $\overline{M} = \int_{t\in(t_i,t')} F(t)F^T(t)dt$

Было проведено моделирование работы алгоритма последовательного планирования для КАДЗЗ с наклонением 65°, высотой круговой орбиты 260 км при сферической погрешности сопряженной ИНС 350 м по положению и 0,26 м/с по скорости. Были построены статистические оценки навигационных определений для критерия трансверсальной составляющей на 16 витке.

Выводы

Анализ результатов моделирования позволяет сделать следующие выводы.

1. Модель дисперсии поверхности отклика в направлении продольной составляющей отклонений позволяет провести линеаризацию модели математического ожидания и разброса на интервалах витка и даже на протяжении суток. Для радиальной и нормальной составляющих отклонений по положению и скорости выводы о линеаризации недопустимы.

2. Анализ результатов показывает, что алгоритм последовательного оптимального планирования стремится парировать падение точности при прогнозировании движения КАДЗЗ за счет повышения интенсивности навигационных определений.

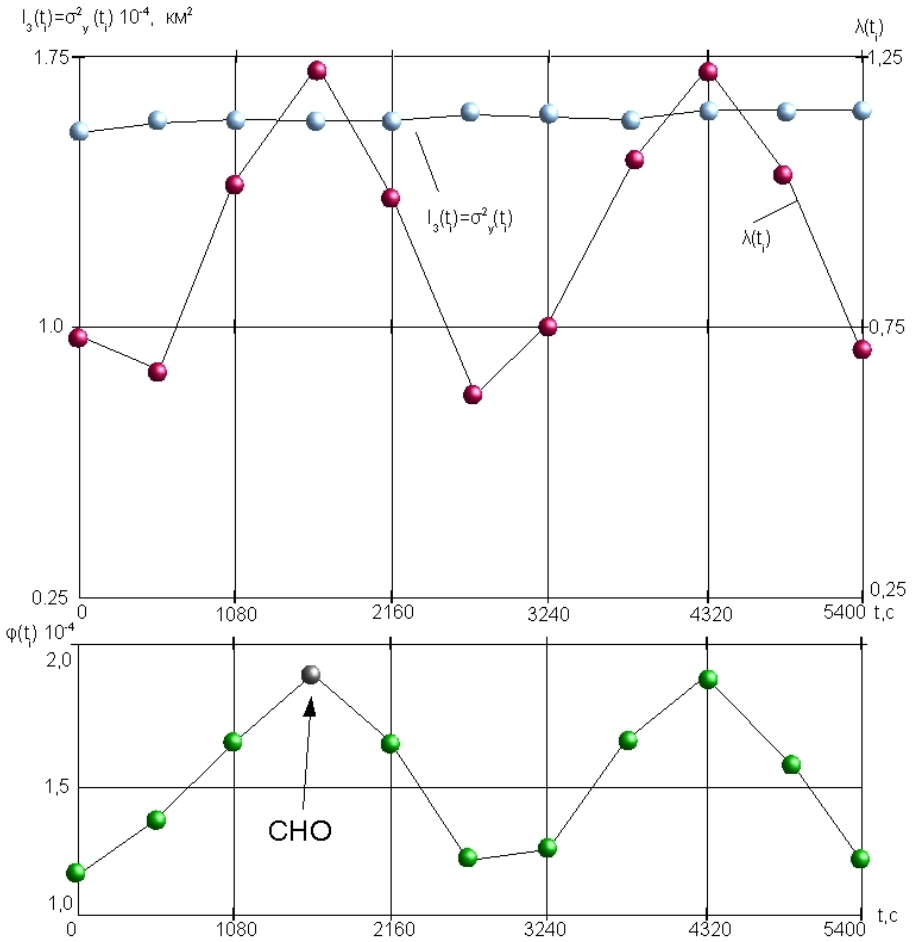


Рис. 1. Результаты последовательного оптимального планирования на 16 витке

Литература

1. Оптимизация процесса наблюдения и управления летательных аппаратов / В.В. Малышев, М.Н. Красильщиков, В.И. Карлов – М.: Машиностроение, 1989. – 312 с.

2. Федоров В.В. Теория оптимального эксперимента (планирование регрессионных экспериментов). М.: Наука, 1971. – 312 с.

Разработка программного комплекса для моделирования движения многофункционального малого космического аппарата

А.А. Масловская, И.С. Ткаченко, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Для решения новых фундаментальных и прикладных задач в космосе в настоящее время широко используются малые космические аппараты (МКА). Процесс проектирования космического аппарата, как сложной технической системы, представляет собой многоуровневый итерационный и оптимизационный процесс, в течение которого рассчитываются характеристики аппарата и его массовая сводка [1]. На начальном этапе формирования облика МКА, как правило, имеется ограниченный набор исходных данных. Однако, к настоящему времени накоплен большой теоретический, практический и статистический материал, позволяющий построить математические модели, описывающие функционирование МКА на орбите при решении различных задач. Для того чтобы процесс проектирования был эффективным, необходимо разработать программный инструментарий, позволяющий проводить вычислительные эксперименты, накапливать и обрабатывать большие массивы данных. Разработке такого инструментального средства посвящена данная работа.

Моделирование задач, решаемых с помощью МКА

Многофункциональный космический аппарат способен решать широкий диапазон задач. *Первая задача* – дистанционное зондирование Земли. Для выполнения данной задачи необходимо обеспечить стабильное движение МКА на орбите. Под действием внешних возмущающих воздействий параметры орбиты МКА значительно изменяются [2]. Включение электрореактивной двигательной установки (ЭРДУ) и коррекция орбиты осуществляются, если величина падения аппарата в перигее превышает допустимое значение. Таким образом, движение МКА представляет собой чередование пассивного и активного режимов. Для моделирования движения МКА на орбите необходимы следующие исходные данные:

- параметры рабочей орбиты (высота перигея и апогея, наклонение орбиты, аргумент перигея, долгота восходящего узла);
- параметры МКА (полная масса, баллистический коэффициент, лимит среднесуточной мощности для ЭРДУ, выбранная ЭРДУ);
- параметры атмосферы (статическая плотность);
- параметры моделирования (время моделирования, шаг интегрирования, допустимая величина падения в перигее).

Результирующими данными должны быть:

- конечные параметры орбиты;
- моторное время;
- количество включений ЭРДУ;
- расход топлива.

В общем случае имитационное моделирование возмущённого движения искусственного спутника Земли, в том числе и процесса коррекции высоты орбиты, основано на решении системы уравнений в оскулирующих элементах (1) методом численного интегрирования уравнений на ЭВМ.

$$\begin{cases} \dfrac{d\Omega}{dt} = W \cdot \dfrac{r}{\sqrt{\mu \cdot p}} \cdot \dfrac{\sin U}{\sin i} \\ \dfrac{di}{dt} = W \cdot \dfrac{r}{\sqrt{\mu \cdot p}} \cdot \cos U \\ \dfrac{dp}{dt} = \sqrt{\dfrac{p}{\mu}} \cdot r \cdot 2 \cdot T \\ \dfrac{d\omega}{dt} = \sqrt{\dfrac{p}{\mu}} \cdot \left[-S \cdot \dfrac{\cos \vartheta}{e} + \left(1 + \dfrac{r}{p}\right) \cdot T \cdot \dfrac{\sin \vartheta}{e} - \dfrac{r}{p} \cdot W \cdot ctgi \cdot \sin U \right] \\ \dfrac{de}{dt} = \sqrt{\dfrac{p}{\mu}} \cdot \left[S \cdot \sin \vartheta + \left(1 + \dfrac{r}{p}\right) \cdot T \cdot \cos \vartheta + e \cdot \dfrac{r}{p} \cdot T \right] \\ \dfrac{d\vartheta}{dt} = \sqrt{\dfrac{p}{\mu}} \cdot \left[\dfrac{\mu}{r^2} + S \cdot \dfrac{\cos \vartheta}{e} - T \cdot \left(1 + \dfrac{r}{p}\right) \cdot \dfrac{\sin \vartheta}{e} \right] \end{cases}, \quad (1)$$

где $\{\Omega, \omega, i, p, e, \vartheta\}$ - элементы орбиты, $\{S, W, T\}$ - проекции возмущающих ускорений в трансверсальном, радиальном и боковом направлениях (рис. 1) [3].

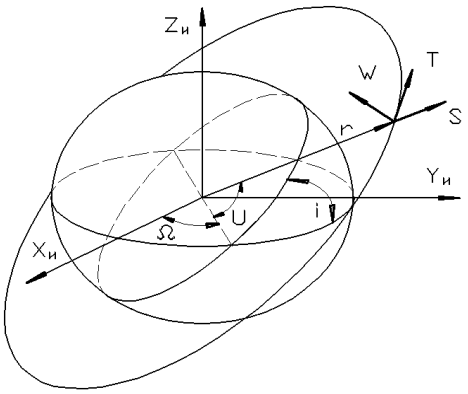


Рис. 1. Элементы орбиты и проекции возмущающих ускорений

Вторая задача – инспектирование космических объектов на орбите (орбитальное сопровождение). Для управления МКА в режиме инспектирования необходимо модели-ровать следующие этапы движения:

1. перевод в плоскость орбиты инспектируемого объекта (компенсация изменения наклонения);
2. компенсация изменения эксцентри-ситета;
3. компенсация изменения большой полуоси.

Входными данными для моделирования являются:

- полная масса МКА;
- выбранная ЭРДУ;
- параметры орбиты КА (высота перигея, приращение высоты орбиты, наклонения, эксцентриситета).

Результирующими данными должны быть:

- характеристическая скорость на каждом из этапов;
- моторное время;
- общий расход топлива.

Таким образом, можно отследить самый энергозатратный этап движения МКА при выполнении маневров.

В работе проводится анализ возможности решения МКА *третьей задачи* – работы в составе группировки малых КА и поддержание ее орбитальной структуры. Принципиальной особенностью группового полета является автономная навигация отдельного аппарата в составе группы и управление их взаимным относительным положением в группе. Это могут быть статические и меняющиеся во времени конфигурации. Возможность определения взаимного положения аппаратов и управления взаимным положением порождает возможность переконфигурирования

системы, маневрирования, сближения и стыковки спутников. Достоинствами группового полета является способность одновременного проведения наблюдений, измерений, других экспериментов в различных точках пространства при заданном и контролируемом взаимном положении спутников, на которых установлена необходимая для решения таких задач полезная нагрузка.

Описание работы системы моделирования движения МКА

В состав программного комплекса входит система моделирования движения МКА [4], реализующая математическую модель оскулирующих элементов (1), и позволяющая моделировать движение МКА в двух режимах: режиме дистанционного зондирования Земли и режиме инспектирования объектов, каждому из которых соотвествует определенная вкладка на главной экранной форме.

Моделирование режима зондирования (вкладка «Зондирование»). Перед началом моделирования необходимо ввести (рис. 2):

– исходные данные (параметры рабочей орбиты, параметры МКА, тип ЭРД, параметры атмосферы);

– параметры моделирования (время моделирования, шаг интегрирования, тип коррекции орбиты: по трансверсали или коррекция в боковом направлении).

После окончания моделирования в правой части формы отображаются рассчитанные значения: конечные параметры орбиты, количество включений ЭРДУ, моторное время, расход топлива.

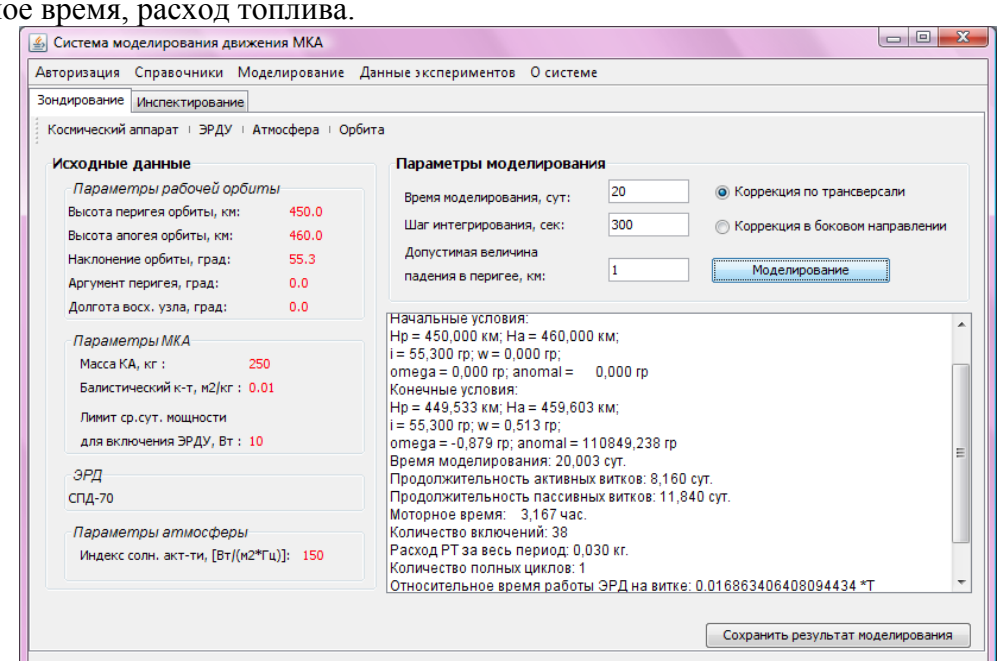


Рис. 2. Экранная форма моделирования режима зондирования

Моделирование режима инспектирования (вкладка «Инспектирование»). Перед началом моделирования необходимо ввести исходные данные для космического аппарата, орбиты и ЭРДУ, - и нажать кнопку «Расчет» (рис. 3).

Результаты моделирования выводятся в правой части формы в текстовое поле и включают рассчитанные значения характеристической скорости и времени полета на каждом этапе коррекции орбиты, ускорения и общего расхода топлива.

В любом режиме имеется возможность сохранить результаты вычилительного эксперимента в базу данных, для этого необходимо «Сохранить результат моделирования».

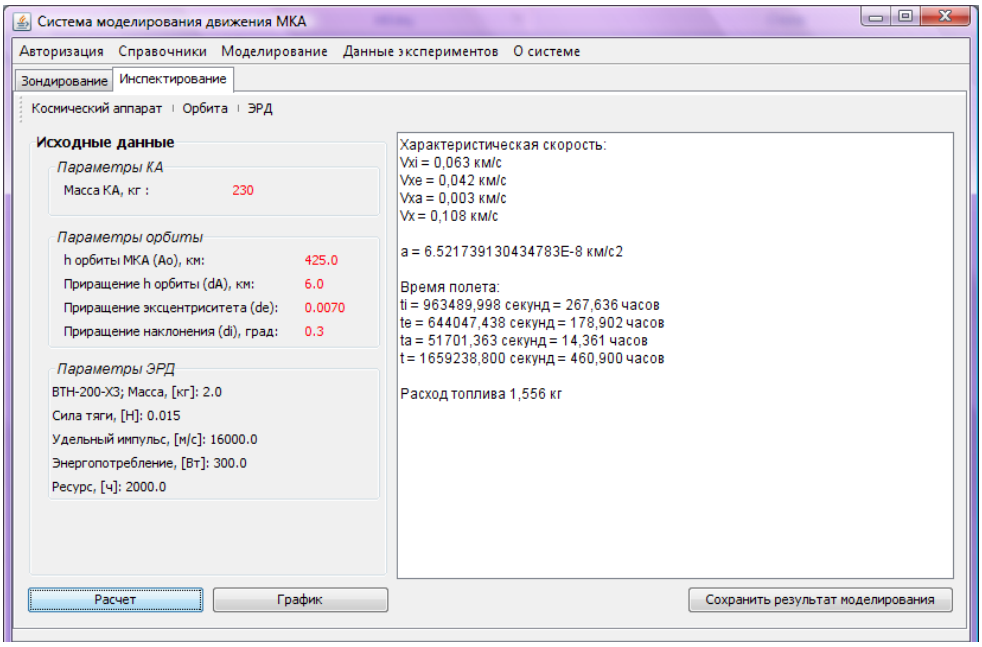


Рис. 3. Главная форма после завершения моделирования

Работа с базой данных. В системе предусмотрено ведение следующих справочников:

- пользователей;
- разрешений для ролей пользователей;
- типов ЭРДУ.

На рис. 4 изображена экранная форма для работы со справочником ЭРДУ.

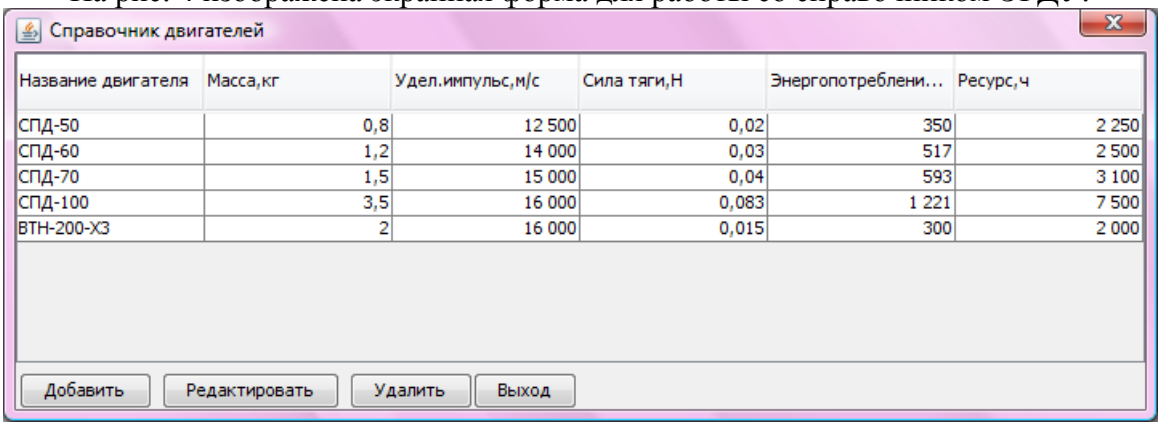


Рис. 4. Экранная форма ведения справочника ЭРДУ

Поиск и просмотр результатов вычислительных экспериментов. В системе предусмотрена возможность поиска результатов моделирования по одному из критериев и просмотра всех сохраненных результатов моделирования в наглядной форме с отображением данных об эксперименте.

Заключение

Разработка программного комплекса для моделирования движения МКА актуальна и обусловлена отсутствием в проектных организациях эффективного интерфейсного механизма, позволяющего проводить оценку параметров функционирования проектируемого космического аппарата на стадии предэскизной разработки. Программный комплекс предоставит возможность инженерам-исследователям получать данные для решения задачи формирования проектного облика МКА.

Литература

1 Куренков, В.И. Основы устройства и моделирования целевого функционирования космических аппаратов наблюдения [Текст]/ В.И. Куренков, В.В. Салмин, Б.А. Абрамов – Самара: изд-во СГАУ, 2006. – 296 с.

2 Салмин, В.В. Приближенный расчёт манёвров формирования орбит спутника Земли с двигателем малой тяги [Текст]/ В.В. Салмин, В.О. Соколов//Космические исследования. - 1989. Т27 - №1. – С. 42-53.

3 Салмин, В.В. Приближенные методы расчета оптимальных перелетов космических аппаратов с двигателями малой тяги. Часть 1./ В.В. Салмин, В.В. Васильев и др.// Вестник СГАУ – Самара, 2007. - №1 - С. 37 – 52.

4 Зеленко, Л.С. Система моделирования движения малого космического аппарата многофункционального назначения. / Зеленко, Л.С., Масловская, А.А., Ткаченко, И.С. // Математика, компьютер, образование. Тезисы. Выпуск 17 – Москва, 2010. – С. 119

Мультиагентная подсистема планирования грузопотока Международной космической станции

А.В. Иващенко[1], А.В. Леонтьев[1], А.Р. Диязитдинова[1],
П.О. Скобелев[1], М.В. Сычева[2], И.И. Хамиц[2]
[1]*Научно-производственная компания «Разумные решения», Самара*
[2]*Ракетно-космическая корпорация «Энергия», Королев*

Введение

Для поддержания функционирования Международной космической станции (МКС) необходимо решить множество взаимосвязанных задач по обеспечению жизнедеятельности экипажа, доставке лабораторного оборудования, различных материалов и инструментов. В связи с этим, актуальной является задача автоматизации интеллектуальной поддержки принятия согласованных решений по планированию доставки грузов на станцию. При этом можно выделить подзадачи интерактивного построения программы полета МКС, автоматизированного планирования грузопотока МКС и планирования ресурсов (топлива, воды, газа и других грузов).

Для решения этой задачи целесообразно использование мультиагентных технологий [1, 2], позволяющих обеспечить согласованное взаимодействие проектантов и кураторов грузов с учетом сложности и многообразия существующих связей в условиях нестабильности и высокой динамики. В данной работе описаны основные особенности мультиагентного алгоритма интерактивного планирования грузопотока МКС, разработанного и реализованного авторами в рамках общей задачи создания мультиагентной системы построения программы полета, грузопотока и расчета ресурсов МКС [3].

1 Задача планирования грузопотока МКС

Планирование грузопотока производится в несколько этапов. Сначала, в процессе стратегического планирования, выполняется построение модели грузопотока по данным о составе экспедиций. Цель разработки модели грузопотока заключается в определении годовых потребностей в грузах на планируемый период (год) работы МКС для оценки количества пусков грузовых транспортных средств, необходимых для доставки на МКС грузов для обеспечения жизнедеятельности человека, воды и топлива. По данным модели грузопотока может быть построена программа полета МКС, в которой определяются–даты стартов кораблей и которая обеспечивает необходимыми входными данными процесс планирования детального (поблочного) грузопотока.

Поблочный грузопоток содержит распределение грузов по полетам транспортных средств на период (обычно на год). Начальное построение поблочного грузопотока формируется на основе данных о фактическом нахождении грузов на станции. С учетом норм расхода и ресурса грузов можно определить дату, до которой необходимо обеспечить доставку новых грузов.

Дальнейшая работа производится на основе заявок, поступающих от кураторов грузов, в которых описывается каждый груз, и указывается полет, на котором этот груз должен быть доставлен. Также в заявке может быть указан интервал времени ожидаемой доставки груза. На основании данных заявок производится распределение грузов по полетам с учетом ограничений транспортного средства по суммарной массе сухих грузов, воды и топлива. Также важной функцией управления грузопотоком является планирование возврата и утилизации грузов.

Основаниями доставки расходуемых грузов, воды, газа, топлива на транспортном средстве являются: информация центра управления полетами (ЦУП) и кураторов о фактических запасах на борту МКС, сохранение требуемого баланса

расходуемых элементов до следующего прилета транспортного средства, обеспечение необходимого запаса расходуемого элемента на момент стыковки транспортного средства на случай нештатных ситуаций и готовность необходимого оборудования (грузов) для доставки.

При работе с грузопотоком проектанты руководствуются приоритетами грузов: сначала планируются жизненно важные грузы для обеспечения жизнедеятельности человека и функционирования станции, затем коммерческие. Кроме этого, существуют заявки, имеющий наивысший приоритет, связанные с уникальностью блока, например, если ценный блок надо срочно доставить на Землю для осуществления ремонта, у него устанавливается самый высокий приоритет.

Доставка топлива и воды на станцию производится на основе данных о расходе, для чего параллельно с планированием грузопотока рассчитывают балансы топлива и воды. Расходы топлива на станции включают расходы топлива на поддержание (коррекции) орбиты станции (компенсация потерь характеристической скорости из-за сопротивления атмосферы) и расходы топлива на управление движением вокруг центра масс станции (по крену, тангажу и рысканью), включая разгрузку гиродинов, то есть на изменение ориентации станции, которое производится, например, перед стыковкой кораблей.

Процедура расчета баланса топлива начинается с момента создания предварительной версии программы полета и модели грузопотока: баллистики выдают проектантам стратегию поддержания высоты станции по годам: сколько коррекций необходимо сделать и сколько ориентировочно потребуется топлива. Коррекции орбиты для поддержания высоты станции в данном контексте называются маневрами. Сведения от баллистиков позволяют сделать корректировки модели грузопотока с учетом топлива, необходимого для поддержания высоты.

Расчет расхода топлива, необходимого для изменения ориентации, производится на основании данных усредненных таблиц, содержащих сведения о расходе топлива по выполнению операций для определенной сборки МКС (конфигурации модулей и пристыкованных транспортных средств). Основными параметрами, влияющими на расход, являются масса станции, расстояние двигателей относительно центра масс и величина угола разворота Поддержание ориентации станции может быть произведено по-разному (можно использовать собственные двигатели станции, или двигатели пристыкованных транспортных средств), расход топлива может быть разным. Работа проектанта в этом случае заключается в анализе всех вариантов и выборе наиболее оптимального. Это необходимо учитывать при перепланировании грузопотока.

Для автоматизации описанных задач планирования требуется создание системы, способной в реальном времени обеспечивать полный цикл управления, от автоматической реакции на события – к планированию грузопотока и контролю исполнения намеченного плана. Разрабатываемое программное обеспечение должно обеспечивать:

- организацию коллективной работы проектантов, ЦУП, баллистиков и кураторов при согласовании грузопотока в рамках единого информационного пространства;
- анализ грузопотока с целью его оптимизации и поиска возможных альтернативных вариантов;
- моделирования различных ситуаций и быстрой (оперативной) проработки разных вариантов программы полета и плана грузопотока в случае необходимости.

2 Применение мультиагентных технологий в планировании грузопотока МКС

В основе мультиагентного подхода планирования грузов МКС лежит создание программных агентов, которые могут действовать от лица и по поручению каждого груза или его владельца (куратора), планируя оптимальную доставку и договариваясь с другими агентами грузов о времени доставки на борт станции или, наоборот, эвакуации со станции на Землю. Особенностью предлагаемого мультиагентного подхода является возможность построения плана грузопотока в результате функционирования динамически формируемой сети потребностей и возможностей [4] ресурсов МКС. Конфигурация (временное равновесие) агентов в такой сети может изменяться в любой момент под действием внешних или внутренних событий, информация о которых поступает от центра управления полетами (ЦУП) и кураторов.

Такая мультиагентная система работает в интерактивном режиме, под управлением пользователя. Автоматически генерируемые в результате работы системы варианты грузопотока становятся доступными проектантам для простой, гибкой и удобной ручной доводки в реальном времени. При необходимости исследования новых возможностей по оптимизации грузопотока проектанты могут создавать новые версии и утверждать их в соответствии с принятым процессом.

Интерактивный характер взаимодействия проектантов и мультиагентной системы подразумевает, что диалог с системой может быть инициирован каждой из сторон в любой момент времени, что позволит адаптивно, т.е. без полного перепланирования «с нуля», достраивать или перестраивать грузопоток по мере возникновения важных событий или с целью оптимизации.

Применение мультиагентных технологий позволяет ввести элементы самоорганизации в систему поддержки принятия решений в области планирования. Тем самым реализуется новый подход к решению поставленной задачи, который позволяет справляться с растущей сложностью планирования грузопотока и частыми изменениями требований, связанными с непредсказуемым появлением новых срочных потребностей в пересылке грузов, переносом стартов, расширением номенклатуры грузов, и т.д.

Область применения системы: подразделения, занимающиеся построением программы полета, планированием грузопотока и расчета грузов для МКС, их взаимодействие с ЦУП и кураторами, а также их обслуживание силами службы информационных технологий предприятия.

3 Мультиагентный алгоритм планирования грузопотока МКС

В системе существуют агенты транспортных средств, а также 4 типа агентов, описывающие топливо, перевозимое на каждом полете, воду, перевозимую на каждом полете и сухие грузы. Возможно добавление дополнительных агентов (например, для коммерческих грузов и грузов партнеров). В результате совместной работы агентов сухих грузов или группы грузов, если грузы связанные (то есть должны планироваться вместе) ищется оптимальное распределение грузов по транспортным средствам.

Подсистема планирования находит и предлагает размещение грузов на грузовых и пилотируемых транспортных средствах, на основе ряда критериев: максимальная наполненность ТС по всем видам грузов, попадание грузов в грузопоток в период между желаемым и крайним сроками удовлетворения соответствующей заявки.

Стратегия мультиагентных переговоров представляет собой описание протокола взаимодействий, определяющего последовательность обмена сообщениями между агентами. В качестве примера можно привести следующую стратегию переговоров для случая планирования доставки нового груза по поступившей заявке от куратора:

- по заявке куратора производится создание нового агента груза;

- агент груза опрашивает агенты транспортных средств, включенные в программу полета МКС, относительно принципиальной возможности планирования доставки в рамках данного полета, и получает от них ответы;
- агент нового груза анализирует и ранжирует возможные варианты по их выгодности для себя (оценка осуществляется в соответствии со степенью соответствия варианта требованиям нового груза, упорядоченных по степени важности);
- если свободного места на выбранном грузом полете транспортного корабля достаточно для размещения (груз проходит по весу и объемам), доставка данного груза включается в план грузопотока. В другом случае, осуществляется следующее:
  - осуществляется опрос агента топлива. В результате, возможно, объем доставляемого в рамках полета ТС топлива уменьшится, и часть планируемого к доставке груза размещается. Если, однако, груз разместился не полностью, несмотря на снижение количества доставляемого топлива, и уменьшилось, опрос будет продолжен. Снижать количество топлива до нуля нельзя, имеется минимальное значение, соответствующее несокращаемому количеству доставляемого при каждом полете к МКС топлива;
  - опрашивается агент воды. В результате, возможно, объем перевозимой воды уменьшится, и груз будет успешно размещен. Если количество воды также достигло уровня несокращаемого остатка, а груз все еще не размещен, то опрашиваются агенты сухих грузов;
  - проводится опрос сухих грузов и поиск тех агентов, кто может «уйти» с данного полета без существенных потерь.

При выборе грузов, которые могут «уйти», возможно использование виртуальных «денег» (аналог затрат энергии на пересмотр связей в системе для саморегуляции активностей), или же выбор производится путем простого сравнения вариантов по главным критериям. Если новый груз для транспортного средства выгоднее (например, по стоимости), «уходит» ранее запланированный груз. Аналогично разрешается конфликт по времени отправки, и тогда более срочный груз попросит освободить место другие грузы, имеющие запас по крайнему сроку доставки.

В ходе планирования учитываются ограничения, влияющие на поведения агентов. Например, по программе полета МКС известны даты старта, стыковки, расстыковки российских кораблей, агенты сухих грузов пытаются занять место на транспортных средствах (с ограничением массы полезной нагрузки, а также баланса сухих грузов/топлива/воды) в наиболее близкому к желаемому времени, либо на определенных транспортных средств, если они указаны, учитываются приоритеты грузов, а агенты воды и топлива стараются заполнить баки кораблей до указанного желаемого значения.

Агенты сухих грузов принимают решения на основе максимизации функции эффективности (KPI), определенной для каждого агента, на основе характеристик грузов и желательного/допустимого интервалов доставки грузов. Функция KPI строится на основе приоритетности грузов – чем более груз «критичнее», тем выше значение. Для агента сухого груза при динамическом распределении, возможны следующие варианта развития событий: груз попал в грузопоток в желаемый промежуток, груз попал в грузопоток в допустимый промежуток, груз не попал в грузопоток в допустимый промежуток (при этом формируется предупреждение пользователю).

Таким образом, каждый агент сухого груза осуществляет поиск полета с наиболее близким временем к оптимальному. При выборе кораблей из списка учитывает желаемые даты (если они заданы), если не заданы либо условия равнозначные, выбирает транспортное средство с более ранней датой полета. Если на полете есть место, груз занимает его (если груз фиксируется пользователем, при

необходимости, он вытесняет грузы с любым приоритетом). Если места не хватает, он производит переговоры с другими агентами сухих грузов (с учетом приоритета и времени, условий «фиксации» груза на данном транспортном средстве), агентами топлива (по массе), воды (по массе) и полетов (по массе).

Агент топлива заполняет объем бака транспортного средства, на котором «хочет полететь», желаемой массой, а при конфликте по массе с другими грузами, начинает с ними договариваться по поводу уменьшения своей массы. При возможности этот агент старается увеличить свою массу до желаемой, вступая при этом в конфликты с другими агентами и разрешая их.

При принятии решений о размещении конфликтных грузов, основной целью является размещение всех грузов, внутри допустимых промежутков доставки, даже если соответствующие агенты будут удовлетворены не полностью. Агенты, вступающие в конфликт, учитывают это правило при принятии решений.

Таким образом, в разработанной подсистеме агенты обладают следующими особенностями:

- «соединенным грузам» назначается один агент, а «зафиксированные» грузы не идут на уступки в случае возникновения конфликтных ситуаций;
- агенты сухих грузов могут вытеснять друг друга при борьбе за один возможный ресурс с учетом приоритетов, а также могут договариваться друг с другом и занимать различные полеты, даже если это не отвечает их личным целям;
- агенты сухих грузов могут вытеснять из полета топливо и воду. Договоренность осуществляется на расчете общей полезности системе;
- агенты грузов борются за освободившееся место (массу) на транспортном средстве. Договоренность может осуществляться как преобладание личных целей (сравнение локальных KPI), так и общей – на основе расчета KPI «совместной полезности»;
- агенты полетов пассивны и используется в качестве хранилища данных о загрузке всеми видами грузов и обладают состоянием удовлетворенности загрузки.

Выводы

Разработанная подсистема планирования грузопотока в составе мультиагентной системы интерактивного построения программы полета, грузопотока и расчета ресурсов МКС позволяет учитывать множество критериев, предпочтений и ограничений, связанных с планированием сухих грузов, воды и топлива. Это обеспечивает интеллектуальную поддержку принятия решений в процессе взаимодействии проектантов, кураторов грузов и руководства предприятия при планировании грузопотока в автоматическом режиме и последующем ручном редактировании в ходе согласования.

Литература

1. Андреев, М.В., Построение адаптивной системы управления предприятием с использованием мультиагентных технологий [Текст]/ М.В. Андреев, А.В. Иващенко, П.О. Скобелев, А.В. Царев. – Вестник Самарского государственного технического университета, Серия «Технические науки».– 2009. – № 1 (23). – с. 5 – 14.

2. Андреев, М.В., Применение мультиагентных технологий динамического планирования персональных задач при организации коллективного взаимодействия в автоматизированных системах управления распределением ресурсов [Текст] / М.В. Андреев, А.В. Иващенко, Д.М. Мартышкин, П.О. Скобелев, Л.В. Уланова, А.В. Царев. – М.: Мехатроника. Автоматизация. Управление. – 2010. – № 7. – с. 21 – 27.

3. Диязитдинова, А.Р., Концепция мультиагентной системы интерактивного построения программы полета и планирования грузопотока Международной космической станции [Текст]/ А.Р. Диязитдинова, А.В. Иващенко, Е.В. Симонова, П.О. Скобелев, М.В. Сычева, И.И. Хамиц, А.В. Царев. – Труды XII Международной конференции «Проблемы управления и моделирования в сложных системах». – Самара: Самарский научный центр РАН. – 2010. – с. 686 – 694.

Мультиагентная подсистема построения программы полета Международной космической станции

А.В. Иващенко[1], К.Ю. Матвеев[1], А.Л. Новиков[1], Е.В. Симонова[1], П.О. Скобелев[1], М.В. Сычева[2], И.И. Хамиц[2]
[1] *Научно-производственная компания «Разумные решения», Самара*
[2] *Ракетно-космическая корпорация «Энергия», Королев*

Введение

Управление обеспечением Международной космической станции (МКС) системами обеспечения жизнедеятельности человека, водой, топливом и другими грузами тесно связано с планированием полетов грузовых и пилотируемых транспортных средств: автономных полетов, стыковок, перестыковок, а также экспедиций и внекорабельной деятельности. Результатом этой работы является программа полета МКС, которая после утверждения становится основой для разработки оперативной документации ЦУПа, также для взаимодействия между проектантами и кураторами по планированию подготовки и транспортировки грузов.

В ходе разработки программы полета необходимо учесть достаточно большое количество требований и ограничений: минимальные интервалы времени между операциями, данные по времени выполнения операций, правила по стыковке транспортных средств к модулям станции и другие. Для того, чтобы выполнить все эти требования, и при этом создать программу полета, наиболее эффективную с точки зрения использования транспортных средств, загрузки экипажа МКС, в том числе для выполнения научной программы экспериментов, целесообразно использование новых подходов по автоматизации интеллектуального планирования ресурсов.

Один из таких подходов, показавший хорошие результаты в транспортной и производственной логистике, состоит в применении мультиагентных технологий, позволяющих обеспечить интерактивное построение программы полета МКС с учетом множества предпочтений и ограничений, заданных в специальной базе знаний – онтологии. В данной работе описаны основные особенности мультиагентного алгоритма построения программы полета МКС, разработанного и реализованного авторами в рамках общей задачи создания мультиагентной системы построения программы полета, грузопотока и расчета ресурсов МКС [3].

1 Задача построения программы полета МКС

Программа полета МКС представляет собой план распределения стыковок транспортных средств (пилотируемых и грузовых космических аппаратов) к основным модулям (портам), составляющим конфигурацию МКС. Для каждой стыковки также учитываются временные периоды автономного полета от старта до стыковки и от отстыковки до приземления (для пилотируемых космических аппаратов). При построении программы полета учитываются разнообразные ограничения (в основном временные), определяющие интервалы между стыковками и отстыковками, а также баллистические данные, определяемые позицией Солнца, освещенностью «мишени» для стыковки и т.п.

Входные данные, используемые для построения программы полета, включают модель грузопотока МКС, данные баллистиков, а также свойства транспортных средств (грузоподъемность сухих грузов, объем баков топлива и воды). При построении программы полета по определенному в модели грузопотока количеству полетов транспортных средств на основании данных предыдущих периодов и с приблизительной равномерностью полеты транспортных средств планируются по времени и распределяются по портам. Результат построения программы полета

представляет собой расписание стартов, стыковок и отстыковок с распределением по модулям МКС на период (год).

В работе находятся (поддерживаются) сразу несколько вариантов программы полета, один вариант актуальный, остальные альтернативные. Процесс интенсивного формирования программы полета идет на протяжении около трех месяцев в конце каждого года. В процесс вовлечены около 50 постоянных участников, курирующих различные подсистемы станции. На протяжении года план уточняется по мере возникновения различных событий (события бывают разной степени важности и в основном разрешаются локальными взаимодействиями участников). Эти события должны приводить к пересмотру программы полета в автоматическом режиме и могут привести к перепланированию стартов и стыковок.

Подготовленная актуальная программа полета направляется на согласование, после которого предоставляется специалистам ЦУП для оперативного планирования и кураторам для формирования заявок на транспортировку грузов. Программа полета утверждается за полгода до исполнения. За четыре месяца до фактического исполнения, в соответствии с программой полета производится планирование грузопотока, т.е. грузы распределяются по пускам.

При построении по программе полета грузопотока определяется масса сухих грузов, затем проверяется план по топливу, воде и газу. Также осуществляются проверки временных ограничений (в том числе на операции космонавтов, которые связаны с использованием грузов). В случае возникновения необходимости (например, при поступлении срочной и важной заявки от кураторов), проводится возврат к плану доставки сухих грузов для его пересмотра. При анализе балансов питания, топлива, газа и воды используются графики баланса (для воды и питания расход рассчитывается по нормативам, для топлива расход определяется в соответствии с планом операций на станции).

Зависимость между программой полета и грузопотоком двусторонняя. При изменении программы полета изменяется грузопоток (переносятся доставки грузов). Если грузы не готовы к запланированному времени, они снимаются. Если в плане грузопотока мало грузов (на конец года накапливается неготовность), некоторый полет может быть отменен, а остаток с него перенесен на более ранние или поздние сроки. Также на грузопоток может влиять осуществление выхода в космос, а изменения грузопотока могут влиять на сроки пусков.

2 Аспекты применения мультиагентных технологий для построения программы полета МКС

Особенность решения задачи интерактивного построения программы полета заключается в том, что необходимо обеспечивать согласованность и исполнимость планов при всех шагах: на этапе прогноза (построения модели грузопотока), этапе построения поблочного грузопотока, этапе расчета балансов топлива, воды, газа и рационов питания. Решить эту задачу, используя стандартные классические методы оптимизации сложно, поскольку приходится учитывать значительное число противоречивых целей и ограничений, количество и суть которых изменяется со временем.

В связи с этим для решения данной задачи предлагается использовать мультиагентный подход к динамическому управлению ресурсами в реальном времени, базирующийся на представлении объекта управления – программы полета – в виде самоорганизующейся системы. Необходимость применения такого подхода обосновывается требованием по обеспечению согласованных изменений программы полета и грузопотока в ответ на поступающие данные от ЦУП и заявки от проектантов,

причем эти изменения всегда должны быть направлены на оптимизацию доставки грузов на МКС.

Это означает, что нужно, во-первых, обеспечивать связь между грузопотоком и программой полета, а во-вторых, в результате эволюционного и согласованного развития программы полета, находить оптимальные варианты. Достаточность применения мультиагентных технологий состоит в известной возможности мультиагентного мира обеспечивать направленный поиск компромиссов при условии существования конфликта интересов, возможности задания частных предпочтений и ограничений агентов и взаимосвязанности сцен.

Сходимость мультиагентного алгоритма построения программы полета обеспечивается, во-первых, конечным количеством вариантов оптимизации полетов и стыковок транспортных средств, а, во-вторых, направленностью поиска агентами возможностей по оптимизации грузопотока. Впрочем, в мультиагентной системе возможны (и предусматриваются в данном решении) ситуации, когда небольшие изменения по локальной оптимизации приводят к разрушению всего согласованного плана. Эти разрушения временные и могут производиться в альтернативных сценах. В данной задаче предусмотрено, что существует одна сцена (программа полета), которая является основной и согласованной, и множество альтернативных, созданных для моделирования различных вариантов изменений и поиска возможностей оптимизации грузопотока.

Таким образом, под построением программы полета следует понимать не единовременно решаемую задачу, а непрерывный процесс корректировки и согласования, инициируемый и стимулируемый различными событиями (обновление сведений о текущей ситуации на МКС от ЦУП, приход новых заявок от кураторов и появление новых идей о необходимости оптимизации от проектантов). При этом согласованность обеспечивается поиском компромиссов (между полетами, экспедициями, транспортными средствами и модулями МКС и грузами).

3 Мультиагентный алгоритм построения программы полета МКС

В процессе автоматизированного создания программы полета используются следующие типы агентов:

− агент порта – это агент ресурса, предоставляет агентам стыковок информацию о свободных интервалах или информацию о том, кто занял интересующее их место. Агенты данного типа наделены логикой, в соответствии с которой они сами могут выбирать агентов полетов, которых лучше присоединить;

− агент полета предоставляет агентам стыковок время, которое они могут занять, исходя из внутренних соображений (несколько агентов стыковок разделяют один ресурс длительности полета). Агент полета создает и контролирует агенты своих стыковок и устанавливает для них правила планирования (например, определяет количество дней между стыковками);

− агент стыковки является основным действующим лицом мира агентов. Именно его активность занимает более половины времени работы планировщика. Для этого агента главная задача состоит в том, чтобы пристыковаться к порту, с учетом множества требований из разных источников;

− мастер-агент контролирует процесс переговоров. В его задачи входит следить за переговорами других агентов и, в случае возникновения длительных циклов, выводить их из этого состояния. Через этого агента все другие агенты получают информацию об активности друг друга.

На мастер-агента возложена задача контролировать взаимное расположение агентов стыковок. В соответствии с правилами планирования, мастер-агент отвечает за

сокращение интервалов между агентами стыковок, не увеличивая их продолжительность.

В системе предусмотрены две стратегии поведения агентов: агрессивная и примирительная. В зависимости от выбранной стратегии поведения агентов меняется скорость сходимости алгоритма и его качество планирования: чем агрессивней стратегия агента, тем быстрее сходится алгоритм, но планирование становится менее качественным. При использовании примирительной стратегии агентов время планирования увеличивается, но в результате переговоров производится поиск лучшего решения. Стоит отметить, что в обоих случаях полностью выполняются условия валидности сцены, т.е. исполнимости найденного плана.

Мультиагентный алгоритм построения программы полета МКС можно разбить на несколько составляющих этапов, каждый их которых выполняется независимо, но использует результаты других этапов. Наиболее важные этапы алгоритма:

- поиск агентов друг друга (тип): в текущей версии агенты «видят» тип друг друга. В процессе поиска агент запрашивает у мастер-агента список активных агентов конкретного типа;
- заключение контракта агента полета с агентом модуля;
- поиск и разрешение конфликтов, связанных с наложением интервалов стыковок и отстыковок;
- выявление и устранение конфликтов между транспортными средствами, которые претендуют на интервал времени на одном и том же порту стыковки;
- заключение контрактов (фиксация договоренностей).

Кроме этого, при планировании удовлетворяется требование по максимальному использованию ресурса транспортного средства, учитываются данные баллистиков, а также специально обрабатывается планирование перестыковок транспортного средства с одного порта на другой.

Выводы

Разработанная подсистема построения программы полета в составе мультиагентной системы интерактивного построения программы полета, грузопотока и расчета ресурсов МКС позволяет учитывать множество критериев, предпочтений и ограничений, связанных с планированием стартов, стыковок и отстыковок транспортных средств. В перспективе предлагаемый подход позволит создавать адаптивные сети динамических планировщиков, способных обеспечить поддержку коллективного согласования и принятия решений в реальном времени на различных этапах разработки, производства, транспортировки и запуска объектов космической техники, в разных подразделениях и организациях, работающих совместно над решением данных задач.

Литература

1. Андреев, М.В., Построение адаптивной системы управления предприятием с использованием мультиагентных технологий [Текст]/ М.В. Андреев, А.В. Иващенко, П.О. Скобелев, А.В. Царев. – Вестник Самарского государственного технического университета, Серия «Технические науки».– 2009. – № 1 (23). – с. 5 – 14.

2. Андреев, М.В., Применение мультиагентных технологий динамического планирования персональных задач при организации коллективного взаимодействия в автоматизированных системах управления распределением ресурсов [Текст] / М.В. Андреев, А.В. Иващенко, Д.М. Мартышкин, П.О. Скобелев, Л.В. Уланова, А.В. Царев. – М.: Мехатроника. Автоматизация. Управление. – 2010. – № 7. – с. 21 – 27.

3. Диязитдинова, А.Р., Концепция мультиагентной системы интерактивного построения программы полета и планирования грузопотока Международной космической станции [Текст]/ А.Р. Диязитдинова, А.В. Иващенко, Е.В. Симонова, П.О. Скобелев, М.В. Сычева, И.И. Хамиц, А.В. Царев. – Труды XII Международной конференции «Проблемы управления и моделирования в сложных системах». – Самара: Самарский научный центр РАН. – 2010. – с. 686 – 694.

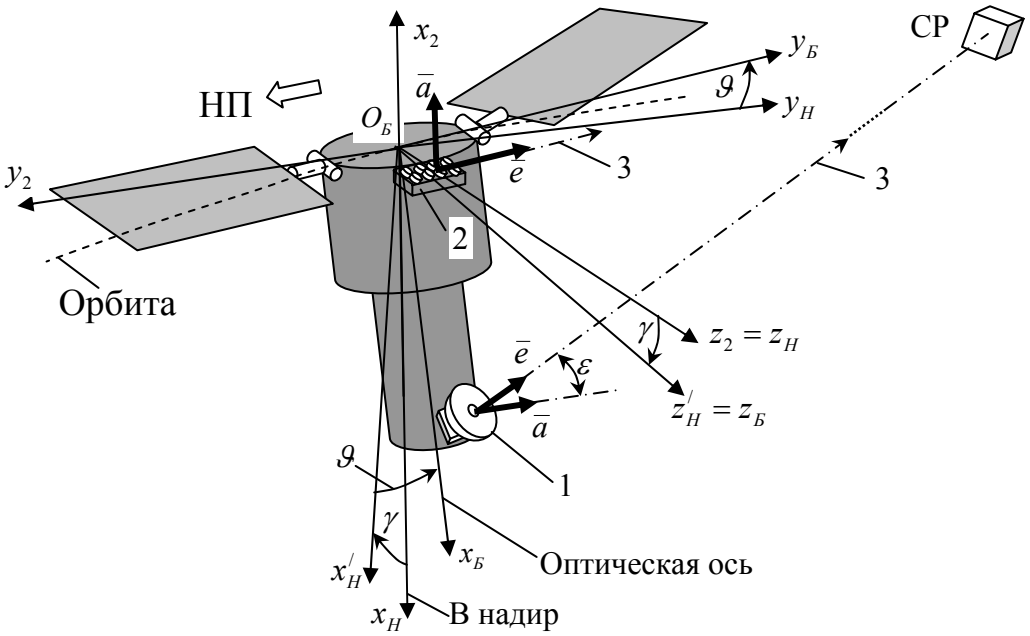
Модели и алгоритмы для оценки среднесуточного времени радиовидимости спутника-ретранслятора с космического аппарата наблюдения с учетом затенения антенн

В.И. Куренков, Н.Р. Стратилатов, О.Г. Федоренко, Л.Б. Шилов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), ГНПРКЦ «ЦСКБ-Прогресс», Самара

Для повышения оперативности доставки видеоинформации с космических аппаратов (КА) наблюдения на наземные пункты приема в дополнение к операции непосредственной передачи информации с передающих антенн КА на Землю может использоваться операция передачи информации через геостационарные спутники-ретрансляторы (СР).

При этом процесс передачи видеоинформации возможен, когда передающая и приемные антенны спутников находятся в условиях взаимной видимости и не затенены корпусом или другими навесными элементами КА (см. рис. 1). Причем, при передаче видеоинформации через СР желательно не прерывать съемку объектов наблюдения.

В этой связи перед проектантами космической техники встает вопрос о предпочтительных местах установки антенных устройств на корпусе КА с целью минимизации времени затенения передающих антенн элементами конструкции КА в процессе целевой работы и одновременного процесса передачи видеоинформации на спутник-ретранслятор.



1, 2 – передающие антенны различных типов; 3 – линия радиосвязи

Рис. 1. Схема направления антенн (векторов $\bar{e}$) на спутник ретранслятор без затенения линии радиосвязи элементами конструкции КА

Для ответа на этот вопрос необходимо сначала решить задачу об оценке среднесуточного времени радиовидимости спутника-ретранслятора передающей антенной, установленной на корпусе КА наблюдения. При этом возможны полная и упрощенная постановки задач. При полной постановке - учитывается степень затенения

передающих антенн элементами конструкции КА наблюдения при направлении этих антенн на спутник-ретранслятор. В данной работе рассматривается упрощенная постановка задачи, а именно, выделяется полупространство вокруг КА наблюдения, в котором установлена рассматриваемая передающая антенна, и производится оценка нахождения спутника-ретранслятора в том или ином полупространстве.

Эту задачу невозможно решить с помощью аналитических зависимостей, так как количество факторов, влияющих на решение велико. В данной работе для решения используется метод имитационного моделирования орбитального полета КА наблюдения и спутника-ретранслятора с учетом программных разворотов КА наблюдения при целевом функционировании.

Суть моделирования

В базовой системе координат КА строится единичный вектор $\bar{a}$ к плоскости полупространства, в котором антенны не затенены элементами конструкции КА (см. рис. 1). В каждый момент времени имитации орбитального движения проверяется условие взаимной видимости (отсутствие экранирования Землей) КА наблюдения и СР без учета степени затенения антенн КА его элементами конструкции. При выполнении этого условия определяются координаты СР и КА наблюдения в геоцентрической гринвичской системе координат и рассчитываются координаты единичного вектора $\bar{e}$ направления луча от КА наблюдения к СР. Далее осуществляется цикл пересчетов из одних координат в другие (подробнее см. в алгоритме) для определения координат вектора $\bar{e}$ в базовой системе координат КА. Потом рассчитывается угол между векторами $\bar{a}$ и $\bar{e}$. Если этот угол менее 90 градусов (косинус этого угла положительный), то СР находится в полупространстве, в котором установлена направленная антенна КА и, следовательно, экранирование передающих антенн корпусом КА наблюдения отсутствует. Осуществляется приращение времени имитации и регистрируется накопленное время видимости антенн КА со спутника-ретранслятора.

Основные допущения

1. Форма Земли – сферическая, однако для расчета параметров орбит используются уравнения эллиптического движения с периодической коррекцией расчетных значений долготы восходящего узла (прецессии орбиты) и аргумента перигея в процессе длительного полета, вызванных несферичностью Земли (учитываются вековые возмущения от второй зональной гармоники в разложении геопотенциала).

2. Для расчета эволюции долготы восходящего узла (прецессии орбиты) и эволюции аргумента перигея в процессе длительного полета учитывается несферичность Земли (учитываются вековые возмущения первого порядка).

3. Влиянием аэродинамических верхних слоев атмосферы на исследуемых орбитах пренебрегаем.

Ниже приведены исходные данные алгоритм и математические модели, используемые для решения задачи.

Исходные данные для расчета

Должна быть задана долгота стояния спутника-ретранслятора λ_{CP} (широта $\varphi_{CP} = 0$). Остальные исходные данные получаются как промежуточный результат вычислений в процессе работы ранее разработанного программного обеспечения [1], предназначенного для оценки целевых показателей эффективности КА наблюдения, а именно: модуль радиуса-вектора КА $r_{KA}(t)$, долгота $\lambda_{KA}(t)$ и широта $\varphi_{KA}(t)$ подспутниковой точки КА на поверхности Земли в каждый момент времени имитационного моделирования. В свою очередь, для запуска этой программы необходимы следующие исходные данные: ω_3 - угловая скорость вращения Земли

относительно своей оси, i - наклонение плоскости орбиты, Ω - долгота восходящего узла, ω_π - аргумент перигея, r_π и r_α - высоты перигея и апогея КА наблюдения.

Алгоритм и используемые математические модели

1. Обнуляется счетчик накопленного времени t_{PB} нахождения направленной антенны КА наблюдения и СР в условиях взаимной видимости.

2. В каждый момент времени имитации орбитального движения проверяется условие экранирования Землей линии взаимной видимости КА наблюдения и СР. Это условие проверяется с использованием моделей [2].

При наличии экранирования Землей линии взаимной видимости КА наблюдения и СР радиосвязь невозможна и сразу переходят к выполнению пункта 12 настоящего алгоритма.

При отсутствии экранирования проверяется условие: находится ли СР в полупространстве базовой системы координат КА наблюдения, в котором установлена направленная антенна. Для этого выполняются пункты 3…11 настоящего алгоритма.

3. Определяются координаты СР и КА наблюдения в геоцентрической гринвичской системе координат (СК) для каждого момента времени имитации полёта

$$x_Г^{KA} = r_{KA} \cos\varphi_{KA} \cos\lambda_{KA};$$

$$y_Г^{KA} = r_{KA} \cos\varphi_{KA} \sin\lambda_{KA};$$

$$z_Г^{KA} = r_{KA} \sin\varphi_{KA}.$$

Схема для определения координат в этом пункте, а также в пунктах 4 и 5 настоящего алгоритма, представлена на рис. 2.

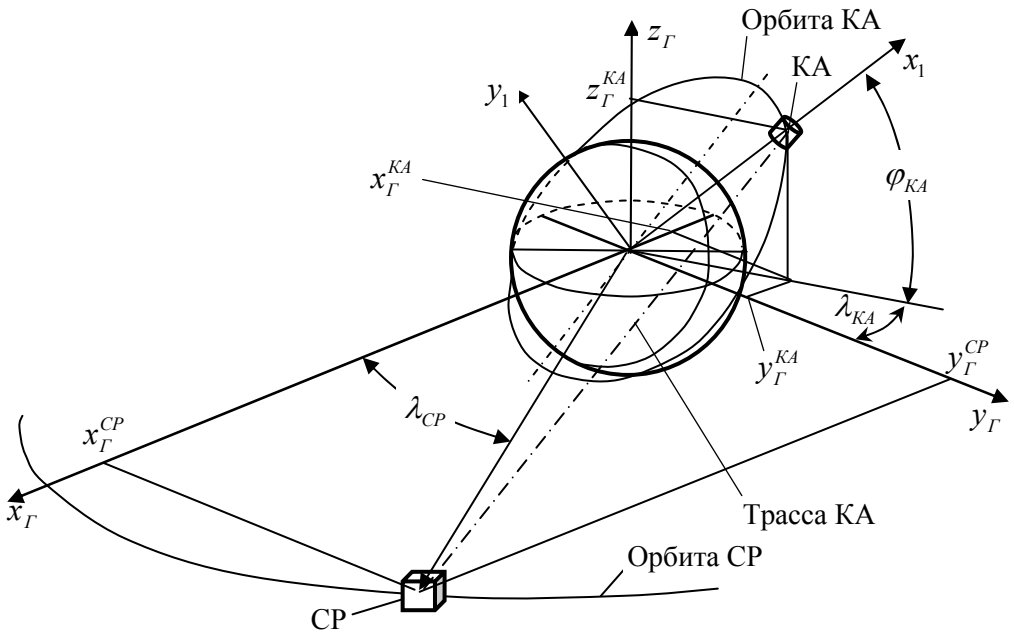


Рис. 2. Схема для определения координат КА наблюдения и спутника-ретранслятора в гринвичской системе координат

4. Рассчитываются координаты спутника-ретранслятора в геоцентрической гринвичской СК.

$$x_Г^{CP} = r_{CP} \cos\lambda_{KA};$$

$$x_Г^{CP} = r_{CP} \sin\lambda_{KA}$$

$$z_Г^{CP} = 0.$$

5. Рассчитываются координаты единичного вектора направления от КА наблюдения к спутнику-ретранслятору (далее для краткости вектора $\bar{e}$) в геоцентрической гринвичской СК

$$x^e_{\Gamma} = \frac{r_{CP}\cos\lambda_{KA} - r_{KA}\cos\varphi_{KA}\cos\lambda_{KA}}{\sqrt{(x_{CP}-x_{KA})^2+(y_{CP}-y_{KA})^2+(z_{CP}-z_{KA})^2}};$$

$$y^e_{\Gamma} = \frac{r_{CP}\sin\lambda_{KA} - r_{KA}\cos\varphi_{KA}\sin\lambda_{KA}}{\sqrt{(x_{CP}-x_{KA})^2+(y_{CP}-y_{KA})^2+(z_{CP}-z_{KA})^2}};$$

$$z^e_{\Gamma} = \frac{-r_{KA}\sin\varphi_{KA}}{\sqrt{(x_{CP}-x_{KA})^2+(y_{CP}-y_{KA})^2+(z_{CP}-z_{KA})^2}}.$$

6. Осуществляется пересчет координат вектора $\bar{e}$ из геоцентрической гринвичской СК в неподвижную геоцентрическую СК

$$\begin{Vmatrix} x^e \\ y^e \\ z^e \end{Vmatrix} = A^T \cdot \begin{Vmatrix} x^e_{\Gamma} \\ y^e_{\Gamma} \\ z^e_{\Gamma} \end{Vmatrix}, \text{ где } A = \begin{Vmatrix} \cos(\omega_З t) & \sin(\omega_З t) & 0 \\ -\sin(\omega_З t) & \cos(\omega_З t) & 0 \\ 0 & 0 & 1 \end{Vmatrix}.$$

7. Осуществляется пересчет координат вектора $\bar{e}$ из неподвижной геоцентрической СК в геоцентрическую орбитальную СК, связанную с перицентром орбиты (на схемах не показана). Далее последовательно: - в барицентрическую СК ($O_2x_2y_2z_2$, см. рис. 1); - в СК, связанную с центром масс КА и центром масс КА ($O_Hx_Hy_Hz_H$); - в базовую СК КА ($O_Бx_Бy_Бz_Б$).

Пересчеты координат вектора $\bar{e}$ в настоящем пункте осуществляются с использованием матриц поворота и формул, аналогичных для пересчета координат единичного вектора направления на Солнце. Эти матрицы и формулы представленны в [3] (см. пункты 3…6 алгоритма в подразделе 5.2.12).

8. Периодически, через заданное время, производится имитация перенацеливания КА с одного объекта наблюдения на другой. Это делается с помощью с помощью задания случайных углов тангажа ϑ и крена γ КА (см. рис. 1) в конусе обзора.

9. Осуществляется пересчет координат вектора $\bar{e}$ из СК, связанной с центром масс КА и центром масс КА, в базовую СК с учетом углов тангажа ϑ и крена γ КА. Пересчет осуществляется с учетом компоновочных схем КА [2];

10. Осуществляется расчет косинуса угла ε между векторами $\bar{e}$ и $\bar{a}$ (см. рис. 1)

$$\cos\varepsilon = a_{xБ}x^e_Б + a_{yБ}y^e_Б + a_{zБ}z^e_Б,$$

где $a_{xБ}, a_{yБ}, a_{zБ}$ – координаты единичного вектора $\bar{a}$ (которые также зависят от компоновочной схемы КА);

$x^e_Б, y^e_Б\ z^e_Б$ – координаты единичного вектора $\bar{e}$ в базовой СК.

11. Если косинус угла ε положительный, то СР находится в полупространстве, в котором установлена направленная антенна КА, и, следовательно, экранирование передающих антенн корпусом КА наблюдения отсутствует.

12. Рассчитывается накопленное время t_{PB} нахождения направленной антенны КА наблюдения и СР в условиях взаимной видимости

$$t_{PB} = t_{PB} + \Delta t,$$

где Δt – шаг расчета при имитационном моделировании.

13. Проверяется условие окончания расчета. Если расчет продолжается, то осуществляется приращение времени имитации полета КА и СР на шаг Δt и переход к п. 1 настоящего алгоритма. В противном случае программа останавливается.

На основе данного алгоритма и представленных моделей разработан программный модуль, с помощью которого модернизировано созданное ранее программное обеспечение для оценки целевых показателей эффективности КА наблюдения [1].

Разработанные модели, алгоритм и программное обеспечение может быть использовано проектантами космической техники для выбора предпочтительных местах установки антенных устройств на корпусе КА по критерию минимума времени затенения передающих антенн элементами конструкции КА в процессе целевой работы и одновременного процесса передачи видеоинформации на спутник-ретранслятор.

Литература

1. Основы устройства и моделирования целевого функционирования космических аппаратов наблюдения: учеб. пособие / В. И. Куренков, В. В. Салмин, Б. А. Абрамов. – Самара: Изд-во Самар. гос. аэрокосм. ун-та, 2006. – 178 с.

2. Шилов Л. Б., Куренков В.И. Оценка оперативности космических аппаратов наблюдения с использованием спутника–ретранслятора. В сб. Наука и технология. Том. 2. - Краткие сообщения XXX Российской школы, посвященной 65-летию Победы (15-17 июня, г. Миасс). Екатеринбург: УрО РАН, 2010, с. 11-13.

3. Моделирование целевого функционирования космических аппаратов наблюдения с учетом энергобаланса: учеб. пособие / В. И. Куренков, В. В. Салмин, Б. А. Абрамов. - Самара: Изд-во Самар. гос. аэрокосм. ун-та, 2007. – 160 с.: ил.

Программное обеспечение проектно-баллистической оптимизации межпланетных КА с ЭРДУ, использующих гравитационные маневры

Д.В. Курочкин, О.Л. Старинова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Программы исследования и освоения межпланетного пространства требуют больших материальных затрат и не дают быстрой отдачи. Одним из возможных путей снижения затрат на доставку космического аппарата (КА) к планете назначения является использование электрореактивных двигательных установок (ЭРДУ), работающих на принципе ускорения рабочего тела в электростатических или электромагнитных полях. Высокая скорость истечения реактивной струи обеспечивает значительно меньший расход рабочего тела по сравнению с двигателями на химическом топливе. Таким образом, можно доставить к планете назначения больше полезной нагрузки, но при этом существенно возрастет длительность перелета.

Другим способом снижения затрат на осуществление межпланетных миссий является использование гравитационного поля массивных тел Солнечной системы. Баллистические схемы перелёта с использованием нескольких гравитационных манёвров позволяют существенно сократить время на осуществление миссии и / или общий расход топлива.

Объединение двух обозначенных способов позволяет существенно повысить эффективность космических миссий за счет снижения стоимости и увеличения массы полезного груза при разумной продолжительности перелета, но требует тщательной оптимизации баллистической схемы перелета с учетом движения нескольких тел (промежуточных планет). Кроме того повышение эффективности перелетов возможно за счет оптимального выбора проектных параметров КА: количества и тяги двигателей, оптимальной скорости истечения, характеристик используемой энергоустановки.

В данной работе рассматривается математическое и программное обеспечение позволяющее проводить проектно-баллистическую оптимизацию межпланетных перелетов КА с ЭРДУ с учетом нескольких гравитационных маневров и проводить приближенное моделирование проектного облика КА. Движение КА описывается в рамках теории сфер действия в гелиоцентрической системе координат под действием малой тяги переменного направления и величины. Движение планет описывается известными законами изменения оскулирующих элементов в зависимости от текущей юлианской даты. Если предполагается использование солнечной энергоустановки, то учитывается падение ее мощности при удалении КА от Солнца.

Необходимые условия оптимальности законов управления получены с использованием формализма принципа максимума Понтрягина получены для оптимальных по быстродействию или расходу рабочего тела участков перелета между планетами. Кроме того, получены краевые условия и соответствующие условия трансверсальности для осуществления гравитационных маневров у промежуточных планет и формирования заданной орбиты у планеты назначения. Для различных критериев оптимальности и граничных условий сформулированы постановки краевых задач и получены соответствующие сопряженные уравнения.

Разработано программное обеспечение, содержащее три взаимодействующих программных продукта.

1. Программный комплекс, предназначенный для численного решения краевых задач и основанный на алгоритме модифицированного метода Ньютона с автоматическим выбором шага вычисления производных. В качестве начального

значения сопряженных переменных, используются результаты решений известных задач об оптимальных гелиоцентрических перелетах хранящиеся во встроенной базе данных. Для оптимизации перелетов с различными граничными условиями и КА с различными проектными параметрами использовался метод продолжения по параметру.

Общая оптимальная баллистическая схема перелета определяется с учетом положения промежуточных планет на каждую рассматриваемую дату старта. Чтобы ускорить поиск, сначала отбираются все даты, которые обеспечивают требуемое взаимное расположение Земли и планеты назначения (в этих случаях перелет осуществим). Далее, из уже отобранного массива дат, отбираются те, на которые близко расположены КА и другие промежуточные планеты.

2. Программный комплекс, предназначенный для отображения результатов баллистической оптимизации. Отображается гелиоцентрическое движение КА с найденным оптимальным законом управления, оптимальными датами старта и пролета промежуточных планет, движение планет Солнечной системы. Пользователь может получать зависимости изменения выбранных фазовых координат и оптимального управления от времени перелета.

3. Программный комплекс, обеспечивающий укрупненное моделирование проектного облика КА. По результатам баллистической оптимизации определяются соответствующие параметры систем КА, приближенно вычисляются массогабаритные характеристики этих систем. Геометрические параметры вычисляются по простым зависимостям исходя из представления систем космического аппарата в виде объемных примитивов. Так, солнечные батареи представлены в виде прямоугольных пластин малой толщины; двигатели, и корпус аппарата имеют форму цилиндров; сферические баки с рабочим телом частично выступают над поверхностью корпуса аппарата. Различные вспомогательные системы, системы крепления, датчики, элементы радиолокационного оборудования не отображаются. Цель такого упрощенного моделирования – создать общее представление о соотношении габаритных размеров и взаиморасположении систем; отработать алгоритм автоматического расчета массогабаритных параметров и визуализации космического аппарата по результатам оптимизации баллистической схемы перелета.

При помощи разработанного математического и программного обеспечения были получены результаты оптимизации и моделирования различных вариантов миссий перелета КА с ЭРДУ с солнечной энергоустановкой к Юпитеру. Сначала осуществляется оптимизация миссий по минимальной длительности перелета и делается вывод о целесообразности гравитационных маневров. Для наиболее целесообразных баллистических схем проводится уточненная оптимизация с целью минимизации расхода рабочего тела. Определяются даты старта, проводится проектная оптимизация (выбор оптимальных значений тяги двигателей, мощности энергоустановки, площади панелей солнечных батарей, объема топливных баков и т.д.). Производится отображение проектного облика КА, моделирование и отображение оптимальных законов управления и соответствующей траектории движения.

Разработка проблемно-ориентированного программного обеспечения для задач проектирования космических аппаратов наблюдения

А.С. Кучеров, В.И. Куренков, В.Б. Белкин, А.А. Якищик
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В процессе проектирования космических аппаратов (КА) используются математические модели для оценки тех или иных проектных параметров бортовых систем. Все эти параметры должны быть увязаны между собой, причем с минимальными запасами по целевому функционированию, так как должны быть удовлетворены ряд условий, вытекающих из ограничений по массе конструкции, мощности электропотребления, времени выполнения отдельных частных операций и т.п.

Количество переменных в таких моделях и соотношений, которым должны удовлетворять характеристики различных составных частей КА, может быть достаточно большим. Это обусловливает возможность различного рода ошибок: несовместность уравнений, составляющих математическую модель проектируемого объекта (КА или его системы), недостаточность исходных данных и т.д.

По этой причине увязка указанных характеристик в настоящее время, как правило, производится на основе многократных циклов итерации. При этом головной проектант (отдел, отделение или конструкторское бюро) на начальных этапах проектирования постоянно уточняет исходные данные или технические задания соисполнителям на проектирование составных частей космических аппаратов. В общем случае, в результате уточнений постоянно меняется и сам проектный облик КА.

Для ускорения разработки проекта на начальных этапах проектирования предлагается разработать единый программный комплекс, в котором расчеты производились бы сразу по всем бортовым системам с одновременной увязкой всех проектных параметров. При этом желательно проверять корректность постановки задачи и выдавать рекомендации исполнителю по дальнейшим действиям. Все это можно сделать, если разработать проблемно-ориентированную систему проектирования, основанную на использовании так называемого непроцедурного программирования [1]. В программах такого рода формируются соотношения (модели), а не последовательность их решения. Порядок поиска решения определяется самой системой в зависимости от соответствующих запросов.

В проблемно-ориентированных системах проектирования открываются и другие возможности. Например, можно предусмотреть постановку нескольких задач проектирования, в зависимости от сочетания известных или неизвестных параметров. При этом может быть учтен специфический тип ограничений на проектные параметры, который формируется требованиями заимствования отдельных элементов, узлов и агрегатов бортовых систем из состава ранее созданных и отработанных бортовых систем КА, в результате чего те или иные проектные параметры принимают фиксированные значения. Кроме того на одном и том же множестве параметров, входящих в состав математической модели рассматриваемого объекта, возможна постановка задач другого рода, например, таких, как минимизация стоимости космической системы наблюдения, минимизация массы КА, улучшение целевых показателей эффективности, оптимизация расписания работы целевой и обеспечивающей аппаратуры, бортовой аппаратуры и т.д.

Логика работы проблемно-ориентированного программного обеспечения, предназначенного для выбора основных проектных харакетистик космических аппаратов, выглядит следующим образом.

1. Пользователь (проектант) вводит в систему множество параметров U, характеризующих проектируемый объект, а также множество отношений V, связывающих между собой различные параметры.

2. В зависимости от постановки проектной задачи, пользователь указывает множество $U_{вх} \subseteq U$ входных параметров задачи (величин, значения которых известны),

и множество $U_{вых} \subseteq U$ выходных переменных (величин, значения которых необходимо определить).

3. Система автоматически проверяет разрешимость и корректность задачи.

Как известно, задача разрешима, если в рамках модели (U,V) существует алгоритм вычисления значений выходных параметров по значениям входных параметров, т.е. если существует и может быть построено отображение $\varphi_i : \overline{U}_{вх} \to \overline{U}_{вых}$, где *i*- номер проектной задачи, $\overline{U}_{вх}$ - множество значений входных параметров, $\overline{U}_{вых}$ - множество значений выходных параметров. В противном случае задача неразрешима. Задача корректна на модели (U,V), если значения всех переменных, вычислимых по значениям $\overline{U}_{вх}$, определяются однозначно. Если хотя бы одна переменная модели может быть определена неоднозначно, то задача является некорректной.

4. Для корректной и разрешимой задачи система строит отображение $\varphi_i : \overline{U}_{вх} \to \overline{U}_{вых}$. Построение отображения включает:

– определение множества $V^* \subseteq V$ отношений модели, необходимых для построения отображения φ_i;

– нахождение схемы решения задачи, то есть разбиение множества V^* на непересекающиеся подмножества: $V^* = \{V_i\}, i = \overline{1,n},$, такие, что реализация отображения φ_i сводится к последовательному применению множеств отношений V_i.

Решение указанных задач основано на использовании теории графов и теории отношений и осуществляется в следующем порядке [1].

По исходной системе уравнений, описывающих рассматриваемый объект, строится двудольный граф $G = (U, V, E)$, где E – множество ребер. Множество вершин такого графа разбивается на два непересекающихся подмножества U и V, и он отражает связи между переменными, используемыми в математической модели, и уравнениями, в которые эти переменные входят.

Далее для каждой конкретной переменной определяется совокупность тех уравнений, из которых она может быть выражена. Это выполняется путем определения в образованном двудольном графе максимального паросочетания, то есть максимально мощного множества ребер, обладающих тем свойством, что каждая вершина графа инцидентна не более чем одному ребру.

После этого осуществляется разделение исходной системы уравнений на подсистемы, которые должны решаться совместно. В терминах теории графов это означает разделение двудольного графа на компоненты сильной связности. Основой алгоритма построения компонент сильной связности является поиск циклов в орграфе: множество вершин простого цикла принадлежит одной и той же сильно связной компоненте.

Проходя последовательно по вершинам двудольного графа, разделенного на компоненты сильной связности, от искомых переменных через вершины отношений к вершинам известных переменных, можно определить последовательность решения данной системы уравнений.

5. Пользователь вводит численные значения входных параметров, и система, используя построенное ею отображение φ_i, находит численные значения выходных параметров задачи.

Рассмотренный алгоритм решения задачи был реализован в программном комплексе, разработанном в среде программирования DELPHI [2]. В разработке комплекса участвовали студенты и аспиранты кафедры летательных аппаратов С.Г. Фомичёв, В.Б. Белкин, В.С. Алиппа, А.А. Якищик.

В качестве иллюстрации работы программного комплекса, рассмотрим работу его модуля, осуществляющего расчет проектных параметров силового гироскопического комплекса (СГК) космического аппарата наблюдения.

Система уравнений, описывающих массогабаритные характеристики силового гироскопического комплекса, имеет следующий вид [3]:

$$M_{КА} = J_{КА}\varepsilon_{КА};\ M_{СГП} = M_{КА};\ K_{СГП} = \frac{M_{СГП}}{\omega_{ПРЕЦ}};\ J_{РГП} = \frac{K_{СГП}}{\omega_{РГП}};\ R = \sqrt[5]{\frac{240 \cdot J_{РГП}}{113 \cdot \pi \cdot \rho_{РГП}}};$$

$$\omega_{пред} = \sqrt{\frac{[\sigma]}{f\rho R^2}};\ m_{РГП} = \frac{5}{6}\pi \cdot \rho_{РГП} \cdot R^3;\ m_{СГП} = \frac{m_{РГП}}{k_1};\ m_{ЭлСГП} = k_2 m_{СГП};$$

$$m_{СГ1} = m_{СГП} + m_{ЭлСГП};\ m_{СГК} = 4m_{СГ1},$$

где $J_{КА}$ – момент инерции космического аппарата; $\varepsilon_{КА}$ – максимальное угловое ускорение КА; $M_{КА}$ – максимальное значение момента силы, необходимого для поворота КА с максимальным ускорением; $M_{СГП}$ – момент силового гироскопического прибора; $\omega_{ПРЕЦ}$ - допустимая угловая скорость прецессии СГП; $K_{СГП}$ – потребный кинетический момент ротора гироприбора; $\omega_{РГП}$ – максимально возможная угловая скорость ротора гироприбора; $J_{РГП}$ – потребный момент инерции ротора гироприбора; f – коэффициент безопасности; ρ – плотность материала ротора; R – радиус ротора гироприбора; $\omega_{пред}$ – предельная угловая скорость вращения ротора гироприбора; $[\sigma]$ – допустимое напряжение в роторе; $m_{СГП}$ – масса силового гироприбора; k_1 – доля массы ротора СГП от массы СГП; k_2 – доля массы электроники СГП от массы СГП; $m_{ЭлСГП}$ – масса электроники СГП; $m_{СГ1}$ – масса одного СГП с блоком электроники; $m_{СГК}$ – масса СГК.

Входными параметрами определим момент инерции КА и угловое ускорение КА, выходным параметром - массу силового гироскопического комплекса.

Введем новые формализованные переменные:

$u_1 = J_{КА};\ u_2 = \varepsilon_{КА};\ u_3 = M_{КА};\ u_4 = M_{СГП};\ u_5 = K_{СГП};$

$u_6 = \omega_{ПРЕЦ};\ u_7 = \omega_{РГП};\ u_8 = J_{РГП};\ u_9 = [\sigma];\ u_{10} = \rho;$

$u_{11} = R;\ u_{12} = m_{СГП};\ u_{12} = m_{СГП};\ u_{13} = m_{РГП};\ u_{14} = k_1;$

$u_{15} = m_{ЭлСГП};\ u_{16} = k_2;\ u_{17} = m_{СГ1};\ u_{18} = m_{СГК}.$

В процессе работы проблемно-ориентированного программного обеспечения для выбора основных проектных характеристик космических аппаратов наблюдения последовательно (самой системой) осуществляется построение графов, необходимых для получения искомой последовательности решений, и выдается эта последовательность в автоматизированном режиме.

На рис. 1 показано диалоговое окно программного модуля, на котором представлена последовательность решения уравнений для нахождения массы силового гироскопического комплекса (выходной переменной u_{18}).

Для верификации программного обеспечения проводилась проверка правильности автоматизированного построения соответствующих графов. Для этого сравнивались результаты построения графов, построенных вручную и полученных в автоматизированном режиме.

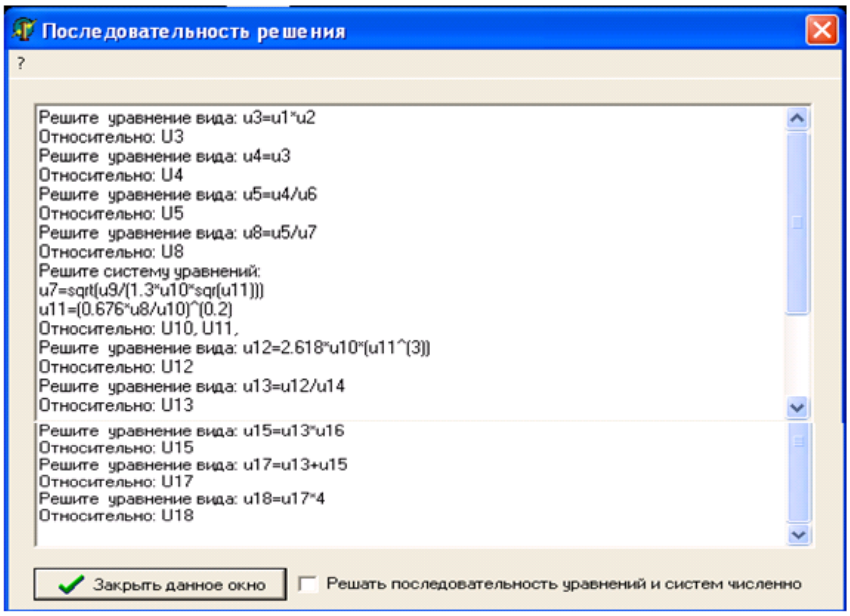


Рис. 1. Диалоговое окно выдачи результатов решения

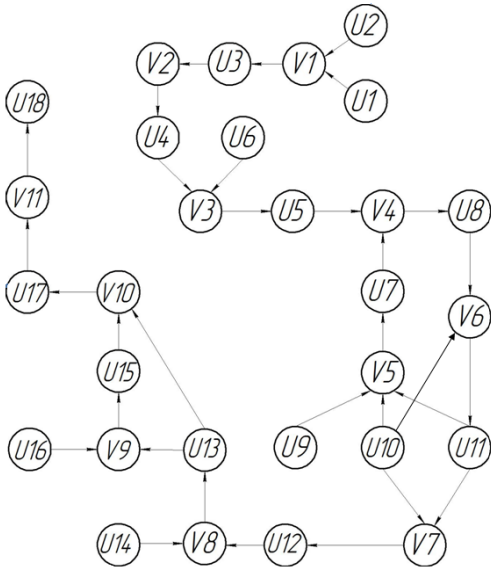


Рис. 2. Двудольный граф связей уравнений и переменных

В частности, на рис. 2 показан исходный двудольный граф, построенный в ручном режиме по приведенной системе уравнений и формализованным переменным.

Результаты сравнения показали, что соответствующие графы совпадают.

Таким образом, в данной работе представлены логика работы и результаты разработки проблемно-ориентированного программного обеспечения, предназначенного для выбора основных проектных характеристик космических аппаратов наблюдения.

Литература

1. Друшляков Ю.И. Теоретические основы программирования: учебное пособие / Ю.И. Друшляков, И.В. Ежова. – М.: МАИ, 1986. – 60 с.

2. Фомичёв С.Г., Куренков В.И., Кучеров А.С. Разработка элементов проблемно-ориентированной системы автоматизированного проектирования / Труды XIII Всероссийского научно-технического семинара по навигации и управлению движением. – Самара, 2007. - С. 305 – 309.

3. Куренков В.И. Методика выбора основных проектных характеристик и конструктивного облика космических аппаратов наблюдения: учеб. пособие [Текст] / В. И. Куренков, В. В. Салмин, А. Г. Прохоров. - Самара: Изд-во Самар. гос. аэрокосм. ун-та, 2007. - 160 с.

Система моделирования вращательного движения микроспутника с магнитной системой ориентации в геомагнитном поле

А.С. Лебедев
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

При разработке микроспутника «Аист» (студенческий спутник СГАУ), требуется смоделировать функционирование его системы пространственной ориентации, обеспечивающей уменьшение начальной угловой скорости, полученной при отделении микроспутника от базового космического аппарата. Система автоматического управления (САУ) микроспутника имеет в качестве исполнительных органов магнитные катушки, которые взаимодействуя с геомагнитным полем, обеспечивают заданную ориентацию микроспутника на орбите. Для проектирования САУ микроспутника необходимо изучить особенности его вращательного движения при различных начальных условиях и алгоритмах управления.

Для решения этой задачи необходимо производить компьютерное моделирование вращательного движения микроспутника с учетом влияния гравитационного и магнитное полей, а также атмосферы планеты. Для решения задачи моделирования движения спутника необходимо решить две подзадачи: моделирование магнитного поля Земли (МПЗ) и реализация алгоритма управления.

Моделирование МПЗ

Сложность моделирования МПЗ заключается в его описании с учётом магнитных аномалий. Поэтому для моделирования МПЗ была использована модель IGRF (International Geomagnetic Reference Field, позволяющая вычислить магнитную индукцию поля в определенных координатах, с учетом источников поля, которые находятся внутри Земли. Модель основана на экспериментальных наблюдениях на геомагнитных обсерваториях, кораблях, самолетах и искусственных спутниках Земли. Математически данная модель может быть представлена в виде формул:

$$B_X(r,\lambda,\theta) = \sum_{n=1}^{L}\sum_{m=0}^{n}\left(\frac{R_\oplus}{r}\right)^{n+2}(I_n^m\cos m\lambda + i_i^m\sin m\lambda)\,P_n^m(\cos\theta),$$

$$B_Y(r,\lambda,\theta) = \sum_{n=1}^{L}\sum_{m=0}^{n}\left(\frac{R_\oplus}{r}\right)^{n+2}(I_n^m\sin m\lambda - i_i^m\cos m\lambda)\,\frac{mP_n^m(\cos\theta)}{\sin\theta},$$

$$B_Z(r,\lambda,\theta) = -\sum_{n=1}^{L}\sum_{m=0}^{n}\left(\frac{R_\oplus}{r}\right)^{n+2}(I_n^m\cos m\lambda + i_i^m\sin m\lambda)\,(n+1)P_n^m(\cos\theta).$$

Здесь λ и θ – географические долгота и дополнение к широте точки наблюдения, r – расстояние от точки наблюдения до центра Земли, I_n^m и i_i^m – постоянные коэффициенты, характеризующие внутренние источники МПЗ, $P_n^m(\cos\theta)$ – квазинормированный по Шмидту присоединенный полином Лежандра первого рода [1].

Выбор алгоритма управления

Система ориентации включает магнитные катушки, создающие вокруг спутника собственное магнитное поле, которое взаимодействует с МПЗ и создает необходимые управляющие моменты. Управляя током в катушках можно добиться обеспечения заданного уровня угловых микроскоростей и микроускорений на борту спутника.

Первый алгоритм управления использует следующее выражения для вычисления магнитного момента катушек:

$$L = -a\frac{dB}{dt}, \quad (1)$$

где B – вектор магнитной индукции МПЗ, a – положительный коэффициент, подбираемый экспериментально.

Для проведения моделирования будем использовать параметры, имеющие приведенные ниже значения (см. таблицу 1).

Таблица 1 .Параметры моделирования

Долгота восходящего узла (λ_Ω), град	60,3
Наклонение орбиты (l), град	62,8
Аргумент перигея (ω), град	50,5
Высота орбиты (r), км	570
Коэффициент a	$5 \cdot 10^6$

Полученная зависимость угловой скорости вращения спутника от времени представлена на рис. 1.

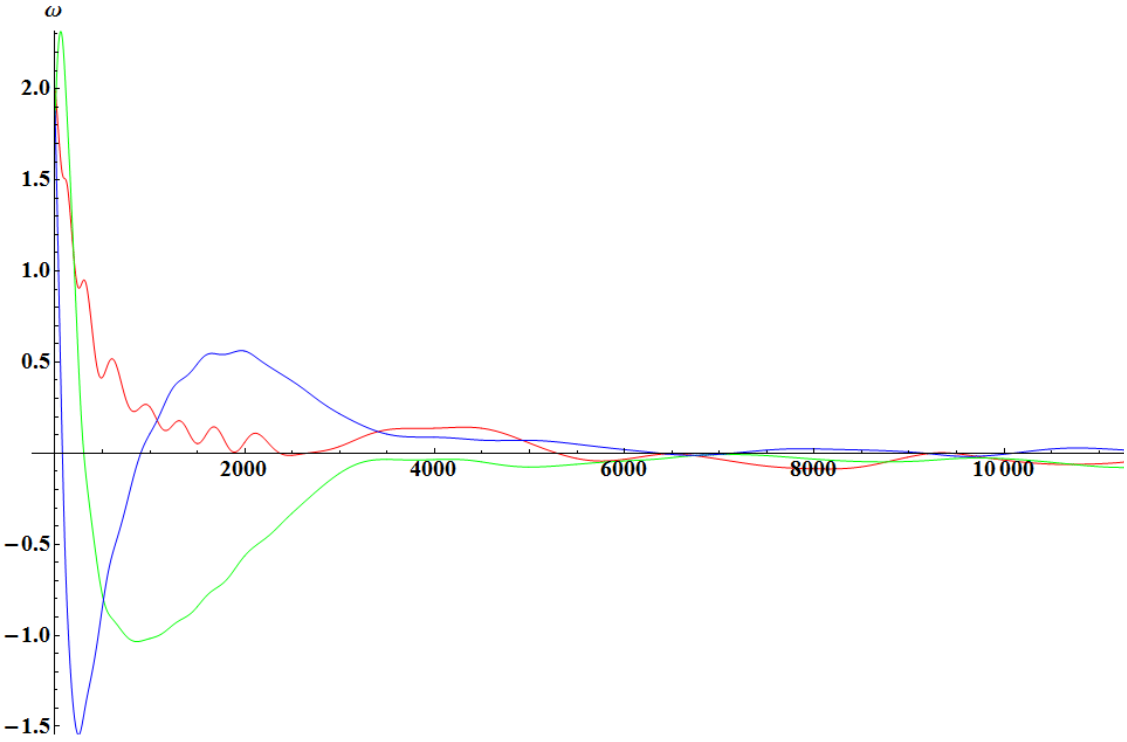


Рис. 1. Уменьшение угловой скорости при первом алгоритме управления

Первый алгоритм управления обеспечивает значительное (в десятки раз) уменьшение угловой скорости, однако, данный результат можно улучшить, применив более общее выражения для нахождения магнитного момента микроспутника:

$$L = a[\vec{K}, \vec{B}], \quad (2)$$

Данное выражение позволяет достичь лучшего результаты, по сравнению с первым алгоритмом, использующим выражение (1), что представлено на рисунке 2.

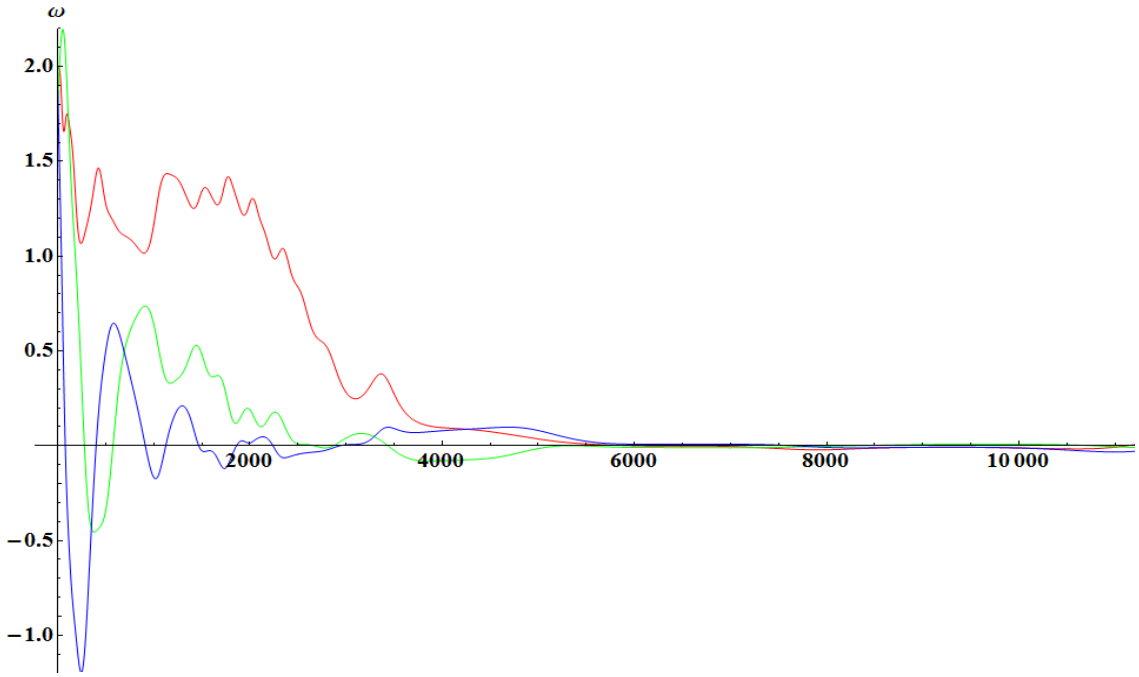


Рис. 2. Уменьшение угловой скорости при втором алгоритме управления

Из данного графика видно, что в результате не удалось значительно уменьшить угловые скорости. Данную проблему можно решить, усложнив закон нахождения магнитного момента (2):

$$L = \begin{cases} a[\vec{K}, \vec{B}], |\vec{\omega}| \geq \varepsilon \\ \frac{a}{\sqrt{2}}[\vec{K}, \vec{B}], |\vec{\omega}| < \varepsilon \end{cases} \quad (3)$$

Здесь ε – пороговое значение модуля вектора полной угловой скорости микроспутника $\vec{\omega}$, при достижении которого необходимо снизить уровень управляющего момента для того, чтобы избежать перерегулирования системы. Для моделирования примем $\varepsilon = 0.04$ рад/с. Результат моделирования представлен на рис. 3.

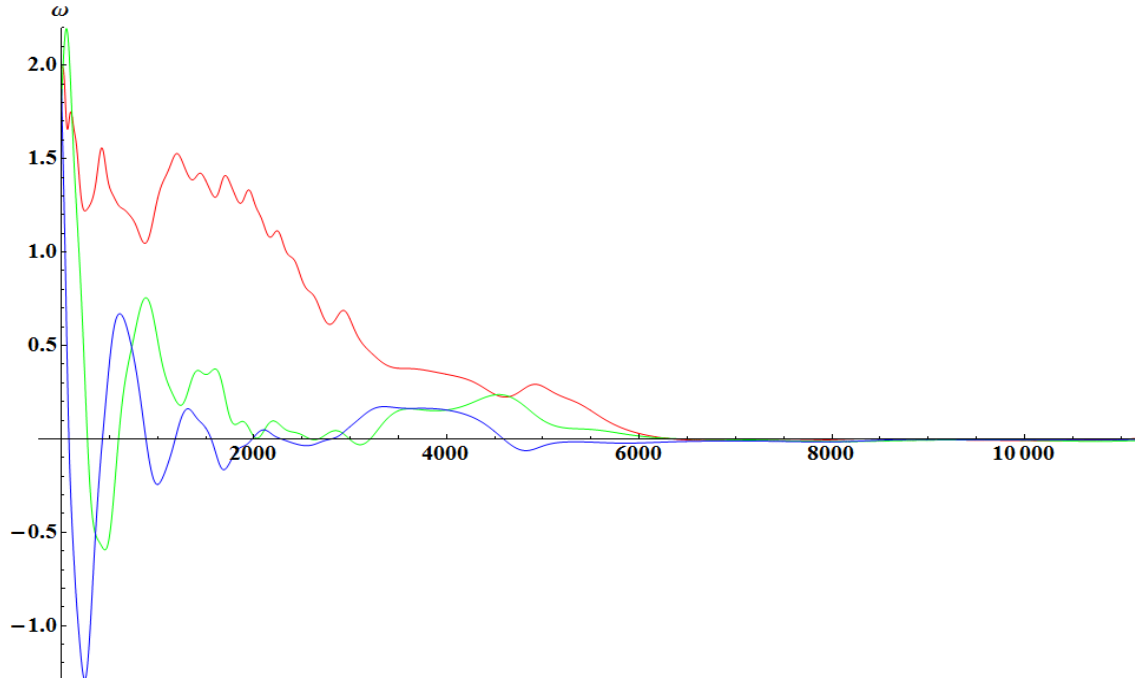


Рис.3. Гашение угловой скорости при законе управления (3)

Выводы

В результате проведения численного моделирования можно сделать вывод о наибольшей эффективности третьего алгоритма, которые использует выражение (3) для вычисления магнитного момента и обеспечивает наилучшее уменьшение начальных угловых скоростей микроспутника.

Для проведения моделирования был разработан программный продукт, позволяющий проводить численное моделирование при широком диапазоне начальных условий и произвольных законах управления. Результаты моделирования могут быть представлены в виде графиков, трехмерной анимации движения спутника или экспортированы в файл в CSV формате.

Литература

Коваленко, А.П. Магнитные системы управления космическими летательными аппаратами [Текст]/А.П. Коваленко. – М.: Машиностроение, 1975. – 248 с.

Об одной системе автоматического управления ориентацией солнечных батарей микроспутника на источник излучения

В.В. Любимов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В современных спутниках для обеспечения требуемой пространственной ориентации применяются различные исполнительные органы [1]. Наиболее универсальным и точным средством управления ориентацией являются микроЖРД. Однако, микроЖРД не применимы для спутников с большим сроком функционирования, поскольку без дополнительной заправки они имеют ограниченный запас ракетного топлива. Для микро и наноспутников, имеющих время функционирования на орбите несколько лет, перспективным является создание систем автоматического управления (САУ), использующих взаимодействие магнитного поля катушек с геомагнитным полем. Такая САУ имеет относительно малые массу (порядка 10 кг.) и энергоёмкость (несколько Вт.). Рассмотрим подробно функционирование такой системы управления.

Пусть микроспутник имеет на борту три одноосных магнитометра, служащих для определения величин индукции геомагнитного поля, располагающихся вдоль осей инерции. Также на микроспутнике вдоль осей инерции установлены три электромагнитных катушки. Кроме того, на спутнике жёстко закреплены один солнечных датчик и несколько солнечных батарей, имеющих нормали к рабочим плоскостям, которые параллельны продольной оси симметрии микроспутника OZ. Будем считать, что после отделения от базового космического аппарата микроспутник приобрел малую угловую скорость, которая слабо влияет на обеспечение ориентации солнечных батарей на Солнце. Ограничимся рассмотрением случая, когда датчик Солнца всё время движения непрерывно функционирует. На практике требуется учитывать случаи, когда датчик не видит Солнце, например, по причине захода спутника в тень Земли. Эти случаи требуют отдельного рассмотрения и в данной работе не рассматриваются.

При разработке САУ ориентацией солнечных батарей на источник излучения требуется смоделировать поведение вектора направления на Солнце в процессе полёта микроспутника по орбите. Для задания вектора направления на Солнце $\vec{S}$ в системе координат, связанной с солнечным датчиком требуется, в начале определить его положение в инерциальной системе координат $O_1X_1Y_1Z_1$ (O_1-центр масс Земли, O_1Z_1-направлена по оси вращения Земли, O_1X_1-направлена в восходящий узел орбиты и лежит в плоскости земного экватора). Координаты вектора $\vec{S}$ в системе координат $O_1X_1Y_1Z_1$ таковы: $\vec{S} = (r\cos\alpha_1, r\sin\alpha_1\cos\varepsilon, r\sin\alpha_1\sin\varepsilon)$, где $r = R_з + h$, $R_з$ - средний радиус Земли, h – высота орбиты; $\alpha_1 = \omega_o t$, ω_o - угловая скорость вращения спутника по круговой орбите, t - время полёта по орбите; ε - угол наклона эклиптики. Затем вектор $\vec{S}$ переопределяется, посредством известных преобразований [2], от инерциальной системы координат $O_1X_1Y_1Z_1$ к связанной с солнечным датчиком системой координат $O_2x_2y_2z_2$ оси которой параллельны соответствующим осям связанной системы координат O*xyz*. В результате, как показано на рисунке 1, он приобретает следующий вид: $\vec{S} = (\cos\gamma_1, \cos\gamma_2, \cos\gamma_3)$. На рис.1 в точке O_2 луч Солнца проникает внутрь датчика и в точке C он отображается на ПЗС-матрице ABCD. Заключительным шагом в определении положения вектора $\vec{S}$ является нахождение углов, выдаваемых солнечным датчиком (α, β). Эти углы связаны с координатами вектора $\vec{S}$ следующим образом: $tg\alpha = -\cos\gamma_2 / \cos\gamma_3$, $tg\beta = -\cos\gamma_1 / \cos\gamma_3$.

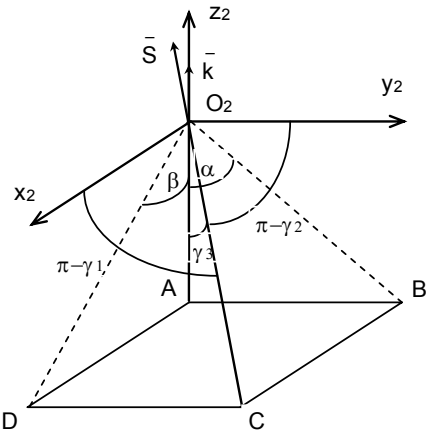


Рис.1. Направление вектора $\vec{S}$ в солнечном датчике

Угловые скорости в связанной системе координат микроспутника O*xyz* будем определять из динамических уравнений Эйлера:

$$I_x \frac{d\omega_x}{dt} + (I_z - I_y)\omega_y\omega_z = M_{xm} + M_{xp},$$
$$I_y \frac{d\omega_y}{dt} + (I_x - I_z)\omega_x\omega_z = M_{ym} + M_{yp}, \qquad (1)$$
$$I_z \frac{d\omega_z}{dt} + (I_y - I_x)\omega_x\omega_y = M_{zm} + M_{zp}.$$

Здесь I_x, I_y, I_z – главные центральные моменты инерции микроспутника; $\omega_x, \omega_y, \omega_z$ – угловые скорости микроспутника в связанной системе координат; M_{xm}, M_{ym}, M_{zm} – управляющие моменты от магнитных исполнительных органов; M_{xp}, M_{yp}, M_{zp} – возмущающие моменты различной природы.

Динамические уравнения (1) в совокупности с кинематическими соотношениями, например, кинематическими уравнениями Эйлера [2] полностью определяют вращательное движение микроспутника. В уравнениях (1) управляющие моменты M_{xm}, M_{ym}, M_{zm} имеют следующий вид: $M_{xm} = L_y B_z - L_z B_y$, $M_{ym} = L_z B_x - L_x B_z$, $M_{zm} = L_x B_y - L_y B_x$. Вектор магнитной индукции геомагнитного поля $\vec{B} = (B_x, B_y, B_z)$ определяется при помощи трех магнитометров и известен во всех точках траектории. Вектор магнитного момента микроспутника $\vec{L} = (L_x, L_y, L_z)$ фактически задаёт закон управления ориентацией спутника.

Для обеспечения ориентации солнечных батарей на источник солнечного излучения требуется переориентировать ось O*z* микроспутника в направлении вектора $\vec{S}$. Данную задачу можно успешно решить, если выбрать, например, следующий выражение для магнитного момента микроспутника: $\vec{L} = (L_x = 0, L_y = 0, L_z = c(\Delta\vec{S} \cdot [\vec{k} \times \vec{B}])$, где c - корректирующий коэффициент, $\Delta\vec{S} = \vec{S} - \vec{k}$, $\Delta\vec{S} = (\cos\gamma_1, \cos\gamma_2, \cos\gamma_3 - 1)$, $\vec{k} = (0,0,1)$ - орт оси O*z*. Аналогичное выражение для вектора магнитного момента $\vec{L}$ использовалось в работе [2]. В отличие от указанной работы будем использовать вектор $\Delta\vec{S} = \vec{S} - \vec{k}$, а не расхождение между текущим и номинальными значениями кинетического момента, что объясняется применением в нашем случае динамических уравнений Эйлера, а не уравнений Белецкого-Черноусько. Раскрывая смешанное произведение векторов находим выражение для третьей ненулевой составляющей магнитного момента $L_z = B_x \cos\gamma_2 - B_y \cos\gamma_1$. Следовательно, при совпадении единичных векторов $\vec{S}$ и $\vec{k}$:

$\gamma_1 = \gamma_2 = \pm\pi/2$, т.е. $L_z = 0$. В результате учета составляющих магнитного момента $L_x = 0, L_y = 0,\ L_z = B_x \cos\gamma_2 - B_y \cos\gamma_1$ в выражениях для механических моментов M_{xm}, M_{ym}, M_{zm} получим с учетом различных знаков углов (α, β) управляющие моменты в виде:

$$M_{xm} = -sgn(\alpha)|B_x||B_y||\cos\gamma_2| - sgn(\beta)B_y^2|\cos\gamma_1|,$$
$$M_{ym} = -sgn(\beta)|B_x||B_y||\cos\gamma_1| - sgn(\alpha)B_x^2|\cos\gamma_2|, \qquad (2)$$
$$M_{zm} = 0.$$

Механические моменты (2) обеспечивают независимое управление по двум каналам вращательного движения. В процессе функционирования САУ требуется знать зависимость моментов M_{xm}, M_{ym}, M_{zm} от углов, выдаваемых солнечным датчиком (α, β). Для этого применяются следующие выражения (3):

$$|\cos\gamma_1| = \sqrt{\frac{tg^2\beta}{1 + tg^2\alpha + tg^2\beta}},\ |\cos\gamma_2| = \sqrt{\frac{tg^2\alpha}{1 + tg^2\alpha + tg^2\beta}} \qquad (3)$$

Ниже на рис. 2 представлена структурная схема работы САУ, составленная для случая положительных углов α, β. В качестве номинальных значений углов, выдаваемых датчиком Солнца, в данной САУ приняты величины $\alpha_н = \beta_н = 1$, которые соответствуют погрешностям датчика. Начальные значения углов α, β также определяемые датчиком Солнца на схеме обозначаются следующим образом: $\alpha 0, \beta 0$.

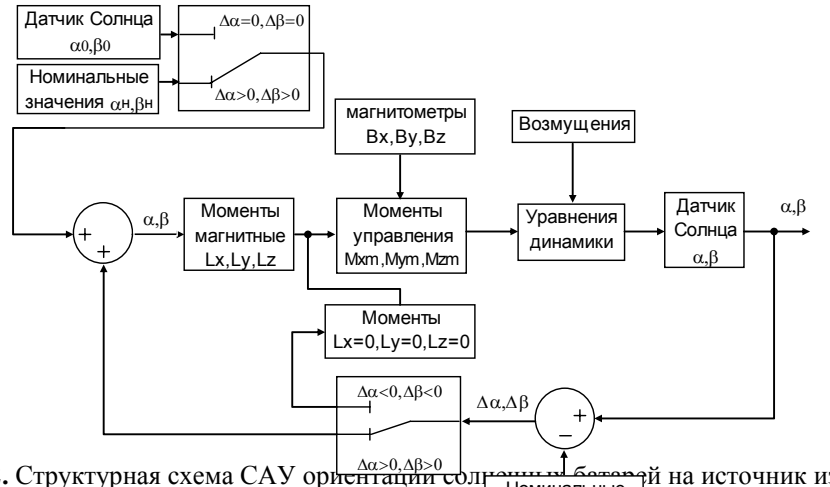


Рис. 2. Структурная схема САУ ориентации солнечных батарей на источник излучения

Таким образом, представленная система автоматического управления решает задачу по обеспечению ориентации солнечных батарей, жестко закрепленных на одной из плоскостей микроспутника, на источник солнечного излучения в процессе его длительного полета по круговой околоземной орбите.

Литература

1. Коваленко, А.П. Магнитные системы управления космическими летательными аппаратами [Текст] / А.П. Коваленко. - М.: Машиностроение, 1975. – 248 с.

2. Овчинников, М.Ю. Ориентирование спутника-гироскопа магнитной системы управления в инерциальном пространстве [Текст] / М.Ю. Овчинников, В.И. Пеньков, Д.С. Ролдугин. - М.: ИПМ им. М.В. Келдыша РАН, 2010, – 28 с. – Препринт. – № 27.

Определение оптимальных мест установки демпферов для снижения вибраций оборудования космических аппаратов на основе конечно-элементного моделирования

К.В. Пересыпкин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

На этапе орбитального полета ряд агрегатов, присутствующих на КА, вызывают вибрации конструкции КА. Это агрегаты с вращающимися частями: вентиляторы; насосы системы терморегулирования; силовые гироскопы. Нормальное функционирование ряда приборов КА может быть нарушено небольшими по амплитуде вибрациями. Примером таких приборов могут служить линзы и зеркала телескопов, колебания которых могут привести к смазу изображения. Другой пример – узконаправленные антенны, колебания которых по угловым степеням свободы способны нарушить связь. Еще один пример – экспериментальные микрогравитационные установки, ускорения которых при вибрациях нарушают условия эксперимента. Колебания датчиков системы ориентации КА также могут препятствовать нормальной работе этой системы. В связи с этим при проектировании КА может оказаться необходимым снизить амплитуды вибраций вышеперечисленных приборов до приемлемых величин.

Рассмотрим возбуждение как силы, обусловленные дисбалансом роторов источников вибрации. Эти силы изменяются по гармоническому закону во времени с частотой, принадлежащей диапазону частот возбуждения.

Решение задачи снижения амплитуд высокочастотных вибраций на борту КА подразумевает выбор метода моделирования вибраций конструкции КА, выбор технических средств, применение которых позволит снизить амплитуды вибраций конструкции и выбор метода поиска такого способа применения этих средств, чтобы удовлетворить требованиям по вибрациям.

В колебаниях КА, вызванных высокочастотным возбуждением участвуют не только податливые элементы конструкции, такие как солнечные батареи и антенны, но и упругие колебания корпуса КА имеют значительные амплитуды. Корпус КА является сложной упругой системой и аналитически найти для него формы вынужденных колебаний затруднительно и остается моделировать его колебания численно. Широкое распространение получили два метода численного моделирования вибраций упругих систем: метод конечных элементов (МКЭ) [2,3,4] и статистический энергетический анализ (СЭА) [5,6]. СЭА не дает подробной картины вибраций и предназначен скорее для анализа вибраций от случайного возбуждения. Его главное достоинство – малые вычислительные затраты. Параметры модели, построенной с помощью СЭА, получают или из натурных испытаний или конечно-элементного расчета. МКЭ позволяет получить подробную картину вибраций, но его применение связано с большими вычислительными затратами. В этой работе используется МКЭ, что связано как с достоинствами этого метода, так и с техническими навыками автора.

Исследователи осуществляют снижение вибраций посредством ряда способов. Это влияние на вибрации с помощью изменения жесткостей элементов конструкции [4], гашение вибраций с помощью пьезоэлектрических актуаторов [2,7], увеличение демпфирования с помощью различных энергопоглощающих материалов, таких как пенистые материалы [8], материалы с памятью формы [9] и слоистые оболочки с вязкими слоями [10].

Для поиска конструкции, приемлемой с точки зрения требований к вибрациям, ряд исследователей применяют различные методы оптимизации. Например, в работе [2] используется градиентный метод оптимизации, а в работе [11] – генетический алгоритм оптимизации.

Задача снижения амплитуд высокочастотных вибраций на борту КА в этой работе решается с применением численной оптимизации градиентным методом. Критерием оптимизации примем минимизацию некоторой целевой функции, являющейся характеристикой амплитуд вибраций мест установки чувствительных к вибрациям приборов или чувствительных элементов (ЧЭ). Определение амплитуд вибраций ЧЭ находится с помощью решения уравнения анализа вынужденных гармонических колебаний:

$$\left[-\omega^2 M + i\omega \cdot C + K\right] \cdot \{u\} = \{P\}, \qquad (1)$$

где M, C и K матрицы масс, демпфирования и жесткости конструкции КА, составленные с помощью метода конечных элементов; ω - круговая частота возбуждения, а $\{P\}$ – вектор амплитуд узловых сил; $\{u\}$ – вектор амплитуд вынужденных колебаний. В этой работе как для моделирования вынужденных колебаний, так и для организации оптимизационного расчета применяется конечно-элементная система MSC.Nastran.

В этой работе в качестве проектных переменных рассматриваются значения характеристик демпфирования, связанного с установкой демпферов (далее – дополнительного демпфирования). Технически дополнительное демпфирование в конечно-элементной модели задавалось в виде коэффициентов демпфирования для материалов элементов конструкции.

Если не ограничить алгоритм оптимизации в возможностях по внесению в конструкцию дополнительного демпфирования, то демпфирование будет увеличено для всех элементов конструкции. Это позволит снизить амплитуду вибраций ЧЭ, однако цель исследования – поиск наиболее эффективного размещения демпферов с точки зрения снижения вибраций – не будет достигнута. Следовательно, возможности по внесению в конструкцию дополнительного демпфирования должны быть ограничены. Данное ограничение предлагается записать в следующем виде :

$$\sum_i \gamma_i \cdot m_i \le D, \qquad (2)$$

где γ_i – коэффициент дополнительного демпфирования i-ого материала элемента конструкции; m_i – масса материала элемента конструкции; D – постоянная величина, ограничивающая демпфирующие возможности конструкции, которые обусловлены дополнительными демпфирующими элементами. Величина D должна выбираться исходя из приемлемого увеличения массы конструкции, связанной с установкой демпферов, а также с весовой эффективностью используемых демпферов.

При выборе целевой функции учитывались следующие соображения:

1. В каждой задаче по снижению вибраций можно выразить негативное влияние вибраций в виде функции от амплитуд вибраций конструкции, *A(u)*;
2. Для простоты полагаем, что частота возбуждения подчиняется плоскому закону распределения вероятностей внутри частотного диапазона возбуждения и вибрации с амплитудами, превышающими допустимые, одинаково опасны для всех частот этого диапазона;
3. В ходе оптимизации необходимо понизить максимальное значение *A(u)* для вибраций с любой частотой из частотного диапазона возбуждения;
4. Амплитуды колебаний отыскиваются численно и могут быть найдены для ограниченного количества частот возбуждения.

Эти соображения учитываются в целевой функции следующего вида:

$$F = \sqrt{\sum_j \left(A(u_j)\right)^2}\ , \qquad (3)$$

где $\{u\}_j$ – вектор амплитуд колебаний при частоте возбуждения f_j. Частоты возбуждения равномерно заполняют диапазон возбуждения. Равномерность распределения f_j соответствует плоскому закону распределения вероятности для частоты возбуждения внутри диапазона.

Рассматривается сумма квадратов значений A для разных частот, а не простая их сумма, для того чтобы наибольшее изменение целевой функции достигалось при изменении наибольших значений A среди всех f_j.

Считаем, что упругость и масса демпферов мала по сравнению с массой и упругостью конструкции. Тогда уравнение (1) принимает вид:

$$\left[-\omega^2 M + i\omega \cdot (C_{констр} + C_{доп}) + K\right] \cdot \{u\} = \{P\}, \qquad (4)$$

где $C_{констр}$ – матрица естественного демпфирования конструкции; $C_{доп}$ – матрица дополнительного демпфирования, коэффициенты которой являются проектными переменными.

Задача численной оптимизации мест установки демпферов для снижения амплитуд высокочастотных вибраций формулируется следующим образом:

1. Критерий оптимизации – минимизация целевой функции (3);
2. Проектные переменные – коэффициенты демпфирования материалов элементов конструкции;
3. Ограничения:

– “Количество“ дополнительного демпфирования не должно превышать максимально допустимого значения (2);

– Коэффициенты демпфирования больше нуля и меньше максимальной величины, определяемой природой устанавливаемых демпферов.

Предложенный метод применен к конструкции реального КА. Требовалось понизить амплитуды вибраций датчиков системы ориентации. Функция A выбрана в виде суммы амплитуд угловых перемещений трех датчиков для рассматриваемого случая нагружения. Обшивки и шпангоуты были разбиты на зоны, для каждой из которых дополнительное демпфирование постоянно и управляется отдельной проектной переменной. Шпангоуты разбиты на зоны по окружному направлению, а обшивки – по окружному и продольному направлениям. Исходное распределение дополнительного демпфирования равномерное – значения всех проектных переменных равны 0,0163. Кроме того, в конструкции действует собственное демпфирование, не обусловленное дополнительными демпферами и характеризующееся коэффициентом демпфирования $\gamma_{собств} = 0{,}0127$. Отметим, что значение коэффициента демпфирования $\gamma = 0{,}0163$ соответствует 5% уменьшению амплитуды колебаний за один период, а $\gamma = 0{,}0127$ – 4% уменьшению.

Проведен оптимизационный расчет для возбуждения от гидроблока. Максимальное “количество” демпфирования D было принято равным значению D при начальных значениях γ, т.е. в ходе оптимизации демпфирование не добавлялось, а перераспределялось. Для приближенной оценки насколько большим является дополнительное демпфирование по сравнению с демпфированием уже имеющимся в конструкции, в таблице 1 приводится отношение “количества” дополнительного демпфирования к “количеству” естественного демпфирования конструкции. Эти отношения показывают, что энергия колебаний, рассеиваемая дополнительным и естественным демпфированием соизмеримы. Изменение целевой функции для этих расчетов показано в столбце 1 таблицы 1. Из этой таблицы видно, что внесение дополнительного демпфирования равномерно по всей силовой конструкции в данных

конкретных примерах снижает целевую функцию на ~20%. При оптимальном распределении дополнительного демпфирования, целевая функция уменьшилась примерно в два раза.

При уменьшении размеров зон с постоянным демпфированием, оптимальное распределение демпфирования может быть найдено более точно, что, вероятно, позволит добиться большего снижения амплитуд вибраций.

Таблица 1. Относительное снижение значения целевой функции для выполненных оптимизационных расчетов

Номер расчета		1	2
Случай нагружения		Гидроблок	
Частотный диапазон		90-115Гц	
Ограничения	Отношение максимального допустимого “количества” дополнительного демпфирования к “количеству” естественного демпфирования конструкции	0.27	0.81
	Максимально допустимый коэффициент дополнительного демпфирования	0.1	0.3
Отношение значения целевой функции для равномерного распределения дополнительного демпфирования к значению целевой функции при отсутствии дополнительного демпфирования		0.783	0.783
Отношение значения целевой функции для оптимального распределения демпфирования к значению целевой функции при отсутствии дополнительного демпфирования		0.515	0.304

На рис. 1 показаны зависимости функции A от частоты для оптимального и равномерного распределений дополнительного демпфирования и при его отсутствии. Сравнение зависимостей показывает, что уменьшение вибраций при переходе от равномерного к оптимальному распределениям происходит главным образом в областях резонансных пиков. Целевая функция рассматриваемого вида была выбрана для достижения выборочного снижения максимальных амплитуд вибраций на частотном диапазоне возбуждения, и зависимости на рис. 1 показывают что этот результат достигнут.



Рис. 1. Зависимости функции A от частоты для оптимального и равномерного распределений дополнительного демпфирования и при его отсутствии.

С целью добиться дальнейшего снижения амплитуд вибраций датчиков были проведен оптимизационный расчеты для больших величин дополнительного демпфирования (см. столбец 2 таблицы 1). Результаты расчета показали, что при

внесении утроенного дополнительного демпфирования происходит снижение целевой функции в три раза для возбуждения от гидроблока.

Выводы

1. Предложен метод определения оптимальных мест установки демпферов с целью снижения амплитуд высокочастотных вибраций на борту КА
2. Метод аппробирован при снижении вибраций датчиков системы управления реального КА
3. Внесение в конструкцию КА дополнительного демпфирования в соответствии с найденным оптимальным распределением позволило снизить амплитуду вибраций датчиков в несколько раз (от двух до пяти для разных расчетов)

Литература

1. Пересыпкин К.В. Оценка влияния отклонений параметров конструкции космических аппаратов на амплитуды вибраций/ Сб. трудов XIV Всероссийского семинара «Управление движением и навигация ЛА», Самара 2009г.
2. Тет32. Favorable locations for piezo actuators in plates with good control effectiveness and minimal change in system dynamics. Dhuri K.D., Seshy P. Smart Mater. and Struct. 2007 16 №6, cnh/ 2526-2542
3. Тет13 Vibroacoustic study of scientific satellites relying on a time domain FEN/BEN method. Jollet D., Giuliano M., Loderean P., Van der Laan T. Proceedings of European Conference on Spacecraft Structures, Materials and Mechanical Testing, Noordwijk, 2005. Noordwijk, ESTEC 2005, стр. 874-880
4. Автоопт. Design sensitivity analysis of structure-induced noise and vibration. Choi K.K., Shim I., Wang S. J. of Vibration and Acoustics, 1997, Vol. 119, стр 173-179
5. Тет14. The prediction of random acoustic vibration of equipment mounted on honeycomb panel. Ando S., Shi Q., Nagahama K., Saitoh M., Saegasa H., Proceedings of European Conference on Spacecraft Structures, Materials and Mechanical Testing, Noordwijk, 2005. Noordwijk, ESTEC 2005, стр. 881-886
6. Тет15. Astudy on vibro-acoustic analysis system for spacecraft design. Sasano H., YoshidaK., Seko H., Nakamura S., Onzuka A., Akiyama J., Murakami K., Shi Q., Tsushihashi M., Nagahama K., Saegusa H. Proceedings of European Conference on Spacecraft Structures, Materials and Mechanical Testing, Noordwijk, 2005. Noordwijk, ESTEC 2005, стр.887-895
7. Тет30. The optimal location of piezoelectric actuators and sensors flor vibration control of plates. Kumar K.Ramesh, Narayanan S. Smart Mater. and Struct. 2007 16 №6 стр.2680-2691
8. Тет31. Damping of a thin-walled honeycomb structure using energy absorbing foam. Woody Sane C., Smith Stuart T. J. Sound and Vibration 2006 № 1-2, Стр.491-502
9. Тет35. Shape memory elastic foundation and supports for passive vibration control of composite plates. Birman Victor. Int. J. Solids and Struct. 2008. 45 №1 стр. 320-335
10. Тет33. Vibration damping using four-layer sandwich. Yadav Binod P. J. Sound and Vibr. 2008. 317, №3-5 стр.576-690
11. Тет39. Optimal design of hysteretic dampers connecting adjacent structures using multi-objective genetic algorithm and stochastic linearization method. Ok Seung-Yong, Song Junho, Park Kwan-Soon. Eng. Struct. 2008, 30, №5, стр.1240-1249
12. Тимошенко, С.П., Колебания в инженерном деле [Текст]/ Тимошенко С.П., Янг Д.Х., Уивер У., пер. с англ. – М., Машиностроение, 1985 – 472 с.

Исследование деформации боковых блоков ракет-носителей при падении в воду с использованием компьютерных конечно-элементных технологий

М.А. Петровичев[1], В.К. Дуплихин[2], Н.Т. Каргин[1], А.М. Петровичев[1]
1 - Самарский государственный аэрокосмический университет,
2- ФГУП ГНП РКЦ «ЦСКБ-Прогресс», Самара

Вопрос устранения последствий пуска РН является важнейшей экологической проблемой. Особенно важно, если отработанные части РН приводняются в международных водах, где по истечении 48 часов после пуска не должно быть никаких следов.

В данной работе рассматривается вопрос естественного затопления боковых блоков (ББ) ракеты-носителя (РН) в результате деформации конструкции в процессе проникания ее в воду.

После отделения бокового блока (штатная ситуация) он движется по инерции до некоторой высоты. После этого начинается спуск с разгоном до высоты ~ 50 км. Кроме поступательного движения боковой блок совершает вращательное движение.

После входа отработанных блоков в плотные слои атмосферы происходит их торможение до дозвуковых скоростей. Особенность ракетных блоков состоит в том, что они состоят их двух основных блоков горючего и окислителя и хвостового отсека, где сосредоточены двигатели и обеспечивающая их работу аппаратура. Поэтому центр масс у ракетных блоков располагается близко к хвостовому отсеку. Это приводит к тому, что в атмосфере ББ движутся хвостовым отсеком вперед под некоторым углом. Касание воды бокового блока происходит хвостовым отсеком.

Такая особенность движения бокового блока значительно осложняет решение задач движения бокового блока в воздухе и воде. Эти задачи были решены экспериментально – аналитическими методами. В результате были получены нагрузки, действующие на конструкцию ББ в процессе проникания в воду.

Решение задачи деформирования конструкции производилось с использованием конечно-элементных технологий.

Общая конечно-элементная модель ББ предназначенная для исследования общей прочности ББ состоит из 9700 узлов и 11500 конечных элементов.

Для моделирования обечаек отсеков и баков а также днищ баков использовались четырехугольные изопараметрические мембранно-изгибные пластинчатые конечные элементы первого порядка h- типа – CQUAD4 . Элементы данного типа можно использовать как для линейных задач статики, так и для исследования геометрически и физически нелинейного поведения конструкции, в том числе и задач устойчивости.

Для моделирования стержневых элементов конструкции таких как стрингеры, стыковые и промежуточные шпангоуты отсеков и баков применяются двухузловые балочные элементы CBEAM, использование которых возможно как в линейных так и нелинейных задачах.

Конечно-элементная модель бака горючего предназначенная для исследования общей прочности ББ состоит из 5127 узлов и 5025 конечных элементов.

Для моделирования обечаек и днищ баков используются четырехугольные изопараметрические мембранно-изгибные пластинчатые конечные элементы первого порядка h- типа – CQUAD4 . Элементы данного типа можно использовать как для линейных задач статики, так и для исследования геометрически и физически нелинейного поведения конструкции, в том числе и задач устойчивости.

Для моделирования, стыковых и промежуточных шпангоутов применяются двухузловые балочные элементы CBEAM, использование которых возможно как в линейных так и нелинейных задачах.

Конечно-элементная модель бака окислителя предназначенная для исследования общей прочности ББ состоит из 4660 узлов и 4580 конечных элементов.

Для моделирования обечаек и днищ баков используются четырехугольные изопараметрические мембранно-изгибные пластинчатые конечные элементы первого порядка h- типа – CQUAD4 . Элементы данного типа можно использовать как для линейных задач статики, так и для исследования геометрически и физически нелинейного поведения конструкции, в том числе и задач устойчивости. Для моделирования, стыковых и промежуточных шпангоутов применяются двухузловые балочные элементы CBEAM, использование которых возможно как в линейных так и нелинейных задачах.

Конечно-элементная модель хвостового отсека, предназначенная для исследования общей прочности ББ состоит из 1554 узлов и 1780 конечных элементов.

Для моделирования обечаек ХО используются четырехугольные изопараметрические мембранно-изгибные пластинчатые конечные элементы первого порядка h- типа – CQUAD4 . Элементы данного типа можно использовать как для линейных задач статики, так и для исследования геометрически и физически нелинейного поведения конструкции, в том числе и задач устойчивости. Для моделирования, стыковых и промежуточных шпангоутов применяются двухузловые балочные элементы CBEAM, использование которых возможно как в линейных так и нелинейных задачах.

В работе рассматривается деформация днища бака горючего под действием удара воды и деформация обечайки бака горючего под действием инерционных сил от частей, расположенных выше днища бака горючего.

В процессе удара о воду на нижнее днище бака горючего действует внешнее избыточное давление, достигающее максимального значения.

Для определения верхних критических нагрузок потери устойчивости конструкции использовалась структурироованная процедура решений SOL 105 (Buckling) – решение задачи устойчивости в линейной постановке. Получены первые три формы потери устойчивости нижнего днища от действия внешнего избыточного давления.

Результаты расчетов закритического деформирования (post-buckling) нижнего днища от действия внешнего избыточного давления в нелинейной постановке получены с помощью структуированной процедуры решения SOL 106 (Nonlinear Static).

Поскольку материал днища бака в процессе изготовления приобретает некоторый наклеп (отсутствует заключительный отжиг) рассмотрены два варианта материала для днища: сплав алюминиевый АМг6-М – отожженный и сплав алюминиевый АМг6-Н – нагартованный.

В расчетах использовалась упругопластическая модель материала (Elastic-plastic) с линейным упрочнением.

Для случая материала АМг6-М процесс развивается следующим образом: при некотором давлении происходит потеря устойчивости сферической стенки шпангоута, далее шпангоут «заваливается» из своей плоскости; затем происходит потеря устойчивости стенки днища в районе шпангоута, которая плавно переходит в процесс «перекатывания» выпуклого сферического днища в вогнутое. В процессе этих деформаций напряжения в днище достигают значений порядка предела прочности и , следовательно, происходит разрушение стенки днища. Максимальные перемещения стенки днища достигают значений порядка 650 мм.

Для случая материала АМг6-Н процесс развивается по другому: при некотором давлении происходит потеря устойчивости сферической стенки в центральной зоне днища, затем происходит плавное «перекатывание» выпуклого сферического днища в вогнутое; далее развивается «заваливание» шпангоута из своей плоскости. При этом напряжения в днище достигают значений предела прочности и, следовательно, происходит разрушение стенки днища. Максимальные перемещения стенки днища достигают значений 550 мм.

Первые три формы потери устойчивости конической обечайки бака горючего от действия осевого сжатия получены для соответствующих верхних критических нагрузок.

Результаты расчетов закритического деформирования (post-buckling) конической обечайки бака горючего в зоне стыка конус-цилиндр в районе нижнего днища полученные с помощью структуированной процедуры решения SOL 106 в нелинейной постановке (Nonlinear Static).

В расчетах использовалась упругопластическая модель материала (Elasto-Plastic) с линейным упрочнением. Рассмотрены расчеты отдельных этапов процесса закритического деформирования конической обечайки бака горючего после потери устойчивости. В процессе этих деформаций напряжения в стенке обечайки достигают значений порядка предела прочности и, следовательно, происходит разрушение стенки обечайки.

Конечно-элементные расчеты показали, что в процессе погружения бокового блока в воду происходит деформация с разрушением днища и обечайки бака горючего, то-есть произойдет затопление бокового блока.

Программный комплекс для решения задач оптимизации перелетов между эллиптическими орбитами с использованием двигателей большой и малой тяги

К.В. Петрухина
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Разработан специализированный программный комплекс, позволяющий производить проектно-баллистические расчеты и на их основе строить траекторию пространственного движения, а также производить формирование проектного облика космического аппарата (КА).

Структурно программный комплекс состоит из шести модулей (рисунок 1):

- модуль расчёта движения КА с химическим разгонным блоком;
- модуль расчёта движения КА с электрореактивной двигательной установкой (ЭРДУ);
- модуль расчёта продолжительности пребывания КА с ЭРДУ в радиационных поясах Земли;
- модуль расчёта продолжительности пребывания КА с ЭРДУ в тени Земли и построения областей равной продолжительности теневых участков;
- модуль формирования проектного облика КА;
- модуль трехмерного моделирования межорбитального перехода.

Программа написана на языке Object Pascal в среде программирования Delphi 7 с использованием программного интерфейса (API) для разработки графических приложений OpenGL и системы твердотельного моделирования Solid Works 2004.

В программном комплексе реализуются принципы COM-технологий, позволяющие передавать расчетные данные между различными программами. В частности, расчет проектно-баллистических параметров ведется в программе, составленной в среде программирования Borland Delphi 7, построение трехмерной модели движения с использованием программного интерфейса OpenGL, а формирование проектного облика космического аппарата в системе твердотельного моделирования Solid Works 2004 с применением параметризации.

Такой подход позволяет минимизировать затраты на исследования рациональных схем выведения, что имеет важное практическое приложение на ранних этапах проектирования, когда возникает необходимость анализа множества альтернативных вариантов.

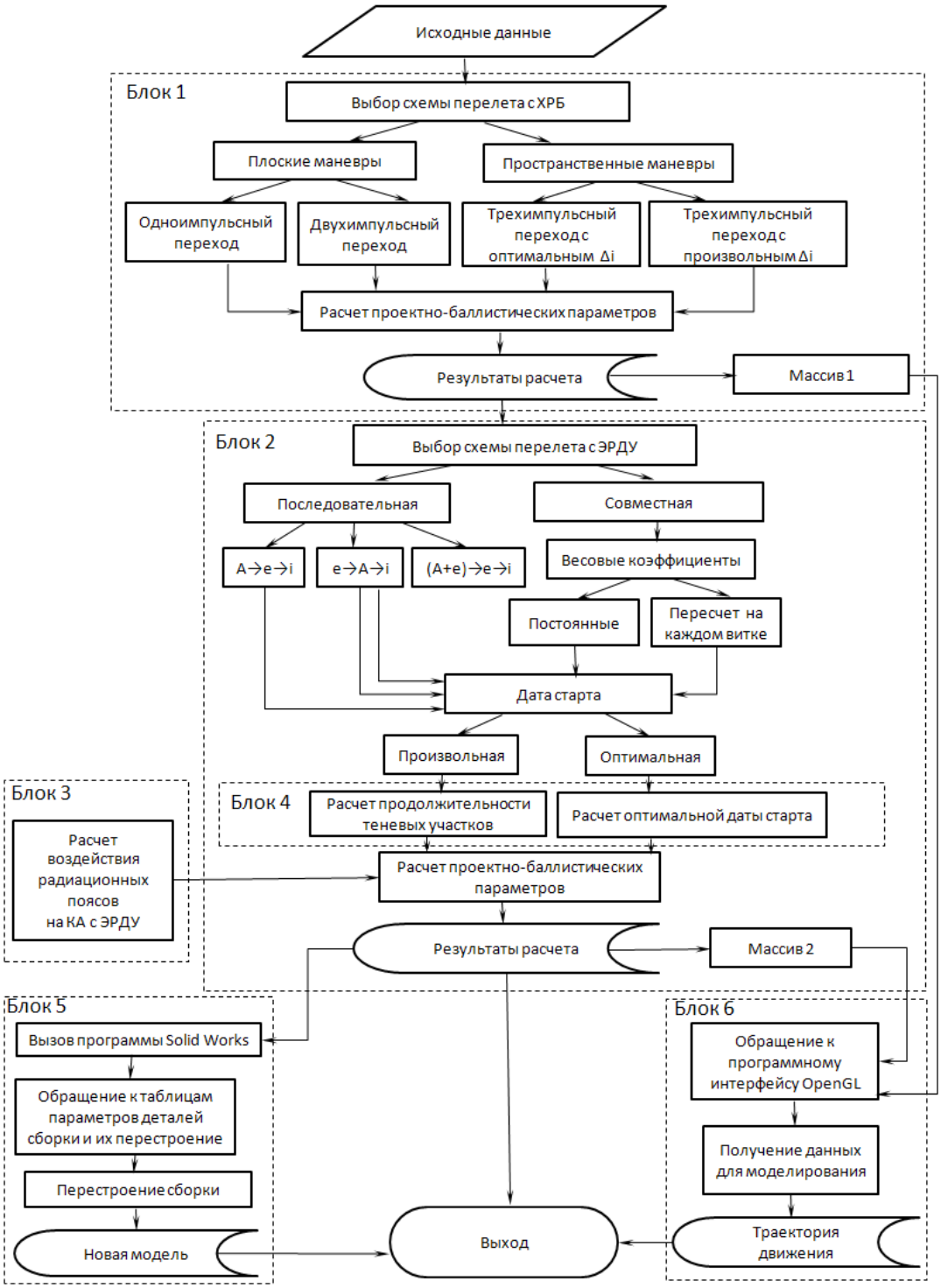


Рис. 1. Структура программного комплекса

Автоматизация расчета проектных характеристик энергодвигательного электроракетного модуля низкоорбитального космического аппарата

В.В. Салмин, В.В. Волоцуев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Для получения эффективных решений задачи проектирования низкоорбитальных космических аппаратов с энергодвигательным электроракетным модулем используется системный подход, в рамках которого проводится моделирование для решения задач анализа и синтеза. Процесс моделирования включает в себя: постановку задачи с выбором целей и критериев поиска решения, разработку математических моделей, численные вычисления, анализ полученных решений и формулировку рекомендаций. Проектное решение по формированию облика низкоорбитальных космического аппарата с энергодвигательным электроракетным модулем описывается относительно широким массивом проектных характеристик. Для уменьшения времени численных вычислений в процессах моделирования удобным является использование специализированного пакета программного обеспечения для проектирования низкоорбитальных космических аппаратов с энергодвигательным электроракетным модулем.

Энергодвигательный электроракетный модуль включает в свой состав: электроракетную двигательную установку; подсистему энергопитания; подсистему обеспечения теплового режима; элементы конструкции модуля. Указанные бортовые подсистемы важно рассматривать как единый модуль, так как параметры элктроракетной двигательной установки будут влиять на алгоритм функционирования, массовые, геометрические, энергетические и тепловые характеристики проектируемого космического аппарата в целом.

Для расчета проектных характеристик низкоорбитального космического аппарата с энергодвигательным электроракетным модулем предлагается использовать способы и методы системного подхода, которые предполагают следующие мероприятия:

- определение целей разрабатываемого низкоорбитального космического аппарата и выбор параметров бортовой аппаратуры;
- учет факторов внешней среды, характерных для низкой орбиты;
- выбор параметров электроракетной двигательной установки, удовлетворяющей требованиям функционирования целевой аппаратуры космического аппарата;
- расчет параметров системы энергопитания, удовлетворяющей требованиям со стороны режима функционирования целевой аппаратуры, режима работы электроракетной двигательной установки, режима работы остальных систем; с учетом факторов внешней среды;
- расчет параметров системы обеспечения теплового режима удовлетворяющей требованиям по соблюдению теплового режима всех бортовых систем.

При выполнении вышеперечисленных мероприятий осуществляется обработка широкого массива данных: численных величин значений проектных характеристик бортовых систем низкоорбитального космического аппарата и электроракетной двигательной установки, численных величин значений параметров внешней среды,

логических величин, характеризующих возможность синтеза проектной схемы космического аппарата и т.п.

Множество математических моделей для расчета соответствующих проектных характеристик или параметров внешней среды описываются сложными системами математических выражений, что, соответственно, затрудняет и замедляет процесс вычисления необходимых для проектирования параметров.

Для автоматизации процесса расчета проектных параметров энергодвигательного электроракетного модуля низкоорбитального космического аппарата разрабатывается пакет специального программного обеспечения расчета проектных характеристик и составления эскизного.

На первом этапе разработаны специальные программы, позволяющие моделировать процесс функционирования низкоорбитального космического аппарата с электроракетной двигательной установкой и проводить выбор оптимальных характеристик по заданным критериям.

На втором этапе предполагается разработать программное обеспечение, позволяющее формировать массив значений проектных характеристик, необходимый для составления эскизного проекта низкоорбитального космического аппарата с энергодвигательным электроракетным модулем. В основу данного программного обеспечения предполагается заложить современные CAD/CAM/CAE технологии проектирования.

Разработанное программное обеспечение позволит автоматизировать процесс разработки эскизных проектов низкоорбитальных космических аппаратов с энергодвигательным электроракетным модулем и может найти применение в проектно-конструкторских организациях ракетно-космической техники.

Программный комплекс для выбора оптимальных проектно-баллистических параметров многоразового межорбитального транспортного аппарата с электрореактивной двигательной установкой

В.В. Салмин, А.С. Четвериков
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В современном мире информационные технологии проникли во все сферы деятельности человека. Без использования информационных и компьютерных технологий сегодня невозможно представить проектирование сложных технических систем, особенно таких как перспективные космические системы. К таким системам относится многоразовый межорбитальный транспортный аппарат (МТА) с электрореактивной двигательной установкой (ЭРДУ), который предназначен для доставки различных полезных нагрузок на удаленные орбиты с возвращением МТА на исходную орбиту. Продолжительность активного участка межорбитального перелета составляет 150-300 суток.

Здесь рассматривается специализированный программный комплекс, разработанный для выбора оптимальных проектных и баллистических параметров многоразового межорбитального транспортного аппарата с электрореактивной двигательной установкой.

При оптимизации проектно-баллистических характеристик МТА возникает проблема взаимосвязи решений параметрической и динамической задач[1].

Динамическая задача – отыскание оптимальной программы управления вектором тяги, обеспечивающей минимальные затраты рабочего тела на прямой и обратный перелеты для заданных параметров исходной и рабочей орбиты и длительности движения. Мера этих затрат называется динамической характеристикой. Этой мерой могут быть характеристическая скорость или моторное время.

Параметрическая задача – это задача выбора оптимальных проектных параметров МТА, обеспечивающих при заданной динамической характеристике перелета максимум полезного груза.

Каждая из этих задач решается в рассматриваемом программном комплексе.

Задачи баллистической и проектной оптимизации могут решаться с использованием моделей различной степени точности. Обычно на начальном этапе оптимизации МТА представляется точкой переменной массы с «идеальным» и «бесплатным» управлением вектора тяги, на последующих этапах используются более сложные модели, учитывающие угловое движение МТА, дополнительные ограничения на проектные параметры, траектории управления, затраты рабочего тела на коррекцию траектории и реализацию заданного перелета[2].

В разработанном программном обеспечении реализованы релейная программа управления вектором тяги, т.е. «идеальная», и программы управления, позволяющие решить проблему ограниченности управляющего момента:

- программа управления с программными разворотами, но с начальным участком разгона;
- программа управления с начальным участком разгона и разворотом на заключительном этапе.

Кроме того, предусмотрено построение приближенного проектного облика МТА.

Таким образом, программный комплекс условно можно разделить на следующие составные блоки:

- оптимизатор проектно-баллистических параметров;
- блок баллистического моделирования;
- блок расчета геометрических характеристик и формирования облика МТА.

На рис. 1 представлена блок-схема работы программного комплекса.



Рис. 1. Блок-схема алгоритма работы программного комплекса

Рассмотрим назначение каждого блока более подробно.

Оптимизатор проектно-баллистических характеристик реализует алгоритм оптимизации проектно-баллистических характеристик МТА по заданным критериям качества и с учетом заданных ограничений (рис. 2). В качестве исходных данных используется время перелета, параметры начальной и конечной орбит, а также в зависимости от выбора режима расчета либо стартовая масса МТА на базовой орбите (зависит от используемой ракеты носителя), либо масса полезной нагрузки. При оптимизации используется математическая модель массы МТА[2]:

$$M_0 = M_{ПН} + M_Э + M_{РТ} + M_{СПХ} + M_К + M_Д,$$

где $M_Э$ - масса энергоустановки, $M_{РТ}$ - масса рабочего тела, $M_{СПХ}$ - масса системы подачи и хранения рабочего тела (баки, трубопроводы и пр.), $M_К$ - масса конструкции, $M_Д$ - масса двигательной установки.

Программный комплекс может производить расчет параметров в нескольких режимах в зависимости от выбора типа перелета и критерия оптимизации. Тип перелета может быть следующим:

- доставка полезного груза $M_{ПГ}$ на целевого орбиту и возвращение МТА на базовую орбиту;
- доставка полезного груза $M_{ПГ1}$ на целевую орбиту и возвращение буксира с некоторой полезной нагрузкой $M_{ПГ2}$ на базовую орбиту.

Критерии оптимизации:

– максимальная полезная нагрузка при фиксированной стартовой массе;

– минимальная стартовая масса при фиксированной полезной нагрузке.

Расчет проектных параметров МТА производится как для «идеально регулируемого», так и для существующих современных ЭРД. При этом используется база российских электрореактивных двигателей, разработанных в ОКБ «Факел» и в центре имени Келдыша. При расчете параметров МТА с учетом характеристик существующих двигателей для каждого из них производится алгоритм подбора потребного количества двигателей. При этом учитывается ресурс двигателя. Если двигатель не удовлетворяет по ресурсу, то расчет количества двигателей и проектных параметров МТА не осуществляется. Также предусмотрен расчет проектных параметров для отдельно выбранного двигателя (двигатель выбирается из базы вручную).

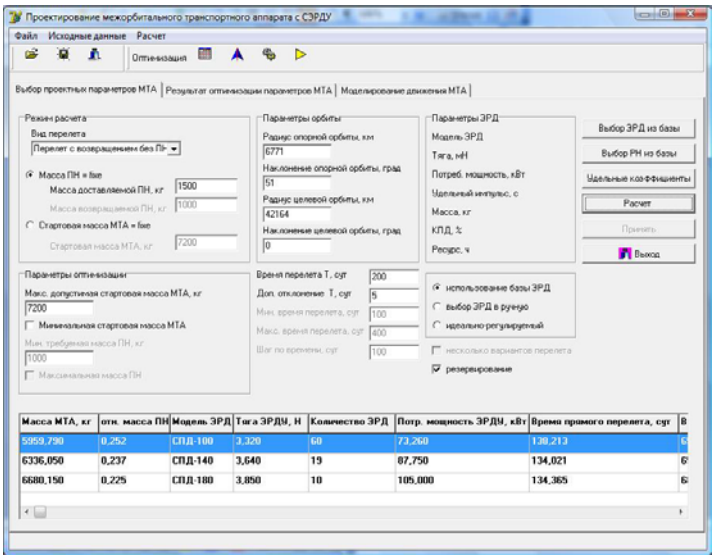
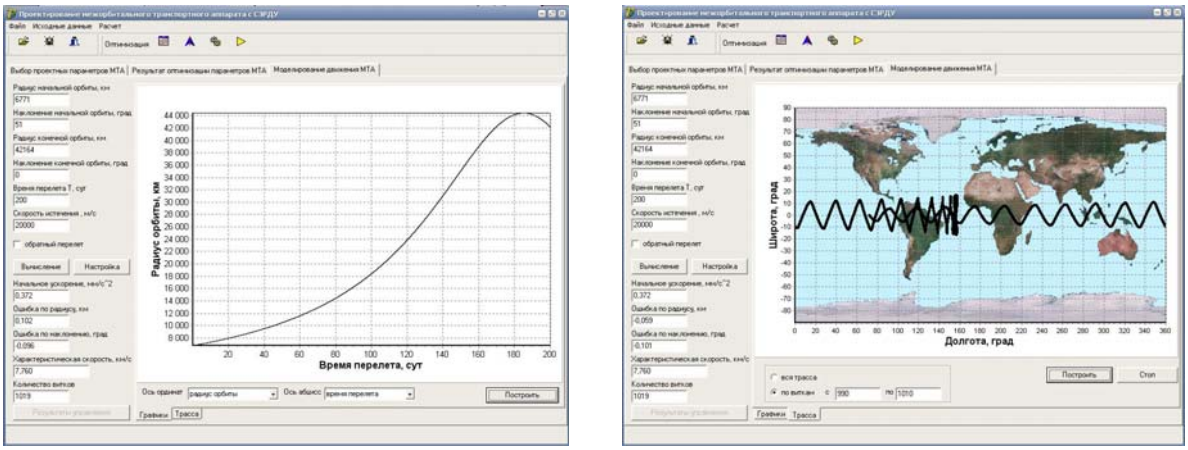


Рис 2. Блок оптимизации проектно-баллистических параметров МТА

Блок баллистического моделирования реализует моделирование процесса полета МТА с учетом возмущающих сил, действующих на МТА, позволяет получить «реальную» траекторию полета, зависимости траекторных параметров (рис. 3, а). Моделирование движения позволяет оценить точность решения задачи на этапе оптимизации. Кроме того, в состав блока входит программный модуль для построения трассы полета космического аппарата (рис. 3, б).



а) б)

Рис. 3. Блок баллистического моделирования

Блок расчета геометрических характеристик представляет собой алгоритмическую реализацию математических моделей связи проектных параметров систем МТА с их геометрическими параметрами. В этом блоке также осуществляется общая компоновка МТА, для чего предусмотрен модуль связи с SolidWorks (рис. 4).

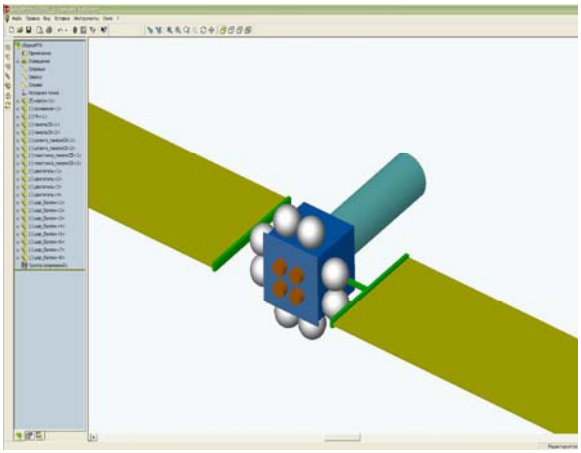


Рис. 4. Проектный облик МТА, построенный в SolidWorks

Разработанный программный комплекс позволяет существенно снизить трудоемкость проектных работ при создании перспективной космической транспортной системы, принять оптимальное решение при выборе баллистических схем перелета и проектного облика космического аппарата.

Литература

1. Салмин, В.В. Оптимизация космических перелетов с малой тягой : Проблемы совместного управления траекторным и угловым движением [Текст] – М.: Машиностроение, 1987 – 208 с.

2. Салмин, В.В. Выбор основных проектных характеристик и конструктивного облика межорбитальных аппаратов с электрореактивными двигательными установками с использованием системы Solid Works [Текст]: учеб. Пособие /В.В. Салмин, С.А. Ишков, О.Л. Старинова – Самара: Изд-во Самар. гос. аэрокосм. ун-та, 2006 – 82 с.

Мультиагентная система для исследования методов взаимодействия космических аппаратов дистанционного зондирования Земли

П.О. Скобелев[1], А.В. Соллогуб[2], А.В. Иващенко[3],
Е.В. Симонова[3], М.Е. Степанов[3], А.В. Царев[3]
[1] *Институт проблем управления сложными системами РАН, Самара;*
[2] *ГНПРКЦ «ЦСКБ-Прогресс», Самара;*
[3] *Научно-производственная компания «Разумные решения», Самара*

Введение

С развитием новых информационных технологий, появлением первых автономных роботов и ростом интеллектуальных возможностей машин всё большее значение приобретает проблема управления коллективами программных систем или роботов, автономно функционирующих в реальном времени в непрерывном цикле восприятия информации из внешней среды, планирования своих действий и контроля их исполнения. Особую сложность данная проблема получает при управлении коллективами подвижных объектов в режиме реального времени.

Для решения рассматриваемой проблемы необходимо обеспечить согласованное взаимодействие подвижных объектов, позволяющее управлять группой мобильных ресурсов с помощью технологий семантического веба для представления знаний, позволяющих принимать согласованные решения в ходе планирования, и мультиагентной технологии [1, 2], активно развивающейся в последние годы на стыке работ по объектному программированию, параллельным системам, искусственному интеллекту и телекоммуникациям. Данный подход состоит в решении задачи согласованного взаимодействия космических аппаратов (КА) дистанционного зондирования Земли (ДЗЗ). Предлагается описание базовых принципов организации взаимодействия таких КА и результатов разработки экспериментального образца мультиагентной системы моделирования взаимодействия КА.

Авторы выражают благодарность Российскому фонду фундаментальных исследований за поддержку в рамках гранта №10-08-01015а.

1 Задача обеспечения согласованного взаимодействия КА ДЗЗ

Обеспечение дистанционного зондирования Земли [3] является весьма актуальной задачей. Области применения ДЗЗ включают обновление топографических карт, отражающих реальное состояние территорий, прогноз и контроль развития наводнений, оценку нанесенного ими ущерба, прогноз урожайности сельскохозяйственных культур, контроль состояния гидротехнических сооружений на каскадах водохранилищ, реальное местонахождение морских судов в той или иной акватории, отслеживание динамики и состояния рубок леса, природоохранный мониторинг, оценку ущерба от лесных пожаров и их последствий, мониторинг разливов нефти и движения нефтяного пятна, контроль несанкционированного строительства, прогноз погоды и мониторинг опасных природных явлений и т.п.

При этом для решения данных задач могут использоваться как различные носители (космические аппараты, самолеты, дирижабли), так и разные средства измерения (датчики) с диапазоном измеряемых электромагнитных волн от долей микрометра (видимое оптическое излучение) до метров (радиоволны). Кроме этого, могут применяться как пассивные методы ДЗЗ (использующие естественное отраженное или вторичное тепловое излучение объектов на поверхности Земли) так и активные (использующие вынужденное излучение объектов, инициированное искусственным источником направленного действия).

С целью повышения эффективности ДЗЗ в настоящее время ведутся работы в области создания методов сигнатурного зондирования Земли (СЗЗ), суть которого заключается в передаче потребителю либо всей информации о зондируемых объектах, либо ее части, которая связана с изменением состояния объектов относительно предшествующих моментов времени. При этом выделяется сигнатура – признак или набор признаков, по которым объект обнаруживается, опознается, классифицируется и идентифицируется.

Разные космические аппараты, которые могут использоваться для ДЗЗ / СЗЗ, как правило, имеют разное оснащение (различные датчики) и разные траектории движения, что накладывает определенные ограничения на время и стоимость решения частной задачи СЗЗ в каждом конкретном случае. Это особенно актуально при выделении сигнатур разных видов (геометрических, пространственно-временных, спектральных, энергетических, динамических или фрактальных), для чего может потребоваться использование разных датчиков, и, соответственно, разных космических аппаратов. Типовые задачи сигнатурного зондирования, для решения которых важно пространственное разрешение, включают обнаружение и распознавание класса, типа и конкретного образа объекта, выделение геометрических характеристик объекта (контуров, размеров, структуры, отдельных элементов), анализ состояния объекта.

Таким образом, задача обеспечения согласованного взаимодействия нескольких КА ДЗЗ /СЗЗ в режиме реального времени является актуальной. Такое взаимодействие может быть централизованным, под управлением с Земли, и распределенным, на основе обмена указаниями о выполняемых действиях между КА. Второй принцип представляется более эффективным с точки зрения временных затрат и адаптивности системы КА. Действительно, самоорганизация подвижных объектов на основе компенсационных взаимодействий даёт принципиально новые возможности и важные преимущества по сравнению с традиционными методами. Вместе с тем, такой подход требует высокой автономности бортовых систем сигнатурного зондирования, а также реализации новых принципов построения программного обеспечения этих систем на основе применения мультиагентных технологий.

Рассматриваемые спутники должны не только самостоятельно и независимо принимать решения, но при необходимости договариваться и гибко формировать команды (организации) различного целевого назначения, в том числе, и новые, изначально не заданные. При этом сигнал, неожиданно запускающий изменение планов всего роя таких спутников, может прийти как с Земли, так и от любого из этих спутников, например, от того, кто первым обнаружил опасный новый объект.

Работа новых систем такого рода, где сложный объект состоит из коллектива автономно функционирующих, но постоянно взаимодействующих частей, требует создания нового класса систем распределённого интеллекта, ключевую роль в которых играет самоорганизация и взаимодействие между подсистемами, обеспечивающее согласованное поведение всего коллектива подсистем в целом при решении поставленных задач. При этом для систем рассматриваемого класса важно обеспечить возможность интерактивного взаимодействия с пользователями, позволяя им не только ставить задачу, но и в любой момент изменять свои предпочтения и ограничения, а также стратегии принятия решений в реальном времени, причем непосредственно в ходе решения задач, адаптивно корректируя планы всех объектов в ходе реализации намеченных целей.

2 Мультиагентный подход к построению системы взаимодействия КА ДЗЗ

Так как задача сигнатурного зондирования является сложной, распределенной пространственно-временной задачей, для ее решения можно эффективно использовать мультиагентные технологии. Мультиагентная система состоит из автономных агентов,

способных воспринимать ситуацию, принимать решения и взаимодействовать с себе подобными. Знания, необходимые для такой системы, могут быть отделены от программного кода системы и храниться в онтологии, представляющей собой сеть понятий и отношений предметной области.

В данной задаче агенты могут быть сопоставлены наземному объекту зондирования, наземному комплексу экспериментальной обработки, тестирования и сопровождения бортовых автономных систем сигнатурного зондирования и КА (спутникам). Цели мультиагентной системы включают определение закономерности (сценария) поведения системы в целом на основе сценариев поведения отдельных агентов и их взаимодействия, а также целеполагание в процессе функционирования системы.

Научная новизна мультиагентного подхода состоит в том, что обеспечивается создание индивидуально планирующих свое поведение интеллектуальных планировщиков реального времени, способных работать как в группе, так и автономно, и поддерживается согласованность коллективного построения индивидуальных планов за счет своевременного выявления конфликтов между «эгоистичными» интересами отдельных подвижных объектов и «коллективными» интересами групп таких объектов.

В ходе выявления и разрешения конфликтов каждое событие может потребовать непредвиденных подвижек других работ или переброски задач с одного ресурса на другой, а в итоге, целой цепной волны изменений в согласованных расписаниях. Такой процесс может осуществляться путем автоматически проводимых переговоров, которые позволят компенсировать взаимные уступки в коллективе рассматриваемых объектов или лиц, принимающих решения, и гармонизировать противоречивые по определению интересы всех участников во имя достижения общей цели или решения общей задачи.

Использование онтологии в качестве базы знаний позволяет описывать особенности КА ДЗЗ и их возможности, а также особенности исследуемых объектов в форме семантической сети и применять эти знания для обеспечения эффективного взаимодействия КА ДЗЗ. Онтология, с одной стороны, обеспечивает устойчивое функционирование системы, а с другой, – развивается и пополняется в реальном масштабе времени новыми знаниями в целях повышения адаптивных свойств системы.

Можно выделить следующие основные задачи создания бортовых автономных систем сигнатурного зондирования (БАСЗЗ):

- обработка видеоданных на борту КА так чтобы, с одной стороны, сократить избыточность передаваемых на землю видеоданных, а с другой стороны, не потерять нужную информацию;
- обеспечение компромисса между требуемым спектральным и пространственным разрешением;
- создание бортовой базы знаний (онтологии), включающей: априорные спектральные характеристики объектов, фонов, помех для различных спектральных диапазонов, условий наблюдения, изменения параметров оптико-электронных систем и систем распознавания, методы, алгоритмы и критерии автоматической спектральной селекции объектов для различной фоно-целевой обстановки и условий наблюдения, минимально необходимый набор эталонных сигнатур, существенный для обнаружения, распознавания класса, типа, размеров, структуры, состояния объектов;
- разработка специализированного программного обеспечение БАССЗ, сокращение информационной избыточности и сжатие видеоданных, принятие решений по передаче видеоданных и их изменений в наземный комплекс приема и обработки информации;
- разработка и совершенствование приемников видеоданных (многоспектральных ОЭС, гиперспектрометров), бортовых спецвычислителей;

− развитие наземного комплекса экспериментальной обработки, тестирования и сопровождения БАССЗ;

− разработка бортового и наземного обучающих комплексов с элементами самокоррекции, самонастройки и самообучения.

3 Особенности мультиагентного взаимодействия КА ДЗЗ

Для исследования особенностей организации взаимодействия КА ДЗЗ (спутников) методом имитационного моделирования была разработана мультиагентная система, позволяющая моделировать рой спутников, их перемещение, факты слежения за целью и передаваемые сообщения. При проведении такого моделирования может быть учтено размещение спутников по орбитам, а также специализация спутников, распознающих тепловое, видимое и радиационное излучение.

При моделировании задачи слежения за множеством движущихся целей каждой цели был поставлен в соответствие спутник – управляющее звено (Шеф) в команде спутников (рабочих), следящих за данной целью. После назначения Шеф рассылает рою спутников информационное сообщение о том, что он является Шефом, а также сообщения с информацией о цели. Если спутник принимает предложение о слежении за целью, он в ответ посылает Шефу сообщение о принятии предложения, после чего Шеф добавляет его в свою базу рабочих.

Если спутник не принимает предложение, Шеф рассылает сообщение другим спутникам-Шефам (их местоположение известно, так как каждый из них оповестил другие спутники о своем назначении) с просьбой произвести перераспределение спутников-рабочих. Произведя перераспределение, каждый Шеф имеет в своем распоряжении примерно одинаковое количество рабочих спутников. С целью оптимизации занятости рабочих вводится система самостоятельного динамического перераспределения рабочих. Каждый рабочий через определенный период времени целенаправленно рассылает по цепи спутников сообщение другим Шефам с просьбой предоставить ему информацию о текущей цели, а затем после выбирает из них наиболее близкую цель.

Была исследована задача обнаружения «неложных» целей с использованием коммуникации между спутниками. Пусть имеется множество объектов, совпадающих по ряду характеристик с искомыми характеристиками, но отличных по другим характеристикам. Назовем такие объекты «ложными» целями. Соответственно, объекты, все характеристики которых совпадают с искомыми, назовем «неложными» целями. Используя разработанную систему можно продемонстрировать, каким образом коммуникация между спутниками позволяет сократить продолжительность поиска искомых объектов среди множества объектов, в котором содержатся «ложные» и «неложные» цели.

Чтобы сократить время, затрачиваемое на исследование «ложных» целей, в мультиагентной системе могут быть реализованы переговоры между агентами спутников, в ходе которых исследование цели прерывается после того, как один из спутников группы обнаружил несовпадение какой-либо характеристики цели с искомыми. После прекращения исследования как отброшенной «ложной» цели, так и полностью исследованной «неложной» цели, Шеф вместе со своими рабочими переходит под управление другого Шефа, чья цель еще не исследована.

Выводы

Проведенные исследования применимости мультиагентных технологий для решения задач дистанционного зондирования Земли позволяют сделать вывод о полезности построения распределенных систем и активного использования базы знаний (онтологии) в рамках поставленной задачи. Данные исследования являются основой разработки нового поколения распределенных интеллектуальных систем

коллективного управления подвижными объектами, которые позволят решать принципиально новые задачи согласованной роботов и существенно повысить эффективность использования мобильных ресурсов.

Литература

1. Виттих, В.А., Метод сопряженных взаимодействий для управления распределением ресурсов в реальном масштабе времени [Текст]/ В.А. Виттих, П.О. Скобелев. – М.: Автометрия. – 2009. – № 2. – с. 78-87

2. Андреев, М.В., Применение мультиагентных технологий динамического планирования персональных задач при организации коллективного взаимодействия в автоматизированных системах управления распределением ресурсов [Текст] / М.В. Андреев, А.В. Иващенко, Д.М. Мартышкин, П.О. Скобелев, Л.В. Уланова, А.В. Царев. – М.: Мехатроника. Автоматизация. Управление. – 2010. – № 7. – с. 21 – 27

3. Соллогуб, А.В., Космические аппараты систем зондирования поверхности Земли / А.В. Соллогуб, Г.П. Аншаков, В.В. Данилов – М.: Машиностроение. – 1993. – 367 с.

Компьютерная технология имитации и анимации углового движения информационного спутника с упругими панелями солнечных батарей

С.Е. Сомов
НИИ ПНМС СамГТУ, Самара

Введение

Колебания конструкции крупногабаритных космических аппаратов (КА) могут оказывать существенное влияние на их пространственное движение, особенно в режиме начального успокоения – после отделения от ракеты-носителя. Современные компьютерные технологии позволяют выполнить динамический анализ и видео-отображение деформаций элементов конструкции КА в процессе его пространственного движения, что очень полезно при проектировании и полетном сопровождении КА. В работе кратко представлено разработанное программное обеспечение для имитации и анимации углового движения упругих КА. Спутник связи *SESAT* (*S*iberia *E*urope *SAT*ellite) с крупногабаритными упругими панелями солнечных батарей (СБ) (рис. 1) разработан ОАО «Информационные спутниковые системы имени академика М.Ф. Решетнева» по контракту с Европейской организацией *Eutelsat* и выведен на геостационарную орбиту в апреле 2000г. Для этого КА в работе созданы необходимые модели пространственного движения при широтно-импульсной модуляции (ШИМ) управления реактивными двигателями, включая изгибно-крутильные колебания панелей СБ, приведены результаты анимации движения как корпуса КА, так и панелей СБ.



Рис. 1. Спутник связи *Sesat*

Моделирование движения деформируемой конструкции КА

Вопросам выбора расчетных динамических схем, методам получения и исследования уравнений движения деформируемых КА посвящено большое количество публикаций. Выбор модели деформируемого КА тесно связан со способом дискретизации его конструкции. При получении приближенных моделей движения упругих КА наиболее распространен метод Релея-Ритца-Галеркина – метод предполагаемых форм колебаний. В задачах построения динамических моделей КА нежесткой конструкции широко применяется метод конечных элементов (МКЭ), представляющий собой локализованный метод предполагаемых форм колебаний. Обладая несомненными достоинствами и развитым программным обеспечением (системы NASTRAN, ASKA, SAP-IV и др.), МКЭ порождает модели весьма высокой размерности, достигающей нескольких тысяч степеней свободы для сложных разветвленных конструкций КА. Применяемый подход основан на представлении упругих колебаний элементов конструкции в виде конечного числа тонов. Расчет форм колебаний выполняется на основе МКЭ с конденсацией (редукцией) по тонам колебаний, на ЭВМ вычисляются также матрицы коэффициентов взаимовлияния движений всех подконструкций как абсолютно твердых, включая корпус КА, так и деформируемых тел. Например, расчетная схема одного крыла панелей СБ КА Sesat (рис. 2) представляется конечно-элементной моделью, состоящей из 129 узловых точек,

из которых 33 являются сосредоточенными массами, и 205 балок с 5 различными геометрическими и 2 различными физическими свойствами [1].

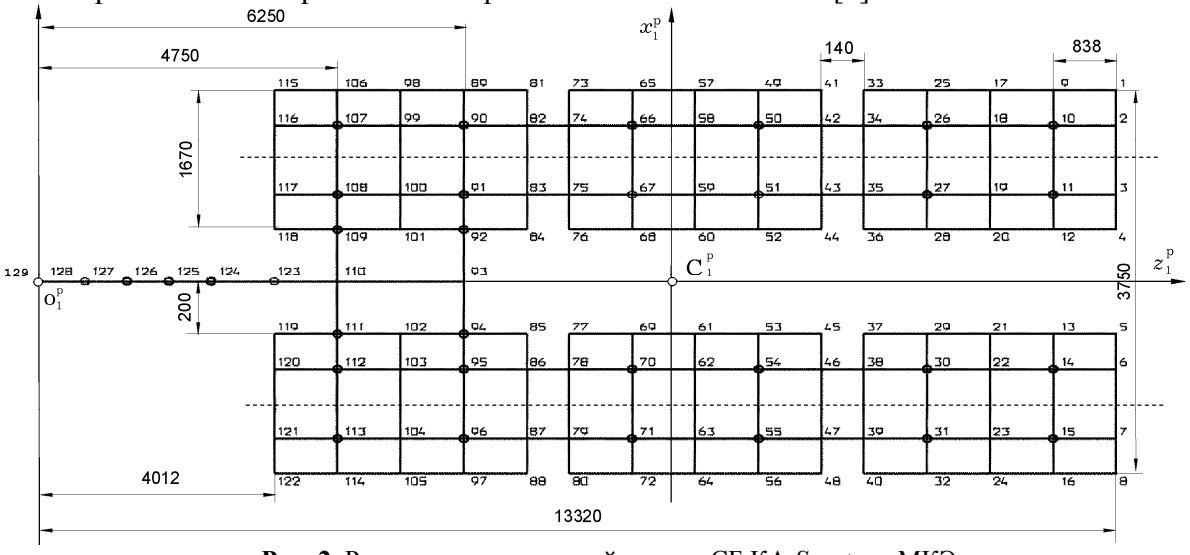


Рис. 2. Расчетная схема первой панели СБ КА *Sesat* по МКЭ

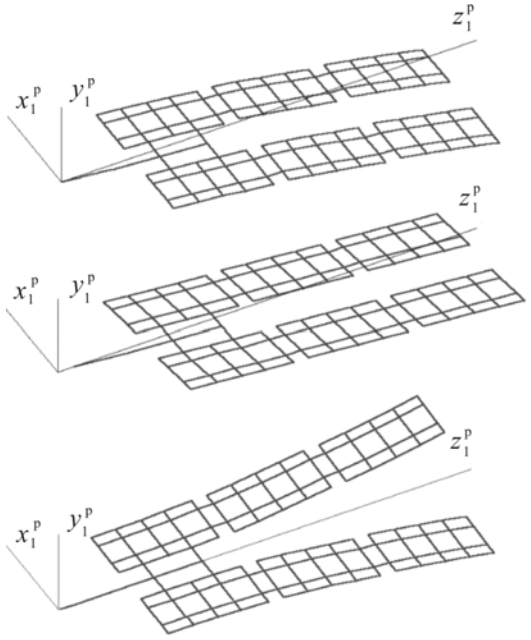


Рис. 3. Формы колебаний первой панели СБ с 3 низшими парциальными частотами

Расчет собственных форм и собственных парциальных частот упругих колебаний каждой панели СБ КА *Sesat* выполнен при учете $n^q = 10$ низших тонов в стандартной нормировке, рис. 3. Модели кинематики и динамики углового движения КА с упругими панелями подвижными панелями СБ подробно представлены в [1].

Имитация движения упругого КА

После отделения КА от разгонного блока и раскрытия панелей СБ в момент времени $t = t_0$ вектор угловой скорости принимает некоторое значение $\boldsymbol{\omega}(t_0) \in \mathbf{S}_\omega$ из ограниченной выпуклой области $\mathbf{S}_\omega$. Пусть заданы командные значения ω_i^c проекций вектора угловой скорости КА на оси связанной с корпусом КА системы координат, которые должны достигаться с точностью $|\omega_i(t) - \omega_i^c| \le \delta_\omega \quad \exists t \ge t_0 + \mathrm{T_r}$ для некоторой приемлемой длительности $\mathrm{T_r}$ режима успокоения (РУ). Особенности выполнения этого режима для КА типа *Sesat* заключаются в одновременном развороте упругих панелей СБ относительно корпуса КА на угол $3\pi/2$. При этом тензор инерции КА изменяется и проявляется вектор момента инерционно-гироскопических сил. Имитация движения КА выполняется с помощью двух программ, созданных в среде *Matlab*:

КА с жесткими панелями СБ (программа RU) имитации движения КА состоит из наборов MDL-файлов и S-функций (DLL-файлов), следующего состава: MDL-файлы: ru.mdl – основная сборка модели; bis3.mdl – модель блока измерителей скорости (БИС); bis.mdl – модель отдельного измерителя угловой скорости; privod.mdl – модель привода панелей СБ; S-функции: aok.dll – модель АОК; contr.dll – модель

алгоритма управления; dyn.dll – динамика объекта управления; kin.dll – кинематика; rdv.dll – модель двигательной установки; shim.dll – модель ШИМ.

КА с упругими панелями СБ (программа RUU). Отличия этой программы от программы RU состоят в том, что все одноименные с RU блоки выполнены в одном из следующих видов: m-функций, S-функций с использованием шаблонов на входном языке *Matlab* и языке структурных схем *Simulink*. При этом язык C и DLL не используются. Вместо модели dyn вставлена новая модель OUU – модель КА с упругими панелями СБ (3 тона колебаний каждой панели СБ). Она реализована в виде S-функции, описывающей динамику упругого спутника на m-языке.

Здесь результаты численной имитации выводятся как в графические окна, так и в файлы uk1bs.mat, uk2bs.mat – упругие отклонения первой и второй панелей СБ по тонам колебаний, а модуль kodqw из значений кватерниона ориентации дополнительно формирует матрицу направляющих косинусов для отображения угловых движений КА. С помощью разработанного программного обеспечения были выполнены численные исследования динамики КА *Sesat* в режиме успокоения, некоторые результаты представлены на рис. 4 –рис. 6.

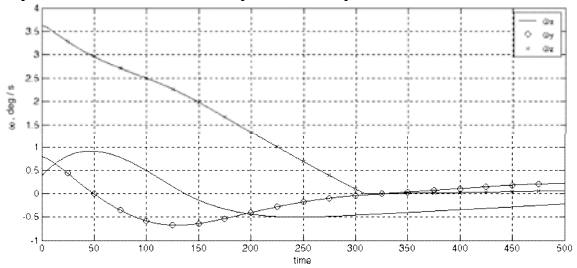


Рис. 4. Угловые скорости КА в РУ

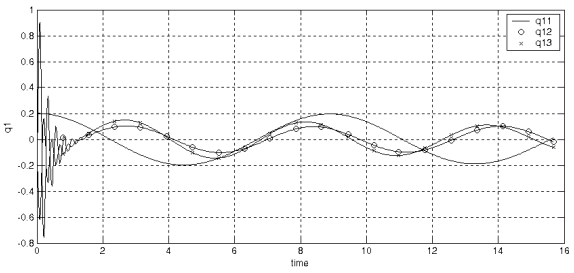


Рис. 5. Колебания 1 панели, первые 16 сек

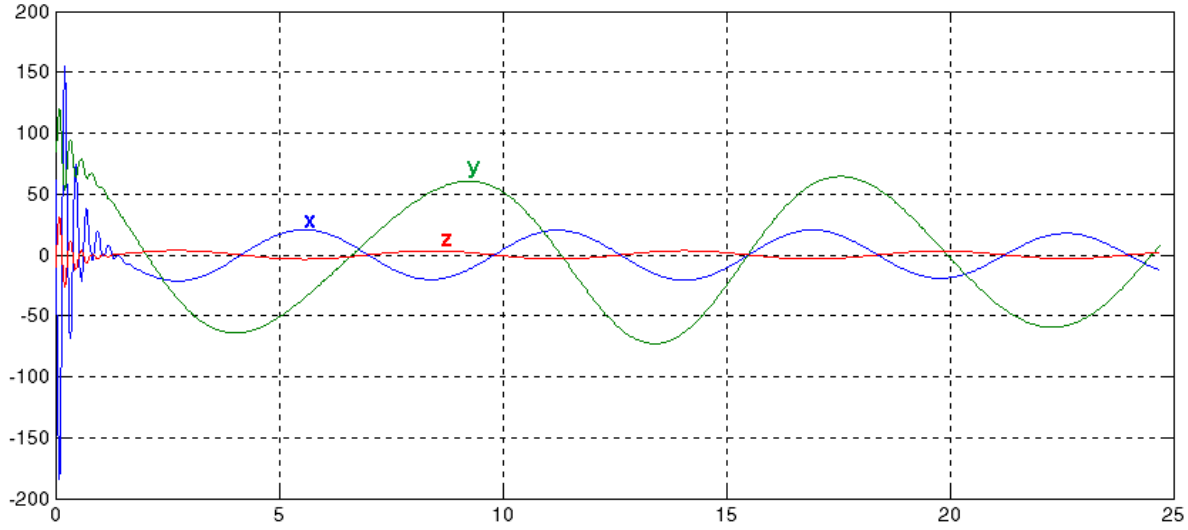


Рис. 6. Упругие отклонения 1-ой точки 1-ой панели СБ в мм от положения равновесия

Анимация пространственного движения упругого спутника

Для анимации движения спутника с деформируемыми подвижными панелями СБ использована российская программная среда *Super Vision* (SVN). В применяемой версии этого программного средства используется [2] спецификация запросов на создание и настройку объектов отображения. Важную роль играет загрузчик данных. Он предназначен для загрузки, буферизации и выдачи данных по запросам от любых графических и анимационных объектов. При загрузке данные «буферизируются» в отдельном файле на локальном диске компьютера и далее осуществляется их «подкачка» для графического отображения. Данные, которые необходимы для

анимации, «буферизируются» непосредственно в ОЗУ компьютера. Панель отображается в виде набора поверхностей и стержневых элементов. Каждая поверхность состоит из множества четырехугольников, каждый стержневой элемент состоит из последовательности прямолинейных стержней. Группировку четырехугольников в поверхности, прямолинейных отрезков в стержневые элементы производит пользователь и вводит уже скомпонованные элементы в программу. При такой организации есть возможность реализации сглаживания изгибов стержней и поверхностей, например сплайнами. Для поверхностей можно использовать встроенную в пакет OpenGL технологию аппроксимации NURBS (неоднородные рациональные бета-сплайны). Отображение резких переломов и разрывов делается путем разнесения соответствующих частей по разным элементам, поскольку сглаживание выполняется лишь в пределах элемента, но не между ними. Интерфейс между компонентами разработанного программного комплекса моделирования и визуализации движения упругой конструкции КА реализуется через файлы на жестком диске. Последовательность взаимодействия и имена файлов представлены на рис. 7.

Этап 1. На первом этапе выполняется моделирование режима успокоения КА. Как указано выше, в комплексе есть 2 модели: RU и RUU. В среде *Matlab* запускается *Simulink* и далее одна из этих моделей. Результатом работы подсистемы являются графики переходных процессов и файлы данных, которые записываются на жесткий диск. В частности создаются 3 файла, содержащие данные для подсистемы визуализации:

1. uk1bs и uk2bs – два файла значений обобщенных координат 1-ой и 2-ой панелей СБ;

2. mnk.txt – файл содержащий матрицу ориентации корпуса КА.

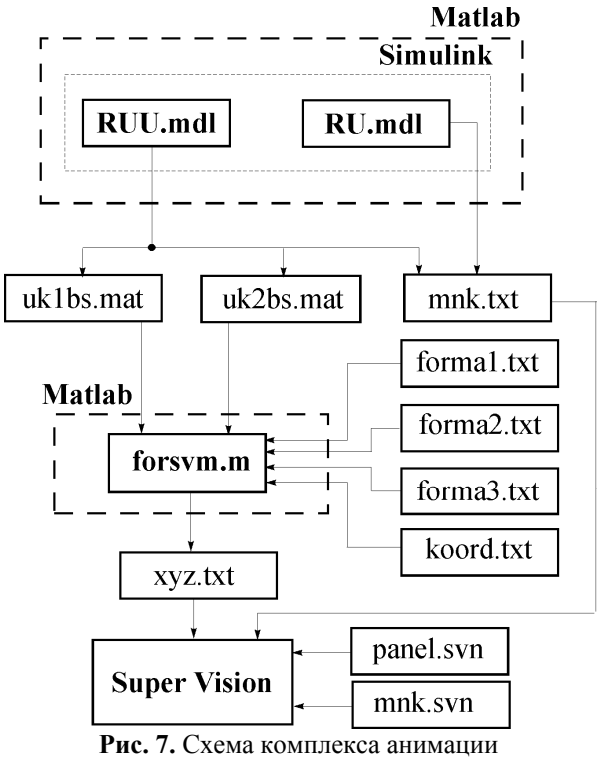


Рис. 7. Схема комплекса анимации

Этап 2. Здесь выполняется подготовка данных для отображения колебаний упругих панелей СБ в среде системы визуализации SVN. При этом в среде *Matlab* запускается на выполнение программа Forsvm. Входом в нее является один из файлов uk1bs или uk21bs. Кроме того, с диска считываются файлы форм forma1 – forma3 по первым 3-м тонам колебаний и файл koord.txt, содержащий координаты узловых точек панелей. Результатом работы является файл xyz.txt, содержащий в функции времени текущие координаты узловых точек панелей в связанной с ней СК. Среда SVN имеет 2 файла настроек: panel.svn и mnk.svn. В случае загрузки panel.svn будут отображаться колебания выбранной панели СБ, а случае загрузки mnk.svn – движение корпуса КА в ОСК. Для отображения колебаний панели СБ используется файл данных xyz.txt, а для случая отображения движения корпуса КА КА – файл mnk.txt. Предусмотрено прореживание для изменения скорости воспроизведения движения на экране дисплея. По умолчанию строки в файлах следуют через 0.04 сек, что при достаточной производительности компьютера дает

воспроизведение в реальном времени – 25 кадров (строк) в сек. Кадры анимации углового движения КА в РУ представлены на рис. 8, а крутильно-изгибных колебаний панелей СБ – на рис. 9.

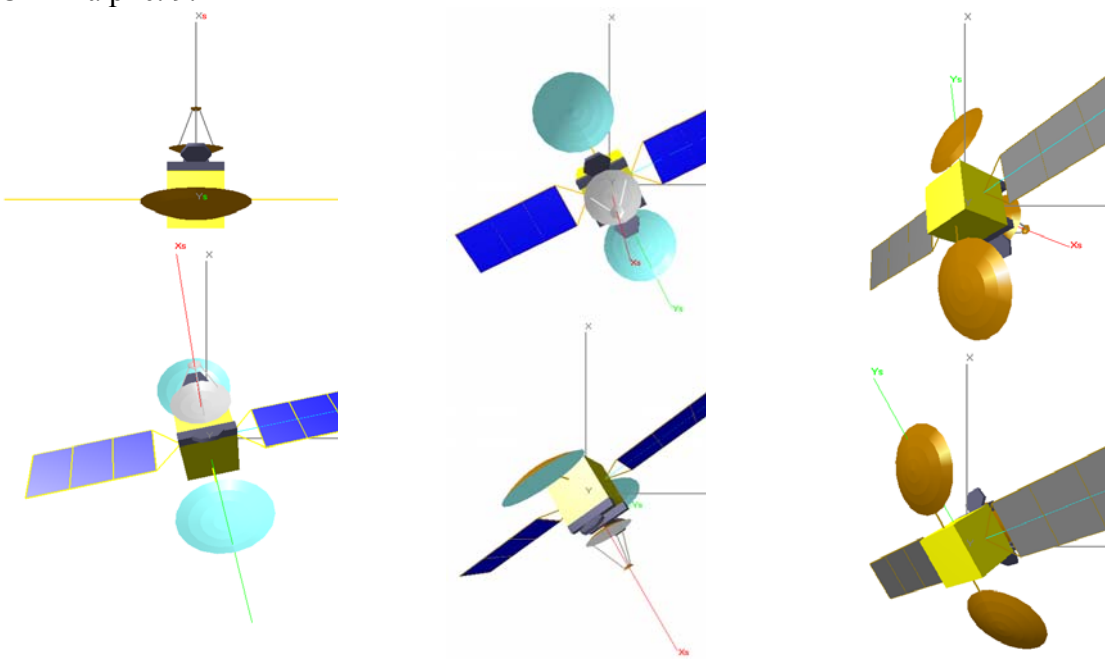


Рис. 8. Кадры анимации углового движения КА в РУ

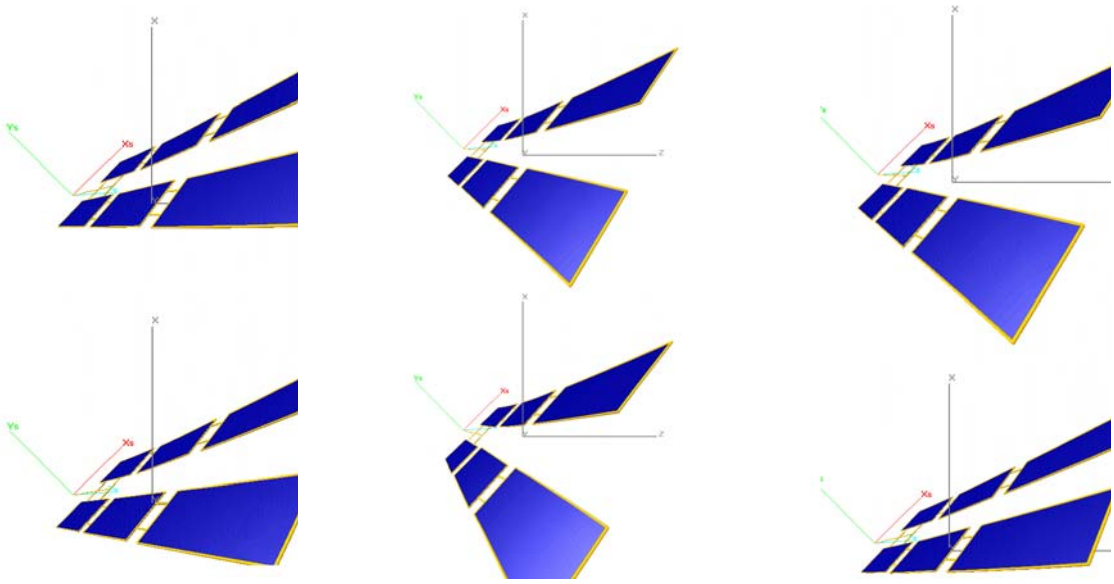


Рис. 9. Кадры анимации колебаний панелей СБ на начальном этапе успокоения КА

Работа поддержана РФФИ (проект 08-08-00512), Отделением энергетики, механики, машиностроения и процессов управления РАН (программа 15, проект 1.3), а также была поддержана Министерством образования и науки Самарской области (Конкурс 2007 г., когда автор был аспирантом СГАУ им. С.П. Королева, грант № 340ТЗ.11К).

Литература

1. Сомов С.Е. Моделирование движения упругого спутника // Актуальные проблемы авиационных и аэрокосмических систем. Казань: КГТУ-КАИ им. А.Н. Туполева. 2007. Том 12. № 3 (25). С. 75-84.

2. Сомов С.Е. Нелинейное моделирование, имитация и анимация движения упругого спутника связи в режиме начального успокоения // Материалы VIII Всероссийской научной конференции «Решетневские чтения». Красноярск: СибГАУ. 2004. С. 53-55.

Компьютерные технологии для анализа устойчивости движения и колебаний упругого спутника при широтно-импульсном управлении ориентацией с физическим запаздыванием

С.Е. Сомов
НИИ ПНМС СамГТУ, Самара

Введение

Методы пространства состояний стационарных линейных непрерывно-дискретных систем [1] с многократной фильтрацией дискретных измерений доступных координат и дискретным динамическим регулятором [2] применяются к анализу устойчивости углового движения упругого космического аппарата (КА) при формировании широтно-импульсного управления реактивными двигателями (РД) с физическим запаздыванием. Представляются разработанные методы и программные средства для анализа динамики упругого КА в начальных режимах, включая подготовку к работе силового гироскопического комплекса (СГК) при завершении режима начальной ориентации КА на Землю, а также инновационный подход к решению задачи наведения КА на Солнце и Землю при определении их ориентации на основе сигналов спутниковых навигационных систем GPS/GLONASS.

Модель линейной многократной системы с запаздыванием

Вводится линейный стационарный объект с кусочно-постоянным управлением

$$\dot{\mathbf{x}}(t) = \mathbf{A}\,\mathbf{x}(t) + \mathbf{B}\,\mathbf{u}_k(t), \quad \dot{\mathbf{u}}_k(t) = \mathbf{0}; \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t - T_{zy}); \quad \mathbf{x}(t_0) = \mathbf{x}_0, , \qquad (1)$$

где $t \in T_0 \equiv [t_0, +\infty), t_0 = 0$, вектор-функция управления $\mathbf{u}_k(t) = \{\mathrm{u}_{jk}(t)\} \in \mathbf{R}^{\mathrm{r}}$ с определением $\mathbf{u}_k(t_k) = \mathbf{u}_k = \mathbf{v}_{k\text{-}1}$ и $\mathbf{u}_k(t_k + T_{zu}) = \mathbf{u}_{k+1} = \mathbf{v}_k$ вычисляется в дискретные моменты времени $t_k + T_{zu}$ и далее при широтно-импульсной модуляции (ШИМ) управления с физическим запаздыванием $\mathrm{T_{uz}}$, где $0 \le T_{zu} < T_u$, формируется как

$$\mathrm{u}_{jk}(t) = \mathrm{U}_{\mathrm{j}}^{\mathrm{m}}\,\mathrm{PWM}(t - T_{zu}, t_k, \mathrm{v}_{jk}); \;\; \mathrm{U}_{\mathrm{j}}^{\mathrm{m}} = \mathrm{const} > 0, \qquad (2)$$

где $\mathrm{PWM}(t, t_k, \mathrm{v}_{jk}) \equiv \begin{cases} \mathrm{sign}\,\mathrm{v}_{jk} & t \in [t_k, t_k + \tau_{jk}) \\ 0 & t \in [t_k + \tau_{jk}, t_{k+1}) \end{cases}$; $\tau_{jk} = \mathrm{sat}(T_u, |\mathrm{v}_{jk}|) \equiv \begin{cases} |\mathrm{v}_{jk}| & |\mathrm{v}_{jk}| \le T_u \\ T_u & |\mathrm{v}_{jk}| > T_u \end{cases}$,

$t_k = k\,T_u, \quad k \in \mathrm{N}_0 \equiv \{0,1,2,3,...\}$. Вектор $\mathbf{x}(\mathrm{t}) \in \mathbf{R}^{\mathrm{n}}$ описывает состояние объекта, а вектор $\mathbf{v}_{\mathrm{k}} = \{\mathrm{v}_{\mathrm{jk}}\} \in \mathbf{R}^{\mathrm{r}}$ представляет дискретную текущую команду – выход дискретного алгоритма управления, формируемый БЦВМ только в дискретные моменты времени t_k. Измерение $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t - T_{zy})$ состояния объекта (1) является неполным и выполняется только в моменты времени $t_s = sT_q, \;\; s \in N_0$ с периодом $\mathrm{T_q} \le \mathrm{T_u}$, кратным периоду управления T_u. Для $\mathrm{T_q} < \mathrm{T_u}$ при вычислении вектора $\mathbf{v}_{k+1} = \mathbf{v}(t_{k+1})$ могут использоваться измерения, не более поздние, чем $\mathbf{y}(t_k') \equiv \mathbf{y}(kT_u + n_{\mathrm{vc}}T_q) = \mathbf{C}\mathbf{x}((k+1)T_u - (n_{zy} + n_{zc})T_q)$, где $n_q = T_u / T_q$; $T_{\mathrm{vc}} = T_u - T_{zc}$; $n_{\mathrm{vc}} = \mathrm{E}[T_{\mathrm{vc}} / T_q]$; $n_{zc} = n_q - n_{\mathrm{vc}}$; $k = \mathrm{E}[s / n_q]$, $\mathrm{E}[\cdot]$ – символ целой части, причем в общем случае $T_{zc} \ne T_{zc}^{'} \equiv n_{zc}T_q$. Пусть при вычислении вектора дискретной команды управления $\mathbf{v}_k$ применяется дискретный фильтр рекуррентного типа $\widetilde{\mathbf{x}}_{s+1} = \widetilde{\mathbf{A}}\widetilde{\mathbf{x}}_s + \widetilde{\mathbf{B}}\mathbf{y}_s$, $\widetilde{\mathbf{x}}_{\mathrm{s}} \in \mathbf{R}^{\mathrm{m}}$; $\widetilde{\mathbf{y}}_s = \widetilde{\mathbf{C}}\widetilde{\mathbf{x}}_s + \widetilde{\mathbf{D}}\mathbf{y}_s$; $\mathbf{y}_s, \widetilde{\mathbf{y}}_s \in \mathbf{R}^l$, $s \in N_0$ с периодом квантования T_q и выходным сигналом $\mathbf{y}(t_k') = \mathbf{y}_k^F = \widetilde{\mathbf{y}}_s\,|_{s = n_q \cdot k^*}$ при $t_k' = k^* T_u$, где $k^* = \mathrm{E}[(s + n_{zc}) / n_q]$ и

$\widetilde{\mathbf{A}}$, $\widetilde{\mathbf{B}}$, $\widetilde{\mathbf{C}}$, $\widetilde{\mathbf{D}}$ – матрицы соответствующей размерности. Сигналы $\mathbf{y}_k^F$ этого фильтра поступают в дискретный динамический регулятор с периодом дискретизации T_u: $\widehat{\mathbf{x}}_{k+1} = \mathbf{A}_{0\mathrm{d}}\widehat{\mathbf{x}}_k + \mathbf{B}_{0\mathrm{d}}\mathbf{v}_k + \mathbf{Q}_{0\mathrm{d}}\mathbf{y}_k^F$, $\widehat{\mathbf{x}}_k \in \mathbf{R}^p$; $\mathbf{v}_{k+1} = \mathbf{K}_\mathrm{u}(\mathbf{r}_{k+1} - \boldsymbol{\chi}_\mathrm{u}\mathbf{C}_0\widehat{\mathbf{x}}_{k+1})$, где $\mathbf{r}_k = \mathbf{C}_0\mathbf{x}_k^\mathrm{r}$ – сигнал команды, $\mathbf{r}_k \in \mathbf{R}^l$, $\mathbf{r}_k = \{r_{ik}\}$; $\mathbf{x}_k^r$ – вектор эталонных переменных состояния системы; $\boldsymbol{\chi}_\mathrm{u}$ – диагональная матрица с элементами, равными 1 либо 0 при замыкании либо размыкании системы по отдельным каналам, соответственно, а $\mathbf{A}_{0\mathrm{d}}, \mathbf{B}_{0\mathrm{d}}, \mathbf{Q}_{0\mathrm{d}}, \mathbf{C}_0$ и $\mathbf{K}_\mathrm{u}$ – постоянные матрицы соответствующей размерности.

Линеаризация непрерывно-дискретной системы с ШИМ управления

При ШИМ управления (2) с запаздыванием возможно лишь приближенное представление объекта (1) в виде линейной дискретной модели. При $T_{zu} = 0$ и обозначениях $\mathbf{x}_k \equiv \mathbf{x}(t_k)$; $\mathbf{u}_k \equiv \mathbf{u}_k(t_k)$ нелинейное разностное уравнение с периодом T_u получается в виде $\mathbf{x}_{k+1} = \mathbf{A}_\mathrm{d}\mathbf{x}_k + \Sigma\mathbf{Q}(\tau_{\mathrm{j}k})\,\mathbf{b}_\mathrm{j}\,\mathrm{U}_\mathrm{j}^\mathrm{m}\,\mathrm{sign}\,\mathrm{v}_{\mathrm{j}k}$, где

$$\tau_{\mathrm{j}k} = \mathrm{sat}(T_u, |\mathrm{v}_{\mathrm{j}k}|); \quad \mathbf{A}_d = \exp(T_u\mathbf{A}); \quad \mathbf{Q}(\tau) \equiv \exp((T_u - \tau)\mathbf{A})\int_0^\tau \exp(t\mathbf{A})\,\mathrm{d}t\,.$$

С использованием известных свойств матричной экспоненты и интеграла от нее имеем $\mathbf{Q}(\tau) = \mathbf{A}_\mathrm{d}(\mathbf{I} - (\mathbf{A}\tau)/2! + (\mathbf{A}\tau)^2/3! - \cdots)\tau$, поэтому предполагая выполнение условий $\tau_{\mathrm{j}k} << T_u$; $T_u << 2\pi/|\lambda_i|$, где λ_i – собственные значения матрицы $\mathbf{A}$ в (1), выделяется линейная часть матриц $\mathbf{Q}(\tau_{ik})$ по отношению к $\tau_{jk} = |\mathrm{v}_{\mathrm{j}k}|$ и выполняется линеаризация ШИМ управления, т.е. получается линейная дискретная модель движения объекта $\mathbf{x}_{k+1} = \mathbf{A}_\mathrm{d}\mathbf{x}_k + \mathbf{B}_\mathrm{d}\mathbf{v}_k$, где матрица $\mathbf{B}_\mathrm{d} = \mathbf{A}_\mathrm{d}\mathbf{B}\,\mathrm{diag}\{\mathrm{U}_\mathrm{j}^\mathrm{m}\}$. Линеаризованная дискретная модель объекта (1) с ШИМ управления (2) и учетом запаздывания имеет представление

$$\mathbf{x}_{k+1} = \mathbf{A}_\mathrm{d}\mathbf{x}_k + \mathbf{A}_\mathrm{d}^{\nu_\mathrm{u}}\mathbf{B}_\mathrm{d}^{\varepsilon_\mathrm{u}}\mathbf{u}_k + \mathbf{B}_\mathrm{d}^{\nu_\mathrm{u}}\mathbf{v}_k; \quad \mathbf{u}_{k+1} = \mathbf{v}_k, \quad k \in N_0,$$

где $\mathrm{T}_{\nu\mathrm{u}} = \mathrm{T}_\mathrm{u} - \mathrm{T}_{\mathrm{zu}}$; $\varepsilon_\mathrm{u} = \mathrm{T}_{\mathrm{zu}}/\mathrm{T}_\mathrm{u}$; $\nu_\mathrm{u} = \mathrm{T}_{\nu\mathrm{u}}/\mathrm{T}_\mathrm{u} = 1 - \varepsilon_\mathrm{u}$; $\mathbf{A}_\mathrm{d} = \exp(\mathrm{T}_\mathrm{u}\mathbf{A}) = \mathbf{A}_\mathrm{d}^{\nu_\mathrm{u}}\mathbf{A}_\mathrm{d}^{\varepsilon_\mathrm{u}}$;

$$\mathbf{A}_\mathrm{d}^{\varepsilon_\mathrm{u}} = \exp(T_{zu}\mathbf{A}); \quad \mathbf{A}_\mathrm{d}^{\nu_\mathrm{u}} = \exp(T_{\nu u}\mathbf{A}); \quad \mathbf{B}_\mathrm{d}^{\varepsilon_\mathrm{u}} = \int_0^{T_{zu}} \exp(\tau\mathbf{A})\,\mathrm{d}\tau\cdot\mathbf{B}; \quad \mathbf{B}_\mathrm{d}^{\nu_\mathrm{u}} = \int_0^{T_{\nu u}} \exp(\tau\mathbf{A})\,\mathrm{d}\tau\cdot\mathbf{B}.$$

Агрегирование и анализ устойчивости

При векторе команды $\mathbf{r}_k = \mathbf{0}$ определение устойчивости нулевого решения $\mathbf{x}(t) = \mathbf{0}$, $t \in T_0$; $\widetilde{\mathbf{x}}_s = \mathbf{0}$; $\widehat{\mathbf{x}}_k = \mathbf{0}$, $s, k \in N_0$ непрерывно-дискретной системы управления понимается как прямая композиция понятий устойчивости ее непрерывной и дискретной частей. Для получения линейной модели замкнутой непрерывно-дискретной системы многократного типа с запаздыванием используются методы пространства состояний [1,2]. Здесь основная задача состоит в построении эквивалентной дискретной модели с главным периодом T_u, как наибольшему из имеющихся периодов квантования. Решение этой задачи представлено в [2], в результате получаются дискретные модели как замкнутой системы, так и разомкнутой по любого из компонентов выходного вектора относительно любого компонента входного вектора $\mathbf{r}_k$ системы. Далее для исследования устойчивости применяются классические частотные и спектральные методы линейной теории дискретных систем в векторно-матричном представлении.

Моделирование движения КА с упругой конструкцией

При получении приближенных моделей движения упругих КА наиболее распространен метод Релея-Ритца-Галеркина в форме метода конечных элементов (МКЭ). Особенность применяемого подхода заключается в представлении упругих колебаний элементов конструкции в виде конечного числа тонов. Здесь расчет форм колебаний выполняется на основе МКЭ с конденсацией (редукцией) по тонам колебаний, на ЭВМ вычисляются матрицы коэффициентов взаимовлияния движений всех подконструкций КА, как абсолютно твердых тел, включая корпус спутника, так и деформируемых тел.

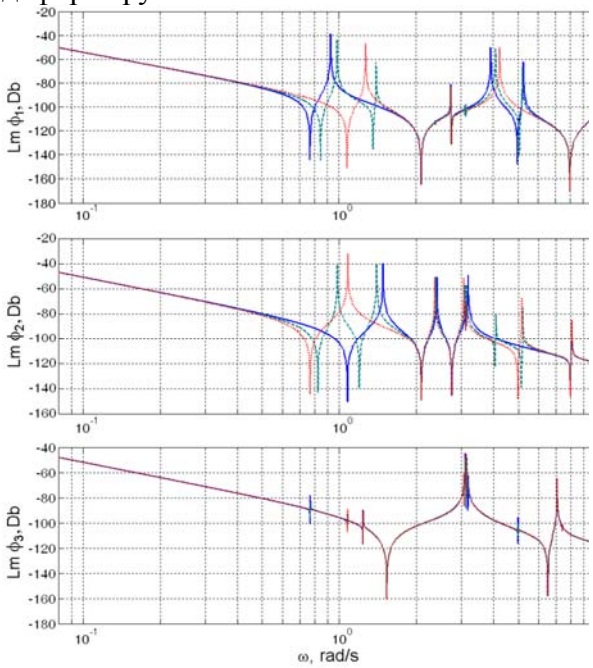


Рис. 1. ЛАХ упругого КА по каналам

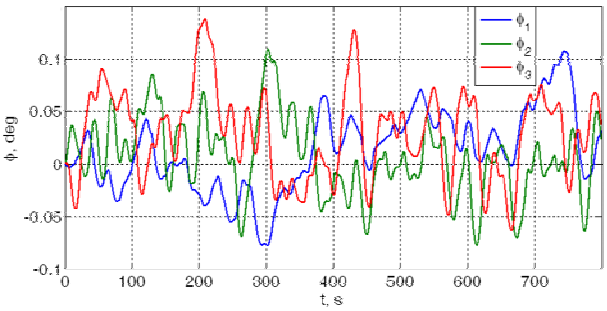


Рис. 2. Погрешности угловой стабилизации

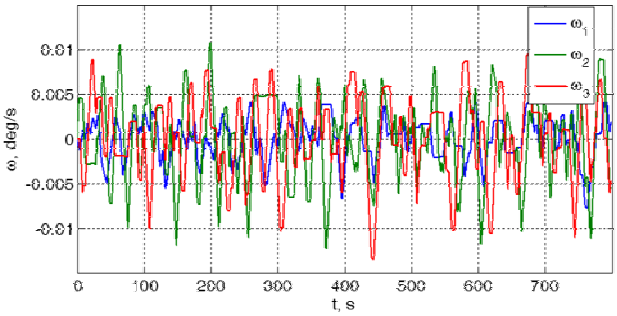


Рис. 3. Ошибки стабилизации по скоростям

Модель пространственного углового движения КА с подвижными крупногабаритными упругими панелями солнечных батарей (СБ) при широтно-импульсном управлении РД ориентации представлена в работах [3 – 6]. Ориентация КА относительно орбитальной системы координат (ОСК) определяется вектором-столбцом $\boldsymbol{\phi} = \{\phi_1, \phi_2, \phi_3\}$ из углов рыскания, крена и тангажа соответственно. Динамические свойства КА существенно зависят от положения панелей СБ, которое определяется углом $\gamma \in [0, 2\pi]$. В качестве примера на рис. 1 представлены логарифмические амплитудные характеристики (ЛАХ) непрерывной модели упругого спутника по каналам управления его ориентацией для различных фиксированных положений панелей СБ с шагом 45°.

Дискретные алгоритмы фильтрации

Применяемый дискретный фильтр измерений с периодом дискретизации T_q имеет дискретную передаточную функции $W_f(z_q) = (1 + b_1)/(1 + b_1 z_q^{-1})$, $z_q \equiv \exp(sT_q)$ с условием $W_f(1) = 1$, где $b_1 \equiv -\exp(-T_q/T_f)$ и T_f – постоянная времени. Частотная функция $\widetilde{W}_f(j\lambda_q)$ в зависимости от псевдочастоты фильтрации $\lambda_q = (2/T_q)\mathrm{tg}(\omega T_q/2)$ имеет вид $\widetilde{W}_f(j\lambda_q) \equiv \widetilde{W}_f(\widetilde{s}_q) = K_f^\lambda \cdot (\widetilde{s}_q - q_q^\lambda)/(\widetilde{s}_q - p_q^\lambda)$, где $K_f^\lambda = (1 + b_1)/(1 - b_1) < 1$; $q_q^\lambda = -(2/T_q)$ и $p_q^\lambda = -K_f^\lambda(2/T_q)$ являются ее нулем и полюсом. При $T_f = 2$с, $T_q = 1$с и

периоде управления $T_u = 4$ с ЛАХ фильтра вносит амплитудное подавление с выходом на постоянный уровень -11 Db, причем вносимый фильтром отрицательный сдвиг фазы является приемлемым в отношении получаемых в результате запасов устойчивости по фазе.

Синтез законов наведения и стабилизации КА в начальных режимах

Представляются разработанные модели направлений на Солнце и Землю для произвольного положения КА на эллиптической орбите, алгоритмы поиска Солнца и Земли. Особенность решаемой задачи синтеза дискретных алгоритмов бортовой обработки информации и управления РД в режимах успокоения, начального наведения на Солнце и Землю состоит в существенном влиянии упругости панелей СБ при слабом собственном демпфировании – декремент колебаний составляет ≈ 0.005 по каждому тону. Предлагается инновационный подход к решению задачи начального наведения упругих КА на Солнце и Землю при определении их ориентации с использованием навигационных систем GPS/GLONASS. Конструктивные базы установки навигационных антенн на корпусе КА информационного назначения обычно не превышают 1 м, поэтому достигаемая точность определения ориентации таких КА с помощью сигналов спутниковых навигационных систем оценивается значением $\approx 0.5^\circ$. Такой точности достаточно как для *одновременного* наведения панелей солнечных батарей КА на Солнце, а полезной нагрузки (антенны спутника связи, телескоп спутника наблюдения) – на Землю, так и последующей орбитальной угловой стабилизации КА в этом режиме для обеспечения начального энергетического баланса КА и разгона роторов 6 гиродинов в составе СГК. Возникающие при этом реактивные моменты являются возмущающими, они успешно нейтрализуются контуром орбитальной стабилизации КА на основе РД с ШИМ управления их тягой. Особенность данного режима заключается в том, что даже при постоянном векторе собственного кинетического момента СГК возникают нутационные колебания спутника. Очевидно, что при наличии достаточных энергетических возможностей бортовых средств КА рационально применять одновременный разгон роторов всех шести ГД. При такой ой схеме разгона роторов ГД не возникают как возмущающие реактивные моменты, так и нутационные колебания – вектор суммарного кинетического момента СГК $\mathbf{H} \equiv \mathbf{0}$. Проблема встает при отсутствии таких энергетических возможностей, когда допустим лишь последовательный разгон роторов пар ГД с учетом существенного различия значений управляющего электромеханического момента и момента сил трения по осям роторов. Представляется решение [7-9] задачи выбора рациональной логики в последовательности разгона роторов ГД по парам в отношении нутационных колебаний спутника, а именно: (i) до каких значений собственного КМ каждого ГД в паре следует разгонять роторы – сразу выводить угловые скорости роторов на их номинальное значение либо с некоторым шагом, последовательно по парам; (ii) имеет ли значение последовательность назначения пар ГД для разгона их роторов. Созданы линеаризованные дискретные модели и выполнен анализ динамических свойств упругого КА с ШИМ тяги РД, включая параметрический анализ миграции резонансных псевдочастот по каналам управления. Выполнен параметрический синтез дискретных алгоритмов управления, включая анализ устойчивости и параметрический синтез нелинейной одноканальной модели КА с ШИМ управления РД в начальных режимах на основе метода функций Ляпунова, параметрический синтез линеаризованной модели КА с упругими панелями СБ на основе частотных и спектральных методов. Научная новизна полученных здесь результатов состоит в детальном учете связи движений корпуса КА и панелей солнечных батарей, в применении нового дискретного векторного закона ШИМ управления тягой РД при пространственном наведении КА на Солнце и Землю, а также в получении гарантированных оценок области

диссипативности углового движения КА по любому одному каналу на основе метода функций Ляпунова.

Параметрический синтез дискретных алгоритмов управления

Завершающий параметрический синтез дискретных алгоритмов фильтрации и широтно-импульсного управления для полной пространственной модели движения упругого спутника при его успокоении, наведении на Солнце и Землю с последующим разгоном роторов гиродинов СГК выполнен на основе компьютерной имитации в среде Matlab. Для физического запаздывания $T_{zu} = 0.7\,\text{с}$ в формировании широтно-импульсного управления тягой РД погрешности угловой стабилизации КА в ОСК при разгоне первой пары гиродинов СГК представлены на рис. 2 и рис. 3.

Выводы

Представлены математические модели, методы и компьютерные технологии для анализа устойчивости и стабилизации углового движения упругих слабо демпфированных информационных спутников в начальных режимах при неполном дискретном измерении состояния, многократной фильтрации и широтно-импульсном управлении реактивными двигателями с временным запаздыванием. Предложен инновационный подход к решению задачи начального наведения упругих КА на Солнце и Землю при определении их ориентации с использованием навигационных систем GPS/GLONASS. Представлены рекомендации по логики последовательного начального разгона роторов ГД по парам в составе силового гироскопического комплекса. Установлено, что даже при декременте $\delta \cong 0.005$ упругих колебаний панелей солнечных батарей синтезированные алгоритмы обеспечивают устойчивость движения КА и приемлемое качество.

Работа поддержана РФФИ (проект 08-08-00512), Отделением энергетики, механики, машиностроения и процессов управления РАН (программа 15, проект 1.3), а также была поддержана Министерством образования и науки Самарской области (Конкурс 2007 г., когда автор был аспирантом СГАУ им. С.П. Королева, грант № 340ТЗ.11К).

Литература

1. Стрейц В. Метод пространства состояний в теории линейных дискретных систем управления. М.: Наука, 1985.

2. Сомов Е.И. Робастная стабилизация упругих космических аппаратов при неполном дискретном измерении и запаздывании в управлении // Известия РАН. Теория и системы управления. 2001. № 2. С. 124-143.

3. Сомов С.Е. Нелинейная динамика упругого спутника при начальном успокоении // Известия Самарского научного центра РАН. 2005. Том 7. № 1. С. 107-117.

4. Сомов С.Е. Динамика успокоения упругого спутника при широтно-импульсной модуляции управления двигателями // Известия ВУЗ. Авиационная техника. 2005. № 4. С. 17-23.

5. Сомов С.Е. Анализ колебаний конструкции спутника при наведении на Солнце и Землю с широтно-импульсной модуляцией управления двигателями // Известия Самарского научного центра РАН. 2007. Том 9. № 3. С. 813-823.

6. Сомов С.Е. Моделирование движения упругого спутника // Актуальные проблемы авиационных и аэрокосмических систем. Казань: КГТУ-КАИ им. А.Н. Туполева. 2007. Том 12. № 3 (25). С. 75-84.

7. Сомов С.Е. Динамика широтно-импульсного управления ориентацией упругого космического аппарата при разгоне роторов гиродинов // Труды IV Международной конференции по проблемам управления. М., ИПУ РАН. 2009. С. 1925-1935.

8. Somov S.Ye. Pulse -width control of a flexible spacecraft // Proceedings of the IFAC Workshop “Aerospace Guidance, Navigation and Flight Control Systems”. Samara: SSC of RAS. 2009. IPACS Electronic Library: http://lib.physcon.ru/?item=1839 P. 1-6.

9. Somov S.Ye. Flexible spacecraft guidance and pulse-width control at initial modes // Online journal “Automatic Control in Aerospace”. 2010. Vol. 3. No. 1. P. 1-9. http://www.aerospace.unibo.it

Программный комплекс моделирования и оптимизации движения космического аппарата с малой тягой в системе Земля-Луна

О.Л. Старинова, И.Л. Матерова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Рассматривается математическое и программное обеспечение позволяющее проводить баллистическую оптимизацию, моделирование и отображение управляемого движения космического аппарата (КА) с двигателем малой тяги в системе Земля – Луна. Математическая модель движения учитывает гравитацию Земли и Луны, воздействие малой тяги переменного направления и величины и возмущения (гравитация Солнца, нецентральность полей Земли и Луны, воздействие атмосферы Земли). Задача об оптимальном управления решается в рамках ограниченной круговой задачи трех тел.

С использованием формализма принципа максимума Понтрягина получены необходимые условия оптимальности законов управления для оптимальных по быстродействию или расходу рабочего тела некомпланарных перелетов с заданной геоцентрической орбиты на заданную селеноцентрическую орбиту. Для различных вариантов исходных и конечных орбит получены условия трансверсальности, сформулированы постановки краевых задач и получены соответствующие сопряженные уравнения.

Разработано программное обеспечение, содержащее два взаимодействующих программных продукта.

1. Программный комплекс, предназначенный для численного решения краевых задач и основанный на алгоритме модифицированного метода Ньютона с автоматическим выбором шага вычисления производных. В качестве начального значения сопряженных переменных, используются результаты решений задач об оптимальных компланарных перелетах в системе Земля-Луна (хранящиеся во встроенной базе данных. Для оптимизации перелетов с различными граничными условиями перелетов и КА с различными проектными параметрами использовался метод продолжения по параметру.

2. Программный комплекс, предназначенный для отображения некомпланарного движения КА с найденным оптимальным законом управления. Пользователь на выбор может получить траектории КА в барицентрической, геоцентрической и селеноцентрической системах координат, а также графики изменения фазовых координат и оптимального управления от времени перелета.

Разработанное программное обеспечение было использовано для решения задач о формировании заданной околокруговой селеноцентрической орбиты и достижения точек либрации L4, L5. для конкретных значений проектных параметров КА. Получены оптимальные законы управления и соответствующие им траектории движения. Проведен анализ зависимости критериев оптимальности от начальных баллистических условий перелетов. Сформулированы рекомендации по выбору оптимальных параметров стартовой орбиты и дат старта КА.

Использование информационных технологий твёрдотельного геометрического проектирования в позиционных задачах проектирования летательных аппаратов

А.И. Шулепов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Новый этап в развитии автоматизированных систем проектирования (АСП) связан с появлением высокоэффективных вычислительных средств, имеющих большую оперативную и внешнюю память, высокие характеристики процессоров и каналов обмена, высокоразвитое специальное и прикладное программное обеспечение, появлением специализированных прикладных пакетов, таких как, например, ACAD, “Solid works”, “Nastran” и другие.

Эти системы позволяют, сохранив эффективность основных алгоритмов, сделать математическую модель проектируемого объекта более адекватную проектируемому изделию, а уточнение модели (главным образом введение описания сложных форм объектов) и разработка дополнительных алгоритмов для выполнения операций с телами сложной формы значительно расширить прикладные возможности системы.

Включение таких прикладных пакетов, как “Solid works” делает вычислительный комплекс эффективным средством для расширения возможностей при решении проектных задач размещения геометрических объектов.

Задача размещения оборудования, узлов и агрегатов внутри отсеков летательных аппаратов (ЛА) является одной из сложных в многочисленном ряду компоновочных задач. Её специфика заключается в разнообразии конструкторского исполнения элементов компоновки и условий нормального функционирования проектируемого объекта.

На языке разработчиков САПР задача внутреннего компонования может быть определена как многокритериальная оптимизационная задача размещения объектов сложной геометрической формы с различными по физическому представлению требованиями к размещению. Решение такой задачи с применением традиционных методов проектирования, основанных на известных методиках, опыте, интуиции и пространственном мышлении проектировщика, связано с многочисленными проблемами. Первая заключается в том, что задача многокритериальная.

Существует несколько подходов к решению многокритериальных оптимизационных задач:

– использование свёрток;

– построение области решений, исходя из списка критериев, и выбор решения, наиболее удалённого от границы этой области;

– выбор главного критерия и перевод остальных в ранг ограничений.

На практике построение и использование комбинированного критерия затруднено в силу его субъективности, различной природы и размерности определяющих его параметров.

Построение области решений, исходя из списка критериев – сложный процесс, поскольку задача размещения имеет большую размерность. Наиболее приемлемым для компоновочных задач оказывается третий подход, он позволяет использовать пошаговый процесс размещения с оптимизацией главного критерия на каждом шаге. При этом снижается размерность задачи, что позволяет:

– разрабатывая прикладное программное обеспечение на функциональных языках высокого уровня, построить быстродействующие алгоритмы размещения;

– просматривать и оценивать по другим, в том числе и неформализуемым критериям несколько вариантов компоновок за короткое время.

Вторая проблема заключается в сложности контроля непересечения пространственных геометрических объектов произвольной формы, качественной и количественной их оценки взаимного расположения. Реализация этих операций требует объёмной графической проработки с использованием трудоёмких методов начертательной геометрии. Существуют аналитические методы подобного контроля, основанные на теории R-функций [1]. Построение R-функций сложных тел также трудоёмкий процесс, но их значения достаточно просто получить алгоритмически. При этом должна быть выбрана соответствующая модель сложного тела, в данном случае логико-аналитическая.

Третья проблема состоит в том, что размещение бортового оборудования – это часть общего процесса проектирования летательного аппарата. Для получения компоновки требуется большое количество исходной информации, которая является результатом предыдущих этапов проектирования. В свою очередь, результаты размещения служат исходной информацией для последующих этапов проектирования. Таким образом, процесс размещения оборудования – это процесс переработки и преобразования информации о проектируемом объекте, форма представления и физическое содержание которой весьма разнообразны.

Решение комплекса проблем, связанных с информационным обеспечением процесса компонования отсеков ЛА, возможно с помощью адаптации в информационную среду системы прикладных графических пакетов.

Четвёртую по счёту проблему компонования бортового оборудования определяет огромное количество графической документации, выпускаемой на различных этапах разработки проекта.

В решении этих проблем эффективными вместе с автоматизированной графикой являются также прикладные графические пакеты.

Решение даже перечисленных далеко не всех проблем требует комплексного подхода, который неизбежно приводит к разработке автоматизированной системы проектирования, в состав которой должны входить кроме программ, написанных на функциональном языке высокого уровня, адаптированные в систему прикладные графические пакеты.

Постановка задачи размещения оборудования, узлов и агрегатов как оптимизационной предполагает выбор функции цели и оределения ограничений. Целевые функции представляют собой интегральные требования к компоновке, которые зависят от положения каждого из компонуемых составляющих. Такими требованиями могут быть:

– минимум отклонения центра масс от заданного положения;

– минимум моментов инерции;

– минимум трудоёмкости монтажных и ремонтных работ (при монтаже - демонтаже и обслуживании бортовой аппаратуры);

– минимум веса кабельной сети;

– минимальная величина собственного магнитного поля;

– минимальная величина зазора между приборами, узлами и агрегатами при получении плотных компоновок (обеспечение монтажа и демонтажа при сборке, обслуживание, соблюдение температурного режима и т.д.).

В общем случае задача размещения в терминах нелинейного математического программирования может быть представлена в следующем виде:

$$\Phi(X_0) = \min \quad \Phi(X)\,|_{X \in G}. \qquad (1)$$

$f(\bar{u}_i) \geq 0, \quad i=n.$ (2)

$\psi(\bar{u}_i, \bar{u}_j) \geq 0, \; i=n, \quad j=n, \quad i \neq j,$ (3)

Здесь $X = \{\bar{u}_1, \bar{u}_2, \bar{u}_3, \ldots, \bar{u}_n\}$ – вектор состояния отсека, где

$\bar{u}_i = \{ x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i \}$ – вектор размещения прибора.

Ограничениями вида (2) и (3) при решении задач могут быть условия двух типов, различающиеся природой их существования:

– традиционные для задач позиционного геометрического проектирования условия непересечения приборов (геометрических тел) и попадание вовнутрь заданного объёма;

– условия установки размещаемых тел внутри отсеков, являющихся следствием общей конструктивной схемы изделий, условий эксплуатации отдельных приборов, узлов и агрегатов в целом, технологических требований и характеристик приборов и систем.

Первый тип ограничений можно записать в виде неравенств, зависящих от положения приборов в системе координат изделия:

$f(\bar{u}_i)$ – уравнение годографа функции плотного размещения оболочки и прибора ***i***;

$\psi(\bar{u}_i, \bar{u}_j)$ – уравнение годографа функции плотного размещения приборов ***i*** и ***j*** ;

$\bar{u}_i$ и $\bar{u}_j$ –вектор параметров размещения приборов.

Второй тип ограничений относится к неравенствам вида (3), определяющих анализ условий установки бортового оборудования. Роль конструктора, работающего с прикладными пакетами проектирования, заключается в том, что он должен указать область внутри отсека, где желательно установить тот или иной объект размещения. Когда область определена, для неё можно построить годограф функции плотного размещения.

Как видно из предыдущего анализа, задача размещения является многокритериальной задачей условной дискретной оптимизации. Для решения подобных задач применяются различные методы пошаговой оптимизации [2, 3]. В соответствии с этими методами на каждом шаге целевая функция оптимизируется только по одному параметру, в нашем случае по положению одного прибора $\bar{u}_i$ (в дальнейшем вектор $\bar{u}_i$ будем обозначать как u_i). Так, в определённой последовательности выбираются положения всех размещаемых объектов от 1 до ***n*** .Такой подход даёт возможность получать только локальный экстремум целевой функции. Для поиска глобального экстремума необходимо обеспечить либо перебор всех локальных экстремумов, либо выбрать оптимальный порядок размещения (в обоих случаях вариантов решений не менее ***n!***). В любом из этих случаев в качестве интеллектуальной поддержки системы, должен быть сделан визуальный контроль конструктором для принятия решения. Таким образом, выбор стратегии оптимизации в задачах геометрического проектирования требует включения в систему интегрированных пакетов твёрдотельного геометрического проектирования таких, как, например, “Solid Works”.

Пусть $X = \{u_1, u_2, \ldots, u_n\}$ – вектор состояния отсека,

где

$u_i = \{x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i \}$ –вектор параметров размещения объекта. Вектор состояния X_0 , оптимизирующий функцию цели

$$\Phi(X_0) = \min_{X \in G} \Phi(X), \quad (4)$$

может быть получен последовательным применением итерационной формулы

$$F_k(u_{10}, u_{20}, \ldots, u_{k0}) = \min_{u_k \in G} \Phi(u_{10}, u_{20}, \ldots, u_k); \quad (5)$$

$$\Phi(X_0) = F_n(\boldsymbol{u}_{10}, \boldsymbol{u}_{20}, \ldots, \boldsymbol{u}_{n0}),\ k = 1,2,\ldots,\ n, \quad (6)$$

где G – множество, характеризующее область допустимых решений, определённых ограничениями вида 2 или 3, наложенными на компоновку отсека. Область допустимых размещений объекта размещения определяется соотношением (рис. 1)

$$G = G_1 \cap G_2 \cap \ \ldots \cap G_k \cap \ldots \cap \ G_n\ . \quad (7)$$

Каждая итерация заключается в оптимизации функции цели по параметрам только одного очередного объекта размещения. Ранее размещённые приборы, узлы и агрегаты при этом считаются неподвижными. Можно задать и угловые параметры размещения прибора

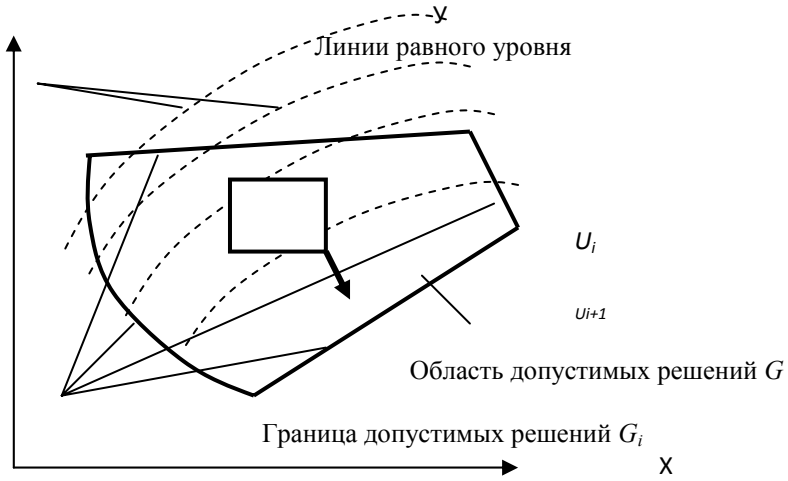


Рис. 1. Поиск допустимого решения

$(\alpha_i, \beta_i, \gamma_i)$, снизив тем самым число переменных до трёх на каждом шаге решения.

Таким образом, снижая трудоёмкость решения задачи, мы снижаем точность её решения. Используя возможности интегрирования прикладного пакета твёрдотельного геометрического проектирования, можно реализовать визуальный контроль над процессом размещения бортового оборудования КА и лицу принимающему решение выбрать необходимое решение.

Литература

1. Гаврилов В.Н. Автоматизированная компоновка приборных отсеков летательных аппаратов. М.: Машиностроение, 1988, 137.

2. Шулепов А.И. Подсистема автоматизированного проектирования компоновок приборных отсеков. Материалы 2-го Всесоюзного совещания по автоматизации проектирования, ч. 2. – М., 1983 , с. 154-155.

3. Shulepov, S. Andreev One approach to the task of soft spacecraft structure elements location. Proceeding of fourth UKRAINE – RUSSIA – CHINA symposium on space science and technology. Kiev, 1996, p.248.

Секция 2
ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ И GRID-ТЕХНОЛОГИИ

Грид-среда Самарского государственного аэрокосмического университета имени академика С.П. Королева (национального исследовательского университета). Состояние и перспективы

А.В. Баскаков, Е.А. Симановский
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

На сегодняшний день термин грид уже не вызывает вопросов, о том, что это такое. Кратко, это распределенная программно-аппаратная компьютерная инфраструктура, объединенная телекоммуникационными каналами передачи данных.

Для чего нужно создавать подобную инфраструктуру? Задачи, появляющиеся перед научными работниками и инженерами с каждым днем становятся все более сложными, требующими огромных ИТ-ресурсов. Сосредоточить такие ресурсы в одной комнате, здании или даже в одном университете не всегда бывает возможно. В связи с этим, разумным, является способ объединения имеющихся компьютерных мощностей в единую управляемую среду с организацией удаленного доступа через глобальные сети связи.

В СГАУ, довольно активно ведутся работы в направлении объединения имеющихся ИТ-ресурсов в интегрированную грид-среду, с целью предоставления студентам и сотрудникам университета возможностей максимально эффективного использования оборудования университета для проведения научных исследований и разработок. Ниже, на рисунке 1 представлена топология грид-среды, где можно увидеть, какие ресурсы задействованы или планируются к включению в грид.

Центральным звеном грид-среды является телекоммуникационная сеть университета ядро которой построено на базе высокопроизводительного модульного коммутатора фирмы Cisco Systems, расположенного в здании межвузовского медиацентра (корпус 15). Модули коммутатора позволяют подключать сетевое оборудование с пропускной способностью до 10Гбит/с. Оборудование в корпусах университета подключено к центральному узлу одномодовым оптоволоконным кабелем. Пропускная способность каналов 1Гбит/с. На рисунке 2 изображена схематичная карта топологии телекоммуникационной сети СГАУ.

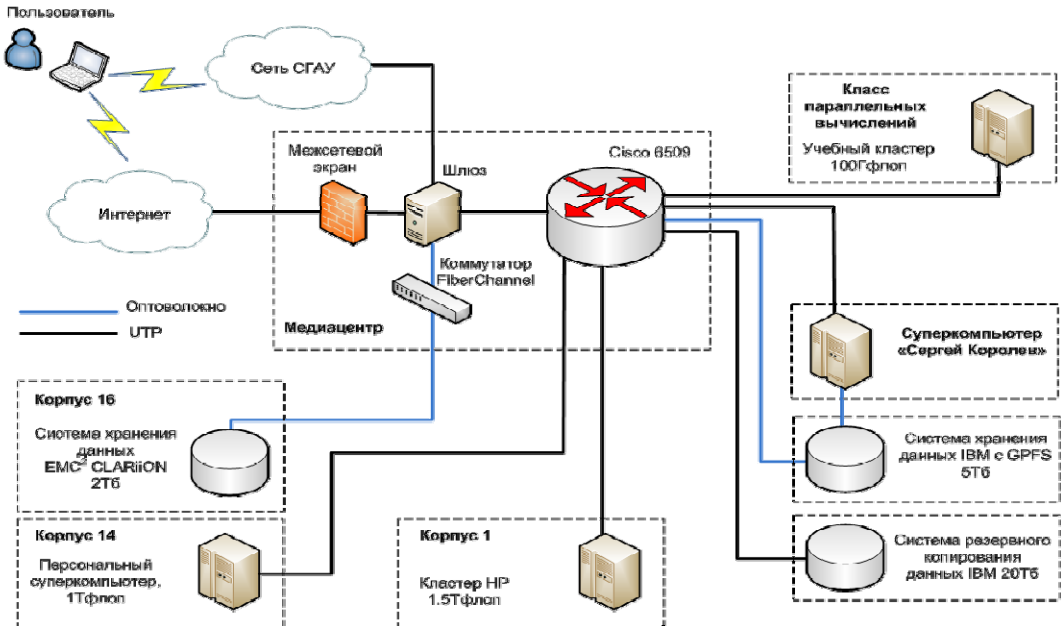


Рисунок 1 – Топология инфотелекоммуникационной грид-среды университета

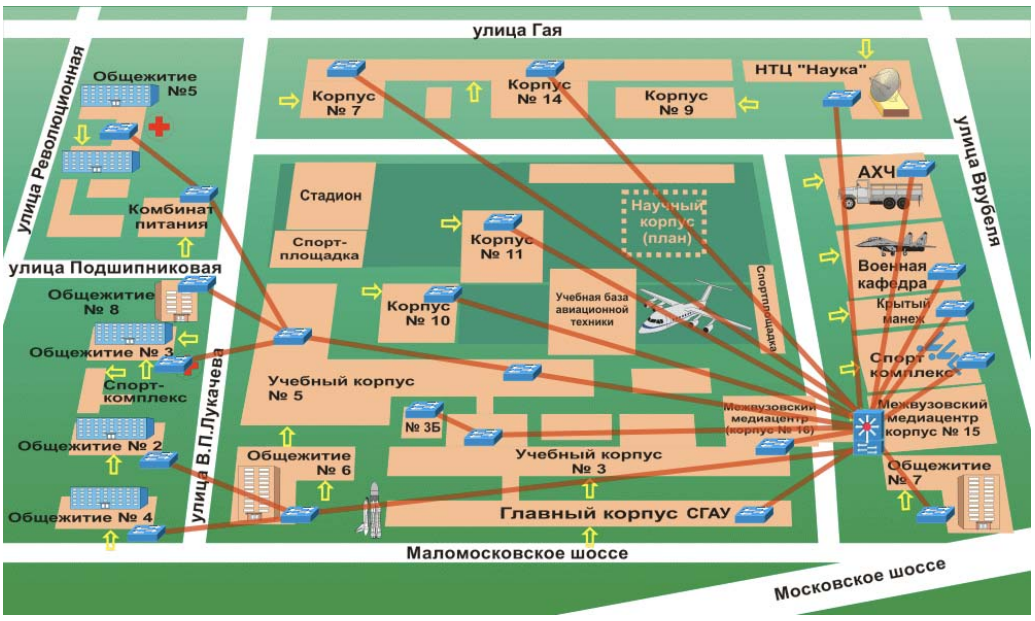


Рисунок 2 – Топология телекоммуникационной сети университета

В качестве ресурсов хранения данных, на сегодняшний момент подключены и используются: система хранения данных EMC CLARiiON ёмкостью 2Тб, расположенная в корпусе научно-технической библиотеки; дисковый массив кластера «Сергей Королев» на 1.2Тб; система хранения на HP на SCSI дисках, емкостью 2x410Гб.

Вычислительные ресурсы, доступные в грид:

1. Суперкомпьютер IBM «Сергей Королев»
   - Модель и количество серверов: Blade-сервер HS22, 112 шт.;
   - Тип и количество процессоров на сервер: 2 x Intel Xeon X5560, 2.80GHz, 8MB Cache 1333MHz;
   - Общее число процессоров / вычислительных ядер: 224/896
   - Модель и количество кластерных шасси: BladeCenter H, 8 шт.;
   - Общая оперативная память: 1,3125 Тб;

- Тип системной сети: InfiniBand DDR, QDR;
- Тип управляющей вспомогательной сети: Gigabit Ethernet;
- Пиковая производительность кластера: 10,0352 Тфлоп;
- Реальная производительность на тесте LinPack: 8,542 Тфлоп.

На рисунке 3 показан внешний вид суперкомпьютера.



Рисунок 3 – Суперкомпьютер «Сергей Королев»

2. Учебный вычислительный кластер на базе класса параллельных вычислений межвузовского медиацентра:
   - Модель и количество компьютеров: Dell Precision 690, 10 шт.;
   - Тип и количество процессоров: Intel Xeon X5355, 2.66GHz, 8MB Cache;
   - Общее число процессоров / вычислительных ядер: 9/36
   - Общая оперативная память: 80 Гб;
   - Тип системной сети и управляющей вспомогательной сети: Gigabit Ethernet;
   - Реальная производительность на тесте LinPack: 97,23 Гфлоп.

3. Кластер фирмы HP производительностью 1.5 Тфлоп, находящийся в первом корпусе университета на значительном расстоянии от основного кампуса, имеющий связь с центральным узлом сети на скорости 1 Гбит\с:
   - Модель и количество серверов: HP Proliant BL220c, 14 шт.;
   - Тип системной сети и управляющей вспомогательной сети: Gigabit Ethernet;
   - Реальная производительность на тесте LinPack: 1,32 Тфлоп.

На рисунке 4 показан внешний вид кластера.



Рисунок 4 – Кластер HP

Программной основой построения грид-среды является операционная система GNU/Linux. Для этой операционной системы разработано огромное количество системного и прикладного программного обеспечения, распространяющегося с открытым исходным кодом.

Одной из ключевых концепций работы грид-среды СГАУ являтся доступность всех сервисов в единой точке входа. Ресурсы хранения данных объединены в общую сетевую файловую систему, так что пользователю не нужно переносить свои данные с одной вычислительной системы на другую, чтобы произвести расчет. Достаточно скопировать данные в свою домашнюю директорию и эти данные становятся доступны для обработки на всех вычислительных элементах грид-среды. Сетевая файловая система поддерживает списки разграничения доступа и квоты, что позволяет обеспечить безопасность хранения данных и контролировать использование ресурсов хранения. Другие элементы хранения данных, такие как системы резервного копирования на ленточных накопителях, имеют интерфейсы сопряжения, легко доступные пользователю.

Второй основной сервис, предоставляемый грид-средой, это сервис высокопроизводительной обработки данных. Этот сервис предоставляется несколькими кластерными системами. Предназначен, в первую очередь, для запуска параллельных программ с использованием программного интерфейса MPI. Для этого на кластерах установлено все необходимое программное обеспечение для компилляции и запуска таких приложений. Оптимизирующие параллельные компиляторы фирмы Intel, библиотека Intel MPI, а также открытые реализации протокола MPI mpich, mvapich и т.д. Для решения инженерных задач используется программное обеспечение компании ANSYS.

Вычислительные ресурсы грид-среды предоставляются всем пользователям на одинаковых условиях. Пользователь, может самостоятельно выбирать, на каком из кластеров запустить свою задачу в зависимости от загруженности систем, наличием тех или иных ресурсов, программного обеспечения и т.д.

Механизм регулирования распределения ресурсов, построен на базе системы пакетной обработки заданий (менеджера ресурсов). На каждом кластере работает серверная часть менеджера, управляющая очередью задач. Клиентская часть, установленная на шлюзе, позволяет подключаться к управляющему серверу и работать с кластером: ставить задания в очередь, останавливать, отменять, модифицировать собственные задания. Задачи на кластерах выполняются в пакетном режиме, т.е. без интерактивного контакта с пользователем, что является одним из сложных моментов для понимания работы с подобными системами для не подготовленного пользователя. В описании задания даются все необходимые параметры, описывающие требования задачи к вычислительным ресурсам: архитектура, тип операционной системы, количество процессоров, объем оперативной памяти и т.д. Затем задается имя программы, которая будет запущена на вычислительных нодах кластера. Вся остальная информация: ключи запуска, входные и выходные файлы, передаются в качестве аргументов программы. Возможно запускать задачи и в интерактивном режиме, например для выполнения отладки параллельной программы.

Для того чтобы начать работать в грид-среде необходимо зарегистрироваться, подав заявку в письменной форме начальнику управления информатизации. Получив логин/пароль для доступа к среде, пользователь может со своего компьютера подключиться к любому управляющему серверу. Так как сетевая файловая система, объединена со всеми кластерными установками суперкомпьютерного центра (далее СКЦ), неважно к какому кластеру подключаться. В целях безопасности, вход на фронтальные машины кластеров ограничевается определенными ip-адресами,

указанными в заявке при регистрации. Используя программы для передачи файлов по протоколам ftp, sftp, например: scp, Winscp, FilZilla и т.п., файлы исходных данных передаются в домашнюю дитекторию.

Для постановки заданий в очередь системы пакетной обработки заданий формируется файл задания и с помощью программы *qsub* отправляется в очередь. В СКЦ СГАУ на всех кластерных установках, используется система с открытым исходным кодом Torque. На рисунке 5 представлена диаграмма работы.

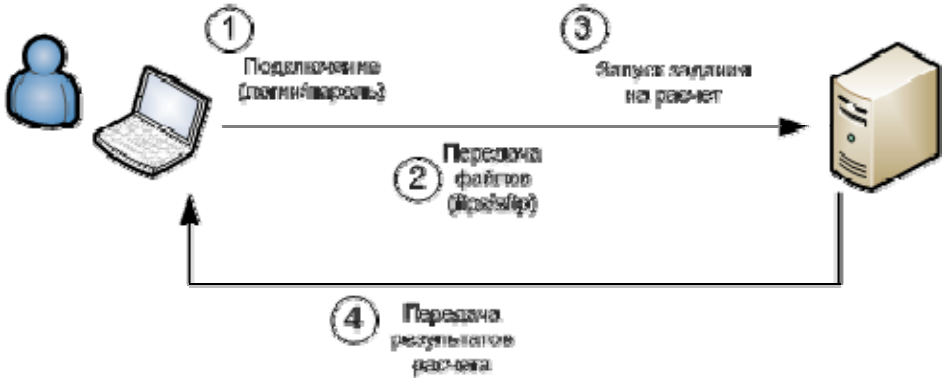


Рисунок 5 – Диаграмма работы с вычислительными системами среды

Развитие грид-среды университета предполагается по нескольким направлениям.

Систему безопасности среды планируется развивать в сторону использования персональных цифровых сертификатов для работы с ресурсами грида. Для начала необходимо создать центр сертификации, который будет подписывать сертификаты пользователей. На системах СКЦ будет установлено программное обеспечение открытого проекта грид-компьютинга Globus Toolkit обеспечивающее необходимую программную поддержку.

Развитие аппаратного обеспечения грид-среды СГАУ будет определяться в первую очередь требованиями задач пользователей. Сейчас необходимо увеличивать емкости систем хранения данных, так как существующие объемы явно недостаточны для проведения серьезных расчетов. Так же, для задач требующих высокоскоростного ввода-вывода необходимо обеспечить высокопроизводительный доступ к системе хранения с помощью технологии паралелльной файловой системы, так как практика показала, что использование сетевой файловой системы NFS является узким местом в производительности таких задач. Наращивание производительности вычислительных ресурсов планируется с суперкомпьютера «Сергей Королев». Дополнительно к нему будут установлены еще 112-168 блейд серверов, что позволит увеличить его производительность до 25-30 Тфлоп.

Телекоммуникационную компоненту среды, также планируется развивать в сторону увеличения пропускной способности каналов передачи данных как внутри грид-среды, так и за пределами её для обеспечения пользователям высокоскоростного доступа к ресурсам. В частности скорость внутренних каналов планируется довети до 10Гбит/с, а внешних до 1Гбит/с.

Все ресурсы будут и дальше интегрироваться в единую среду с простым доступом к каждой системе.

Автоматизация параллельного программирования с использованием системы Graphplus templet

С.В. Востокин
Самарский государственный аэрокосмический университет им. академика С.П Королева (национальный исследовательский университет), Самара

В работе представлена новая система автоматизации параллельного и распределенного программирования Graphplus templet, разработанная в рамках исследовательского проекта «Граф Плюс» (graphplus.ssau.ru) Самарского государственного аэрокосмического университета (национального исследовательского университета).

Актуальность, постановка задачи, цели разработки

На основании известного и собственного опыта параллельного программирования можно выделить следующие актуальные проблемы, на преодоление которых направлена предлагаемая система автоматизации программирования.

Практика показывает, что программист средней квалификации обычно не может самостоятельно разработать параллельный алгоритм решения задачи. Более того, при наличии точного описания алгоритма возникают проблемы его реализации требуемыми средствами. Это кардинально отличается от последовательного программирования, где процесс перехода от спецификации к реализации в настоящее время достаточно формализован. Поэтому разработка параллельных программ должна быть основана на максимальном повторном использовании имеющихся решений. Известно около двух десятков типовых решений в области параллельного программирования. Требуется предложить способ удобного описания и конструирования программ на их основе, пригодный для автоматического использования. Причем такие типовые решения должны быть по возможности языково-нейтральными и предусматривать реализацию кода на языках, используемых в коммерческой разработке Java, C#, C++.

Существует ряд проблем, обусловленных неудобством базовых средств программирования параллельных вычислений. В первую очередь - надежность кода. Известно, что недетерминированный характер вычислительного процесса затрудняет выявление ошибок методами отладки. Также возникают трудности при переносе программ в различные параллельные архитектуры. Многопоточные программы обычно не удается сделать распределенными, если это не предусматривалось изначально при проектировании. Представляет большой интерес и разработка масштабируемых программ, не теряющих эффективность при увеличении аппаратных ресурсов. Общий подход в решении перечисленных проблем – использование формальной модели вычислений, которая позволяет выполнить строгую декомпозицию кода, более удобную по сравнению с аппаратно-ориентированной моделью многопоточности.

При наличии понимания особенностей параллельного алгоритма, методов его программирования, встает вопрос стоимости и сроков разработки программного продукта. Современные программисты привыкли работать в интегрированных средах (IDE). Процесс разработки и отладки параллельных программ в настоящее время отличается более низкой автоматизацией. Очевидный способ решения этой проблемы – встраивание системы параллельного программирования в известные процессы разработки программного обеспечения и интегрированные среды разработки, поддерживающие их. То есть такая система должна быть реализована как стандартное расширение известной среды разработки. В противном случае ее практическое использование становится проблематичным. Отладка также должна выполняться стандартными средствами среды разработки в последовательном режиме.

Таким образом, проектирование системы Graphplus templet выполнено с учетом следующих требований: (1) возможность работы с типовыми решениями на разных языках программирования (C++, C#, Java); (2) независимость прикладного кода, разрабатываемого в системе, от API управления вычислениями, а также взаимозаменяемость API без переработки приложений (например, API многопоточности Win32 на POSIX Threads); (3) совместимость с коммерческими интегрированными средами разработки (например, IDE Microsoft Visual Studio), в частности, с целью удобства отладки параллельных приложений стандартными средствами IDE.

К классам приложений, на которые в первую очередь ориентирована система, относятся приложения для научных вычислений (численного моделирования), оптимизированные для работы как в распределенной, так и разделяемой памяти. Система также предназначена для разработки прикладных протоколов асинхронного взаимодействия в коммерческих распределенных приложениях.

В работе описывается новая архитектура системы Graphplus templet и рассматривается, каким образом выбор архитектуры обеспечивает удовлетворение перечисленных требований к системе.

Основные результаты и научная новизна

Семантическая основа языка templet - модель процессов диффузного типа. В данной модели конструктивными элементами являются пассивные объекты – процессы и активные объекты – двунаправленные каналы, соединяющие их. Объекты выполняют свои методы при поступлении сообщений по каналам и в ответ могут сформировать произвольное число сообщений (от 0 до некоторого N).

Семантику отражает код на традиционном языке программирования (C++, C#, Java), построенный с учетом ограничений, диктуемых моделью. В отличие от традиционных подходов, соответствие кода и модели исполнения обеспечивается препроцессором templet в интерактивном режиме. Для этого в проект IDE помимо кода на целевом языке (C++, C#, Java) в терминах модели исполнения включается XML-файл, описывающий структуру этого кода. В тех случаях, когда требуется изменить структуру модели, редактируется ее XML-спецификация. Если требуется изменить код, не затрагивающий структуру модели, то редактируется код на целевом языке. Фактически код, отражающий семантику модели исполнения, представляет собой чередующиеся блоки кода, построенного препроцессором по XML спецификации, и блоки кода, написанные пользователем вручную. Контроль целостности осуществляется каждый раз при сборке проекта. Препроцессор templet запускается перед штатным компилятором, что в IDE Microsoft Visual Studio обеспечивается настройкой атрибутов включенного в проект файла XML-спецификации.

Код, аннотированный XML-спецификацией, является последовательным, исполняется только одним процессором и отлаживается методами, похожими на методы отладки оконного интерфейса, управляемого событиями. При этом, для трассировки открыт и последовательный механизм исполнения модели (он поставляется отдельным модулем в исходных кодах).

Для получения собственно параллельного приложения на заданном API для параллельных вычислений выполняется так называемая процедура отображения на программно-аппаратную архитектуру. Эта процедура делает «распараллеливание» логически отлаженного последовательного кода, ориентируясь на его XML-спецификацию. В докладе описывается расширяемая архитектура templet-препроцессора, обеспечивающая настройку на разные типы отображений. Также показан принцип описания семантики XML-спецификации, принятый в системе Graphplus templet.

Сравнение с известными подходами

В основе представленной системы автоматизации параллельного программирования лежит система визуального программирования GraphPlus [1]. Усовершенствования коснулись модели программирования, в которую была добавлена концепция каналов, используемая, например, в операционной системе Singularity и языке программирования Sing# [2]. Подвергся пересмотру метод определения семантики XML-спецификации. В отличие от работы [3], где семантика модели описана в терминах темпоральной логики, применен метод эталонной реализации [4], упрощающий понимание системы программистом. В язык XML-спецификации введена возможность описания типовых решений (паттернов, шаблонов) [5].

В основе концепции данной системы лежит парадигма языково-ориентированного программирования [6], в рамках которой система может рассматриваться как пример доменно-ориентированного языка для параллельных вычислений. Система предназначена для реализации и использования паттернов параллельного программирования [7] в более гибкой форме, чем в соответствующих каркасных библиотеках. В частности, как в работе [8], единая вычислительная модель служит целям композиции типовых решений. Параллелизм в представленной системе вводится не как расширение языка или собственный язык, а в форме библиотеки времени исполнения [9], однако для обеспечения переносимости определяется лишь ее эталонная реализация. Система занимает промежуточное положение между языком и каркасной библиотекой, построена на основе препроцессора, и в этом похожа на систему программирования Qt [10]. Однако она не имеет жесткой привязки к языку программирования C++, не выполняет даже лексический разбор языковых конструкций, что значительно упрощает дизайн.

Общеупотребительными являются несколько способов введения семантики параллельного исполнения в язык программирования. Каркасная библиотека, например TBB на языке C++[11], жестко связывает решение с языком программирования. Расширение языка требует написания как минимум лексического анализатора языка, что может оказаться неприемлемо сложным для практического применения, хотя в некоторых случаях удается построить компактные расширения, например T++ [12]. Использование распараллеливающих директив в форме комментариев, например OpenMP [13] или DVM [14], тоже предполагает наличие сложного компилятора, но кроме этого, заставляет пользователя вручную приводить структуру кода в соответствие с моделью программирования этих средств. Использование приема генерации исходного кода вместо его анализа снимает данную проблему.

Семантика кода в представленной системе основана на модели процессов. Данный выбор обусловлен широкими описательными возможностями, императивностью, удобством декомпозиции сложных систем на ее основе. Несмотря на достоинства функциональной модели, и в том числе возможности описания некоторых паттернов параллелизма в ее терминах [15], мы считаем наш подход более универсальным. Использование модели процессов диффузного типа [16] позволяет выполнить ее реализацию базовыми средствами процедурно-ориентированных языков, например, избегать манипулирования со стеком вызовов.

Заключение

Эксперименты с созданной нами системой программирования Graphplus templet подтверждают возможность практического применения изложенных в работе решений. Проверка функциональности проведена при проектировании и кодировании наиболее распространенных типовых схем управления вычислениями: «применить ко всем» (MAP); «портфель задач» (TASKBAG); «конвейер / цепь асинхронных процессов» (PIPELINE / CHAIN). Разработана система отображения программ на SMP архитектуру,

использующая синхропримитивы интерфейса Win32 для многопроцессорных и многоядерных компьютеров под управлением операционных систем семейства Microsoft Windows на ядре NT. Примеры иллюстрируют снижение трудоемкости и повышение надежности параллельного и распределенного программирования с использованием системы Graphplus templet.

В комплект программ Graphplus templet входит транслятор модели / препроцессор, модуль с отладочным последовательным кодом системы исполнения, модуль многопоточного исполнения для Win32 API, примеры тестовых приложений и опционный графический редактор файлов XML-спецификации. Система реализована на языке C/C++, использует SAX-парсер Expat 2.0.1 для разбора файлов XML-спецификации. В настоящий момент Graphplus templet поддерживает программирование приложений на языке C++. В прототипах системы также тестировались варианты реализации типовых схем на языках Java и C#. Код системы зарегистрирован Федеральным институтом промышленной собственности (ФИПС), регистрационный номер 2010615139 . Предполагается использование данной системы автоматизации на суперкомпьютере «Сергей Королев» типа IBM BladeCenter HS22 производительностью 10 TFLOPS, установленном в Самарском государственном аэрокосмическом университете.

Литература

1. Востокин С.В. Графическая объектная модель параллельных процессов и ее применение в задачах численного моделирования. Самара: Издательство Самарского научного центра РАН, 2007. 286 с. (http://graphplus.ssau.ru/docs//GraphicalObjectModel.pdf)
2. Galen C. Hunt and James R. Larus, Singularity: Rethinking the Software Stack // ACM SIGOPS Operating Systems Review, vol. 41, no. 2, pp. 37-49, Association for Computing Machinery, Inc., April 2007.
3. Востокин С.В. Спецификация визуальной модели параллельных и распределенных вычислений на основе логики TLA // Труды Четвертой Международной конференции «Параллельные вычисления и задачи управления» PACO'2008. Москва, 27-29 октября 2008 г. М.: Институт проблем управления им. В.А. Трапезникова РАН, 2008, С. 1338-1348.
4. Dalci, Eric; Fong, Elizabeth; Goldfine, Alan. Requirements for GSC-IS Reference Implementations. National Institute of Standards and Technology, Information Technology Laboratory, 2003.
5. Гамма Э., Хелм Р., Джонсон Р., Влиссидес Дж. Приемы объектно-ориентированного проектирования. Паттерны проектирования. Спб: Питер, 2003. 368 с.: ил.
6. Дмитриев С. Языково-ориентированное программирование: следующая парадигма // RSDN Magazine, №5. 2005.
7. Шмидт Д., Хьюстон С. Программирование сетевых приложений на C++. Том 2. М.:ООО "Бином-Пресс", 2004. 400 с.: ил.
8. Бергизияров П.К. Программирование на типовых алгоритмических структурах с массивным параллелизмом // Вычислительные методы и программирование. 2001. Т.2, Разд. 2, С.1-16.
9. Страуструп Б. Язык программирования C++, 3-е изд. Спб.: М:. «Невский диалект» - «Издательство БИНОМ», 1999.
10. Земсков Ю.В. Qt 4 на примерах. СПб.: «БХВ-Петербург», 2008.
11. Reinders J. Intel Threading Building Blocks: Outfitting C++ for Multi-core Processor Parallelism. O'Reilly Media, 2007.
12. С.М. Абрамов, А.И. Адамович, А.В. Инюхин, А.А. Московский, В.А. Роганов, Ю.В. Шевчук Т-система с открытой архитектурой // Суперкомпьютерные системы и их применение SSA'2004: Труды международной научной конференции, 26-28 октября 2004 г., Минск, ОИПИ НАН Беларуси. Минск, 2004. С. 18-22.
13. OpenMP API specification for parallel programming. http://openmp.org. 2010.
14. Крюков В.А., Удовиченко Р.В. Отладка DVM-программ // Программирование. 2001. №3. С. 19-29.
15. Московский А.А., Первин А.Ю., Сергеева Е.О. Первый опыт реализации шаблона параллельного программирования на основе Т-подхода // Программные системы: теория и приложения: тр. Междунар. конф. М.: Наука. Физматлит. 2006. Т. 1.
16. Dijkstra E.W., Scholten C.S. Termination detection for diffusing computations // Information Processing Letters. Augest 1980. No 11(1). P. 1-4.

Графический редактор системы визуального моделирования параллельных вычислений (PGRAPH)

П.В. Аболмасов, А.Н. Коварцев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Одной из проблем как последовательных, так и параллельных вычислений, является наглядное представление вычислительных алгоритмов с целью создания такой нотации (желательно графической), которая упрощает разработку алгоритмов конечным пользователям – специалистам в различных предметных областях.

Использование графических моделей позволяет ускорить разработку алгоритмов параллельных вычислений и повысить качество параллельных программ, поскольку допускает более компактное, чем текст, интуитивно понятное и легко формализуемое представление алгоритма, исключающее множество ошибок при его разработке и удобное для построения методов автоматического анализа модели.

В области описания вычислительных алгоритмов наиболее удобной представляется форма, близкая к традиционным блок-схемам. Проектирование вычислительных алгоритмов в модели, близкой к блок-схемам, реализовано в технологии графо-символического программирования (ГСП-технологии), разработанной на кафедре программных систем Самарского государственного аэрокосмического университета [1].

Идея ГСП-технологии заключается в замене текстового описания программы, написанной на языке высокого уровня, на графическое путём представления вычислительного процесса в виде помеченного ориентированного графа. Вершины графа соответствуют обработке данных, а дуги определяют передачу управления от вершины к вершине. В графе имеется специальным образом помеченная корневая вершина, в которой начинается вычислительный процесс, и одна или несколько тупиковых (концевых) вершин, где процесс завершается. В рамках ГСП-технологии создание программы представляется как процесс ее конструирования в виде графа из набора заранее подготовленных программных блоков [2].

Настоящая работа посвящена программной реализации теоретической модели ГСП-технологии с использованием современных средств разработки.

Требования к подсистеме редактирования

Разрабатываемый граф-редактор должен отвечать следующим требованиям:

- граф-редактор должен быть удобным для пользователя,
- исходный код граф-редактора должен быть более простым для сопровождения и модернизации,
- граф-редактора должен быть легко переносимым и кроссплатформенным,
- работать с СУБД MySQL.

Для решения этих задач был выбран язык C++ и дополнительная библиотека классов Qt [3], предоставляющая широкие возможности для работы с графикой.

Описание подсистемы редактирования

Разработанная подсистема редактирования граф-моделей состоит из трех модулей:

- менеджер работы с БД
- модуль взаимодействия с пользователем
- и сам граф-редактор – модуль, отвечающий за создание граф-модели.

Менеджер работы с БД отвечает за сохранение граф-модели и данных предметной области в базу данных. Описание граф-моделей, создаваемых в системе ГСП «PGRAPH», хранится в информационном фонде, который реализован средствами реляционной базы данных (БД).

Укрупнено структура информационного фонда системы ГСП «PGRAPH» представлена на рисунке 1.

Из рисунка 1 видно, что информационное обеспечение состоит из двух групп БД: из баз данных предметной области и из системных баз данных. Группа баз данных предметной области содержит БД с информацией об исходных модулях; БД с информацией об акторах и предикатах; БД с информацией об агрегатах; БД с информацией о переменных ПО. Отдельно хранится информация о графическом образе агрегатов. Группа системных баз данных содержит информацию, необходимую для работы компилятора граф-объектов, информацию о системных ошибках.

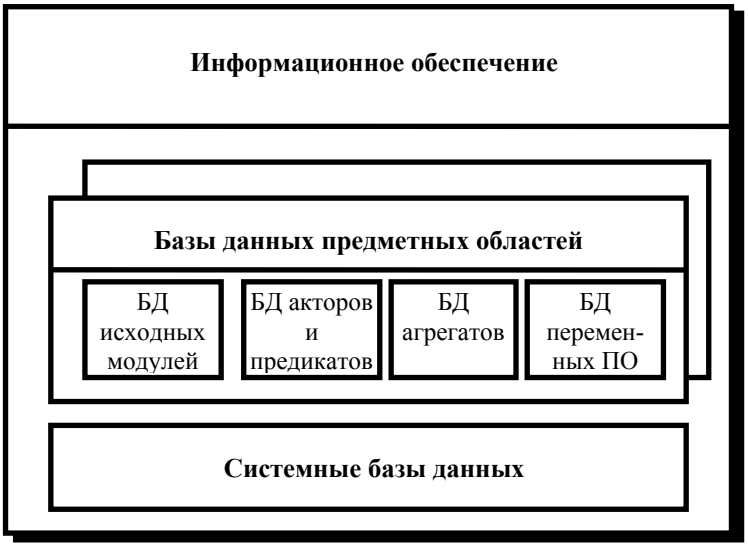


Рисунок 1 - Структура информационного обеспечения системы PGRAPH

Информационный фонд системы организован в виде таблиц, управляемых СУБД MySQL. Для поддержки информационного фонда используются плагин Qt для работы с этой базой данных.

Модуль взаимодействия с пользователем представляет собой окно с набором инструментов и различных меню и является контейнером для граф-редактора. Через пункты главного меню можно вызвать необходимый диалог для работы с соответствующим справочником ПО (редактор справочника акторов, предикатов, переменных, типов и т.д.). Главная панель инструментов содержит ряд наиболее часто вызываемых операция таких как: сохранить, открыть, вызвать редактор объектов, компилировать и т.д. Левая панель инструментов содержит набор кнопок с иконками, обозначающими графические примитивы, нажав на одну из них можно установить соответствующий режим граф-редактора. Центральное место в окне приложения занимает граф-редактор, о возможностях которого пойдет речь ниже. Внешний вид приложения показан на рисунке 2.

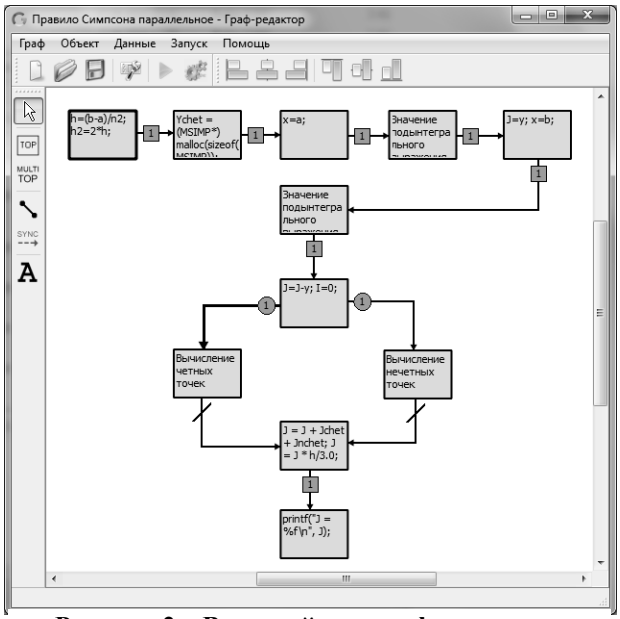


Рисунок 2 – Внешний вид граф-редактора

Описание граф-редактора

Благодаря структуре приложения и возможностям Qt граф-редактор может быть использован не только в рамках разработанного интерфейса взаимодействия с пользователем, описанного выше, но и встроен в другое приложение или даже использован как элемент ActiveX на web-странице в браузере. Хотя сам по себе граф-редактор так же может существовать, как окно приложения, например, это используется при работе с иерархической структурой графа, когда пользователь может посмотреть, какой подграф используется в той или иной вершине граф-модели.

В каждый момент времени редактор находится в одном из 6-х состояний: добавление вершины, рисование дуги, добавление комментария, добавления многопоточной вершины, рисования дуги синхронизации или указателя, позволяющий выделять и перемещать объекты. Базовым элементом управления редактора является сцена, на которой располагаются графические примитивы. Сцена самостоятельно умеет определять столкновения, изменение состояния объектов и вызывать их перерисовку при необходимости. В зависимости от установленного режима редактора, сцена обрабатывает действия мыши над ней. Здесь наиболее сложным, конечно, является алгоритм автоматической перерисовки дуг при перемещении вершины, т.к. для достижения большей логичности и простоты граф-модели в ней допускаются только вертикальные и горизонтальные составляющие управляющих дуг. Также для сохранения внешнего вида модели необходимо как можно дольше сохранять максимальное число вершин дуги при перемещении вершины.

Описание графических примитивов

Для построения граф-редактора необходимо разработать физическую модель данных, которая обеспечивала бы возможность эффективного выполнения его функций и отвечала требованиям теоретической модели ГСП-технологии. UML-диаграмма классов, реализующих основные графические примитивы граф-редактора, представлена на рисунке 3.

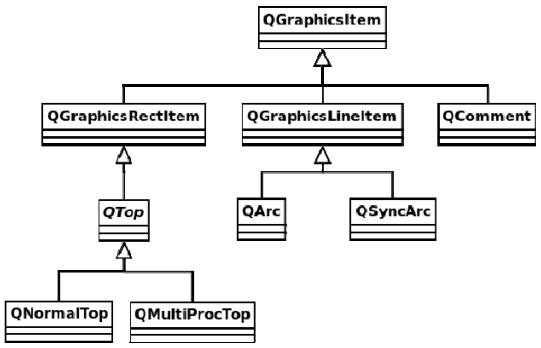


Рисунок 3 – UML-диаграмма классов граф-редактора

Класс QGraphicsItem является базовым классом для всех графических элементов на сцене. Он предоставляет удобную основу для написания своих собственных элементов.

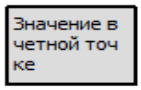
Один из основных элементов модели – это вершина. QTop является абстрактным классом описывающим основные свойства вершины такие как: положение, актор, номер, признак корневой вершины и.т.д., но не содержит реализации отрисовки элемента. Вершина знает обо всех инцидентных ей дугах. Вершина имеет контекстное меню, позволяющее изменять ее свойства, а так же в меню содержится пункт, позволяющий удалить вершину, и пункт, позволяющий сделать вершину корневой. В графе может быть только одна корневая вершина, с которой начинается работа алгоритма

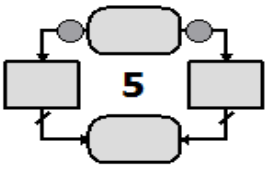
Дочерние классы QNormalTop и QMultiProcTop реализуют функционал соответственно обычной или многопоточной вершины и содержат метод отрисовки элемента.

Многопоточная вершина отличается от обычной вершины видом (см. рис. 4). Она изображается как граф с 2-мя параллельными ветвями, а в центре пишется фактическое количество параллельных процессов, скрывающихся в вершине. Для каждого типа вершины имеется свой диалог редактирования свойств вершины. У

обычной вершины можно изменять размер и актор. В свойствах многопоточной вершины можно задать актор и число параллельных процессов. Пользователь указывает, как использовать выбранный актор: одинаково для всех процессов или создать для каждого процесса свой актор на основе выбранного, при этом будет создано необходимое число переменных. Если в вершине выбран актор, то граф-редактор автоматически создаст подграф, состоящий из n+2 вершины, где n – число параллельных процессов. Топология данного подграфа проста: одна начальная вершина, из которой выходит n параллельных дуг, соединенных с n вершинами, столько же терминирующих дуг, и одна конечная вершина. По-умолчанию для начальной и конечной вершины назначается пустой актор (актор ничего не делает), однако пользователь может открыть сгенерированный граф и отредактировать его.

Другой основной элемент графа – дуга управления. Класс QArc реализует представление дуги в программе. Из статических свойств дуга имеет: приоритет, указатель на начальную и конечную вершину, имя предиката и тип. Объект дуги состоит из набора примитивов: линий, объекта-вершины и полигона-стрелки. Объект-вершина имеет различную форму в зависимости от типа дуги (см. рисунок 5) и имеет контекстное меню, в котором есть пункты, позволяющие удалить дугу и изменить ее свойства. Полигон-стрелка имеет декоративное назначение. Толщина дуги определяется ее приоритетом. Самая приоритетная дуга (с приоритетом 1) – самая толстая. Дуга хранит в себе список линий – отрезков, из которых она состоит.

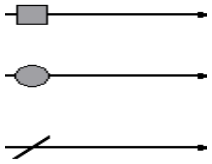




а) Обычная вершина

б) Многопоточная вершина

Рисунок 4 – Внешний вид элемента “вершина”



- последовательная дуга

- параллельная дуга

-терминирующая дуга

Рисунок 5 – Внешний вид элемента “дуга”

В граф-модель вводится дополнительный элемент – комментарий, который служит для пояснения граф-модели и может быть вставлен в произвольное место.

Заключение

Таким образом, в результате работы создана программа граф-редактор, реализующая указанные методы и средства, предназначенная для построения моделей алгоритмов параллельных вычислений и автоматического синтеза параллельных программ на их основе. Результаты работы представляют большую практическую значимость, поскольку позволяют упростить и ускорить разработку алгоритмов параллельных вычислений специалистами в различных предметных областях, предоставляя интуитивно понятную модель описания вычислений, средства автоматического анализа модели и возможность повторного использования обширного фонда алгоритмов и программ, созданных в технологии ГСП, при переходе к параллелизму вычислений.

Литература

1. Коварцев А.Н. Автоматизация разработки и тестирования программных средств. - Самарский государственный аэрокосмический университет, Самара, 1999. - 150 с.

2. Коварцев А.Н., Жидченко В.В. Программный комплекс разработки алгоритмов и программ параллельных вычислений. - Материалы Девятой международной конференции-семинара Высокопроизводительные параллельные вычисления на кластерных системах, , Владимир, 2009. - С. 225-229

3. International. Products – Qt - A cross-platform application and UI framework, http://qt.nokia.com/products 1.08.2010.

Идентификация с последовательной схемой отбора согласованных оценок

С.А. Бибиков, А.В. Гаврилов, В.А. Фурсов
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара
Институт систем обработки изображений РАН, Самара

1. Формулировка задачи согласованной идентификации

Рассматривается задача идентификации линейной по параметрам модели вида

$$z(t_i) = \sum_{j=1}^{M} c_j \cdot s_j(t_i) + \xi(t_i), \tag{1}$$

где $z(t_i)$, $s_j(t_i), j = \overline{1,M}$ – заданные и/или наблюдаемые процессы на входе и выходе идентифицируемого объекта, соответствующие значению независимой переменной t_i, c_j, $j = \overline{1,M}$ – искомые параметры модели, а $\xi(t_i)$ – ошибки, связанные с неточностью задания модели, шумами наблюдений и др.

Предположим, что в рамках модели (1) для N различных значений аргумента t_i $(N > M)$ заданы значения $s_j(t_i), j = \overline{1,M}, i = \overline{1,N}$ и получены в результате наблюдения значения $z(t_i), i = \overline{1,N}$. Из N векторов-строк

$$\mathbf{s}_i = \left[s_1(t_i), s_2(t_i), \ldots, s_M(t_i) \right], i = \overline{1,N}$$

составим $N \times M$-матрицу $\mathbf{S}$, а из N значений $z(t_i), i = \overline{1,N}$ и N соответствующих им ошибок $\xi(t_i)$ составим $N \times 1$-векторы $\mathbf{Z}$ и $\boldsymbol{\Xi}$ соответственно. Если вектор параметров модели $\mathbf{c}$ остается неизменным для всех N наблюдений, с использованием введенных обозначений можно записать следующее матричное равенство:

$$\mathbf{Z} = \mathbf{S}\mathbf{c} + \boldsymbol{\Xi}. \tag{2}$$

Задача идентификации заключается в том, чтобы по $N \times M$-матрице $\mathbf{S}$ и $N \times 1$-вектору $\mathbf{Z}$ $(N > M)$, при неизвестном $N \times 1$-векторе ошибок $\boldsymbol{\Xi}$ построить оценку $\hat{\mathbf{c}}$ вектора параметров $\mathbf{c}$.

Обычно эту задачу решают в рамках обычных предположений линейной регрессии, ключевыми среди которых являются: $M\{\boldsymbol{\Xi}\} = 0$ и $\mathrm{cov}\{\boldsymbol{\Xi}\} = \sigma^2 \mathbf{E}_N$, где σ^2 - дисперсия отклонений, $\mathrm{cov}\{\cdot\}$ – $N \times N$-матрица ковариаций отклонений, а E_N - единичная $N \times N$-матрица. При этом оценка метода наименьших квадратов (МНК)

$$\hat{\mathbf{c}} = \left[\mathbf{S}^T \mathbf{S} \right]^{-1} \mathbf{S}^T \mathbf{Z} \tag{3}$$

является несмещенной и эффективной.

К сожалению, указанные предположения при малом числе наблюдений оказываются ненадежными, а использование методов статистического оценивания – малообоснованным. В данном случае должны быть предусмотрены меры по робастизации оценок параметров моделей. В работах [1, 2] предложен метод согласованной идентификации, в котором не используются априорные вероятностные модели полезных сигналов и шумов.

2. Метод согласованной идентификации с полным перебором

При построении алгоритмов согласованной идентификации с полным перебором [1] используются следующие предположения.

1.Матрица $\mathbf{S}$ и вектор $\mathbf{Z}$ фиксированы, т. е. $\mathbf{s}_{i,j}$, $\mathbf{z}_i$, $i=\overline{1,N}$, $j=\overline{1,M}$ известны в результате измерений на одной реализации.

2.Число наблюдений мало, поэтому имеет место неопределенность свойств обусловленности матрицы $\mathbf{S}$ и статистических характеристик вектора ξ.

3.Норма вектора ошибок $\Xi=\left[\xi_1,\xi_2,...,\xi_N\right]^T$ ограничена, а его направление в шаре

$$\Xi=\left\{\xi:\left(\xi^T\xi\right)^{1/2}=\left\|\xi\right\|_2\le R_\xi=\text{Const}\right\} \tag{4}$$

является произвольным и неопределенным.

4.В исходных данных, несмотря на возможные ошибки, все же содержится достаточное число наблюдений (подсистема, наиболее свободная от шума), по которым оценки ϵ могут быть вычислены с требуемой точностью.

5.Существует соответствующая уравнению (2) точная модель:

$$\mathbf{Z}^*=\mathbf{Sc}, \tag{5}$$

где $\mathbf{Z}^*=\mathbf{Z}-\Xi$.

Указанные предположения более реалистичны и менее обременительны, чем те, которые приведены выше и обычно используются в теории статистического оценивания.

Идея метода согласованной идентификации заключается в том, чтобы для исходной системы построить множество оценок на всех возможных подсистемах небольшой размерности и отобрать среди них подмножество наиболее согласованных между собой по критерию взаимной близости. На этом подмножестве затем строится искомая точечная оценка. Для достижения достаточно высокой точности в данном случае достаточно существования среди данных некоторого существенно менее зашумленного подмножества наблюдений.

Решение ищется среди *подсистем верхнего уровня* получаемых из P строк исходной системы. Всего формируется C_N^P систем верхнего уровня. Из строк, входящих в подсистему верхнего уровня, формируются все возможные *подсистемы нижнего уровня*, содержащие M строк. Таким образом, каждой подсистеме верхнего уровня соответствует множество систем нижнего уровня и соответствующих им промежуточных МНК-оценок. Для построения оценки выбирается подсистема верхнего уровня, обладающая минимальным значением критерия:

$$W_l\left[l\right]=\sum_{\substack{i,j=1,\\ i\ne j}}^{K_l}\left\|\epsilon_{l,i}-\epsilon_{l,j}\right\|_E^2,\ l=\overline{1,L}, \tag{6}$$

где $\epsilon_{l,i}$, $\epsilon_{l,j}$, $i=\overline{1,K_l}$, $j=\overline{1,K_l}$ оценки, полученные на подсистемах нижнего уровня каждой (l-й) подсистемы верхнего уровня.

Основная проблема построения согласованных оценок состоит в том, что для их вычисления в соответствии с критерием (6) необходимо осуществлять полный перебор возможных комбинаций оценок на подсистемах нижнего уровня для всех подсистем верхнего уровня. Даже при сравнительно небольшой размерности задачи такой перебор становится неприемлемым для обычных вычислительных средств. В данной работе предлагается схема последовательного отбора наиболее согласованных оценок из общего пула всех возможных оценок, полученных на подсистемах нижнего уровня.

При этом не строятся подсистемы верхнего уровня, а для остановки процедуры задают лишь допустимое число используемых оценок, взятых из пула.

3. Последовательная схема отбора согласованных оценок

Для реализации предлагаемой схемы формирования множества согласованных оценок задается начальная оценка, обозначаемая далее как $\boldsymbol{\epsilon}_1$: $\boldsymbol{\epsilon}_1 \in \Phi_1 \subset \Omega$, вычисленная на некоторой подсистеме нижнего уровня. Далее множество согласованных оценок формируется путем последовательного присоединения к созданному на очередном шаге промежуточному множеству $\Phi_q, q = 1,2,...$ оценки $\boldsymbol{\epsilon}_{q+1}$ из дополнения к формируемому множеству $\left(\boldsymbol{\epsilon}_{q+1} \notin \Phi_q\right)$, удовлетворяющей условию

$$\boldsymbol{\epsilon}_{q+1}: \quad w\left[\boldsymbol{\epsilon}_{q+1}\right] = \min_{\forall \boldsymbol{\epsilon}_{q+1} \notin \Phi_q} \sum_{i=1}^{q} \left\| \mathbf{c}_i - \mathbf{c}_{q+1} \right\|_E^2 . \tag{7}$$

В качестве критерия остановки задается допустимое число присоединяемых оценок.

Имеет место следующее

Утверждение 1:

Пусть $\boldsymbol{\epsilon}_1$ – любая оценка из множества согласованных оценок: $\boldsymbol{\epsilon}_1 \in \Phi_1 \subset \Omega$, удовлетворяющих критерию (6). Тогда множество согласованных оценок, сформированное по правилу (7) и содержащее C_n^M оценок, совпадает с множеством согласованных оценок для подсистемы верхнего уровня, удовлетворяющей критерию (7) и построенной методом полного перебора.

Доказательство.

Для всего множества последовательно присоединяемых векторов оценок выпишем значения сумм (7):

$$w\left[\boldsymbol{\epsilon}_2\right] = \min_{\forall \boldsymbol{\epsilon}_i \notin \Phi_1} \left\| \mathbf{c}_i - \mathbf{c}_1 \right\|_E^2 , \tag{8}$$

$$w\left[\boldsymbol{\epsilon}_3\right] = \min_{\forall \boldsymbol{\epsilon}_i \notin \Phi_2} \left\{ \left\| \mathbf{c}_i - \mathbf{c}_1 \right\|_E^2 + \left\| \mathbf{c}_i - \mathbf{c}_2 \right\|_E^2 \right\}, \tag{9}$$

…

$$w\left[\boldsymbol{\epsilon}_{C_S^M}\right] = \min_{\forall \boldsymbol{\epsilon}_i \notin \Phi_{C_S^M - 1}} \sum_{j=1}^{C_S^M - 1} \left\| \boldsymbol{\epsilon}_i - \boldsymbol{\epsilon}_j \right\|_E^2$$

Если число последовательно присоединяемых по указанной схеме векторов оценок равно C_n^M, соответствующая их объединению сумма принимает вид:

$$W\left[C_S^M\right] = \sum_{q=2}^{C_S^M} w\left[\boldsymbol{\epsilon}_q\right] = \sum_{q=2}^{C_S^M} \left\{ \min_{\forall \boldsymbol{\epsilon}_i \notin \Phi_q} \sum_{j=1}^{q} \left\| \boldsymbol{\epsilon}_i - \boldsymbol{\epsilon}_j \right\|_E^2 \right\}.$$

Нетрудно заметить, что $W\left[C_S^M\right]$ является минимальным значением суммы квадратов разности всех возможных парных сочетаний множества C_n^M оценок, для подсистемы, содержащей n уравнений. Следовательно $W\left[C_n^M\right]$ суть функция взаимной близости оценок для подсистемы верхнего уровня, удовлетворяющей критерию (7), а множество входящих в эти частные суммы векторов оценок совпадает с множеством согласованных оценок, сформированным путем полного перебора всех подсистем верхнего уровня. ■

Схема последовательного отбора согласованных оценок дает существенный выигрыш в вычислительном отношении. Однако, достижение качества оценок, сравнимого со случаем полного перебора, возможно лишь при условии, что начальная

оценка $\mathbf{\epsilon}_1$ принадлежит множеству согласованных оценок: $\mathbf{\epsilon}_1 \in \Phi_1 \subset \Omega$, удовлетворяющих критерию (7). Поэтому представляет самостоятельный интерес задача построения начальной оценки $\mathbf{\epsilon}_1$.

Пусть

$$\Delta\mathbf{\epsilon}_k = \left[\mathbf{S}^T\mathbf{G}_k\mathbf{S}\right]^{-1}\mathbf{S}^T\mathbf{G}_k\Xi$$

ошибка идентификации на k-й подсистеме нижнего уровня.

Утверждение 2:

Для того чтобы оценка $\mathbf{\epsilon}_1$ метода наименьших модулей, построенная на множестве всех C_N^M подсистем нижнего уровня, принадлежала множеству согласованных оценок $\Omega, |\Omega| = C_n^M$, достаточно, чтобы существовало множество $]n[$ наблюдений, где $]n[$ ближайшее (в сторону увеличения) целое число, удовлетворяющее неравенству

$$2\left(]n[-M+1\right)...\left(]n[-1\right)]n[> \left(N-M+1\right)...\left(N-1\right)N, \tag{10}$$

такое, что для всех сформированных на нем подсистем нижнего уровня

$$\max_{\forall\Delta\mathbf{c}_k\in\Omega}\left\{\lambda_{\min}^{-1}\left(k\right)R_{\xi}^2\left(k\right)\right\} < \min_{\forall\Delta\mathbf{c}_k\notin\Omega}\left\{\lambda_{\max}^{-1}\left(k\right)R_{\xi}^2\left(k\right)\right\} \tag{11}$$

где $R_{\xi}^2\left(k\right) = \left\|\xi\left(k\right)\right\|^2 = \xi^T\left(k\right)\xi\left(k\right)$, $\xi\left(k\right)$ – проекция вектора ошибок измерений подсистемы нижнего уровня на пространство столбцов матрицы $\mathbf{G}_k\mathbf{S}$, $\lambda_{\min}\left(k\right)$ и $\lambda_{\max}\left(k\right)$ – минимальное и максимальное собственные значения матрицы $\mathbf{S}^T\mathbf{G}_k\mathbf{S}$ подсистемы нижнего уровня с номером k.

Доказательство.

Для доказательства утверждения, во-первых, необходимо показать, что

$$2C_n^M > C_N^M \text{ или } C_{\frac{N+1}{2}}^M > \frac{1}{2}\cdot C_N^M.$$

Условие (10) вытекает из известных оценок [2] для квадрата евклидовой нормы вектора ошибок оценивания:

$$\lambda_{\max}^{-1}\left(k\right)R_{\xi}^2\left(k\right) \le \left\|\Delta\mathbf{c}\left(k\right)\right\|_2^2 \le \lambda_{\min}^{-1}\left(k\right)R_{\xi}^2\left(k\right),$$

где $\lambda_{\min}\left(k\right), \lambda_{\max}^{-1}\left(k\right), R_{\xi}^2\left(k\right)$ – те же, что и в (11). ■

Поскольку оценка метода наименьших модулей является средним вариационного ряда, построенного на множестве всех оценок нижнего уровня, приведенные условия гарантируют, что эта оценка будет принадлежать множеству согласованных оценок.

4. Реализация последовательного алгоритма согласованной идентификации

Опираясь на указанные утверждения можно построить следующий алгоритм определения согласованных оценок:

Шаг 1. Вычисление множества Σ всех возможных C_N^M оценок на подсистемах нижнего уровня.

Шаг 2. Вычисление на множестве Σ оценки, являющейся средним членом вариационного ряда (начального элемента множества Φ_1).

Шаг 3. Формирование множества согласованных оценок Ω, путем поочередного присоединения к множеству Φ_q векторов оценок из множества Σ, удовлетворяющих условию (7).

Шаг 4. Вычисление точечной оценки на множестве согласованных оценок.

Сравнение предложенного последовательного метода согласованной идентификации (ПСИ) с исходным методом согласованной идентификации (СИ) проводилось следующим образом. Для каждого эксперимента формировалась точная модель (5) размерности $M = 4$. С помощью точной модели формировались $N = 10$ значений откликов на случайный сигнал. После этого ко всем откликам добавлялся аддитивный гауссов шум. К двум произвольным откликам добавлялся дополнительный аддитивный гауссов шум, обладающий большей дисперсией (средний шума – в 3 раза, сильный – в 5 раз). По построенным уравнениям проводилась идентификация вектора параметров тремя методами: МНК, СИ и ПСИ. Для формирования оценки в исходном и модифицированном методах и использовалось P уравнений. Для каждого фиксированного набора параметров проводилось по 10000 экспериментов. Для каждого эксперимента сравнивались относительные отклонения от точных значений параметров модели МНК-оценок с отклонениями оценок, полученных методами СИ и ПСИ. На рисунках 1-4 представлены гистограммы распределения полученных отношений: сплошной линией отношение ошибок МНК к ошибкам СИ, пунктирной ошибки МНК к ошибкам ПСИ. Как видно из графиков, МСИ-оценки, оказываются лучше МНК-оценок в большем количестве случаев, чем СИ-оценки.

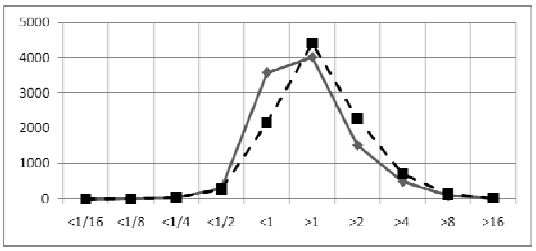


Рис. 1. – Количество используемых уравнений $P = 9$. Шум средний.

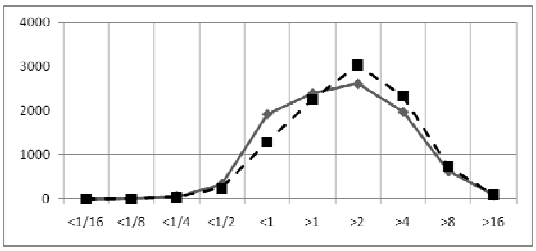


Рис. 2. – Количество используемых уравнений $P = 8$. Шум средний.

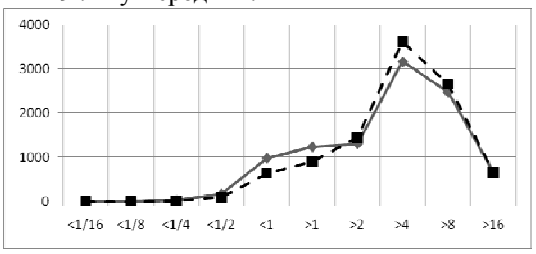


Рис. 3. – Количество используемых уравнений $P = 8$. Шум сильный.

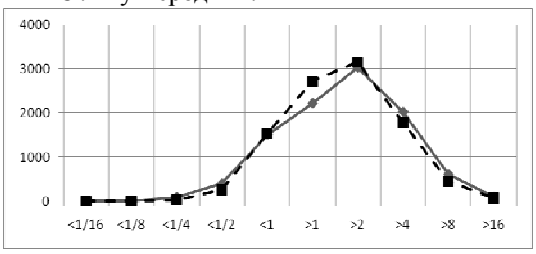


Рис. 4. – Количество используемых уравнений $P = 7$. Шум средний.

Литература

1. Фурсов В.А. Согласованная идентификация управляемого объекта по малому числу наблюдений [текст] / В.А. Фурсов / Мехатроника, автоматизация, управление. – №3 (108). – 2010. –С.2-8.

2. Фурсов В.А. Идентификация моделей систем формирования изображений по малому числу наблюдений [текст] / В.А. Фурсов. – Самар. гос. аэрокос. ун-т. Самара, 1998. – 218с.

Некоторые аспекты разработки гибридных параллельных эвристических алгоритмов оптимизации

О.И. Булычов
Ульяновский государственный университет, Ульяновск

В статье рассматриваются аспекты разработки параллельных эвристических алгоритмов оптимизации, реализованных в библиотеке параллельных метаэвристик HeO. Библиотека основана на использовании шаблонов проектирования, что позволяет получать универсальные и гибкие алгоритмы с возможностью гибридизации.

Оптимизационные задачи возникают в различных областях науки и техники, и для их решения разработан богатый арсенал методов. Однако нахождение точного решения многих задач на практике оказывается невозможным. Это может быть вызвано разными причинами: сложностью решаемой задачи, погрешностями входных данных, временными и аппаратными ограничениями и др. В связи с этим актуальным остается вопрос разработки эффективных приближенных методов решения таких задач.

Благодаря бурному развитию компьютерной техники вот уже несколько десятилетий для нахождения приближенного решения многих трудных оптимизационных задач (прежде всего задач дискретной оптимизации) с успехом применяются метаэвристические алгоритмы (см., например, [1]). На сегодняшний день разработано большое количество программных средств, в которых реализованы те или иные параллельные метаэвристические алгоритмы для решения конкретных задач. Одной из тенденций последнего десятилетия является создание универсальных библиотек параллельных метаэвристических алгоритмов, которые можно применять для решения широкого круга оптимизационных задач.

Кроме того, в последние годы большое внимание уделяется гибридизации метаэвристик, что позволяет на основе базовых алгоритмов получать новые, более эффективные методы оптимизации. Вопросы гибридизации метаэвристических алгоритмов в разных программных продуктах решаются по-разному. Учитывая большое разнообразие метаэвристических алгоритмов, универсального метода их гибридизации просто не существует, однако можно выделить некоторые общие подходы к гибридизации, применяемые на практике. Хороший обзор методов гибридизации метаэвристик приведен в [2].

В настоящей статье рассматриваются практические вопросы программной реализации и гибридизации параллельных метаэвристических алгоритмов в библиотеке HeO, разработанной с использованием современных технологий программирования, метапрограммирования и параллельных вычислений на языке C++. Подробное описание библиотеки можно найти в [3].

Библиотека HeO (**He**uristic **O**ptimization) является проектом с открытым исходным кодом и распространяется на основе лицензии MIT. Цель проекта — обеспечить исследователей современными и простыми в использовании средствами для решения широкого круга оптимизационных задач.

Ключевыми особенностями библиотеки являются:

кроссплатформенность, поддержка архитектур x86 и x64;

реализация методов оптимизации в виде проблемно-независимых алгоритмических каркасов;

реализация каждого метода оптимизации с использованием двух технологий параллельного программирования: MPI и OpenMP;

наличие мастеров (wizards), облегчающих создание пользовательских проектов;

прозрачность параллелизма для пользователей библиотеки.

прозрачность создания гибридных алгоритмов.

Официальная страница проекта располагается по адресу: http://www.code.google.com/p/heo.

Библиотека представляет собой набор классов, которые можно разделить на две большие группы: основные классы и вспомогательные классы.

Основные классы участвуют в непосредственной реализации того или иного метода оптимизации. Ядро основных классов библиотеки составляют так называемые *решатели* (solvers). Каждый решатель реализует определенный метод оптимизации в соответствии с определенной технологией параллельного программирования (MPI или OpenMP).

Решатели выполнены в виде шаблонов языка C++ в соответствии с идеологией алгоритмических каркасов. Это означает, что в них:

- реализована только проблемно-независимая часть алгоритма;
- скрыты все аспекты параллельной реализации.

Эти особенности позволяют применять решатели к широкому кругу задач и делают параллелизм прозрачным для пользователя библиотеки. Для использования решателей пользователем должны быть реализованы проблемно-зависимые классы, описывающие конкретную задачу оптимизации и ее решения.

Вспомогательные классы библиотеки выполняют всевозможные служебные функции (работа с конфигурационными файлами, запуск решателей, передача данных по сети, генерация псевдослучайных чисел и т.п.).

Для демонстрации возможностей библиотеки разработаны несколько проектов, в которых реализованы проблемно-зависимые классы для нескольких задач дискретной и непрерывной оптимизации.

В настоящий момент в библиотеке HeO реализованы три решателя: генетический алгоритм (GA — Genetic Algorithm), метод имитационной нормализации или имитации отжига (SA — Simulated Annealing), метод ветвей и границ (BnB – Branch and Bound) (MPI-версия BnB находится в стадии разработки), а также их гибриды (например, GA+SA, SA+BnB). Описания последовательных версий этих алгоритмов можно найти, например, в [4]. Ниже будет рассмотрен использованный в библиотеке подход к параллельной реализации этих методов, основанный на технологии шаблонного проектирования. Подробное изложение методов шаблонного проектирования приведено в [5].

Для разных типов метаэвристических алгоритмов могут быть предложены разные способы гибридизации. В частности, для алгоритмов GA и SA, наиболее популярными способами являются интеграционная (integrative) и кооперационная (cooperative) гибридизация. При интеграционной гибридизации один из алгоритмов является частью другого, например, SA может рассматриваться как один из генетических операторов, выполняемых в основном цикле GA. При кооперационной гибридизации алгоритмы обмениваются друг с другом решениями.

Наибольшие трудности возникают при практической реализации первой стратегии. Как правило, в существующих реализациях гибридного метода GA+SA жестко фиксируется структура алгоритма, а для ее модификации приходится вносить изменения в исходный код программы. Для решения этой проблемы в библиотеке HeO используется шаблон проектирования Factory (Фабрика), позволяющий сделать код создания объектов более универсальным, не привязываясь к конкретным классам, а оперируя лишь общим интерфейсом.

Все операторы гибридного алгоритма наследуются от абстрактного интерфейса abstract_operator. В исходном коде гибрида и фабрики имена классов-операторов нигде

не упоминаются, что упрощает добавление новых операторов. Каждый оператор регистрируется на фабрике с помощью вспомогательной функции.

Поскольку для правильной работы программы требуется лишь один экземпляр конкретной фабрики, то для ее конструирования используется шаблонный класс Singleton (Одиночка).

При обмене данными между решателями, реализующими различные алгоритмы, может потребоваться необходимость преобразования проблемно-зависимых классов, поскольку представления данных в разных алгоритмах могут быть различными. Для решения этой проблемы в библиотеке HeO используется шаблон проектирования Adapter (Адаптер), позволяющий не принимать во внимание различия между проблемно-зависимыми классами разных алгоритмов.

Для хранения и конвейерной обработки решений используется класс Solution_Pool – пул решений. В случае построения комбинаций различных метаэвристических алгоритмов, обрабатывающих решения по конвейеру (например, гибрида SA+BnB), используется шаблонный класс weak_hybrid («слабый» гибрид), обеспечивающий запуск решателя SA до запуска BnB. При этом решатель BnB инициализирует из пула решений начальную верхнюю границу, что может сократить пространство поиска.

Предлагаемый подход полностью избавляет пользователя от изменения исходного кода гибридного метода и позволяет получать гибридные алгоритмы, наиболее подходящие для решения конкретной задачи. Однако использование такого подхода требует высокой квалификации программиста и знакомство с основами технологии шаблонного проектирования.

Литература

1. Handbook of metaheuristics / Ed. by F. Glover, G. A. Kohenberger. – Kluwer Academic Publishers, 2003.

2. Raidl G.R., A unified view on hybrid metaheuristics // Proceedings of Hybrid Metaheuristics, Third International Workshop, volume 4030 of LNCS / Ed. by F.Almeida et al. – Springer, 2006. – P. 1-12.

3. Цыганов А.В., Булычов О.И. HeO: библиотека метаэвристик для задач дискретной оптимизации // Программные продукты и системы. – 2009. – № 4. – С. 148-151.

4. Hromkovič J, Algorithmics for hard problems: introduction to combinatorial optimization, randomization, approximation, and heuristics. - 2nd edition. - Springer, 2004.

5. Влиссидес Дж., Джонсон Р., Гамма Э., Хелм Р. Приемы объектно-ориентированного проектирования. Паттерны проектирования, – СПб.: Питер, 2007 г.

Исследование аэродинамических характеристик снаряда при активном старте

Ю.В. Ганзий, А.Ш. Нагуманов, Н.В. Митюков
Ижевский государственный технический университет, Ижевск

Существование высокопроизводительных вычислительных средств, наличие необходимого математического и программного обеспечения позволяют считать математическое моделирование одним из наиболее эффективных способов оценки прочности, прогнозирования долговечности и оптимизации конструкций и технологических процессов производства.

Традиционно, моделируя динамику полета снарядов и ракет, первостепенное значение уделялось области высоких скоростей. Низкие скорости, отсутствующие у артиллерийских снарядов, характерны для начальной стадии полета ракет. При этом, аэродинамические параметры низкоскоростных летательных аппаратов, как правило, определяются при аппроксимации более высокоскоростных данных, то есть работают на пределе или за пределом их научнообоснованной применимости. Этот подход полностью себя оправдывал, когда все погрешности начального участка разгона ракеты компенсировались впоследствии [1].

Однако в последнее время появилось множество низкоскоростных летательных аппаратов: стрелы и болты для бесшумного оружия антитеррористических подразделений, поражающие элементы разрывных частей и т.п., для которых дозвуковой режим является единственным режимом полета. Да и как показывает практика, период полета более сложных изделий при выходе из транспортно-пускового контейнера под действием вышибного заряда до начала работы маршевого двигателя, также является немаловажным источником системных ошибок, вследствие, например, «провала» траектории при сходе с направляющих.

Проведение точного и апробированного аэродинамического моделирования снаряда для определения аэродинамических коэффициентов на дозвуковых скоростях позволит оптимизировать конструкцию и технологические процессы уже на ранних стадиях проектирования нового изделия, когда проведение натурного эксперимента еще невозможно.

В связи с вышеизложенным, целью данной работы являлось определение аэродинамических характеристик снаряда на начальном этапе полета при дозвуковых скоростях. В качестве объекта был выбран аналог противотанковой ракеты «Штурм» в начальной стадии полета, летящий под действием вышибного заряда до включения маршевого двигателя. Определение коэффициента лобового сопротивления проводилось в пакете CFX ANSYS и сравнивалось с результатами расчетов по существующим аналитическим моделям.

Для достижения поставленной цели было необходимо решить следующие задачи:

- построение трехмерной модели ракеты «Штурм» и ее экспорт в среду CFX ANSYS;
- расчет внешнего вязкого обтекания равномерным набегающим потоком воздуха с числами M в диапазоне 0,03…0,30, при угле атаки $\alpha = 0°$;
- оптимизация расчетной сетки по критерию «точность – затраты машинного времени»;
- сравнение результатов расчета коэффициента аэродинамического сопротивления с существующей методикой аэродинамического расчета.

При экспорте трехмерной модели в среду CFX ANSYS было обнаружено, что не все узлы расчетной работали с одинаковой «информационной загрузкой» в связи с чем

было произведено дробление сетки в области головной части и стабилизаторов, а также укрупнение в областях отстоящих от поверхности ракеты на рсстояние свыше 4...5 диаметров миделя (т.е. на границе и вне пограничного слоя). Внешний вид трехмерной модели ракеты «Штурм» и оптимизированной расчетной сетки представлен на рис. 1.

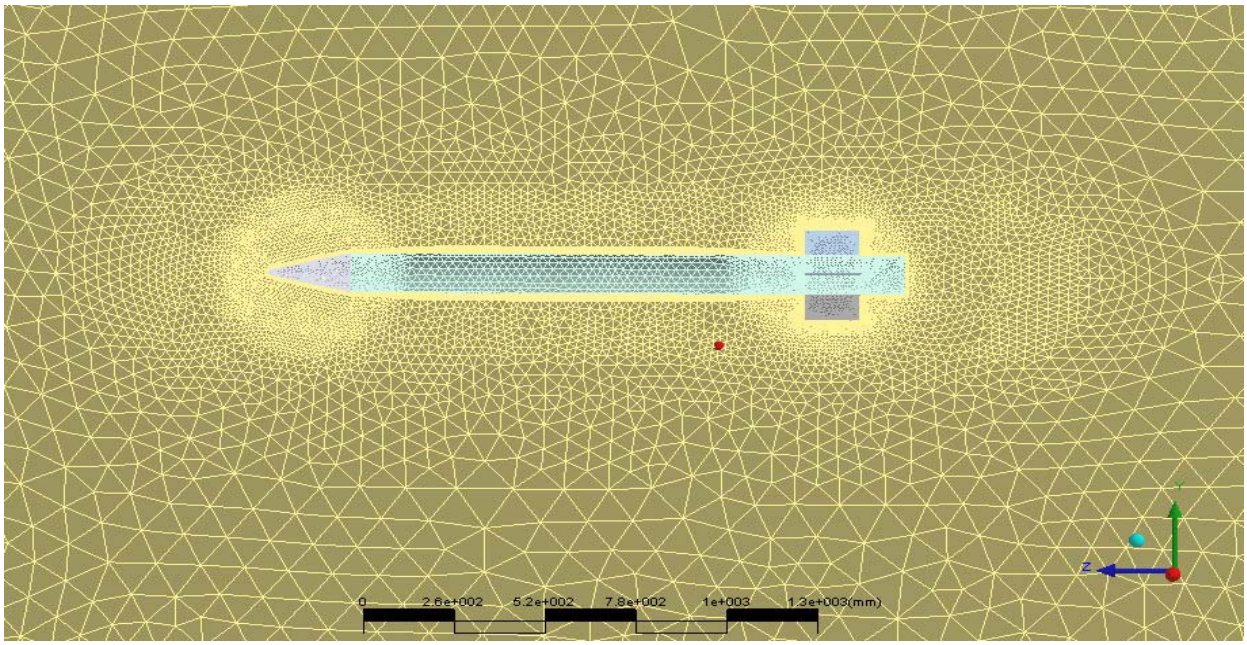


Рис. 1 – Трехмерная модель ракеты «Штурм»

Далее, в соответствии с решаемыми задачами определялись аэродинамические параметры: сила сопротивления Rx, по которой рассчитывался коэффициент аэродинамического сопротивления Cx для заданного диапазона чисел маха: M = 0,03...0,30 (см. табл.), и производилось сравнение с расчетами по классическим формулам для определения Cx (рис. 2, расчет проводился по методике [2]).

Значения сил лобового сопротивления и коэффициентов сил лобового сопротивления на разных скоростях

M	0,03	0,06	0,12	0,18	0,21	0,24	0,27	0,30
Rx, H	0,44	1,68	6,48	14,34	19,36	25,20	31,78	39,10
Cx	0,520	0,497	0,479	0,471	0,467	0,466	0,464	0,462

Расчеты показали, что имеется несоответствие в результатах, полученных в CFX ANSYS и определенных по методике [2]. При этом, как оказалось, относительное расхождение является функцией от M: оно плавно увеличивается на всем диапазоне расчетных значений. При M = 0,03 расхождение составляет Δ = 0,27, а при M = 0,30 оно увеличивается до Δ = 0,40. Пробные расчеты на более высоких скоростях, показали, что впоследствии расхождение понижается, достигая максимума при M ~ 0,4...0,6.

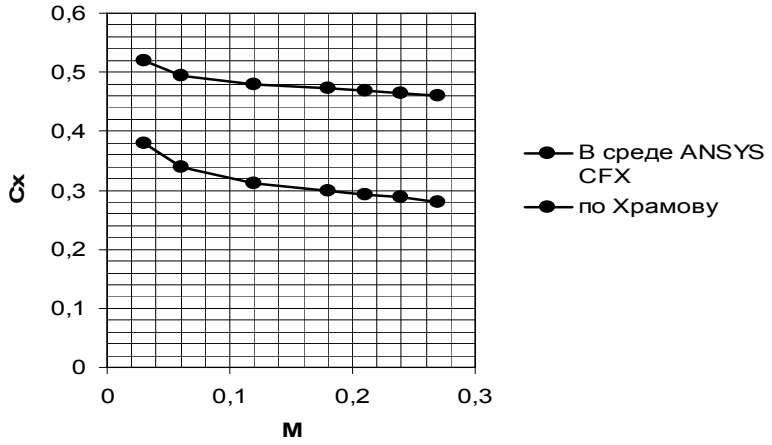


Рис.2 – График зависимости коэффициента лобового сопротивления от числа Маха в сред ANSYS CFX и по методике [2]

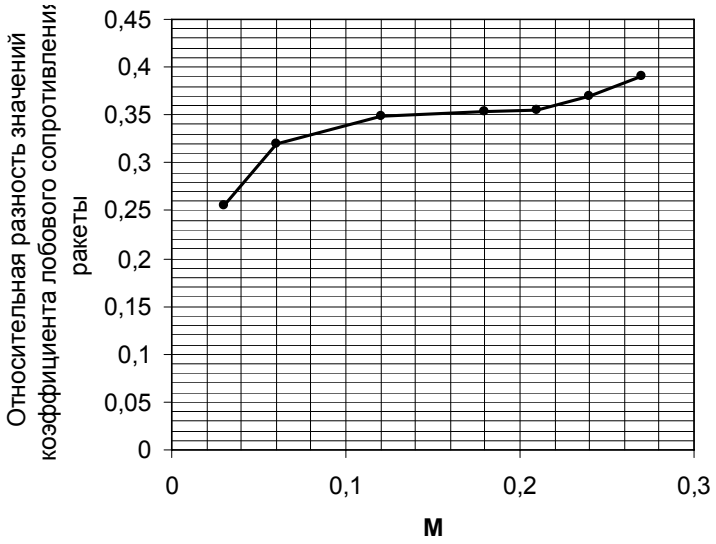


Рис.3 – График относительной разности значений коэффициента лобового сопротивления ракеты, полученных разными способами

Более детальный анализ причин расхождения показал, что, вероятно, оно связано с погрешностью эмпирической формулы А.А. Дородницына при небольших турбулентных числах Re (порядка 10^6) и при небольших числах М.

Литература

1. Краснов Н.Д. Аэродинамика. М.: Высшая школа, 1971. 632 с.
2. Баженов Е.К., Храмов С.Н. Методические указания к курсовому и дипломному проектированию оперативно-тактических ракет с ЖРД. Ижевск, Изд-во ИМИ, 1981. 23с.

Автоматический синтез параллельных программ на основе их графических моделей в системе PGRAPH

В.В. Жидченко, А.Н. Коварцев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Современные учебники по параллельному программированию, как правило, описывают две наиболее распространенные технологии: MPI и OpenMP. Знать эти технологии, безусловно, должен каждый разработчик параллельных программ. Однако насколько эффективно их использование при обучении параллельному программированию?

Применение технологии MPI подразумевает использование специализированных функций из библиотеки MPI. Большинство из этих функций имеют от 6 до 12 аргументов. В результате исходный текст программы загромождается объемными конструкциями, для восприятия которых приходится внимательно изучать каждый параметр функции. Кроме того, использование указателей на буферы приема и передачи, явное указание типа передаваемых данных, создание пользовательских типов на основе карт типов (type maps) приводят к тому, что начинающие программировать на MPI тратят большинство времени на поиск ошибок в работе с типами и указателями.

Противоположный подход используется в OpenMP. Для распараллеливания программы в простейшем случае достаточно двух строк: подключение заголовочного файла, и директива компилятору. Это вызывает определенную эйфорию после MPI, которая, однако, исчезает сразу же, как что-то пойдет не так. Если результат работы программы не соответствует ожиданиям начинающего разработчика, то определить причину без специализированных средств отладки ему зачастую невозможно.

В итоге у студентов создается впечатление, что параллельное программирование – это либо раздел системного программирования, либо автоматическая работа компилятора, которую сложно предсказать, и поэтому лучше ограничить распараллеливанием простейших циклов.

Авторам статьи представляется, что для изучения параллельного программирования необходимо начинать с некоторого промежуточного варианта. При создании параллельной программы с явным параллелизмом разработчик должен ясно видеть, какие фрагменты программы распараллеливаются, какие действия выполняет каждый фрагмент и над какими данными. При этом описание параллелизма не должно быть громоздким, не должно затруднять восприятие основной (смысловой) части программы.

Наиболее удобной представляется графическая форма описания параллельной программы. В отличие от текста, последовательного по своей сути, графическое описание позволяет компактно и в то же время интуитивно понятно изобразить параллелизм. Желательно использовать абстрактные элементы, не привязанные к конкретной технологии параллельного программирования. Обязательным требованием должна быть легкость восприятия таких элементов, чтобы их не нужно было отдельно изучать. Параллельная программа должна синтезироваться автоматически по ее графическому описанию. Для обеспечения переносимости программа должна использовать универсальные технологии, такие как MPI и OpenMP.

С учетом приведенных выше требований на кафедре программных систем Самарского государственного аэрокосмического университета разрабатывается система визуального параллельного программирования PGRAPH [1]. В настоящей статье

описывается способ автоматического синтеза программ по их графическому описанию, применяемый в указанной системе.

Графическая модель параллельной программы в системе PGRAPH

В основу графического описания параллельных программ в системе PGRAPH положена граф-модель технологии графо-символического программирования (ГСП-технологии) [2]. Параллельная программа в этой модели представляется ориентированным графом, вершины которого – это функции, выполняющие заданные разработчиком действия над данными. Порядок запуска вершин на исполнение определяется дугами: вершина, в которую ведет дуга, исполняется после вершины, из которой эта дуга исходит. Для последовательного и параллельного запуска вершин введены дуги различных типов: параллельная дуга и последовательная дуга. Если из вершины исходит несколько параллельных дуг, то вершины, в которые они ведут, будут исполняться параллельно. Последовательные дуги помечаются метками-предикатами, которые необходимы для организации ветвления. Переход по последовательной дуге возможен лишь при истинности ее предиката.

Формально граф-модель представляется четверкой <D,F,P,G>, где D – множество данных некоторой предметной области, F – множество функций, определенных над данными предметной области, P – множество предикатов над данными из множества D, G – множество ориентированных помеченных графов, называемых агрегатами. Каждый агрегат представляет собой четверку $G_k = \{A, \Psi, \Phi, R\}$, где $A = \{A_1, A_2, \ldots, A_n\}$ – множество вершин графа. Каждая вершина A_i помечена локальной функцией $f_i(D) \in F$ или другим агрегатом $G_j \in G$, т.е. агрегаты могут быть вложенными. На графе задано множество дуг управления $\Psi = \{\Psi_{1,i1}, \Psi_{1,i2}, \ldots, \Psi_{j,m}\}$ и множество дуг синхронизации $\Phi = \{\Phi_{1,i1}, \Phi_{1,i2}, \ldots, \Phi_{j,l}\}$. R – отношение над множествами вершин и дуг, определяющее способ их связи. Дуга управления, соединяющая любые две вершины A_i и A_j, имеет три метки: предикат $p_{ij}(D) \in P$, приоритет $k(\Psi_{ij}) \in N$ и тип дуги $T(\Psi_{ij}) \in N$.

Граф G_k инициальный – работа программы начинается с выполнения функции $f_0(D)$, помечающей начальную вершину A_0. Дальнейшее развитие вычислений ассоциируется с переходами из вершины в вершину по дугам управления. Переход по дуге управления возможен лишь в случае истинности предиката, которым она помечена. Если несколько предикатов, помечающих исходящие из вершины дуги, одновременно являются истинными, переход выполняется по самой приоритетной дуге.

Тип дуги Ψ_{ij} определяется как функция $T(\Psi_{ij}) \in \{1,2,3\}$, значения которой имеют следующую семантику:

$T(\Psi_{ij}) = 1$ – *последовательная дуга* (описывает передачу управления на последовательных участках алгоритма);

$T(\Psi_{ij}) = 2$ – *параллельная дуга* (обозначает начало нового параллельного фрагмента алгоритма);

$T(\Psi_{ij}) = 3$ – *терминирующая дуга* (завершает параллельный фрагмент алгоритма).

Для описания параллелизма вводится понятие параллельной ветви β - подграфа графа G_k, начинающегося параллельной дугой и заканчивающегося терминирующей дугой. $\beta = <A_\beta, \Psi_\beta, R_\beta>$, где A_β – множество вершин ветви, Ψ_β – множество дуг управления ветви, R_β – отношение над множествами вершин и дуг ветви, определяющее способ их связи. На рисунке 1 параллельные ветви обведены пунктиром. Ветви, начинающиеся в одной вершине некоторой ветви, должны терминироваться также в одной вершине этой же ветви.

Число параллельных ветвей в модели фиксируется при ее построении, при этом максимальное количество ветвей не ограничивается.

Функционирование модели начинается с запуска единственной ветви, называемой мастер-ветвью β_0. Мастер-ветвь начинается с корневой вершины (на рис. 1 корневая вершина обозначена жирной границей) и заканчивается в одной из конечных вершин. В вершинах мастер-ветви, имеющих исходящие параллельные дуги, порождаются новые параллельные ветви. Допускается вложенность параллельных ветвей.

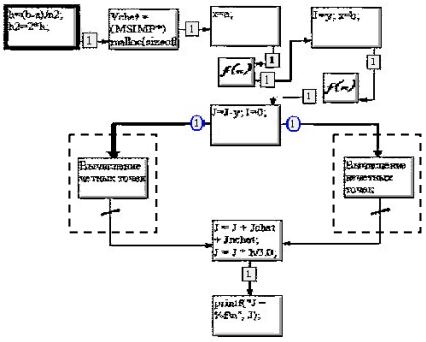


Рисунок 1 – Пример графической модели параллельной программы

Функции над данными предметной области, которыми помечаются вершины графа, описываются в виде небольших фрагментов исходного текста на языке C++ и называются базовыми модулями.

Генерация исходного текста параллельной программы на языке C++ по ее графической модели

Для построения исполнимых файлов на основе созданной пользователем графической модели система PGRAPH генерирует исходный текст программы на языке C++. В процессе генерации используется описание объектов технологии ГСП, хранящееся в информационном фонде системы. Полученные исходные тексты подаются на вход компилятора C++, который автоматически вызывается системой PGRAPH. Генерируемая программа на языке C++ состоит из нескольких модулей (рисунок 2).

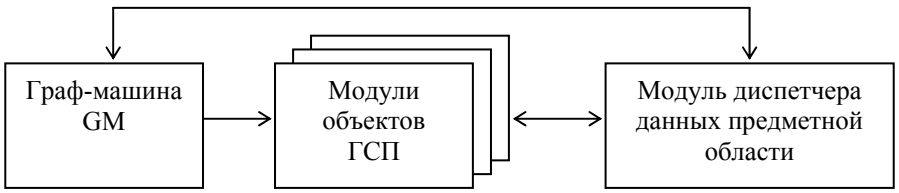


Рисунок 2 – Структура программы, автоматически генерируемой на основе граф-модели параллельной программы

Модуль объекта ГСП – это отдельный файл с исходным текстом, создаваемый системой для каждого объекта технологии ГСП (актора, предиката, агрегата). Все обращения к объектам ГСП реализованы в виде функций, поэтому указанный файл содержит описание соответствующей функции. Для акторов и предикатов это исходный текст базового модуля, в котором все обращения к данным предметной области заменены на специальные конструкции в соответствии с межмодульным информационным интерфейсом параллельной программы. Для агрегатов функция состоит из двух частей:

1) заголовка со структурами данных, описывающими вершины и дуги агрегата;

2) вызова граф-машины с передачей ей в качестве параметров этих структур.

Граф-машина представляет собой универсальную для всех агрегатов программу, которая управляет вычислительным процессом в соответствии со структурой агрегата. Это центральная часть синтезируемой параллельной программы, ее "движок" ("engine") – отсюда название. Граф-машина запускается при каждой передаче управления в вершину, помеченную агрегатом. Поскольку мастер-ветвь – это тоже агрегат,

выполнение всей параллельной программы начинается с запуска граф-машины. Передаваемые в граф-машину структуры данных описывают агрегат и содержат указатели на функции, описанные в модулях объектов ГСП, используемых агрегатом. Работа граф-машины заключается в просмотре передаваемых ей структур и передаче управления по хранящимся в них указателям. Структуры организованы таким образом, что их просмотр осуществляется в соответствии с направлением дуг агрегата.

Работа параллельных ветвей с использованием MPI организована следующим образом. При компиляции граф-модели подсчитывается общее количество параллельных ветвей Np. Каждая ветвь преобразуется системой в отдельный агрегат, представляющий собой подграф, образуемый этой ветвью. Созданному агрегату ставится в соответствие отдельный процесс, ранг которого запоминается в этом агрегате. При запуске параллельной программы создается Np ее экземпляров, работа каждого из которых определяется его рангом. Вначале получает управление экземпляр мастер-ветви. Остальные экземпляры, соответствующие другим параллельным ветвям, ожидают команду на запуск. Такой командой служит сообщение заданного вида, которое генерирует граф-машина при встрече параллельной дуги. После отправки команд на запуск всех ветвей, исходящих из некоторой вершины, граф-машина переходит к вершине, в которой эти ветви терминируются. Перед запуском этой вершины граф-машина ожидает прихода сообщений о завершении работы параллельных ветвей, которые генерируются каждой ветвью отдельно. После получения сообщений от всех ветвей, терминируемых в данной вершине, граф-машина продолжает свою работу.

Межмодульный информационный интерфейс

Модуль диспетчера данных предметной области выполняет все функции по инициализации, хранению, изменению и межмодульному обмену данными. Для этого словарь данных представляется в синтезируемой программе в виде класса языка C++.

Подсистема компиляции объектов граф-модели использует свойство классов, обеспечивающее абстракцию данных: скрытие деталей представления объекта и предоставление доступа к содержащимся в нем данным только посредством функций и операций, описанных как часть этого класса.

Данные предметной области оформляются в виде членов–данных специализированного класса для хранения данных (назовем его TPOData). Доступ к этим членам возможен лишь с помощью специальных функций, которые обеспечивают чтение и модификацию данных предметной области:

```
class TPOData {      int A, B;
                          int Get_A(); Put_A(int a);
                          int Get_B(); Put_B(int b);          }
```

Все обращения к данным в базовых модулях автоматически заменяются на вызовы функций в соответствии со способом использования каждого данного. Если данное используется для чтения, то вызывается функция чтения. Если значение данного изменяется, то вызывается функция записи соответствующего члена–данного.

На рисунке 3 приведен пример модификации исходного текста базового модуля в соответствии с описанным правилом. Базовый модуль, приведенный на рисунке 3, а), приводит к порождению исходного текста, приведенного на рисунке 3, б).

```
A=A+B;                      TPOData→Put_A(TPOData→Get_A()+TPOData→Get_B());
printf("A=%d\n", A);        printf(TPOData→Get_A( ));
```

а) б)

Рисунок 3 – Порождение исходного текста на основе базового модуля

Если некоторые данные используются в качестве параметра функции, то в общем случае затруднено автоматическое определение способа их использования. В этом случае система требует вмешательства пользователя, который явно указывает способ использования каждого данного.

Функции чтения и записи членов–данных класса могут выполнять дополнительные операции, например проверку диапазонов значений, ведение статистики по частоте использования данного, дамп значения данного в журнальный файл или на экран, передача данного между различными параллельными процессами.

Рассмотрим последнюю возможность более подробно. В распределенной вычислительной системе любой узел работает с локальной областью памяти, следовательно, необходимо обеспечить пересылку данных между процессами, выполняющимися на различных узлах. Ниже описывается один из способов организации подобной пересылки, реализованный в системе PGRAPH.

Обмен данными между параллельными процессами

Предлагаемый способ организации хранения и обмена данными между параллельными процессами ориентирован на использование технологии MPI.

Диспетчер данных исполняется в отдельном процессе параллельной программы. Он порождает объект, описываемый классом TPOData, который хранит значения данных предметной области. В каждом из процессов, содержащих параллельные ветви граф-модели, также порождается объект класса TPOData. Однако функции доступа к членам-данным у объекта диспетчера данных и у объектов параллельных ветвей различаются. Диспетчер данных хранит все данные в локальной памяти и для обращения к ним использует обычные указатели. В параллельных ветвях граф-модели для чтения или записи каждого данного осуществляется обращение к диспетчеру памяти с помощью совокупности сообщений. В первом сообщении пересылается запрос на чтение или запись данного. При чтени параллельная ветвь переходит к ожиданию ответа от диспетчера данных. При записи во втором сообщении пересылается новое значение данного. Диспетчер данных циклически принимает и обрабатывает запросы параллельных ветвей (рисунок 4).

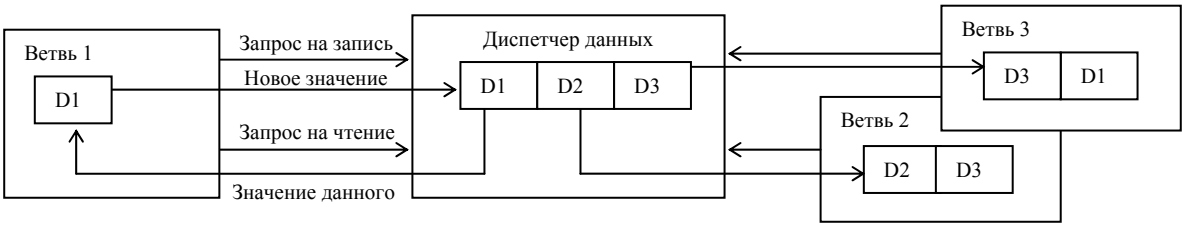


Рисунок 4 – Организация обмена данными в распределенной вычислительной системе

Паспортизация объектов технологии ГСП позволяет выявить данные, монопольно используемые некоторыми ветвями. Поэтому в общем случае возможно более эффективное использование данных, при котором диспетчер памяти содержит лишь совместно используемые данные, а параллельные ветви обращаются к нему только для чтения или записи этих данных.

Литература

1. А.Н. Коварцев, В.В. Жидченко Программный комплекс разработки алгоритмов и программ параллельных вычислений. - Материалы Девятой международной конференции-семинара Высокопроизводительные параллельные вычисления на кластерных системах, Владимир, 2009. - С. 225-229.

2. Коварцев А.Н. Автоматизация разработки и тестирования программных средств. - Самарский государственный аэрокосмический университет, Самара, 1999. - 150 с.

Численное решение уравнений газодинамики неравновесных сред на GPU с использованием технологии CUDA

В.Г. Макарян

Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В последние годы рост производительности графических ускорителей (GPU) значительно превосходил данный показатель у центральных процессоров (CPU), что привело к использованию GPU для неграфических, чаще всего научных вычислений. Предложенная компанией NVidia технология CUDA предназначена для написания неграфических приложений для GPU [1]. Основным достоинством этой технологии является ее простота – все программы пишутся на «расширенном» языке С что сильно облегчает перенос на GPU хорошо распараллеливаемых алгоритмов, реализованных на традиционных языках программирования, таких как С, С++, Паскаль и др. [2]

В настоящей работе проведен перенос программы для численного решения уравнений газодинамики неравновесных сред, разработанной автором ранее, на GPU с использованием технологии CUDA.

Исходная система уравнений газодинамики неравновесной среды, включает уравнения непрерывности, Навье-Стокса, состояния газа, переноса тепла и, в простейшем виде, уравнение релаксации внутренних степеней свободы среды (в дальнейшем, для определенности, будем говорить о колебательной релаксации молекул газа):

$$\frac{dE_K}{dt} = \frac{E_{Ke} - E_K}{\tau_K(T,\rho)} + Q,$$

где E_K – колебательная энергия в расчёте на одну молекулу, E_{Ke} – её равновесное значение (то есть при равенстве стационарных колебательной и поступательной температур $T_V = T_0$), τ_K – время колебательной релаксации, Q – мощность внешнего источника накачки (на одну молекулу), необходимая для поддержания неравновесности $E_K > E_{Ke}$.

В безразмерных переменных, определяемых следующим образом:

$$\widetilde{\rho} = \rho/\rho_0, \quad \vec{\widetilde{\mathrm{v}}} = \vec{\mathrm{v}}/c_\infty, \quad \widetilde{P} = P/P_0, \quad \widetilde{T} = T/T_0, \quad S = Q\tau_0/T_0, \quad \widetilde{E}_K = E_K/T_0,$$

$$\widetilde{E}_{Ke} = E_{Ke}/T_0, \ \widetilde{\tau} = \tau/\tau_0, \ \widetilde{t} = t/\tau_0, \ \widetilde{x}_i = x_i/(c_\infty\tau_0),$$

где $\vec{\mathrm{v}}$ – скорость, ρ, P, T – плотность, давление и температура (температура в единицах энергии); ρ_0, P_0, T_0 их невозмущенные значения; $\tau_0 = \tau_K(T_0,\rho_0)$, $c_\infty = \sqrt{\gamma_\infty P_0/\rho_0}$ – высокочастотная скорость звука, γ_∞ – высокочастотный показатель адиабаты; S – степень неравновесности среды; исходная система уравнений примет вид:

$$\frac{\partial\widetilde{\rho}}{\partial\widetilde{t}} = -\widetilde{\mathrm{v}}_k\frac{\partial\widetilde{\rho}}{\partial\widetilde{x}_k} - \widetilde{\rho}\frac{\partial\widetilde{\mathrm{v}}_k}{\partial\widetilde{x}_k},$$

$$\frac{\partial\widetilde{\mathrm{v}}_\mathrm{i}}{\partial\widetilde{t}} = -\widetilde{\mathrm{v}}_\mathrm{k}\frac{\partial\widetilde{\mathrm{v}}_\mathrm{i}}{\partial\widetilde{x}_k} - \frac{1}{\gamma_\infty}\left(\frac{\partial\widetilde{T}}{\partial\widetilde{x}_i} + \frac{\widetilde{T}}{\widetilde{\rho}}\frac{\partial\widetilde{\rho}}{\partial x_i}\right) + \frac{\eta}{\widetilde{\rho}\rho_0\tau_0 c_\infty^2}\left(\frac{\partial^2\widetilde{\mathrm{v}}_i}{\partial x_k^2} + \frac{\partial^2\widetilde{\mathrm{v}}_k}{\partial x_i\partial x_k}\right),$$

$$\frac{\partial \widetilde{T}}{\partial \widetilde{t}} = -\widetilde{\mathrm{v}}_{\mathrm{i}} \frac{\partial \widetilde{T}}{\partial \widetilde{x}_i} - \frac{1}{C_{V\infty}} \left[\frac{\widetilde{E}_{Ke} - \widetilde{E}_K}{\widetilde{\tau}\left(\widetilde{\rho}, \widetilde{T}\right)} + S + \widetilde{T} \frac{\partial \widetilde{\mathrm{v}}_i}{\partial \widetilde{\mathrm{x}}_i} - \right.$$
$$\left. - \frac{\chi}{\tau_0 c_\infty^2} \frac{\partial^2 \widetilde{T}}{\partial \widetilde{x}_i^2} - \frac{\eta \gamma_\infty}{\widetilde{\rho} \rho_0 \tau_0 c_\infty^2} \frac{\partial \widetilde{\mathrm{v}}_i}{\partial \widetilde{x}_k} \left(\frac{\partial \widetilde{\mathrm{v}}_i}{\partial \widetilde{x}_k} + \frac{\partial \widetilde{\mathrm{v}}_k}{\partial \widetilde{x}_i} - \frac{2}{3} \delta_{ik} \frac{\partial \widetilde{\mathrm{v}}_l}{\partial \widetilde{x}_l} \right) \right], \quad (1)$$
$$\frac{\partial \widetilde{E}_K}{\partial \widetilde{t}} = -\widetilde{\mathrm{v}}_{\mathrm{i}} \frac{\partial \widetilde{E}_K}{\partial \widetilde{x}_i} + \frac{\widetilde{E}_{Ke} - \widetilde{E}_K}{\widetilde{\tau}\left(\widetilde{\rho}, \widetilde{T}\right)} + S,$$

где по повторяющимся индексам подразумевается суммирование; η и χ − коэффициенты сдвиговой вязкости и теплопроводности; $C_{V\infty}$ − высокочастотная теплоемкость при постоянном объеме. В (1) из уравнений Навье-Стокса исключено давление при помощи уравнения состояния $\widetilde{P} = \widetilde{\rho}\widetilde{T}$.

1. Алгоритм решения

Для численного интегрирования в правой части системы уравнений (1) первые пространственные производные представлялись пятиточечными разностными функциями четвертого порядка точности, а вторые производные − разностными функциями второго порядка, что является достаточным из-за присутствия перед ними малых множителей η и χ.

$$\frac{\partial \overline{\mathbf{\Sigma}}_{i,j}}{\partial \widetilde{x}} = \frac{-\overline{\mathbf{\Sigma}}_{i+2,j} + 8\overline{\mathbf{\Sigma}}_{i+1,j} - 8\overline{\mathbf{\Sigma}}_{i-1,j} + \overline{\mathbf{\Sigma}}_{i-2,j}}{12h_x} + O\left(h_x^4\right)$$
$$\frac{\partial \overline{\mathbf{\Sigma}}_{i,j}}{\partial \widetilde{y}} = \frac{-\overline{\mathbf{\Sigma}}_{i,j+2} + 8\overline{\mathbf{\Sigma}}_{i,j+1} - 8\overline{\mathbf{\Sigma}}_{i,j-1} + \overline{\mathbf{\Sigma}}_{i,j-2}}{12h_y} + O\left(h_y^4\right)$$
$$\frac{\partial^2 \overline{\mathbf{\Sigma}}_{i,j}}{\partial \widetilde{x}^2} = \frac{\overline{\mathbf{\Sigma}}_{i+1,j} - 2\overline{\mathbf{\Sigma}}_{i,j} + \overline{\mathbf{\Sigma}}_{i-1,j}}{h_x^{\,2}} + O\left(h_x^2\right)$$
$$\frac{\partial^2 \overline{\mathbf{\Sigma}}_{i,j}}{\partial \widetilde{y}^2} = \frac{\overline{\mathbf{\Sigma}}_{i,j+1} - 2\overline{\mathbf{\Sigma}}_{i,j} + \overline{\mathbf{\Sigma}}_{i,j-1}}{h_y^{\,2}} + O\left(h_y^2\right)$$
$$\frac{\partial^2 \overline{\mathbf{\Sigma}}_{i,j}}{\partial \widetilde{x} \partial \widetilde{y}} = \frac{\overline{\mathbf{\Sigma}}_{i+1,j+1} - \overline{\mathbf{\Sigma}}_{i-1,j+1} - \overline{\mathbf{\Sigma}}_{i+1,j-1} + \overline{\mathbf{\Sigma}}_{i-1,j-1}}{4h_x h_y} + O\left(h_x h_y\right)$$

где $\overline{\mathbf{\Sigma}} = (\widetilde{\rho}, \widetilde{\mathrm{v}}_{\mathrm{x}}, \widetilde{\mathrm{v}}_{\mathrm{y}}, \widetilde{T}, \widetilde{E}_K)$ − вектор значений параметров газа в узле сетки, h_x, h_y − величина шага пространственной сетки по координате x и y, соответственно, i, j − индексы узла сетки по x и y.

Таким образом, система уравнений в частных производных (1) сводится к системе обыкновенных дифференциальных уравнений вида

$$\frac{d\overline{\mathbf{\Sigma}}_{i,j}}{d\widetilde{t}} = \overline{F}(\mathbf{\Sigma}_{i,j}, \mathbf{\Sigma}_{i+1,j}, \mathbf{\Sigma}_{i+2,j}, \mathbf{\Sigma}_{i-1,j}, \mathbf{\Sigma}_{i-2,j}, \mathbf{\Sigma}_{i,j-1},$$
$$\mathbf{\Sigma}_{i,j-2}, \mathbf{\Sigma}_{i,j-1}, \mathbf{\Sigma}_{i,j-2}, \mathbf{\Sigma}_{i+1,j-1}, \mathbf{\Sigma}_{i+1,j+1}, \mathbf{\Sigma}_{i-1,j+1}, \mathbf{\Sigma}_{i-1,j-1}).$$

Интегрирование полученной системы обыкновенных дифференциальных уравнений по времени велось по методу Рунге-Кутта 4-го порядка [3]:

$$\frac{\overline{\mathbf{\Sigma}}(t+\tau) - \overline{\mathbf{\Sigma}}(t)}{\tau} = \frac{1}{6}[\bar{k}_1 + 2\bar{k}_2 + 2\bar{k}_3 + \bar{k}_4]. \quad (2)$$

Здесь для простоты записи индексы i, j опущены; $\bar{k}_1 = \bar{F}(\bar{\Sigma}(t))$, $\bar{k}_2 = \bar{F}(\bar{\Sigma}(t) + \tau \bar{k}_1 / 2)$, $\bar{k}_3 = \bar{F}(\bar{\Sigma}(t) + \tau \bar{k}_2 / 2)$, $\bar{k}_4 = \bar{F}(\bar{\Sigma}(t) + \tau \bar{k}_3)$, τ – шаг интегрирования по времени.

2. CUDA-реализация алгоритма

Реализация этого алгоритма с использованием технологии CUDA представляет собой программный модуль, написанный на расширении языка C. Он включает в себя три функции, выполняющиеся на GPU. Это функция Calc_k (Calc_k_Ex1 в оптимизированном варианте программы), служащая для вычисления массивов коэффициентов $\bar{k}_m$ в методе Рунге-Кутта. Функция Calc_drr служит для вычисления значений $\bar{\Sigma}$ на следующем временном слое согласно формуле (2). Наконец функция Cikl реализует циклические граничные условия по одной из пространственных координат. Необходимость использования нескольких функций-ядер проистекает из невозможности выполнить в CUDA синхронизацию выполнения команд внутри одного ядра, если оно содержит более одного блока. Отсутствие же синхронизации внутри ядра неминуемо привело бы к потере целостности массивов $\bar{\Sigma}(t)$ и др., хранимых в памяти GPU.

В простейшем случае программа на CUDA являлась почти точной копией программы для CPU лишь с небольшими изменениями. Никакой специальной оптимизации кода для GPU не проводилось. На рис. 1 показано распределение времени работы GPU между различными функциями. Оно получено с помощью визуального профилировщика, входящего в пакет CUDA toolkit. Как хорошо видно из рисунка, подавляющее время работы GPU (85%) занимает вычисление массивов коэффициентов $\bar{k}_m$.

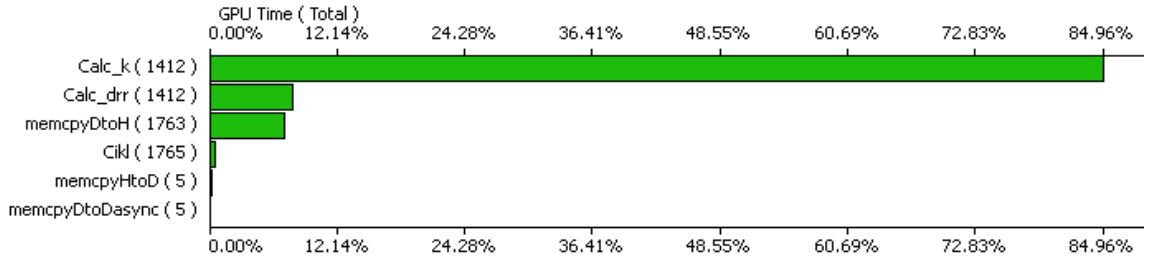


Рис. 1 Распределение времени работы GPU между различными функциями (до оптимизации)

На рис. 2 показано относительное количество вызовов различных типов

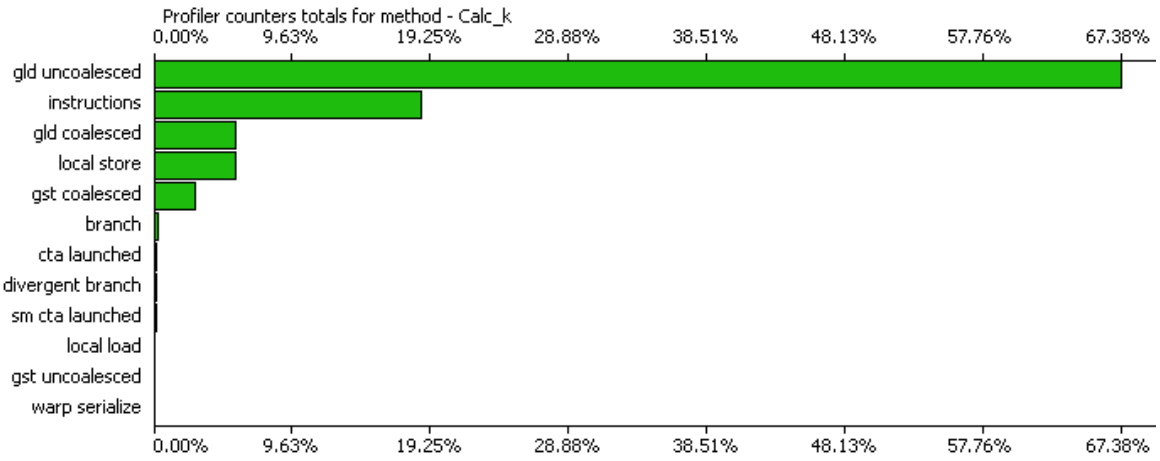


Рис. 2 Относительное количество вызовов различных типов инструкций для функции Calc_k

инструкций для функции Calc_k. Видно, что основное количество инструкций (более 67%) приходится на чтения из глобальной памяти видеокарты (gld), при этом большинство из них происходят с нарушением коалесинга (gld uncoalesced), т.е. инструкции чтения из соседних ячеек памяти не объединяются в один запрос [1, 2]. Все это с учетом отсутствия кэширования на графическом чипе данных, находящихся в глобальной памяти видеокарты, приводит к тому, что подавляющая часть времени выполнения программы уходит на работу с глобальной памятью GPU.

Для повышения быстродействия была оптимизирована функция Calc_k (Calc_k_Ex1 в оптимизированном варианте). Теперь, в начале выполнения функции, все необходимые данные загружаются из медленной глобальной памяти в быструю разделяемую память, а затем чтение данных ведется только из нее. Это позволило в несколько раз уменьшить количество обращений к глобальной памяти. Дополнительно была проведена оптимизация части программы, выполняемой на CPU, что позволило снизить количество передаваемой информации с видеокарты в память компьютера.

Результаты профилирования оптимизированного варианта программы показаны на рис. 3 и 4. Видно (рис. 4), что количество чтений из глобальной памяти уменьшилось более чем в 10 раз, причем теперь для всех обращений к глобальной памяти выполнены условия коалесинга. Соответственно время выполнения оптимизированной функции Calc_k_Ex1 уменьшилось более чем в восемь раз (рис. 3)!

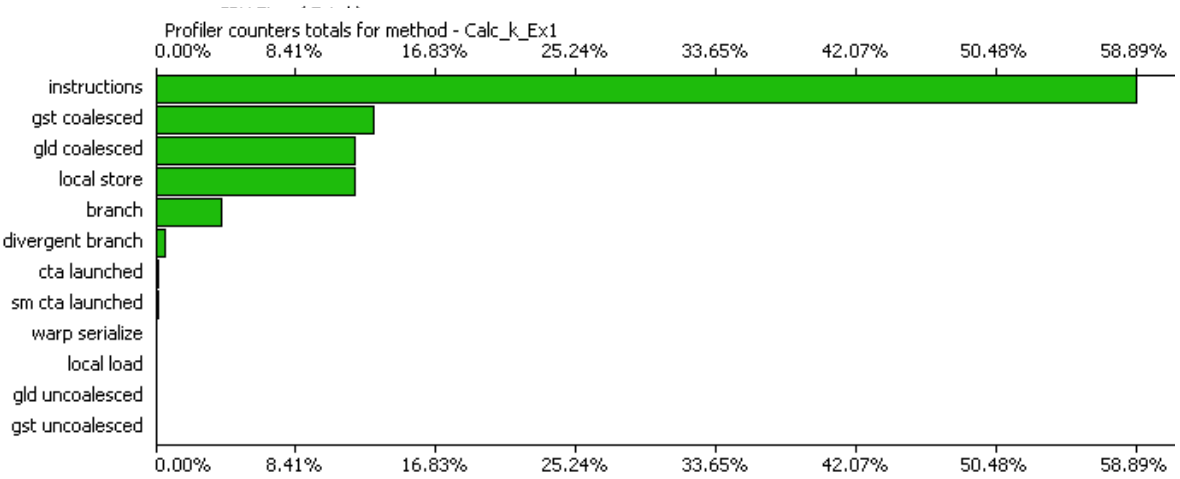


Рис. 4 Относительное количество вызовов различных типов инструкций для функции Calc_k_Ex1 (после оптимизации)

В таблице 1 приведены сравнительные данные по скорости работы описываемой программы на CPU и GPU. В первом столбце даны сведения об оборудовании, на котором тестировалась программа. Использовался двуядерный процессор Intel Core 2 Duo E8400 3.0 ГГц и видеокарта GF8800 GT. Во втором столбце приведена скорость счета программы, выраженная в количестве узлов сетки, обрабатываемых программой в секунду. В третьем столбце приведено ускорение выполнения программы относительно скорости ее выполнения на одном ядре CPU.

Таблица 1. Сравнение скорости работы программы на CPU и GPU

Оборудование и программа	Скорость счета (млн. узлов/с)	Ускорение
Intel Core 2 Duo E8400 3.0 ГГц 1 ядро	0.56	1.0
Intel Core 2 Duo E8400 3.0 ГГц 2 ядра	1.04	1.85

GF8800 GT без оптимизации	5.9	10.5
GF8800 GT с оптимизацией	36.7	65.5

Заключение

Как видно из таблицы 1 перенос программы на GPU даже безо всякой оптимизации кода обеспечивает рост скорости счета более чем в 10 раз в сравнении с одним ядром CPU и почти в 6 раз по сравнению с двумя ядрами. Простейшая же оптимизация программы, практически не меняющая основной код, повышает скорость работы программы еще более чем в 6 раз. Все это еще раз свидетельствует о большой эффективности применения GPU и технологии CUDA для численного решения задач математической физики.

Работа частично поддержана аналитической целевой программой Министерства образования и науки РФ «Развитие научного потенциала Высшей Школы (2009-2010 гг)» проект 2.1.1/309, Федеральной целевой программой «Научные и научно-педагогические кадры инновационной России» на 2009-2013 гг. (проекты НК-410П/38 и НК-422П/59) Министерства образования и науки РФ.

Литература

1. NVIDIA: Nvidia CUDA Compute Unified Device Architecture. Programming Guide version 2.3. (2007)

2. Боресков, А.В. Основы работы с технологией CUDA / А.В. Боресков, А.А. Харламов – ДМК Пресс, 2010. – 232 с.

3. Самарский, А.А. Введение в численные методы. / А.А. Самарский. М.: Наука, 1982. 272 с.

Алгоритм структурного преобразования графа управления параллельными вычислениями для систем с распределенной памятью MPI

Д.А. Попова-Коварцева, А.Н. Коварцев
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

Быстрое развитие многопроцессорных вычислительных систем кластерного типа с распределенной памятью привело к необходимости не столько к решению задачи распределения вычислительной нагрузки параллельного алгоритма, но в большей степени, к необходимости разрешения проблемы организации информационного взаимодействия между процессорами при передаче данных. Решение этих проблем обеспечивает интерфейс передачи данных MPI. В рамках MPI проблема эффективной организации параллельных вычислений неизбежно связана с решением проблемы рациональной организации передачи данных между параллельными процессами. Неслучайно, что подавляющее большинство функций стандарта MPI посвящено передаче данных между процессами.

Более естественным для решения задачи построения алгоритмов параллельных вычислений представляется использование многопроцессорных вычислительных систем с общей памятью (симметричных мультипроцессоров SMP), где отсутствует необходимость организации управления данными. Графическая модель алгоритмов параллельных вычислений, созданная на основе модели ГСП-технологии [1], в первую очередь ориентирована на подобную конфигурацию.

Для систем с общей памятью при разработке модели параллельных вычислений в среде PGRAPH на структуру граф-программы не накладываются сколь-нибудь серьезных ограничений. Однако в стандарте MPI возникает необходимость структурных преобразований граф-программы к виду приемлемому для использования данного стандарта.

Основные положения

Модель описания параллельных алгоритмов в системе PGRAPH [2] представляется четверкой <D, F, P, G>, где D – множество данных некоторой предметной области, F – множество операторов, определенных над данными предметной области, P – множество предикатов, действующих над структурами данных предметной области, $G = \{ A, \Psi \}$ – ориентированный помеченный граф, называемый агрегатом. $A = \{ A_1, A_2, ..., A_n \}$ – множество вершин графа. Каждая вершина A_i помечена локальным оператором $f_i(D) \in F$. На графе задано множество дуг управления $\Psi = \{ \Psi_1, \Psi_2, ..., \Psi_m \}$.

Дуга управления, соединяющая любые две вершины A_i и A_j, имеет три метки: предикат - $p_{ij}(D) \in P$; приоритет - $k(\Psi_i) \in N$ и тип дуги - $T(\Psi_i) \in \{1,2,3\}$.

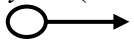
Граф G – иниициальный. Работа алгоритма начинается с выполнения оператора $f_0(D)$, помечающего начальную вершину A_0. Дальнейшее развитие вычислений ассоциируется с переходами из вершины в вершину по дугам управления. При этом переход по дуге управления возможен лишь в случае истинности предиката, которым она помечена. Если несколько предикатов, помечающих исходящие из вершины дуги,

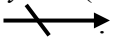
одновременно становятся истинными, переход осуществляется по наиболее приоритетной дуге.

Типы дуг имеют следующую семантику:

T(Ψi) = 1 – *последовательная дуга* (описывает передачу управления на последовательных участках алгоритма). Имеет обозначение 

;

T(Ψi) = 2 – *параллельная дуга* (обозначает начало нового параллельного фрагмента алгоритма). Обозначается  ;

T(Ψi) = 3 – *терминирующая дуга* (завершает параллельный фрагмент алгоритма). Имеет обозначение на схеме  .

На рисунке 1 представлен пример граф модели параллельного алгоритма.

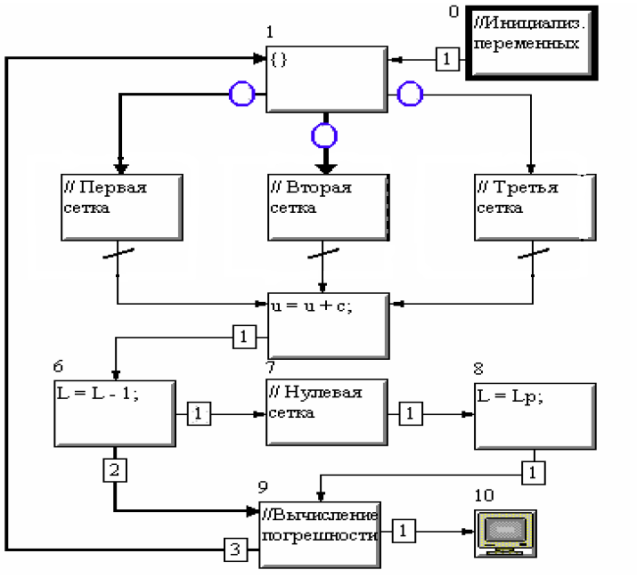


Рисунок 1.

В общем случае для описания параллелизма вводится понятие **параллельной ветви** β – подграфа графа G, начинающегося параллельной дугой и заканчивающегося терминирующей дугой. $\beta = \langle A_\beta, \Psi_\beta \rangle$, где A_β – множество вершин ветви, Ψ_β – множество дуг управления ветви.

В соответствии с определением, каждая ветвь имеет ровно один вход и один выход. Число параллельных ветвей в модели фиксируется при ее построении, при этом максимальное количество ветвей не ограничивается. При кодировании алгоритма, описанного с помощью предлагаемой модели, каждая параллельная ветвь порождает отдельный процесс – совокупность операторов, исполняемых последовательно на одном из процессоров параллельной вычислительной системы.

Функционирование модели начинается с запуска единственной ветви, называемой **мастер-ветвью** β_0. Мастер-ветвь начинается с корневой вершины и заканчивается в одной из конечных вершин алгоритма. В вершинах мастер-ветви, имеющих исходящие параллельные дуги, могут порождаться новые параллельные ветви. Вершины этих ветвей также имеют параллельные дуги, таким образом, допускается вложенность параллельных ветвей. Переход между двумя вершинами, принадлежащими различным параллельным ветвям, возможен только по параллельным и терминирующим дугам. Запрещены условные переходы между вершинами различных параллельных ветвей.

В системе PGRAPH используется стандартная схема управления вычислительным процессом. Синтез агрегата происходит на основе его графического изображения, представленного соответствующими структурами данных, которые

хранятся в информационном фонде системы. Программа агрегата состоит из двух частей: структур данных, описывающих граф агрегата и стандартного для всех агрегатов модуля - граф-машины (GM), управляющей развитием вычислительного процесса на графе агрегата (см. рисунок 2).

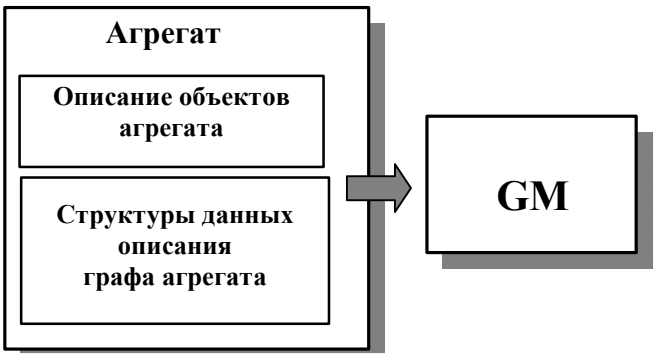


Рисунок 2. Функциональная схема агрегата

Процедурная часть агрегата содержит декларирующую часть, в которой описываются используемые в нем объекты. Алгоритм управления вычислительным процессом реализован в отдельном модуле GM. Процедурная часть агрегата индивидуальна для каждого агрегата, в то время как граф-машина является универсальной программой, ответственной за управление ходом вычислений для всех агрегатов.

Для описания структуры данных графа используется оригинальный способ представления графа с помощью, так называемых, структур смежности графа на смежной памяти [1]. Введем две структуры, предназначенные для описания вершин графа и дуг переходов на графе:

```
typedef struct _ListTop
        {     char NameT[9];
        int  FirstDef;
              int  LastDef;
        } DEFTOP;
```

```
typedef struct _ListGraf
        {char NameA[9];
      int NambArc;
            int NambTop;
       int ArcType;
        } DEFGRAF;
    DEFTOP    ListTop[6];
    DEFGRAF  ListGraph[8];
```

В массиве ListTop каждой вершине массива ставится в соответствие участок массива ListGraph, на котором последовательно перечислены смежные с ней вершины графа. Начало и конец участка определяют параметры FirstDef и LastDef. В массиве ListGraph хранится описание дуг графа (NameA – имени дуги, ArcType – типа дуги) и номера вершины графа NambTop смежной с родительской вершиной. При этом номер вершины совпадает с номером элемента массива ListTop, где приводится полное описание вершин графа. Такой способ представления графа позволяет организовать быструю навигацию по структуре ориентированного графа в граф-машине GM. Например, для графа, представленного на рисунке 3, соответствующие массивы показаны на рисунке 4. Кодом -77 отмечена концевая вершина.

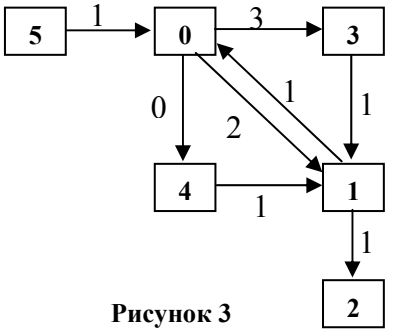


Рисунок 3

ListTop

	FirstDef	LastDef
0	0	2
1	3	4
2	-77	-77
3	5	5
4	6	6
5	7	7

ListGraph

	Arc	Top
0	0	4
1	3	3
2	1	1
3	2	0
4	1	2
5	1	1
6	1	1
7	1	0
8		

Рисунок 4

Постановка задачи

Вершины, из которых исходят параллельные дуги, будем называть вершинами ветвления и на схемах обозначим их символом “H”. Вершины, в которых “сходятся” терминирующие дуги, назовем терминирующими вершинами и обозначим символом “E”. Вершины, принадлежащие параллельным ветвям, обозначим символом “#”.

Введем понятие *правильно построенного* агрегата (ППА), схематично представленного на рисунке 5, где элементами 1 и 2 обозначены последовательные подграфы, имеющие дуги только первого типа. Правильно построенный агрегат имеет только одну вершину ветвления, одну терминирующую вершину и множество параллельных вершин, связанных с вершинами ветвления и терминации.

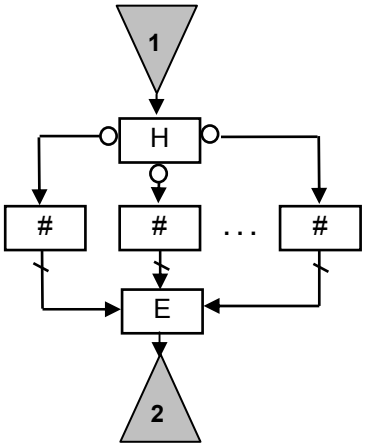


Рисунок 5. Правильно построенный агрегат

Введение понятия ППА вызвано особенностями организации универсальной граф-машины, построенной с учетом специфики стандарта MPI.

К правильно построенному агрегату отнесем граф, составленный из последовательного соединения ППА, который можно назвать “гирляндой” или многофазным агрегатом.

Очевидно, что при разработке модели параллельных вычислений пользователь не должен руководствоваться введенными ограничениями, связанных с особенностями организации граф-машины в MPI. Поэтому исходная модель организации параллельных вычислений может быть построена произвольным способом в соответствии с правилами формирования агрегатов, принятых в системе PGRAPH, например, так как это показано на рисунке 6.

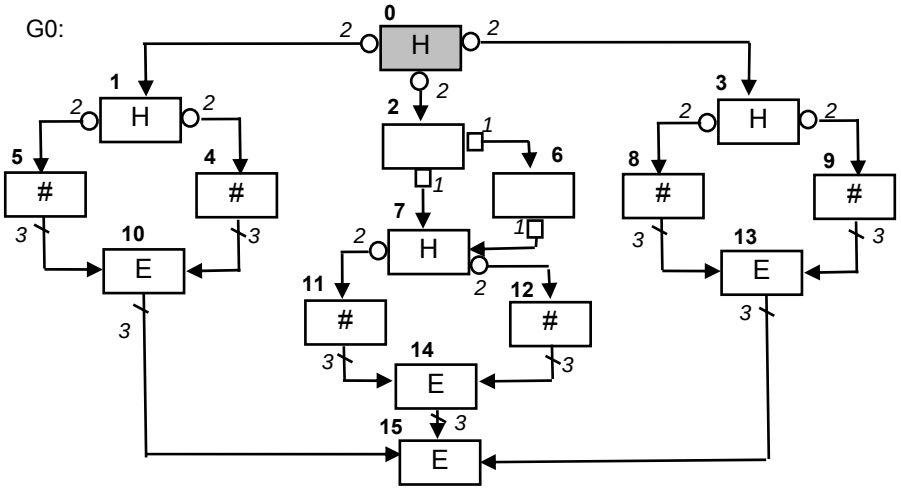


Рисунок 6. Исходный агрегат

Мастер ветвь агрегата G0 содержит 3 параллельные ветви, в каждую из которых вложены еще по две параллельные ветви, и в этом смысле G0 не является ППА. Для выполнения параллельных вычислений в системе PGRAPH необходимо произвести разбиение структуры графа G0 на совокупность правильно построенных агрегатов, например, как это показано на рисунке 7.

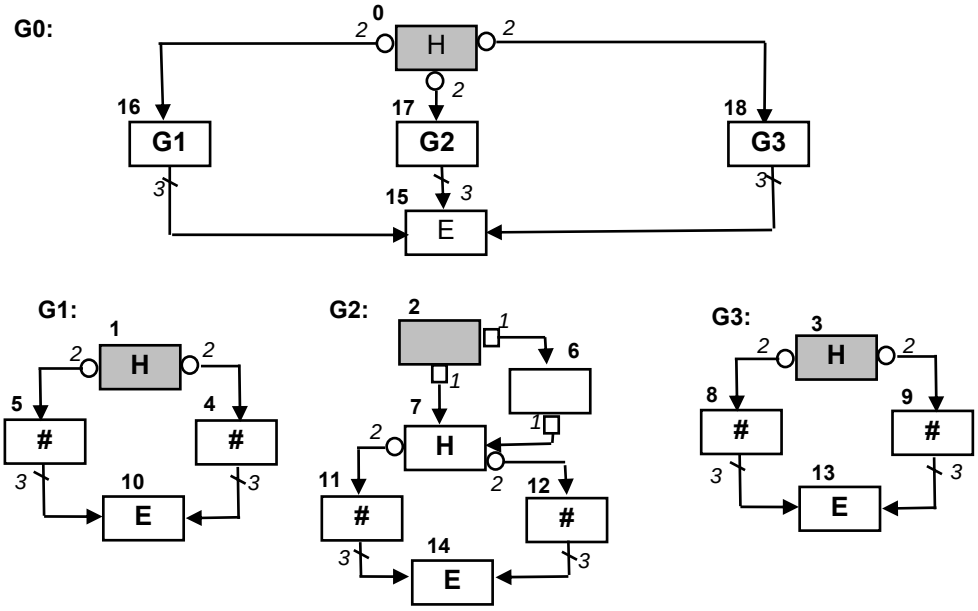


Рисунок 7. Разбиение агрегата G0 на правильные компоненты

Можно сформулировать следующую формальную постановку задачи:

Пусть задан ориентированный граф $G_0 = (A, \Psi), |A| = n$ и пусть $\widetilde{A} = \{A_1, ..., A_n, G_1, ..., G_r\}$. Требуется найти разрезание $R(\widetilde{A}) = (\widetilde{A}^{(1)}, \widetilde{A}^{(2)}, ..., \widetilde{A}^{(r)})$ (здесь $\widetilde{A}^{(i)}$ - подмножество множества вершин $\widetilde{A}$) графа G_0 на r *правильно построенных подграфов* G_i, таких что $\widetilde{A} = \bigcup_{k=1}^{r} \widetilde{A}^{(k)}, \widetilde{A}^{(i)} \bigcap \widetilde{A}^{(j)} = \varnothing, i \neq j$

Описание алгоритма

Поставленная задача предполагает выполнение предварительного анализа графа с целью изучения его строения и выявления требуемых подструктур с уточнением инфраструктуры графа относительно выделенных подструктур. Обычно этим занимается теория нумераций графов [3].

Вершины графа будем помечать метками, имеющие следующую семантику:

$M = < N_{ur}.Type.N_1.N_2.\cdots.N_{Nur} >$, где N_{ur} - количество уровней иерархии при вложенности параллельных структур друг в друга; N_i - номер вершины на i-м уровне иерархии (если *i=Nur*) или номер иерархии вложения; *Type* – тип вершины (*Type=H* для вершины ветвления; *Type=E* для терминальных вершин; для остальных *Type=V*).

С помощью алгоритма поиска в глубину [3] на первом этапе мы получаем разметку вершин графа, представленную на рисунке 8. При этом использовалось описанное выше представление графа с помощью структур смежности графа на смежной памяти (массивы ListTop и ListGraph), которое автоматически получается после предварительной обработки исходного графа G0.

Несложный анализ меток вершин разрезаемого графа позволяет определить число правильно построенных и вложенных компонент графа. На самом деле, количество ППА равно числу вершин типа «H» минус 1 (мастер ветвь G0 не в счет). Для нашего примера, где тип «H» имеют вершины <0.H>, <1.H.1>, <1.H.2> и <1.H.3>, необходимо построить как минимум 3 правильные структуры: G1, G2 и G3.

Следует отметить, что если в графе между метками одного уровня, например, <1.H.1> и <1.E.1> на параллельных ветках встречаются последовательно расположенные вершины с одинаковыми метками, например, <2.V.1.1>, то их необходимо агрегировать в одну вершину. Новые вершины дописываются в конец массива ListTop.

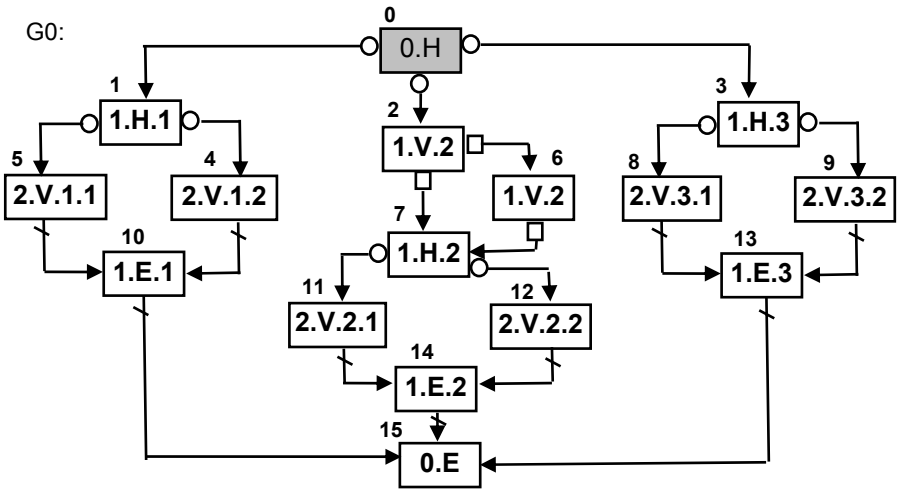


Рисунок 8. Размеченный граф

На втором этапе алгоритма несложными преобразованиями массивов ListTop и ListGraph формируется мастер ветвь алгоритма G0. Для оформления ППА, вложенных в мастер ветвь G0, необходимо для каждого агрегата (Gi) найти корневую вершину, которая является смежной вершине <0.H>, и концевую вершину, стоящую в данной параллельной ветке перед вершиной <0.E>. При этом входная вершина помечается как корневая, а для концевой вершины ППА в массиве ListTop записывается соответствующий код (-77). Аналогично обрабатываются параллельные структуры, вложенные в любые параллельные ветки исходного графа.

Заключение

В статье предложен алгоритм структурного преобразования модели параллельного алгоритма, представленного графом управления. Данный алгоритм позволяет без участия разработчика программы произвести все необходимые структурные преобразования графа управления модели к виду пригодному для реализации вычислений в среде MPI. Можно показать, что алгоритм имеет линейную сложность $O(n)$, где n – число вершин в графе.

Литература

1. Коварцев А.Н. Автоматизация разработки и тестирования программных средств. - Самарский государственный аэрокосмический университет, Самара, 1999. - 150 с.

2. Жидченко В.В., Коварцев А.Н. Моделирование синхронных параллельных вычислений при построении математических моделей сложных систем // Первая международная конференция «Системный анализ и информационные технологии» САИТ-2005: Труды конференции. В 2т. Т.2. – М.: КомКнига, 2005. – С. 154-160.

3. Евстигнеев В.А. Применение теории графов в программировании. – М.: Наука, 1985. -352 с.

Параллельная реализация алгоритма Камеды-Вейнера

А.В. Цыганов
Ульяновский государственный педагогический университет им. И. Н. Ульянова, г. Ульяновск

Введение

Конечные автоматы (КА) находят самое широкое применение в системах обработки информации. В качестве наиболее показательных примеров можно привести системы обработки больших массивов текста, машинного перевода и др. Возникающие при этом автоматы могут иметь десятки и сотни тысяч состояний и переходов.

Во многих приложениях важную роль играет способ представления КА в памяти компьютера. Прежде всего, перед разработчиками встает вопрос о выборе между двумя эквивалентными (с точки зрения описываемых ими языков) типами автоматов: недетерминированными конечными автоматами (НКА) и детерминированными конечными автоматами (ДКА). Как правило, НКА позволяют повысить наглядность представления данных и сэкономить память, но они сложнее моделируются на ЭВМ. Напротив, ДКА обычно являются более громоздкими, чем эквивалентные им НКА, но моделировать их работу на ЭВМ гораздо проще.

В процессе работы с обоими типами КА может потребоваться их минимизация, которая может выполняться, например, по числу состояний (вершинная минимизация), по числу переходов (дуговая минимизация) или по другим критериям. Для ДКА минимизация числа состояний может быть выполнена за полиномиальное время, в то время как для НКА аналогичная задача является NP-трудной.

В настоящей работе описывается параллельная реализация алгоритма вершинной минимизации НКА Камеды-Вейнера, предложенного в [1]. Несмотря на то, что данный алгоритм был сформулирован около 40 лет назад, описания его практических реализаций в литературе практически отсутствуют. Более подробно с теорией НКА можно ознакомиться, например, в [2].

1. Постановка задачи и описание алгоритма Камеды-Вейнера

НКА называется пятерка вида: $A=(Q, \Sigma, \delta, S, F)$, где Q – некоторое конечное множество состояний (вершин), Σ – рассматриваемый алфавит, δ – функция переходов вида $\delta: Q\times\Sigma\to 2^Q$ (2^Q – булеан множества Q), $S\subseteq Q$ – множество стартовых состояний (входов), $F\subseteq Q$ – множество финальных состояний (выходов).

Задача вершинной минимизации НКА состоит в построении такого автомата, который описывал бы тот же самый регулярный язык, что и исходный автомат, но имел бы при этом как можно меньше состояний.

Введем следующие обозначения. Пусть A – некоторый КА, тогда через $\overline{A}$ будем обозначать автомат, зеркальный к A, а через $D(A)$ будем обозначать автомат, получаемый из A процедурой детерминизации (построения подмножеств). Для ДКА A через $\widetilde{A}$ будем обозначать редуцированный автомат (т.е. автомат, получаемый из A объединением эквивалентных состояний).

Пусть A – некоторый НКА, $B=D(A)$, $C = D(\overline{A})$, а $\widetilde{B}$ и $\widetilde{C}$ – соответствующие им редуцированные автоматы. Состояниями автоматов B, C, $\widetilde{B}$ и $\widetilde{C}$ являются подмножества множества состояний исходного автомата A (заметим, что после выполнения процедуры детерминизации число состояний в автомате может экспоненциально возрасти).

Матрицей RAM (Reduced Automaton Matrix) назовем матрицу, состоящую из нулей и единиц, число строк в которой совпадает с числом состояний автомата $\widetilde{B}$, а

число столбцов – с числом состояний автомата $\mathcal{C}$. Пусть $p_1,p_2,\ldots,p_m$ – состояния автомата $\mathcal{B}$, а $q_1,q_2,\ldots,q_n$ – состояния автомата $\mathcal{C}$, тогда элемент r_{ij} матрицы RAM равен 0, если $p_i \cap q_j = \varnothing$ и 1 в противном случае.

Некоторую пару подмножеств строк и столбцов матрицы RAM назовём *блоком* (гридом), если, во-первых, на всех их пересечениях стоят 1, и, во-вторых, это множество нельзя пополнить ни строкой, ни столбцом без нарушения первого свойства. *Покрытием* назовем такое множество блоков, что каждая единица матрицы RAM принадлежит, по крайней мере, одному из них. *Минимальным покрытием* назовем покрытие, содержащее наименьшее возможное количество блоков.

По редуцированному автомату $\mathcal{B}$ и данному покрытию Z матрицы RAM с помощью специальной процедуры, называемой в [1] *правилом пересечений* (intersection rule), можно построить НКА $I(Z,\mathcal{B})$, который, при выполнении дополнительных требований, может оказаться эквивалентным исходному.

Алгоритм вершинной минимизации НКА A, предложенный в [1], можно сформулировать следующим образом:

1. Построить автоматы $B=D(A)$ и $C = D(\overline{A})$.
2. Для автоматов B и C построить редуцированные автоматы $\mathcal{B}$, $\mathcal{C}$.
3. По автоматам $\mathcal{B}$ и $\mathcal{C}$ построить матрицу RAM.
4. Найти минимальное покрытие матрицы RAM. Пусть i_0 – число блоков в минимальном покрытии. Положим $i=i_0$.

Для каждого покрытия Z, состоящего из i блоков построить НКА $I(Z,\mathcal{B})$ и проверить его эквивалентность исходному автомату A.

Если эквивалентного НКА не найдено, то положить $i=i+1$ и перейти к пункту А.

Заметим, что проверка эквивалентности автомата $I(Z,\mathcal{B})$ исходному автомату A может быть выполнена разными способами (в работе [1] их приведено два). Вспомогательная задача, решаемая на шаге 4 алгоритма, является разновидностью задачи о покрытии множества (SCP – Set Covering Problem). Как известно, SCP в общем случае является NP-трудной задачей.

2. Практическая реализация алгоритма Камеды-Вейнера

Программная реализация алгоритма вершинной минимизации НКА Камеды-Вейнера была выполнена на языке C++ с использованием кросс-платформенной библиотеки HeO. С подробным описанием библиотеки можно ознакомиться в работе [3] (официальная страница проекта расположена по адресу: http://code.google.com/p/heo/, в настоящее время для скачивания доступна версия 1.1 библиотеки).

Для реализации алгоритма минимизации были написаны все необходимые классы, методы и вспомогательные функции, а для проведения численных экспериментов – процедуры случайной генерации НКА с заданными свойствами.

Результаты численных экспериментов подтвердили, что наиболее трудоемкой операцией в данном алгоритме является шаг 4, так как даже для НКА с небольшим числом состояний размеры соответствующих матриц RAM и число блоков в них могут достигать очень больших значений. Поэтому решено было распараллелить именно его. Распараллеливание шага 4 было выполнено с помощью двух технологий параллельного программирования: MPI и OpenMP.

Поиск оптимального покрытия Z, для которого автомат $I(Z,\mathcal{B})$ является эквивалентным исходному автомату A, выполняется полным перебором для всех i =1,2,…n–1, где n – число состояний исходного автомата. Все пространство поиска при этом делится поровну между всеми доступными процессорами.

Поскольку из-за полного перебора нахождение точного решения рассматриваемой задачи может потребовать значительных временных затрат, в программе имеется возможность нахождения приближенного решения рассматриваемой задачи с помощью параллельных метаэвристических алгоритмов библиотеки HeO (см., [4]).

Заключение

В работе описана параллельная реализация алгоритма вершинной минимизации НКА Камеды-Вейнера. Точное решение данной задачи даже для НКА с небольшим числом состояний может потребовать больших вычислительных затрат, поэтому актуальным остается вопрос разработки эффективных приближенных методов ее решения.

В настоящее время автором ведутся работы по реализации для подзадачи поиска минимального покрытия параллельной версии мультиэвристического алгоритма [5], основанного на использовании метода ветвей и границ в сочетании с комплексом эвристик для выбора разделяющего элемента (заметим, что данный алгоритм может быть применен для достаточно широкого круга задач дискретной оптимизации).

Литература

1. *Kameda T., Weiner P.* On the state minimization of nondeterministic finite automata // IEEE Transactions on Computers. 1970. Vol. C-19, no. 7. P. 617–627.

2. *Мельников Б.Ф.* Недетерминированные конечные автоматы. – Тольятти: Изд-во ТГУ, 2009. – 160 с.

3. *Цыганов А.В., Булычов О.И.* HeO: библиотека метаэвристик для задач дискретной оптимизации // Программные продукты и системы. – 2009. – №4. – С. 148–151.

4. *Цыганов А. В., Булычов О.И.* Параллельные эвристические алгоритмы для задачи вершинной минимизации недетерминированных конечных автоматов // Вестник ВГУ. Серия: Системный анализ и информационные технологии. – 2010. – №1. – С. 60–63.

5. *Melnikov B.F., Tsyganov A.V., Bulychov O.I.* A multi-heuristic algorithmic skeleton for hard combinatorial optimization problems. // CSO'09: Proceedings of the 2009 International Joint Conference on Computational Sciences and Optimization. – Washington, DC, USA: IEEE Computer Society, 2009. – P. 33–36.

Интеграция контроля доступа для распределенной системы управления физическим экспериментом

И.С. Ястребов
Европейский Институт Ядерных Исследований (ЦЕРН), Женева

Введение

На сегодняшний день одними из самых крупных физических экспериментов являются ускорители элементарных частиц. Ускоритель элементарных частиц – класс устройств, созданный для получения элементарных частиц высоких энергий. В основе работы ускорителя заложено взаимодействие заряженных частиц с электрическим и магнитным полями [1]. Огромное число датчиков, измерительных и исполнительных устройств в современных физических экспериментах (в частности, акселераторов) требуют развертывания распределенной системы управления экспериментом.

Энергия, хранящаяся в физических устройствах, колоссальна. Поэтому возникновение возможных внештатных ситуаций и повреждений оборудования вызывает серьёзную обеспокоенность. В связи с этим, для физических экспериментов разрабатываются многоступенчатые системы защиты, включающие в себя помимо прочего комплекс управления доступом.

Система управления доступом работает превентивно и поэтому является простым и экономичным способом защиты оборудования. Другие же системы защиты, такие как блокировки, срабатывают уже после возникновения ошибки, и обычно для продолжения функционирования устройства требуется вмешательство операторов. Предотвращение таких ситуаций намного более эффективно [2].

В крупных распределенных системах управление доступом связано с рядом проблем. Одной из таких проблем является сложность интеграции контроля доступа с существующей системой управления. Основная проблема заключается в том, что в рамках единой гетерогенной системы может одновременно использоваться несколько коммуникационных протоколов, каждый из которых предоставляет собственный интерфейс управления. При этом расхождения в интерфейсах коммуникационных протоколов приводят к снижению интероперабельности, что осложняет создание единой системы контроля доступа. Наиболее полную интероперабельность интегрируемых систем обеспечивает семантическая интеграция [3]. Инструментом семантической интеграции являются онтологии. Онтология – это точная спецификация концептуализации [4]. Онтология включает машинно-интерпретируемые формулировки основных понятий предметной области и отношения между ними. В данной статье предлагается метод семантической интеграции комплекса контроля доступа для распределенной системы управления физическим экспериментом на основе онтологической модели.

Распределенная система управления физическим экспериментом

Рассмотрим систему управления физическим экспериментом на примере Большого Адронного Коллайдера, для которого автором был разработан комплекс контроля доступа. Большой Адронный Коллайдер (БАК) — самая сложная экспериментальная установка, когда-либо созданная человеком. В экспериментах на этом ускорителе высоких энергий будут изучаться самые экстремальные состояния материи. Для своего функционирования БАК использует кольцевой туннель длиной 27 км, а также одни из самых мощных на сегодняшний день сверхпроводящих магнитов и радиочастотных резонаторов [5].

Для управления физическими устройствами в ЦЕРНе используются различные промышленные компьютеры, взаимодействующие с программно-электронными

модулями. Специализированные модули программной электроники предназначены для записи и считывания информации с устройств. Взаимодействие между большинством низкоуровневых процессов обеспечивается с использованием разработанной в ЦЕРНе системы управления Controls Middleware (CMW), общая архитектура которой представлена на рис. 1.

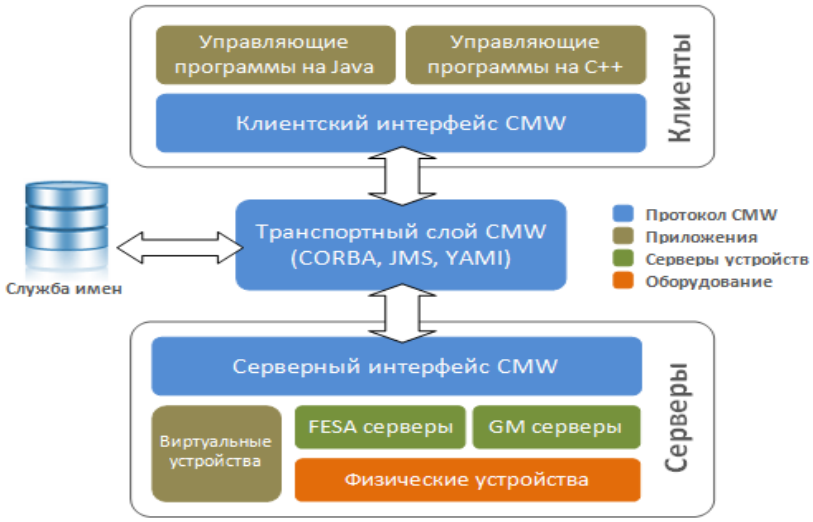


Рис. 1. Архитектура CMW

В основе реализации системы управления используются несколько коммуникационных протоколов:

1. Общая архитектура брокера объектных запросов CORBA. Используемые реализации: Remote Device Access (RDA), созданная на базе omniORB и JacORB.
2. YAMI – простой коммуникационный протокол, позволяющий выполнять все базовые операции удаленного доступа.
3. JMS (Java Message Service) – стандарт программного обеспечения для рассылки сообщений в среде Java. Используемые реализации: SonicMQ и ActiveMQ.

Семантическая модель системы контроля доступа

Наличие нескольких коммуникационных протоколов затрудняет интеграцию системы контроля доступа. Для повышения интероперабельности автором предлагается семантическая модель, описывающая базовые понятия и операции в предметной области контроля доступа для распределенных систем управления физическими экспериментами. Семантическая модель создана на базе онтологии высокого уровня BWW [6]. Статическое описание структуры системы в терминах онтологии BWW включает концепты, существующие в системе и их свойства. Основные элементы статической модели, описывающей базовые понятия и операции системы, представлены на рис. 2.

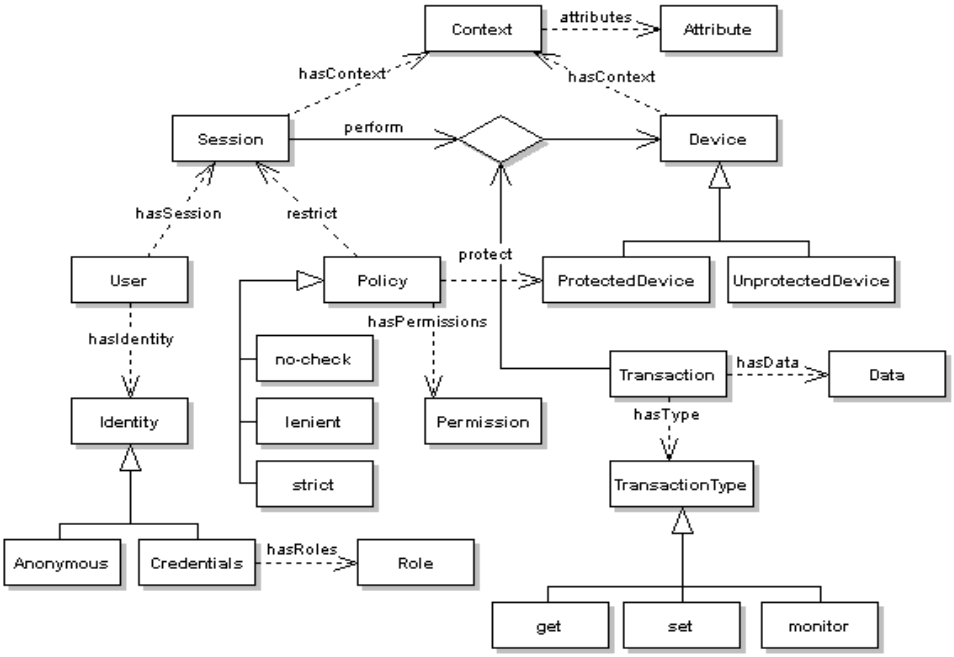


Рис. 2. Статическая модель системы контроля доступа

Ключевым элементом онтологии является пользователь (*User*), который может работать в системе как анонимно (*Anonymous*), так и будучи идентифицированным (*Credentials*). Для идентификации в системе пользователь использует назначенные ему идентификаторы, для каждого из которых мы задаем список доступных ролей (*Role*).

Для работы с системой управления пользователь создаёт объект сессии (Session). Пользователь может создать сессию и выбрать из доступных ему ролей одну или несколько в качестве активных.

Объектами доступа в системе управления являются физические устройства (*Device*). В рамках работы с системой управления пользователь выполняет определенные транзакции (*Transaction*) с оборудованием. Задачей используемых политик безопасности (*Policy*) является ограничение доступа, то есть предотвращение выполнения неавторизованных транзакций при работе с защищенными устройствами (*ProtectedDevice*).

Политики безопасности выполняют авторизацию транзакций, используя множество правил доступа (*Permission*), а также контекстную информацию о субъектах и объектах доступа (*Context*).

Для описания динамического поведения системы в онтологии BWW используются состояния, события (рис. 3) и операции (рис. 4). Изначально пользователь, работающий в системе управления, не имеет созданной сессии, при этом какая-либо информация о субъекте пока отсутствует (*SessionUnknown*). С целью выполнения защищенных транзакций в системе управления, пользователь выполняет аутентификацию – процесс, позволяющий идентифицировать субъекта по его данным. Когда субъект доступа известен, создается сессия (действие *Authentication*, переход в состояние *SessionExists*).

В состоянии SessionExists субъект может активировать роль, расширив таким образом свои права доступа (*ActivateRole*). После завершения выполнения операций, требующих прав доступа роли, роль можно деактивировать (*DeactivateRole*).

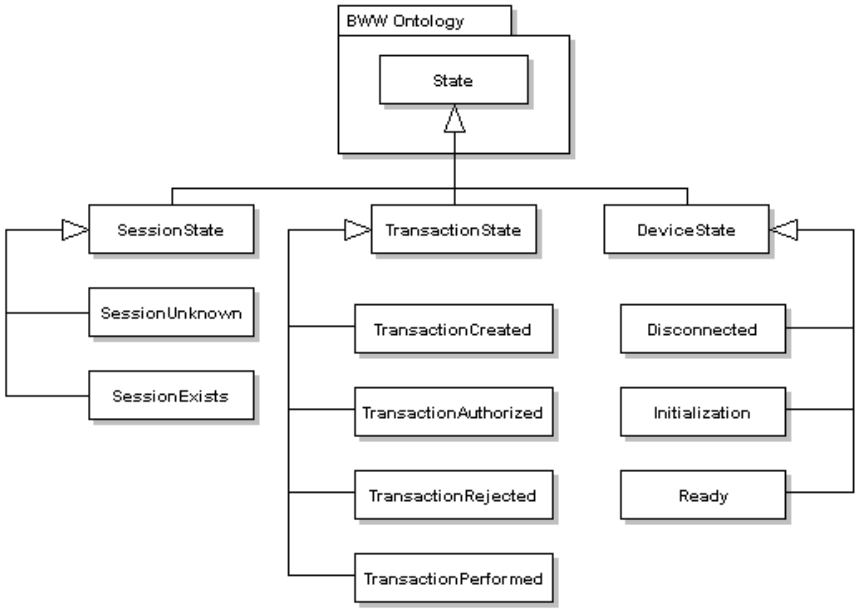


Рис. 3. Состояния системы контроля доступа

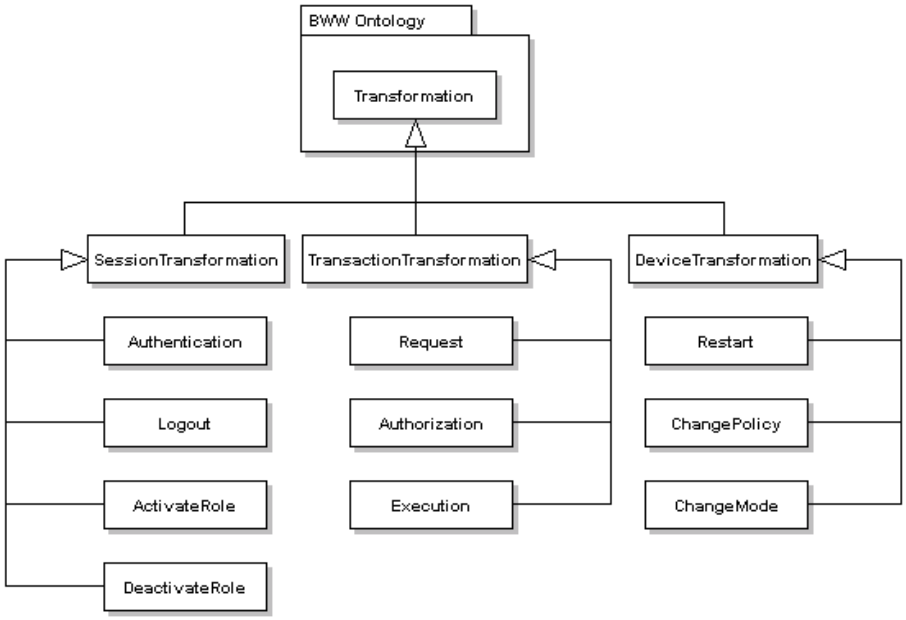


Рис. 4. Операции системы контроля доступа

Работа пользователя с системой управления заключается в выполнении транзакций с оборудованием. Пользователь выполняет действие *Request*, в результате которого создается транзакция (*TransactionCreated*). Транзакция представляет собой запрос пользователя на обмен информацией с устройством. Затем система контроля доступа выполняет авторизацию, то есть проверку прав доступа субъекта на выполнение запроса (*Authorization*). В результате авторизации запрос либо блокируется (*TransactionRejected*), либо успешно выполняется (*Execution*, состояние *TransactionPerformed*).

Заключение

Проблема обеспечения интероперабельности информационных систем имеет фундаментальный характер. Она актуальна как для унаследованных систем, которые

требуется связать с вновь создаваемыми, так и для вновь проектируемых, в которых необходимо предусмотреть возможности реализации взаимодействия с другими информационными системами в перспективе, при изменении требований к ним.

Предложенная онтологическая модель контроля доступа для распределенной системы управления служит основой для семантической интеграции, поскольку задает семантику базовых понятий и операции в данной предметной области. За счет использования концептуальной модели методика обеспечивает повышение интероперабельности: значительно расширяет круг поддерживаемых коммуникационных протоколов и методов управления доступом.

На основе предложенных моделей автором реализован комплекс программного обеспечения CMW-RBAC. В настоящее время данный комплекс программ используется для контроля доступа ко всему оборудованию Большого Адронного Коллайдера. В дальнейшем планируется расширение области применения системы.

В целом, полученные результаты подтвердили состоятельность предлагаемой модели с точки зрения решения задачи интеграции контроля доступа в распределенную систему управления физическим экспериментом.

Литература

1. Бояркин О.М. Введение в физику элементарных частиц. – М.: URSS, 2006. – 260 с.
2. Yastrebov I. Role-Based Access Control for the Large Hadron Collider at CERN. // International Journal of Computers, Communications & Control, №3. – Romania: Agora University Editing House, 2010. – V.5. – P.398–409.
3. Obrst L. Ontologies for semantically interoperable systems. // Conference on Information and Knowledge Management. – New Orleans, USA, 2003. – P.366–369.
4. Gruber T.R. A Translation Approach to Portable Ontology Specification. // Knowledge Acquisition. – Academic Press Ltd. London, UK, 1993. – V.5. №2. – P.199–220.
5. Why the LHC. CERN. 2008. http://public.web.cern.ch/public/en/LHC/WhyLHC-en.html
6. Weber R. Ontological Foundations of Information Systems // Monograph №4. Australia: Melbourne, Vic., Coopers & Lybrand and the Accounting Association of Australia and New Zealand. – 1997.

Построение быстрых алгоритмов преобразований в разностно-упорядоченных системах дискретных функций Уолша

М.Ю. Васильева
Казанский государственный технологический университет, г. Казань

Введение

В настоящее время актуальным направлением в информационных технологиях является разработка эффективных методов и средств цифровой обработки сигналов и изображений. На развитие этого направления значительное влияние оказал и продолжает оказывать аппарат дискретных ортогональных преобразований (ДОП).

Один из основных факторов, способствующих широкому использованию ряда ДОП в прикладных задачах, заключается в возможности их вычисления посредством быстрых алгоритмов, которые обладают существенно меньшей вычислительной сложностью по сравнению с прямыми (классическими) алгоритмами преобразований [1].

Особое место в ряду практически важных ДОП занимают преобразования в различных упорядочениях систем дискретных функций Уолша (ДФУ), для реализации которых используются только операции сложения и вычитания. К наиболее известным в обработке сигналов упорядочениям ДФУ в системе относятся следующие: Уолша-Качмажа, Уолша-Пэли, Уолша-Адамара. Известны также новые фиксированные способы упорядочения ДФУ в системе, получаемые определенной перестановкой базисных функций опорных матриц [2].

На практике для известных дискретных систем используется лишь незначительное количество фиксированных способов упорядочения базисных векторов, хотя известно, что эффективность решения отдельных задач цифровой обработки сигналов зависит от упорядочения функций в системе. Например, разностно-упорядоченные системы ДФУ нашли эффективное применение при решении задач сжатия цифровых изображений [3]. Таким образом, актуальной задачей является построение быстрых алгоритмов преобразований по синтезированным системам ДФУ.

Быстрые алгоритмы одномерных и двумерных преобразований

Вычисление спектральных коэффициентов двумерного дискретного преобразования Уолша (ДПУ) возможно несколькими способами: классическим способом и используя алгоритмы, построенные на основе быстрого двумерного ДПУ построчно-столбцовым способом [4]. Например, вычисление может быть сведено к двум последовательно выполняемым операциям перемножения трех матриц:

$$\mathbf{F} = \mathbf{HAD}_N^{\,T}\,\mathbf{f}\,\mathbf{HAD}_N,$$

где f – матрица исходных данных; F – матрица спектральных коэффициентов; $\mathbf{HAD}_N$ - матрица Уолша-Адамара.

Классический способ вычисления операции перемножения матриц 8×8 требует $N^2(N-1)=448$ операций сложения и вычитания. С учетом этого для вычисления двумерного ДПУ размерности 8×8 потребуется 896 операций сложения и вычитания.

Построчно-столбцовый способ реализации быстрого двумерного ДПУ требует всего $2(N^2\log_2 N)=384$ операций сложения и вычитания. При этом уменьшение числа операций достигается путем факторизованного разложения матрицы Уолша на разряженные множители в виде:

$$\mathbf{HAD}_N = \prod_{i=0}^{n-1}\mathbf{HAD}_N^{(i)},$$

где n – количество матриц факторизации ($N = 2^n$);

$\mathbf{HAD}_N^{(i)}$ - слабозаполненная матрица вида

$$\mathbf{HAD}_N^{(i)} = (\mathbf{I}_{2^{n-i-1}} \otimes \mathbf{HAD}_2 \otimes \mathbf{I}_{2^i}),$$

где $\mathbf{I}_k$ - единичная матрица порядка *к***.**

При расмотрении алгоритмов быстрого ДПУ ограничимся синтезом алгоритмов прямого преобразования, так как они однозначно определяют и алгоритмы обратного преобразования.

Запишем формулы одномерного преобразования в разностно-упорядоченных системах ДФУ. В матричном виде они будут иметь вид:

$$\mathbf{F} = \mathbf{S}_N^D \mathbf{HAD}_N \mathbf{f}, \qquad (1)$$

где f – вектор исходных данных; F – вектор спектральных коэффициентов; $\mathbf{S}_N^D$ – матрица *D*-перестановок.

Элементы матрицы *D*-перестановок $\mathbf{S}_N^D = \left[s_{ij}\right]_{i,j=\overline{0,N-1}}$ формируются следующим образом:

$$s_{ij} = \begin{cases} 1, & \textit{если } j = p_i, \\ 0, & \textit{в остальных случаях}, \end{cases}$$

$$s_{ij} = 0, \textit{тогда } s_{i,p_i} = 1.$$

где p_i - элементы вектора перестановочной последовательности [2].

Для вектора исходных данных при *N*=8 матрица $\mathbf{HAD}_N$ может быть представлена как произведение трех разреженных множителей:

$$\mathbf{HAD}_8 = \mathbf{HAD}_8^{(2)} \mathbf{HAD}_8^{(1)} \mathbf{HAD}_8^{(0)},$$

где $\mathbf{HAD}_8^{(0)} = \mathbf{I}_4 \otimes \mathbf{HAD}_2 = diag(\mathbf{HAD}_2, \mathbf{HAD}_2, \mathbf{HAD}_2, \mathbf{HAD}_2)$;

$$\mathbf{HAD}_8^{(1)} = \mathbf{I}_2 \otimes \mathbf{HAD}_2 \otimes \mathbf{I}_2 = \begin{bmatrix} \mathbf{HAD}_2 \otimes \mathbf{I}_2 & \\ & \mathbf{HAD}_2 \otimes \mathbf{I}_2 \end{bmatrix};$$

$$\mathbf{HAD}_8^{(2)} = 1 \otimes \mathbf{HAD}_2 \otimes \mathbf{I}_4 = \begin{bmatrix} \mathbf{I}_4 & \mathbf{I}_4 \\ \mathbf{I}_4 & -\mathbf{I}_4 \end{bmatrix}.$$

Матрица *D*-перестановок $\mathbf{S}_8^D$ имеет следующий вид:

$$\mathbf{S}_8^D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

В работе также проведено построение графа вычислительных процедур 8-и точечного быстрого преобразования в базисе разностно-упорядоченных системах ДФУ с учетом формулы (1).

Приведем формулы вычисления двумерных преобразований в разностно-упорядоченных системах ДФУ при размерности 8×8. Двумерные преобразования реализуем на основе быстрых одномерных алгоритмов ДПУ. Предварительно преобразуем входную матрицу размером 8×8 в вектор размерности 64, используя построчное сканирование [5].

Запишем формулы вычисления одномерного преобразования в разностно-упорядоченных системах ДФУ вектора размерности 64. В матричном виде будут они иметь следующий вид:

$$\mathbf{F} = \mathbf{S}_{64}^{\mathbf{D}(2)}\,\mathbf{HAD}_{64}^{(2)}\,\mathbf{S}_{64}^{\mathbf{D}(1)}\,\mathbf{HAD}_{64}^{(1)}\,\mathbf{f}\,, \tag{2}$$

где $\mathbf{HAD}_{64}^{(1)} = \mathbf{I}_8 \otimes \mathbf{HAD}_8 = diag(\mathbf{HAD}_8, \mathbf{HAD}_8, \mathbf{HAD}_8, \mathbf{HAD}_8, \mathbf{HAD}_8, \mathbf{HAD}_8, \mathbf{HAD}_8, \mathbf{HAD}_8)$;

$$\mathbf{HAD}_{64}^{(2)} = \mathbf{HAD}_8 \otimes \mathbf{I}_8 = \begin{bmatrix} \mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 \\ \mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 \\ \mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 \\ \mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 \\ \mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 \\ \mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 \\ \mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 \\ \mathbf{I}_8 & -\mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 & \mathbf{I}_8 & \mathbf{I}_8 & -\mathbf{I}_8 \end{bmatrix};$$

$\mathbf{S}_{64}^{\mathbf{D}(1)} = \mathbf{I}_8 \otimes \mathbf{S}_8^{\mathbf{D}} = diag\,(\mathbf{S}_8^{\mathbf{D}}, \mathbf{S}_8^{\mathbf{D}}, \mathbf{S}_8^{\mathbf{D}}, \mathbf{S}_8^{\mathbf{D}}, \mathbf{S}_8^{\mathbf{D}}, \mathbf{S}_8^{\mathbf{D}}, \mathbf{S}_8^{\mathbf{D}}, \mathbf{S}_8^{\mathbf{D}})$;

$$\mathbf{S}_{64}^{\mathbf{D}(2)} = \mathbf{S}_8^{\mathbf{D}} \otimes \mathbf{I}_8 = \begin{bmatrix} \mathbf{I}_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I}_8 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{I}_8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}_8 & 0 \\ 0 & \mathbf{I}_8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{I}_8 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{I}_8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{I}_8 \end{bmatrix}.$$

В работе предлагается также второй вариант вычисления 64-точечного быстрого преобразования в разностно-упорядоченных системах ДФУ, при котором перестановка вектора спектральных коэффициентов размерности 64 выполняется только на последнем этапе.

В этом случае получим следующее представление формулы (2.38):

$$\mathbf{F} = \mathbf{S}_{64}^{\mathbf{D}}\,\mathbf{HAD}_{64}^{(2)}\,\mathbf{HAD}_{64}^{(1)}\,\mathbf{f}\,, \tag{3}$$

где матрица D-перестановок $\mathbf{S}_{64}^{D}$ имеет вид:

$$\mathbf{S}_{64}^{D} = \mathbf{S}_{8}^{D} \otimes \mathbf{S}_{8}^{D} = \begin{bmatrix} \mathbf{S}_{8}^{D} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{S}_{8}^{D} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{S}_{8}^{D} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{S}_{8}^{D} & 0 \\ 0 & \mathbf{S}_{8}^{D} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{S}_{8}^{D} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{S}_{8}^{D} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{S}_{8}^{D} \end{bmatrix}.$$

В работе проведено построение соответствующих графов вычислительных процедур 64-точечного быстрого преобразования в разностно-упорядоченных системах ДФУ (2, 3).

Заключение

При выполнении быстрых алгоритмов двумерных ДПУ размерности 8×8 [1] и предложенных алгоритмов быстрого преобразования в разностно-упорядоченных системах ДФУ количество операций сложения и вычитания одинаково и составляет 384 операции. Также требуется 64 перестановки при выполнении первого варианта предложенного алгоритма и 48 при выполнении второго. Отметим, что в предложенных алгоритмах выигрыш достигается на уровне организаций вычислений (отсутствие операций транспонирования, работа с одномерным массивом).

Это позволяет сделать вывод, что алгоритмы быстрых преобразований в разностно-упорядоченных системах ДФУ представляют не только теоретический интерес, но и могут найти эффективное применение при решении ряда задач цифровой обработки сигналов.

Литература

1. Ахмед Н., Рао К.Р. Ортогональные преобразования при обработке цифровых сигналов: Пер. с англ. - М.: Связь, 1980. - 248 с.

2. Исмагилов И.И., Васильева М.Ю. Разностно-упорядоченные системы дискретных функций Уолша // Инфокоммуникационные технологии глобального информационного общества. Сб. науч. тр. 2-й науч.- прак. конф., Казань, 2004 г. - М: Новые технологии, 2004. – С. 298-302.

3. Исмагилов И.И., Васильева М.Ю. Сжатие цифровых изображений с использованием преобразований Уолша: алгоритмы и сравнительный анализ их эффективности // Известия вузов. Проблемы энергетики. – 2008. - № 9-10. - С. 91-99.

4. Ярославский, Л.П. Введение в цифровую обработку изображений / Л.П. Ярославский. – М.: Сов. Радио, 1979. –312 с.

5. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ-МИФИ, 2002, - 384 с.

Применение схем вычислительных процессов в системе GraphPlus templet на примере схемы Pipeline

С.В. Востокин, А.Р. Хайрутдинов
Самарский государственный аэрокосмический университет им. академика С.П Королева (национальный исследовательский университет), Самара

Введение

Накопленный опыт применения высокопроизводительных компьютеров показывает, что при решении разных по своему содержанию задач используются похожие способы организации вычислений. В связи с этим в настоящее время актуальной является задача формализации типовых алгоритмических структур параллельных вычислительных процессов [1], что позволяет снизить сложность программирования и расширить область применения высокопроизводительных вычислительных систем. В докладе рассматривается последовательность разработки схемы вычислительного процесса Pipeline [2] в системе GraphPlus templet. Данная система программирования разработана в рамках исследовательского проекта Graphplus (graphplus.ssau.ru) [3] Самарского государственного аэрокосмического университета.

Схема Pipeline

Графическое представление схемы вычислительного процесса Pipeline (конвейер) в системе Graphplus templet приведено на рисунке 1.

Схема представлена процессами трех типов. Процесс Left формирует поток данных для обработки, передаваемый по цепочке процессов. Процесс Main выполняет обработку проходящих через него данных. Данный процесс сигнализирует своего левого соседа об окончании обработки и передает результат по цепочке правому соседу. Процесс Right замыкает цепь процессов. Он получает обработанные Main данные и сигнализирует процесс Left о готовности получить следующую порцию. Схема масштабируется путем добавления дополнительных процессов типа Main в цепочку.

По описанной схеме могут распараллеливаться численные алгоритмы цифровой фильтрации. При этом процессы Main интерпретируются как фильтры, а Left и Right, соответственно, как источник и приемник потока данных. Также по схеме Pipeline могут работать алгоритмы сеточных методов, когда область расчета подвергается одномерной декомпозиции. Примером могут являться термодинамические расчеты и задачи моделирования распространения электромагнитного излучения [3]. В данных задачах процессы Left и Right вычисляют области данных, примыкающих к границам среды, а процессы Main внутреннюю область.

Задача усреднения массива

В качестве упрощенного примера приведём реализацию процесса многократного усреднения элементов одномерного массива методом скользящего среднего. Этот процесс аналогичен по методу управления вычислительным алгоритмам, использующим сеточный итеративный метод. Он используется для проверки функциональности параллельной программы, представленной на рисунке 1 и в целях отладки.

Для получения параллельной программы рисунок 1 преобразуется в представленный ниже XML-файл со спецификацией схемы процесса Pipeline.

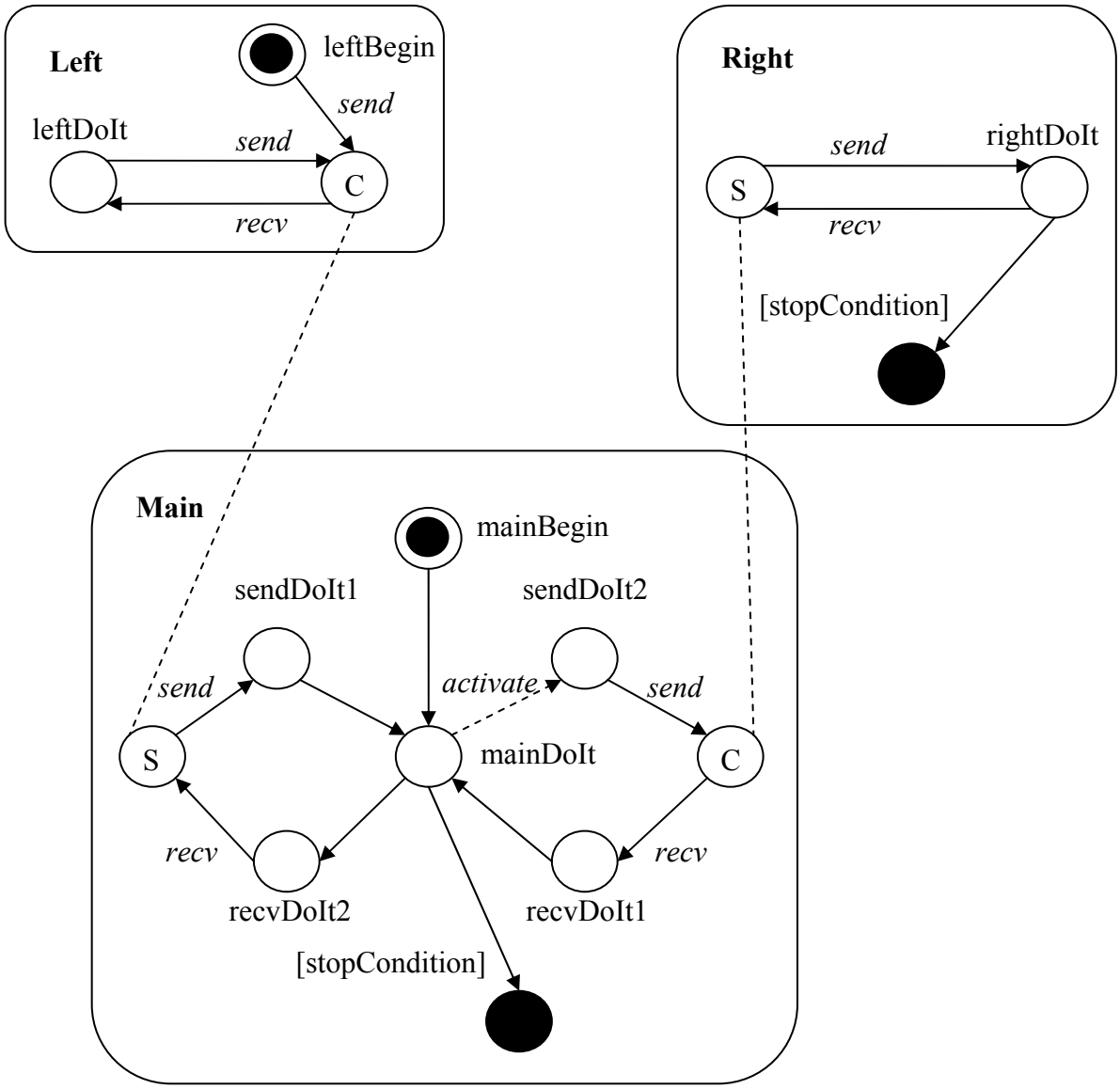


Рисунок 1. Схема Pipeline

```
<module id="Pipeline" ver="1" rem="">

<include file="gprt.h" module="GPRT"/>

<channel id="Chan" entry="S1" templet="" rem="">
   <state id="S1" type="cli" x="0" y="0" rem="">
            <message id="send" state="S2" x="0" y="0" rem=""/>
   </state>
   <state id="S2" type="srv" x="0" y="0" rem="">
            <message id="recv" state="S1" x="0" y="0" rem=""/>
   </state>
</channel>

<process id="Left" entry="leftBegin" templet="" rem="">
   <port id="LeftPort" channel="Chan" module="" type="cli" x="0" y="0" rem="">
            <receive id="recv" method="leftDoIt" x="0" y="0" rem=""/>
   </port>
   <method id="leftBegin" count="1" x="0" y="0" rem="">
      <send id="send" port="LeftPort" x="0" y="0" rem=""/>
```

```
    </method>
    <method id="leftDoIt" count="1" x="0" y="0" rem="">
            <send id="send" port="LeftPort" x="0" y="0" rem=""/>
    </method>
</process>

<process id="Right" entry="" templet="" rem="">
    <port id="RightPort" channel="Chan" module="" type="srv" x="0" y="0" rem="">
            <receive id="send" method="rightDoIt" x="0" y="0" rem=""/>
    </port>
    <method id="rightDoIt" count="1" x="0" y="0" rem="">
      <condition id="stopCondition" method="0"/>
            <send id="recv" port="RightPort" x="0" y="0" rem=""/>
    </method>
    <method id="0" count="1">
    </method>
</process>
<process id="Main" entry="mainBegin" templet="" rem="">
    <port id="MainSrv" channel="Chan" module="" type="srv" x="0" y="0" rem="">
            <receive id="send" method="sendDoIt1" x="0" y="0" rem=""/>
    </port>
    <port id="MainCli" channel="Chan" module="" type="cli" x="0" y="0" rem="">
            <receive id="recv" method="recvDoIt1" x="0" y="0" rem=""/>
    </port>
    <method id="mainBegin" count="1" x="0" y="0" rem="">
      <condition id="cond" method="mainDoIt" x="0" y="0" rem=""/>
    </method>
    <method id="sendDoIt1" count="1" x="0" y="0" rem="">
            <condition id="cond" method="mainDoIt" x="0" y="0" rem=""/>
    </method>
    <method id="mainDoIt" count="2">
      <activate method="sendDoIt2"/>
      <condition id="stopCondition" method="0"/>
      <condition method="recvDoIt2"/>
    </method>
    <method id="0" count="1">
    </method>
    <method id="sendDoIt2" count="1" x="0" y="0" rem="">
            <send id="send" port="MainCli" x="0" y="0" rem=""/>
    </method>
    <method id="recvDoIt1" count="1" x="0" y="0" rem="">
            <condition id="cond" method="mainDoIt" x="0" y="0" rem=""/>
    </method>
    <method id="recvDoIt2" count="1" x="0" y="0" rem="">
            <send id="recv" port="MainSrv" x="0" y="0" rem=""/>
    </method>
</process>

<assemble id="PipelineAsm" templet="" rem="">
    <process id="Main" module=""/>
    <process id="Left" module=""/>
    <process id="Right" module=""/>
    <channel id="Chan" module=""/>
</assemble>

</module>
```

Представленный файл может быть сформирован вручную в текстовом или специализированном XML-редакторе. Для целей автоматизации также может быть использован специализированный графический редактор. Данная схема обрабатывается транслятором gpt.exe и отображается в программный код. В настоящий момент

используется трансляция в код на языке C++. Транслятор создаёт два файла: файл заголовков gppipeline.h и файл с реализацией gppipeline.cpp. Файл заголовков отражает структуру, заданную в модели и описанную XML-файлом. Например, для класса, представляющего процесс Left, будет сгенерировано:

```
class Left:public GPRT::Process{
public:
  Left(GPRT::Assemble*a);
  ~Left();
  void setM(int left, int center, int right);
  int* getM();
private:
  bool leftBegin(/*out*/Chan::send*p1);//
  bool leftDoIt(/*in*/Chan::recv*p1,/*out*/Chan::send*p2);//

/*$GPBS$Left$!userdata!*/
  int* m;
/*$GPBS$*/

public:
  enum{
        _port_LeftPort,//
        _method_leftBegin,//
        _method_leftDoIt//
  };
public:
  void _bindPortToChan(int port,GPRT::Channel*c,GPRT::Channel::ChanActivity t);
protected:
  virtual bool _run(int _selector);
private:
  Chan* _LeftPort;//
  GPRT::Activator* _entry;
};
```

При необходимости расширить описание класса под нужды конкретной задачи программный код следует добавлять в местах, обозначенных «// place your code here». Места вставки пользовательского кода ограничены специальными тэгами, разграничивающими код, построенный автоматически по XML-описанию и добавленный вручную.

Расширим описание классов Main, Left и Right, добавив объявления обрабатываемых типов данных и методов обработки, как показано ниже.

```
struct send{//<S1>:
/*$GPBS$Chan$send*/
        int value;
/*$GPBS$*/
  };
  struct recv{//<S2>:
/*$GPBS$Chan$recv*/
        int value;
/*$GPBS$*/
  };
```

Изменим реализации соответствующих методов в классах Main, Left и Right для решения поставленной задачи.

В зависимости от подключенного модуля системы времени исполнения полученный код может исполняться по-разному. В нашем примере был подключен модуль GPRT для отладки и модуль GPRTM для многопоточного исполнения в операционной системе Windows. Эти модули входят в комплект программ Graphplus templet.

Выводы

Был продемонстрирован пример использования системы GraphPlus templet для построения алгоритма параллельных вычислений по схеме Pipeline. По схеме был автоматически сгенерирован программный код. После написания соответствующих функций тестового алгоритма усреднения методом скользящего среднего алгоритм был запущен в отладочном режиме и с использованием потоков исполнения Windows. Пример иллюстрирует технику разработки приложений в системе Graphplus templet и возможности ее практического использования для автоматизации параллельного программирования.

Литература

1. Бергизияров П.К. Программирование на типовых алгоритмических структурах с массивным параллелизмом // Вычислительные методы и программирование. 2001. Т.2, Разд. 2, С.1-16.

2. Востокин С.В., Прохоров С.А. Пример визуальной модели распределенного вычислительного процесса. — Известия СНЦ РАН, том 6,№1(10) 2004, с.180-185.

3. Востокин С.В. Графическая объектная модель параллельных процессов и ее применение в задачах численного моделирования. Самара: Издательство Самарского научного центра РАН, 2007. 286 с. ISBN 978-5-93424-284-9

Обработка изображений в распределенной массивно-многопоточной CUDA-среде

А.В. Никоноров, П.Ю. Якимов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение. Постановка задачи

В последние годы наряду с совершенствованием и ростом производительности мощных суперкомпьютеров наметилась устойчивая тенденция развития технологий параллельных вычислений на персональных компьютерах. Эта тенденция обусловлена существенным технологическим прорывом, по крайней мере, в трех направлениях.

Во-первых, это появление настольных CPU с возможностью параллельного выполнения программ. В настоящее время для персональных компьютеров доступны процессоры *i*7 с параллельным исполнением 8-ми потоков. Вторым технологическим прорывом является представленное в 2007 году компанией Nvidia альтернативное направление развития массивно-параллельных вычислений для настольных компьютеров с использованием GPU. Третий фактор, позволяющий всерьез говорить о настольных суперкомпьютерных системах, – это повышение скорости коммуникаций по системной шине для платформы x86 с развитием стандарта PCI Express. Скорость коммуникаций между узлами всегда была узким местом кластерных систем, снижающим их эффективность. Принятие подавляющим большинством производителей стандарта *Infiniband* позволяет строить весьма недорогие вычислительные системы со скоростью обмена данными между узлами до 5 GB/s.

Совокупность указанных факторов стала мотивом возрастающего интереса к созданию компактных, недорогих и, вместе с тем, высокопроизводительных систем обработки и анализа изображений нового поколения для систем видеоконтроля, видеонаблюдения, подготовки цифровых изображений к печати и др. В работах [1-3] рассматривались примеры создания компактных распределенных систем обработки изображений. В частности, в работе [1] описана реализация алгоритма коррекции цветных цифровых изображений в локальной сети компьютеров посредством специально разработанной GRID-системы, реализованной на Java с native-модулями. В этой работе показана возможность существенного ускорения одного из наиболее затратных по времени процессов подготовки цветных изображений к печати. В работах [4-5] приведены примеры дальнейшего ускорения предпечатной подготовки изображений за счет реализации алгоритмов обработки изображений в CUDA-среде.

Наиболее популярным и широко используемым во многих технологиях обработки изображений является метод скользящего окна. Известны быстрые рекуррентные (например, параллельно-рекурсивные) алгоритмы обработки изображений скользящим окном, реализованные на традиционных аппаратных средствах [6]. Создание таких же эффективных процедур на основе CUDA-технологий, реализуемых с использованием графических ускорителей NVIDIA, к сожалению, пока еще остается проблемной задачей. Связано это с тем, что вычислительная среда CUDA имеет более сложную, с точки зрения программирования, структуру по сравнению с многопоточной CPU средой или параллельной средой MPI. Основные структурные особенности среды CUDA следующие.

Во-первых, CUDA-среда имеет двухуровневую многомерную топологию исполнения потоков, которая может описывать десятки тысяч потоков. По аналогии с термином *массивно-параллельная* среда такую среду можно назвать *массивно-многопоточной*.

Во-вторых, потоки могут обмениваться данными только посредством сравнительного небольшого объема общей памяти. Такой обмен возможен в пределах небольшой области локальной связности – блока, включающего (для выпускаемых в настоящее время GPU) не более 512 потоков. Кроме того, когерентность данных в общей памяти достигается только путем приостановки всех потоков, т.е. упорядоченный обмен данными невозможен. Указанные ограничения создают серьезные проблемы при реализации в среде CUDA рекурсивных алгоритмов (сжатия данных, фильтрации и др.) [7].

Многие базовые алгоритмы обработки данных требуют существенной переработки с учетом указанных особенностей. Например, в работе [10] рассмотрен адаптированный для CUDA-среды алгоритм префиксного суммирования. В настоящей статье, являющейся развитием работ [1-5], из всего комплекса рассмотренных в них задач выделены наиболее характерные и сложные для реализации на GPU-процессорах задачи: обработка изображений скользящим окном и организация передачи видеоданных в реальном времени. В частности, рассмотрены основные особенности построения указанных алгоритмов и приводится краткое описание созданного в массивно-многопоточной CUDA-среде программного комплекса.

Рекуррентная CUDA-реализация обработки изображений скользящим окном

Продемонстрируем особенности реализации рекуррентных вычислительных алгоритмов в CUDA-среде на примере простейшего локального оконного фильтра, выполняющего расчет среднего значения. Простота этого алгоритма позволяет наглядно продемонстрировать все основные особенности среды CUDA. Вместе с тем, иллюстрируемый на этом простом примере подход может быть легко применен к более сложным локальным оконным фильтрам, применяемым в различных технологиях обработки изображений.

Пусть алгоритм обработки скользящим окном состоит в вычислении для каждой точки изображения суммы значений отсчетов в некоторой окрестности (окне):

$$s_{x,y} = \sum_{i,j=1}^{m} L_{x+i,y+j}, \quad 0 < x, y \le N, \tag{1}$$

где N – размер изображения, m – размер окна (для простоты полагаем, что изображение и окно квадратные). При этом потребуется

$$O_1 = N^2 m^2 \tag{2}$$

арифметических операций.

При рекуррентной реализации этих вычислений в CUDA-среде существенное значение играет согласование адреса памяти с номером активного потока. Будем считать обращение потока к глобальной памяти согласованным, если оно происходит к словам памяти размером 32 бита с адресами

$$I(t,k) = 32 * n * k + t, \quad k = 0,1,2..., \quad t = 0,1,2,..., \tag{3}$$

где $I(k,t)$ - адрес ячейки глобальной памяти, к которой производится обращение, n – натуральное число, характеризующее особенности конкретной задачи, связанные с количеством потоков в блоке; t - номер потока внутри блока, k – номер шага некоторой итерации алгоритма, одинаковый для всех потоков внутри блока. Приведенные в спецификации CUDA [8] требования согласованности несколько более сложные и к тому же различаются для различных реализаций стандарта CUDA. Тем не менее, легко проверить, что выполнение требования (3) для всех потоков внутри блока позволяют получить согласованный доступ к памяти в смысле спецификации [8].

Рекуррентный алгоритм, удовлетворяющий условию согласованности (3), может быть построен, например, как двухпроходный. При этом сначала выполняется вычисление частичных сумм по столбцам:

$$s^1_{x,y} = s^1_{x,y-1} - L_{x,y-m} + L_{x,y}, \quad s_{x,1} = \sum_{j=1}^{m} L_{x,y+j}, \tag{4}$$

а окончательная сумма получается суммированием по строкам:

$$s_{x,y} = s_{x-1,y} - s^1{}_{x-m,y} + s^1{}_{x,y}, \quad s_{1,y} = \sum_{j=1}^{m} s^1{}_{x+j,y}. \tag{5}$$

Вычисления по такой схеме требуют

$$O_1 = (2N - m)^2 \tag{6}$$

арифметических операций.

Ниже приводятся результаты сравнения эффективности предложенных реализаций. Эксперименты проводились на тестовом изображении размером 1024×1024 для окон различных размеров. В первом случае декомпозиция проводилась на 2048 блоков по 512 потоков, т.е. всего 10 048 576 потоков, часть которых запускалась не одновременно. Во втором и третьем случае запускалось одинаковое количество потоков – 1024, при этом они группировались в 32 блока по 32 потока в каждом. Тестирование проводилось для CPU PIV 3Ghz и GPU Nvidia GF 9500. Результаты представлены в таблице 1.

Таблица 1. Время выполнения различных алгоритмов для разных размеров окна (*m*).

	m = 5	m = 9	m = 15
1. CPU, не рекуррентный	283 мс	922 мс	2 531 мс
2. CPU, рекуррентный	142 мс	297 мс	733 мс
3. GPU, не рекуррентный	78 мс	262 мс	473 мс
4. GPU, рекуррентный	23 мс	25 мс	27 мс
5. GPU, рекуррентный с оптимизацией обращений к памяти	12 мс	12 мс	12 мс

Как видно из таблицы 1, результаты для рекуррентной CUDA-реализации практически не зависят от *m*, а зависят только от стратегии использования памяти.

Реализация программного комплекса и результаты экспериментов

Описанная выше схема взаимодействия потоков с памятью при реализации рекуррентного алгоритма использовалась при создании последних версий программного комплекса локализации и коррекции бликов на цифровых изображениях репродукций произведений живописи. Первые версии этого программного комплекса описаны в работах [2-5]. Программный комплекс имеет широкий набор функций для работы с цифровыми изображениями. Приложение работает в режиме *stand alone*, т. е. не требует установки дополнительного программного обеспечения. Благодаря использованию кроссплатформенных библиотек wxWidgets комплекс работает как на платформе Windows, так и на Unix-подобных операционных системах.

На Рисунке 1 представлен вид GUI, разработанного приложения в режиме автоматизированного поиска точечных артефактов (красные точки указывают найденные блики). В последней версии параметры поиска могут задаваться как вручную, так и автоматически путем перехода в режим адаптивного определения параметров. Кроме того, имеется возможность выполнять масштабирование

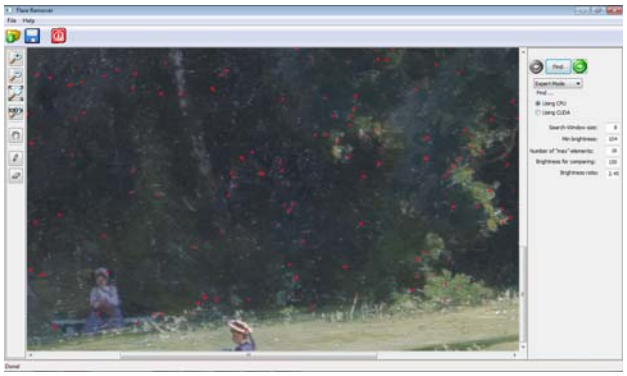


Рис. 1. - Вид GUI.

изображения, выбор области поиска, а также модифицировать маску с отмеченными артефактами, сформированную алгоритмом, вручную. На рисунке 2 приведены изображения, иллюстрирующие результаты устранения бликов с использованием разработанного программного комплекса.

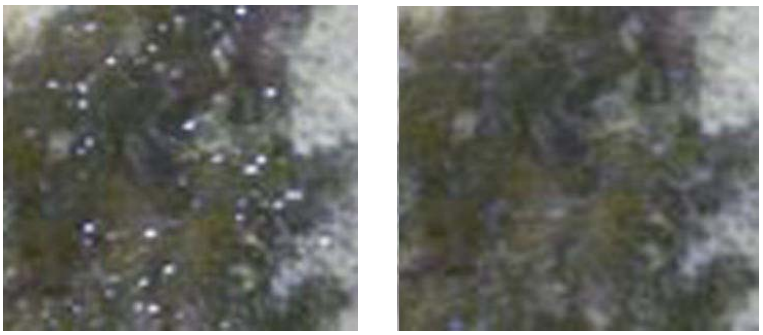


Рис.2. Пример работы программной системы по устранению бликов на изображении «лес»: а) исходное, б) обработанное.

В работах [4] и [5] авторами была показана возможность существенного ускорения алгоритмов обработки изображений в программной системе локализации и устранения бликов за счет использования универсальных графических процессоров NVIDIA и технологии CUDA. В частности, на графических ускорителях NVIDIA GeForce 8400 и 9500 с 8 мультипроцессорами, было продемонстрировано более чем десятикратное ускорение CUDA реализации по сравнению с реализацией на CPU. В настоящей работе мы демонстрируем возможность дальнейшего повышения ускорения за счет использования в программной системе оптимизированных рекуррентных алгоритмов.

Для экспериментальной проверки скорости использовались три изображения размером 3.2, 5.8 и 18.7 мегапикселей. Для расчетов использовалась видеокарта Nvidia GeForce 9500. На рисунке 3, а приведено время вычислений для алгоритма из работы [4], на рисунке 5, б приведено время вычислений для алгоритма поиска бликов на изображении, использующего оптимизированную рекуррентную процедуру обработки скользящим окном. Из приведенных диаграмм видно, что в среднем, по трем экспериментам, достигнуто ускорение более чем в 45 раз.

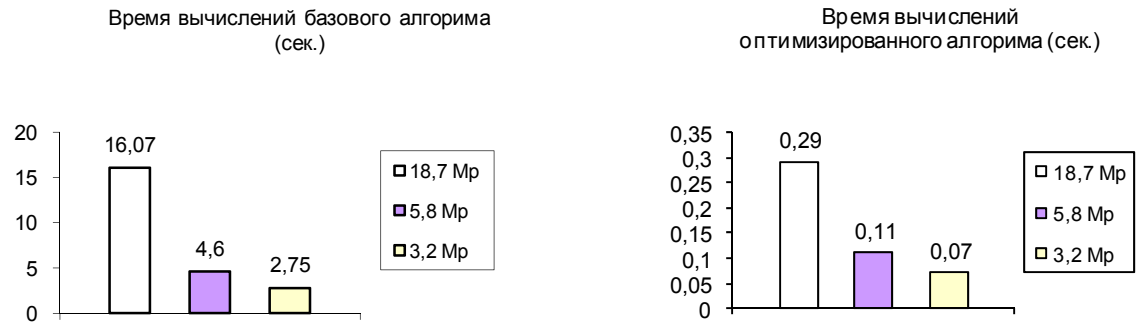


Рис.3. Время работы CUDA алгоритма: а) алгоритм, опубликованный в работе 4; б) результаты, оптимизированного рекуррентного алгоритма

Подчеркнем, что полученное в эксперименте ускорение получено исключительно за счет оптимизации CUDA алгоритма. Достигнутые показатели

скорости обработки с использованием оптимизированной версии алгоритма поиска бликов открывают возможности для их более широкого применения. В частности, наряду с использованием этого алгоритма в задачах предпечатной обработки изображений, как это было сделано в [4], по-видимому, удастся использовать эту технологию также для обработки видео данных в реальном времени.

Литература

1. Никоноров А.В., Фурсов В.А. Предоставление сервиса управления цветовоспроизведением цветных изображений в сети интернет. Труды XIV Всероссийской научно-методической конференции "Телематика 2007", С.-Петербург, 18-21 июня, 2007, с 412-413.
2. Никоноров А.В., Фурсов В.А. Распределенная вычислительная среда коррекции цветных изображений. Труды XV Всероссийской научно-методической конференции "Телематика 2008", С.-Петербург, 23-26 июня, 2008, с 88-89.
3. Фурсов В.А., Никоноров А.В. Распределенный алгоритм коррекции точечных артефактов на цветных изображениях // Труды четвертой междунеродной конференции «Параллельные вычисления и задачи управления» (PACO'2008), Москва, 27-29 октября, 2008, 1087.
4. Бибиков С.А., Никоноров А.В., Фурсов В.А., Якимов П.Ю. Исследование эффективности технологии CUDA в задаче распределенной предпечатной подготовки цифровых изображений. Труды Всероссийской суперкомпьютерной конференции «Научный сервис в сети Интернет: масштабируемость, параллельность, эффективность», 21-26 сентября 2009 г., г. Новороссийск. – М.: Изд-во МГУ, 2009. – С.204-207.
5. Бибиков С.А., Никоноров А.В., Фурсов В.А., Якимов П.Ю..CUDA-технология цветовой коррекции теневых искажений на цифровых фотокопиях произведений живописи. Сб. трудов международной научной конференции «Параллельные вычислительные технологии' 2010» Уфимский государственный авиационный технический университет (УГАТУ), г.Уфа, 29 марта – 2 апреля 2010 г., с. 656.
6. Сойфер, В.А. и др., Методы компьютерной обработки изображений. – М.: Физматлит, 2003. – 784 с.
7. Nickolls J., Buck I., Skadron K., Scalable Parallel Programming with CUDA, // ACM Queue, VOL 6, No. 2 (March/April 2008) pp. 40-53 http://mags.acm.org/queue/20080304/?u1=texterity.
8. NVIDIA CUDA Programming Guide, 2010, 130p. http://developer.download.nvidia.com/compute/cuda/3_0/toolkit/docs/NVIDIA_CUDA_ProgrammingGuide.pdf .
9. NVIDIA CUDA Best Practices Guide, 2009, 68p.
10. Harris M., Parallel Prefix Sum (Scan) with CUDA, 2008, 21p., http://developer.download.nvidia.com/compute/cuda/sdk/website/projects/scan/doc/scan.pdf.

Программный комплекс имитационного моделирования алгоритмов балансировки

И.Д. Зекцер
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В настоящее время для решения всё большего числа научных задач вычислительных ресурсов одной рабочей станции становится недостаточно. Поэтому разрабатываются системы, среды и модели программирования для организации вычислительной системы, развёрнутой на нескольких рабочих станциях, соединённых высокоскоростной сетью [1].

В организованных таким образом распределенных системах возникают проблемы эффективного использования ресурсов вычислительных узлов сети:

−может измениться вычислительная среда, где происходит выполнение приложения, какой либо вычислительный узел может выйти из строя;

−вычислительный узел, на котором выполняется распределенное приложение, занят другими вычислениями, и их доля со временем может возрасти.

В связи с этим необходимо использовать балансировку нагрузки, т.е. обеспечение равномерной нагрузки вычислительных узлов. Балансировка реализуется путем переноса части вычислений с наиболее нагруженных вычислительных узлов на менее нагруженные узлы. Цель балансировки загрузки может быть сформулирована следующим образом: исходя из набора задач, включающих вычисления и передачу данных, и сети компьютеров определенной топологии, необходимо найти такое распределение задач по компьютерам, которое обеспечивает примерно равную вычислительную нагрузку компьютеров и минимальные затраты на передачу данных между ними.

Различают статическую и динамическую балансировки.

Статическая балансировка выполняется до начала выполнения распределенного приложения. Однако предварительное размещение логических процессов по процессорам (компьютерам) не даёт эффекта в случае увеличения нагрузки на вычислительные узлы в процессе исполнения распределенного вычисления. Следовательно, необходима динамическая балансировка, предусматривающая перераспределение вычислительной нагрузки на узлы во время выполнения приложения.

Управление ресурсами и балансировка нагрузки являются ключевыми проблемами эффективного использования ресурсов. Существует множество стратегий и алгоритмов балансировки нагрузки. Встает вопрос о необходимости анализа алгоритмов балансировки для конкретной топологии и интенсивности нагрузки сети для выбора наиболее эффективного алгоритма при определенных условиях.

При исследовании эффективности алгоритмов балансировки встаёт ряд проблем:

−натурные эксперименты имеют высокую стоимость,

−подсистема измерения оказывает влияние на ход вычислительного процесса,

−системы общего назначения не подходят по ряду критериев.

Поэтому возникает потребность в создании специализированной системы для имитации процесса балансировки загрузки, которая, кроме того, позволит упростить процесс исследования путём автоматизации имитационных экспериментов с алгоритмами динамической балансировки нагрузки.

Задача подобной системы предоставить инструментальные средства

моделирования и анализа процесса балансировки нагрузки для различных сетевых топологий, которые могут быть использованы для анализа стратегий балансировки нагрузки в вычислительной сети.

Метод исследования

В качестве основной модели распределенных вычислений выбрана модель цепь из асинхронно взаимодействующих процессов, реализованная в программном комплексе GraphPlus. В данной модели процессы, входящие в сеть, взаимодействуют за счет обмена сообщениями. Основа модели – библиотека времени исполнения, являющаяся каркасом для написания распределенных приложений. В данном каркасе алгоритм управления вычислительным процессом отделен от прикладного кода задачи, т.е. используется инвертированный поток управления. С точки зрения прикладного программиста программа представляет собой взаимодействие процессов, выполняющих вычисления, описанные прикладным кодом и некоторым управляющим алгоритмом. Детали реализации управляющего алгоритма скрыты от программиста. В более детальном рассмотрении модели распределенных вычислений - процессы, выполняющие вычисления, являются резидентами некоторых объектов библиотеки времени исполнения – комнат, являющихся вычислительными узлами. Т.е. используется модель процессов, причем количество процессов много больше количества узлов. Передача сообщений между резидентами осуществляется через очередь сообщений комнат. Управление комнат осуществляет API операционной системы с помощью системных потоков управления [2].

В связи с тем, что программа определяется моделью GraphPlus, т.е. прикладной код и данные отделены от алгоритма управления вычислительным процессом, используется модель процессов, а также «справедливая в слабом смысле» стратегия планирования, можно внедрить код имитационной подсистемы в готовую систему времени исполнения GraphPlus [3].

Объекты библиотеки времени исполнения (комнаты, резиденты) могут быть задействованы в имитационном моделировании без внесения изменений в их реализацию. Стандартная система времени исполнения может быть заменена системой имитационного моделирования, состоящей из монитора моделирования, управляющего объектами RTL через модели системных потоков управления; подсистемы балансировки; подсистемы отображения результатов, хранения и настройки параметров экспериментов, важной частью которой является подсистема сбора статистики. Принцип, произведения замены стандартной системы времени исполнения на систему имитационного моделирования, изображён на рис. 1 и 2.

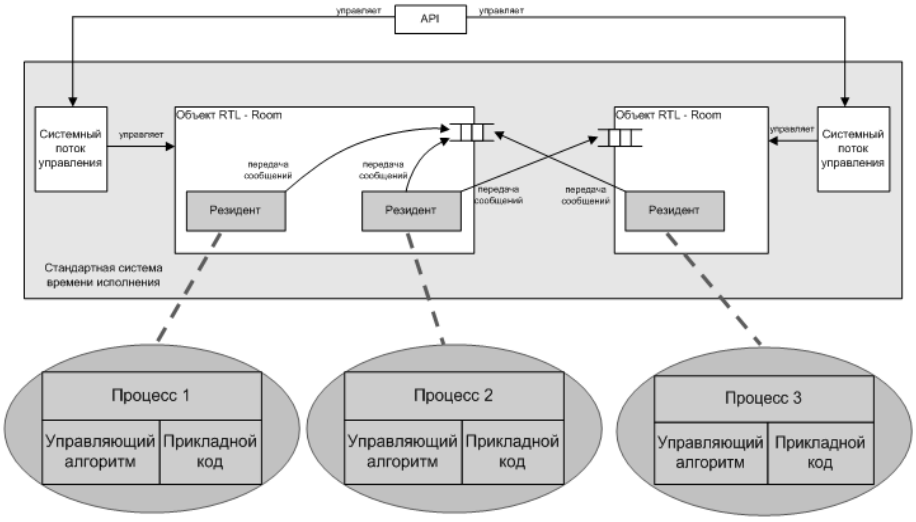


Рис. 1. Концептуальная модель системы. Реальное исполнение

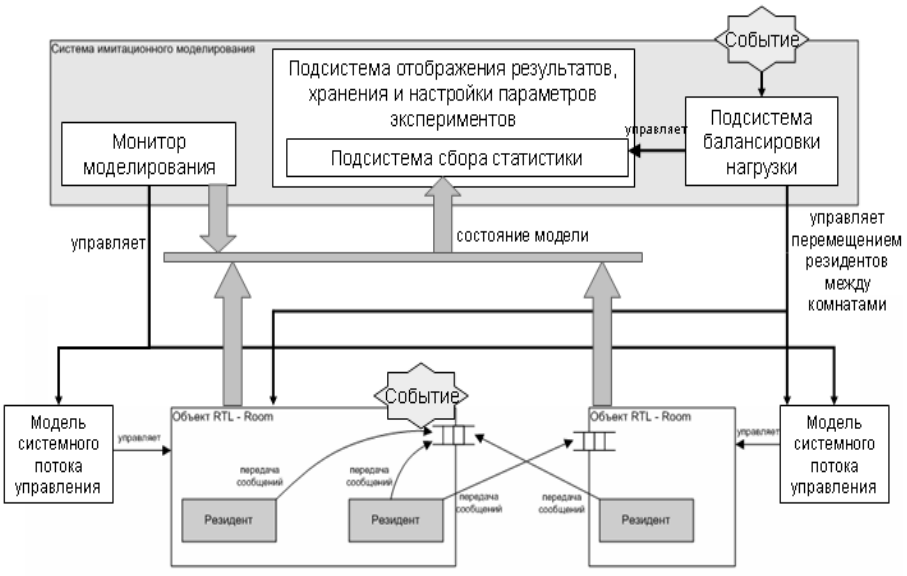


Рис. 2. Концептуальная модель системы. Имитационное исполнение

В ходе имитационного эксперимента при добавлении сообщения в очередь комнаты возникает событие разбалансировки, активизирующее подсистему балансировки, которая анализирует нагрузку на узлы – комнаты и в случае необходимости производит перемещение резидентов – задач между узлами. Нагрузка на узел характеризуется длиной очереди сообщений. При возникновении события происходит сбор статистики о нагрузке на узлы, которая в дальнейшем используется для построения графиков; вычисления эффективности алгоритма; вычисления характеристик узлов: средней нагрузки (длины очереди сообщений), среднего квадратичного отклонения нагрузки каждого узла.

Для сравнения различных алгоритмов балансировки был введён критерий эффективности. Эффективность алгоритма балансировки – это процент времени, в течение которого система выполняла полезные вычисления. Эффективной нагрузкой вычислительной сети будем считать процент задействованных в данный момент в вычислениях узлов. Так как нагрузка вычислительного узла измеряется длиной очереди сообщений, узел является задействованным в вычислениях (не простаивающим), если очередь его сообщений не пуста. Эффективная нагрузка вычислительной сети является нормализованной по количеству узлов результата сложения эффективностей нагрузки каждого узла.

Таким образом, эффективность алгоритма балансировки вычисляется по следующим формулам:

$f_{\text{общ}}(t) = \dfrac{\sum_{i=1}^{n} f_i(t)}{n}$ – эффективная загрузка вычислительной сети, где

$f_i(t)$ – мгновенное значение эффективности нагрузки i-го узла

n – количество узлов сети

$\varepsilon = \dfrac{1}{T}\int_0^T f_{\text{общ}}(t)dt$ – эффективность алгоритма балансировки нагрузки вычислительной сети

T – продолжительность выполнения вычислений в модельном времени.

Экспериментальное обоснование

Целью проводимых экспериментов была проверка работоспособности системы на примере двух алгоритмов балансировки при имитации метода с одномерной

декомпозицией данных. Для достижения данной цели было проведено три вида экспериментов.

Проведение серии имитационных экспериментов

Рассмотрим работу системы на примере исследования эффективности статического и динамического алгоритма балансировки – это доля задействованных в вычислениях узлов сети на время эксперимента. Соответственно, в случае одного узла эффективность алгоритма должна быть равна 1. В случае двух узлов при использовании любого алгоритма - не менее 0,5. Для проверки этих предположений была проведена серия имитационных экспериментов для случая разбиения задачи на три резидента и исполнения на сети, состоящей из одного, двух, трёх узлов. Распределять задачу, разбитую на три резидента, на четыре узла нецелесообразно, так как один узел будет простаивать, а эффективность алгоритма не может быть больше 0,75.

Для сравнения времени выполнения вычислений был проведён имитационный эксперимент для одного вычислительного узла без алгоритма балансировки.

Таблица 1. Результаты исследования алгоритма балансировки

Количество узлов	Тип балансировки	Время выполнения вычислений, с	Эффективность алгоритма
1	-	480,3	1
2	Без балансировки	480,3	0,5
2	Статическая	330,5	0,898966
2	Динамическая	330,5	0,899234
3	Статическая	300,2	0,599489
3	Динамическая	300,2	0,599311

Результаты исследования, приведённые в таблице 1, подтверждают сделанные предположения, а также показывают, что эффективность динамической балансировки не хуже статической, а иногда даже лучше.

Оценка эффективности алгоритма динамической балансировки

Оценка эффективности алгоритма динамической балансировки была проведена с помощью серии имитационных экспериментов для случая разбиения задачи на 10 резидентов. Результаты приведены в таблице 2.

Таблица 2. Результаты исследования алгоритмов балансировки для случая разбиения на 10 резидентов

Количество узлов	Тип балансировки	Время выполнения вычислений, с	Эффективность алгоритма
2	Без балансировки	1741	0,5
2	Статическая	1025,8	0,962328
2	Динамическая	873,9	0,997158
3	Статическая	664,6	0,782973
3	Динамическая	870,8	0,643395
4	Статическая	547,7	0,7189551
4	Динамическая	870,8	0,4825468

На рис. 3 представлена зависимость эффективности алгоритмов балансировки от типа балансировки для случая сети из двух узлов. Динамическая балансировка в данном случае более эффективна, однако, при увеличении узлов наблюдается снижение эффективности динамической балансировки. Это объясняется тем, что в первые моменты времени все резиденты находятся на одном узле и лишь при добавлении сообщений в их очередь происходит перенос вычислений на остальные узлы сети. Этот процесс занимает некоторое время, за счёт чего эффективность снижается.



Рис. 3. Эффективность алгоритмов балансировки

Следовательно, необходим комбинированный алгоритм, осуществляющий балансировку перед началом эксперимента и производящий миграцию резидентов в его ходе.

Моделирование комбинированного алгоритма балансировки

Для моделирования случая комбинированного алгоритма балансировки выберем динамическую балансировку и проведём распределение резидентов по узлам сети вручную. Результаты приведены в таблице 3.

На рисунке 4 представлены графики, показывающие зависимость эффективности алгоритмов балансировки от конфигурации вычислительной сети.

Таблица 3. Результаты исследования комбинированного алгоритма балансировки для случая разбиения на 10 резидентов

Количество узлов	Тип балансировки	Время выполнения вычислений, с	Эффективность алгоритма
2	Комбинированная	873	0,9974454
3	Комбинированная	688,1	0,8121242
4	Комбинированная	542,8	0,7190263

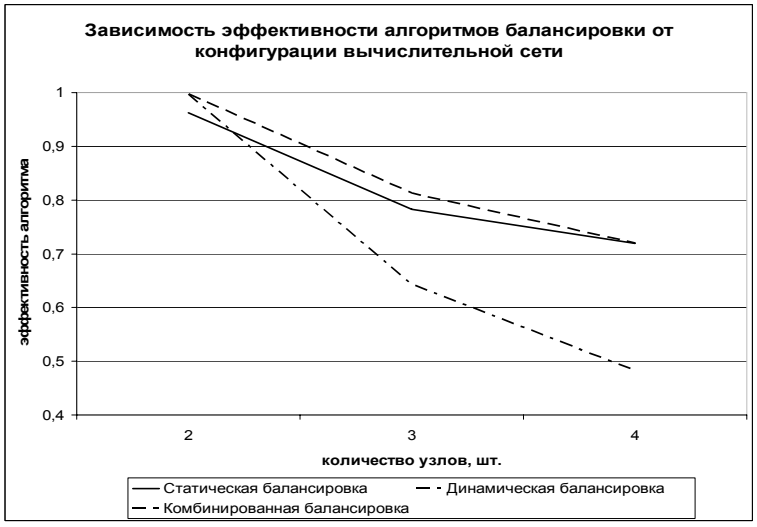


Рис. 4. Эффективность алгоритмов балансировки

Таким образом, эффективность алгоритмов балансировки зависит от разбиения исходной последовательной задачи на подзадачи и соотношения количества вычислительных узлов и подзадач. Алгоритм статической балансировки оказался эффективнее алгоритма динамической балансировки при выбранных условиях экспериментов, что согласуется с известными данными об их функционировании и свидетельствует о работоспособности системы моделирования.

Сравнение различных алгоритмов балансировки

Для сравнения различных видов балансировки сведём результаты экспериментов с использованием статической, динамической и комбинированной балансировки в одной таблице.

Из данной таблицы можно сделать вывод, что комбинированный вид балансировки является наиболее эффективным.

Заключение

Руководствуясь описанным выше методом имитационного моделирования, была разработана система автоматизации имитационных экспериментов с алгоритмами динамической балансировки нагрузки и произведено исследование работы алгоритмов балансировки на цепи из асинхронно работающих параллельных процессов.

Таблица 4. Сравнение результатов различных видов балансировки

Количество узлов	Тип балансировки	Время выполнения вычислений, с	Эффективность алгоритма
2	Без балансировки	1741	0,5
2	Статическая	1025,8	0,9623
2	Динамическая	873,9	0,9972
2	Комбинированная	873,7	0,9974
3	Статическая	664,6	0,7829
3	Динамическая	870,8	0,6434
3	Комбинированная	688,1	0,8121
4	Статическая	547,7	0,7180
4	Динамическая	870,8	0,4825
4	Комбинированная	542,8	0,7190

Проведённые эксперименты позволяют нам сделать следующие выводы:

1. Необходим комбинированный алгоритм, осуществляющий балансировку перед началом эксперимента и производящий миграцию резидентов в его ходе.
2. Эффективность алгоритмов балансировки зависит от разбиения исходной последовательной задачи на подзадачи и соотношения количества вычислительных узлов и подзадач.
3. Алгоритм статической балансировки оказался эффективнее алгоритма динамической балансировки при выбранных условиях экспериментов, что согласуется с известными данными об их функционировании и свидетельствует о работоспособности системы моделирования.

Данные выводы позволяют сделать заключение о работоспособности и пригодности системы имитационного моделирования алгоритмов балансировки, из чего следует, что цель данного исследования достигнута.

Литература

1. Востокин С.В. Применение метода парного взаимодействия объектов для построения сред разработки распределенных приложений // Весник СамГТУ, №38, 2005 – С.26-28.
2. Востокин, С.В. Графическая объектная модель параллельных процессов и ее применение в программных комплексах численного моделирования: дисс. доктора технических наук: 05.13.18: защищена 12.02.02: утв. 24.06.02 – Самара, 2007. – 307с.
3. Эндрюс, Г. Р. Основы многопоточного, параллельного и распределенного программирования. Пер. с англ. – М.: Издательский дом «Вильямс», 2003.

Континуальный и дискретный подходы при имитационном моделировании соударений высокоскоростных частиц с герметичными конструкциями космического аппарата

А.А. Потапкин, Л.С. Зеленко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Эффективных практических мер по уничтожению космического мусора на орбитах более 600 км (где не сказывается очищающий эффект от торможения об атмосферу) на настоящем уровне технического развития Человечества не существует. Вместе с тем актуальность задачи обеспечения безопасности космических полетов в условиях техногенного загрязнения околоземного космического пространства стремительно растет. Возникает потребность в разработке мероприятий, направленных на защиту проектируемых космических аппаратов (КА) от воздействия частиц космического мусора, и методов оценки опасности катастрофического разрушения конструкций КА.

Для решения перечисленных задач требуются знания о количественных характеристиках нестационарных явлений, возникающих при интенсивных нагрузках, которые инициированы процессами физического или химического взрыва и высокоскоростного удара. Проведение натурных экспериментов для исследования рассматриваемых явлений требует крупных экономических затрат и подчас не дает полной картины происходящих явлений. Поэтому методы имитационного моделирования в задачах физики взрыва и удара в настоящее время являются важным и весьма эффективным инструментом изучения процессов, протекающих в экстремальных условиях нагружения газов, жидкостей и твердых деформируемых сред.

При построении имитационной модели нужно анализировать возможность применения континуального и дискретного подходов. Континуальные методы механики сплошных сред в общем виде позволяют исследовать области, где деформации и напряжения относительно невелики. Чистый дискретный подход, основанный на симуляции «молекулярной» динамики следует применять в областях больших деформаций и нагрузок, поскольку он естественным образом позволяет моделировать такие сложнейшие эффекты как образование трещин, пластичность, разрушение и фазовые переходы [1].

Континуальный подход

При континуальном подходе первоочередной задачей является определение физической модели поведения системы. В механике сплошных сред математическое описание физической модели представляет собой замкнутую систему дифференциальных уравнений в частных производных, которая в самом общем виде имеет следующий вид [2]:

$$\frac{d\rho}{dt} + \rho \operatorname{div} \bar{u} = 0 \qquad (1)$$

– *уравнение непрерывности*; ρ- плотность среды, $\bar{u}$ -вектор скорости;

$$\frac{d\bar{u}}{dt} - \frac{1}{\rho} \operatorname{div} \sigma^{ij} - \frac{\bar{F}}{\rho} = 0 \qquad (2)$$

– *уравнение движения*; σ^{ij} - компоненты тензора напряжений, $\bar{F}$ – внешняя сила;

$$\frac{d\bar{u}}{dt} - \frac{1}{\rho}\,\sigma^{ij}\dot{\varepsilon}_{ij} - \mathrm{div}\bar{q} = 0 \qquad (3)$$

– *уравнение энергии*, (для адиабатного приближения $\mathrm{div}\bar{q} = 0$);

$$\dot{\varepsilon}_{ij} = \frac{1}{2}\left(\frac{\partial u^i}{\partial x_j} + \frac{\partial u^j}{\partial x_i}\right) \qquad (4)$$

– *компоненты тензора скоростей деформации*;

$$\frac{d\bar{x}}{dt} = \bar{u} \qquad (5)$$

– *кинематическое соотношение*;

$$\sigma^{ij} = f(\varepsilon, \dot{\varepsilon}, T) \qquad (6)$$

– *физическое соотношение*, определяющее сопротивление среды изменениям.

Таким образом, система содержит общие для всех сред *дифференциальные уравнения механики*, выражающие фундаментальные законы сохранения массы (1), импульса (2) и энергии (3), а также общие для всех сред *кинематические соотношения* (4) и (5). Индивидуальные особенности среды по отношению к сопротивлению деформациям учитываются соотношением (6). Для упрощения расчетов в модели не учитываются внешние силы, действующие на объекты. Считается, что процесс является адиабатным, так как является быстропротекающим в замкнутой системе.

Тензор напряжений определяется уравнением $\sigma_{ij} = -p\delta_{ij} + D_{ij}$, где δ_{ij} – символ Кронекера (метрика пространства, в декартовом пространстве – единичная матрица), p – давление, D – девиатор тензора напряжения.

Сила сопротивления среды формоизменениям определяется реологическим соотношением $\frac{dD_{ij}}{dt} = 2G(\dot{\varepsilon}_{ij} - \dot{\varepsilon}_{ij}\,\delta_{ij})$, где G – модуль сдвига (модуль упругости второго рода), δ_{ij} – символ Кронекера. Пластические свойства среды учитываются с помощью условия текучести Мизеса $\frac{1}{2}D^{ij}D_{ij} < \sigma_T^2$. Девиатор тензора напряжений после каждого шага интегрирования масштабируется до попадания под условие Мизеса (т.е. в круг текучести). Таким образом, сила сопротивления среды формоизменениям увеличивается лишь до σ_T – предела текучести материала на сдвиг.

Рассмотренное выше реологическое соотношение справедливо лишь для твердых тел. В идеальных жидкостях и газах сопротивление формоизменению отсутствует, поэтому девиатор тензора напряжения равен нулю. Для учета фазового перехода из твердого в жидкое состояние, необходимо учесть плавление материала, при этом переход должен быть плавным, а не скачкообразным. С этой целью действующие константы материала σ_T и G умножаются на коэффициент K.

$$K = \begin{cases} 1 - \dfrac{E}{E_{пл}}, & при\ E < E_{пл} \\ 0, & при\ E > E_{пл} \end{cases}.$$

Для описания изменения давления во всех фазах среды требуется построение широкодиапазонного уравнения состояния [3]. Так в области твердой фазы работает уравнение Ми-Грюнайзена (7):

$$p = p_S(\rho) + \gamma(\rho)\rho[E - E_S(\rho)]. \qquad (7)$$

Здесь E и p – энергия и давления на изэнтропе сжатия. Сильно сжатое и нагретое вещество, когда $E >> \Lambda_0$ (Λ_0 - энергия сублимации по модулю), описывается квантово-

статической моделью с уравнением состояния $p = \beta\rho E$. При $E \sim \Lambda_0$ и сильном разрушении работает модель идеального газа $p = 2/3\rho E$. Нахождение интерполяционной формулы, плавно переходящей в указанные асимптотические зависимости и достаточно точно отражающей суть дела в промежуточных состояниях, обеспечивает построение требуемого уравнения состояния. Предлагаемая интерполяционная формула имеет вид

$$p(\rho, E) = p_S(\rho) + \gamma(\rho, E)\rho[E - E_S(\rho)],$$

$$E_S(\rho) = \frac{9B_0}{2\rho_0 L_0}\left\{\exp\left[L(1-\sqrt[3]{x})\right]-1\right\}^2,$$

$$\gamma(\rho, E) = \frac{\frac{2}{3}x + \beta y}{x + y} + \frac{\gamma_0 - \frac{2}{3}x(1-\nu y x^2)}{(1+x^2 y)^2},$$

$$p_S(\rho) = \rho^2 \frac{dE_S}{d\rho},$$

$$L = \begin{cases} L1, & при \quad x \geq 1 \\ L2, & при \quad x < 1 \end{cases},$$

$$x = \frac{\rho_0}{\rho},$$

$$y = \frac{E}{\Lambda_0},$$

где $\beta_0 = \rho_0 c_0^2$ – модуль объемной адиабатной сжимаемости; γ_0 - термодинамический коэффициент Грюнайзена; L_1, L_2, β, ν – параметры формулы. Индекс "0" означает начальное состояние. Параметр L_1 определяется привязкой к энергии сублимации $\Lambda_0 = \frac{9B_0}{2\rho_0 L_1^2}$. Параметр L2 - по производной от модуля сжимаемости по давлению $L_2 = \frac{dB}{dp} - 1$. Величина предельного сжатия $\frac{\rho}{\rho_0} = 1 + \frac{2}{\beta}$.

Так же при моделировании разрушения учитываются два механизма повреждения твердых тел – отрывной и сдвиговой, работающие в зависимости выполнения соответствующих условий при нагружении. Разрушение по типу отрыва рассматривается как накопление повреждений (микропор и микротрещин), и при достижении некоторого порога материал считается разрушенным. Для этого в систему вводится кинетическое уравнение повреждения, описанное в [4].

Для численного интегрирования физической моделей применяется *метод гладких частиц*, в котором среда представляется в виде совокупности точек, являющимися узлами интерполяции. И производные по пространству находятся путем аналитического дифференцирования функции сглаживания и построения конечно-разностных аналогов дифференциальных уравнений.

Дискретный подход

Метод динамики частиц основан на представлении материала совокупностью взаимодействующих частиц (материальных точек или твердых тел), для которых записываются классические уравнения динамики. Взаимодействие частиц описывается посредством потенциалов взаимодействия, основным свойством которых является отталкивание при сближении и притяжение при удалении. Перед началом моделирования задается некоторое начальное распределение частиц в пространстве

(исходная структура материала) и начальное распределение скоростей частиц (механическое и тепловое движение системы в исходном состоянии). Далее задача сводится к решению задачи Коши для системы обыкновенных дифференциальных уравнений.

Потенциал взаимодействия в динамике частиц играет такую же роль, что и определяющие уравнения в механике сплошной среды. Однако структура потенциала неизмеримо проще, чем у определяющих уравнений, так как он представляет собой скалярную функцию расстояния, в то время как определяющие уравнения представляют собой операторы, в которые входят тензорные характеристики напряженного состояния и деформирования, а также термодинамические величины. Конкретный вид потенциала взаимодействия частиц определяется из сравнения механических свойств компьютерного и реального материалов. Для простейших характеристик, таких как, например, упругие модули, это сравнение может быть проведено аналитически [5, 6]. В остальных же случаях соответствие устанавливается на основе тестовых компьютерных экспериментов [7].

Техника моделирования методом динамики частиц описана в работах [8, 9]. Уравнения движения частиц в простейшем случае имеют вид

$$m\ddot{\mathbf{r}}_k = \sum_{n=1}^{N} \frac{f\left(\left|\mathbf{r}_k - \mathbf{r}_n\right|\right)}{\left|\mathbf{r}_k - \mathbf{r}_n\right|}\left(\mathbf{r}_k - \mathbf{r}_n\right), \tag{8}$$

где r_k - радиус-вектор k-ой частицы, m - масса частицы, N - общее число частиц, $f(r) = -P'(r)$ - сила взаимодействия между частицами, соответствующая потенциалу взаимодействия $P(r)$. Одним из наиболее простых и, вместе с тем, часто используемых, является потенциал Леннарда-Джонса

$$P(r) = D\left[\left(\frac{a}{r}\right)^{12} - 2\left(\frac{a}{r}\right)^{6}\right],$$

где a - равновесное расстояние для двух частиц, D - энергия связи.

Моделирование методом частиц с математической точки зрения представляет собой решение задачи Коши для уравнений (8). Начальные условия включают в себя координаты и скорости каждой частицы. Генерация начальных условий является самостоятельной и весьма нетривиальной задачей, так как начальное расположение частиц и их скорости существенно влияют на свойства полученного компьютерного материала. Задание начальных условий происходит на двух масштабных уровнях, которые условно можно назвать макро- и микроскопическим. На макроскопическом уровне задается внешняя форма объектов моделирования и их макроскопические скорости. На микроуровне задается вид упаковки частиц (структура материала) и скорости хаотического движения (тепловое движение). Скорость каждой частицы в начальный момент времени складывается из макроскопической скорости, мало изменяющейся от частицы к частице, и случайной компоненты, получаемой при помощи генератора случайных чисел.

В процессе моделирования, усредняя значения координат и скоростей частиц можно получить как континуальные уравнения, описывающие динамику с точки зрения механики сплошной среды, так и исследовать кинетику разрушения.

Заключение

В задаче исследования надежности герметичных конструкций космического аппарата моделирование всей конструкции при помощи метода динамики частиц не представляется возможным в силу крупных вычислительных затрат. Поэтому для мультимасштабного моделирования особую роль играет гибридный подход, основанный на совместном использовании континуальных и дискретных методов. В

этом случае метод гладких частиц используется в области, где деформации и напряжения относительно невелики, а в зоне, где происходит нарушение континуальности за счет сильного деформирования и разрушения, используется метод динамики частиц.

Литература

1. Hoover, W. G. Isomorphism linking smooth particles and embedded atoms [Текст]// Physica, 1998. -V. 260. - №3–4. - P. 44–254.
2. Бабкин, А.В. Прикладная механика сплошных сред [Текст]. Том 1 «Основы механики сплошных сред»/ А.В. Бабкин, В.И. Колпаков, В.Н. Охитин, В.В. Селиванов. – М.: издательство МГТУ имени Н.Э. Баумана, 1996.
3. Афанасьев, С.А. Моделирование задач экранной противометеоритной защиты [Текст]/ С.А. Афанасьев, В.Ф.Толкачев, В.Г. Трушков. – Томск: изд-во НИИ ПММ.
4. Бабкин, А.В. Прикладная механика сплошных сред [Текст]. Том 2 «Механика разрушения деформируемого тела»/ А.В. Бабкин, В.И. Колпаков, В.Н. Охитин, В.В. Селиванов. – М.: издательство МГТУ имени Н.Э. Баумана, 1998.
5. Krivtsov, A. I. From nonlinear oscillations to equation of state in simple discrete systems [Текст] // Chaos, Solitons & Fractals. 2002. - V. 17. - №1. - P. 79–87.
6. Krivtsov, A. I. Relation between spall strength and mesoparticle velocity dispersion [Текст]// International Journal of Impact Engineering. 1999. V. 23. - №1. - P. 466–476.
7. Кривцов, А.М. Метод частиц и его использование в механике деформируемого тела [Текст]// Дальневосточный мат. журнал ДВО РАН. – 2002. - Т.3, №2. - с 254-276.
8. Allen M. P. Computer simulation of liquids [Текст]/ M.P. Allen, A.K. Tildesley. – Oxford: Clarendon Press. 1987.
9. Hockney, R.W. Computer simulation using particles [Текст]/ R.W. Hockney, J.W.Eastwood– IOP Publishing. 1988.

Реализация FDTD-метода по технологии CUDA

С.А. Малышева
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Дифракционная компьютерная оптика развивается более 25 лет. За прошедшие годы решены фундаментальные задачи: создание технологической базы, алгоритмов расчета и методик проведения экспериментов [1]. Дифракционные оптические элементы нашли широкое применение в лазерных установках, оптических приборах, устройствах хранения информации. Последние годы наблюдается тенденция к миниатюризации ДОЭ и их интеграции с другими оптическими компонентами различных устройств. При этом методы расчета, основанные на геометрическом и скалярном приближениях, становятся неадекватными, что приводит к необходимости решения уравнений Максвелла.

Одним из наиболее распространенных методов решения системы уравнений Максвелла является разностный метод, предложенный в 1966 г. S.K. Yee [2] и развиты в работах A. Taflove [3].

FDTD является универсальным методом решения уравнений Максвелла и применим как для решения одномерных, так и двух- и трехмерных задач. Однако его существенным недостатком является высокая вычислительная сложность. Так, трехмерные реализации FDTD характеризуются большим числом арифметических операций и высокими требованиями к объему оперативной памяти, в силу чего расчет может занимать длительное время.

Существуют различные подходы к снижению вычислительной сложности в данном случае. Одним из них является наложение подвижной сеточной области (B. Fidel, E. Heyman, R. Kastner and R.W. Zioklowski) [4], применяемое при исследовании распространения одиночного импульса. Другим распространенным способом снижения длительности вычислений признается использование параллельных алгоритмов. Они впервые были предложены A.T. Perlik [5] применительно к явным разностным схемам Yee и развиты в современной монографии W. Yu, R. Mittra, T. Su, Y. Liu, H. Yang [6].

Альтернативный способ снижения длительности расчетов основан на векторизации алгоритма. До последнего времени векторные алгоритмы не получили широкого распространения из-за недоступности аппаратной базы широкому кругу исследователей. Развитие графических процессоров (Graphics Processing Unit, GPU) позволило использовать их в неграфических векторных вычислениях.

Исследователями D.K. Price, J.R. Humphrey, E.J. Kelmelis [7] предложены общие принципы реализации FDTD-метода на GPU с использованием текстур (каждому узлу сеточной области ставится в соответствие 1 пиксел). Ими проведены расчеты для двумерных и трехмерных задач, произведено сравнение с результатами, полученными на CPU. В работе [8] S. Adams разработана общая схема реализации алгоритма, дана оценка вычислительной сложности и требований к объему памяти, проведен подробный анализ условий, необходимых для получения ускорения на различных GPU и CPU. Однако, использование текстур существенно затрудняет написание и верификацию программного кода, в силу чего не получило широкого распространения.

Появление технологии CUDA (Compute Unified Device Architecture)[9] открыло новые возможности реализации векторных алгоритмов на GPU, не требуя более использования графического API (Application Programming Interface). В работах Volcarce с соавторами [10,11] рассматривается двумерная вычислительная область с

граничным условием CMPL (Convolutional Perfectly Matched Layer). Авторами предложено использовать отдельные вычислительные ядра для каждого из 4 CMPL-слоев и основной области вычислений и общая схема ядра, а также рассмотрен вопрос об объеме блока для достижения максимальной загрузки GPU.

Однако все вышеперечисленные работы объединяет отсутствие исходного кода либо его платность, фрагментарность описания реализации алгоритма на GPU. В силу этого невозможно восстановить условия экспериментов и провести их повторно, т.к. получаемые результаты существенно зависят от способа реализации.

Цель данной публикации - предоставить доступ к технологии векторизации FDTD метода широкому кругу пользователей.

1. FDTD-алгоритм

Ограничиваясь двумерным случаем распространения H-волны, запишем систему уравнений Максвелла [12]:

$$\begin{aligned} \mu_0 \frac{\partial H_y}{\partial t} &= -\frac{\partial E_x}{\partial z}; \\ \mu_0 \frac{\partial H_z}{\partial t} &= \frac{\partial E_x}{\partial y}; \\ \varepsilon_0 \varepsilon \frac{\partial E_x}{\partial t} &= \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}. \end{aligned} \qquad (1)$$

где t [c] - время, H_i [А/м] – проекция вектора напряженности магнитного поля на направление i, E_i [В/м] – проекция вектора напряженности электрического поля на направление i, ε - относительная диэлектрическая проницаемости, ε_0 [Ф/м] и μ_0 [Гн/м] - электрическая и магнитная постоянные [3].

Решая представленную систему, традиционно моделируют распространение TE-волны в цилиндрических оптических элементах (линзах, фокусаторах в отрезок, двумерных фотонных кристаллах и т. д.).

Переходя к рассмотрению решения системы (1), на двумерную область вычислительного эксперимента Ω ($0<t\le T$, $0\le y\le L_y$, $0\le z\le L_z$) обычно накладывают сеточную структуру Ω_h, в узлах которой $\{(t_m, y_j, z_k)$: $t_m=mh_t$, $m=0,..,M=T/h_t$, $y_j=jh_y$, $j=0,..,J=L_y/h_y$, $z_k=kh_z$, $k=0,..,K=L_z/h_z\}$ определена сеточная проекция электрического поля на ось X - $E^m_{x_{j,k}}$. Сеточная проекция магнитного поля на ось Z - $H^{m+0,5}_{z_{j+0,5,k}}$ определена в узлах $\{(t_{m+0,5}, y_{j+0,5}, z_k)$: $t_{m+0,5}=(m+0,5)h_t$, $m=0,..,M-1$, $y_{j+0,5}=(j+0,5)h_y$, $j=0,..,J-1$, $z_k=kh_z$, $k=1,..,K-1\}$ и проекция магнитного поля на Y - $H^{m+0,5}_{y_{j,k+0,5}}$ в узлах $\{(t_{m+0,5}, y_j, z_{k+0,5})$: $t_{m+0,5}=(m+0,5)h_t$, $m=0,..,M-1$, $y_j=jh_y$, $j=1,..,J-1$, $z_{k+0,5}=(k+0,5)h_z$, $k=0,..,K-1\}$. Индексы j, k обозначают узлы по пространству (направления Y и Z), m – по времени. Расстояния между узлами задаются пространственными (hy и hz) и временным (ht) шагами сетки. Сеточное значение диэлектрической проницаемости ($\varepsilon_{j,k}$) характеризует изучаемый оптический элемент.

Тогда, следуя работе S.K. Yee [2], запишем для системы (1) разностный аналог:

$$\begin{aligned} \mu_0 \frac{H^{m+0,5}_{y_{j,k+0,5}} - H^{m-0,5}_{y_{j,k+0,5}}}{h_t} &= -\frac{E^m_{x_{j,k+1}} - E^m_{x_{j,k}}}{h_z}; \\ \mu_0 \frac{H^{m+0,5}_{z_{j+0,5,k}} - H^{m-0,5}_{z_{j+0,5,k}}}{h_t} &= \frac{E^m_{x_{j+1,k}} - E^m_{x_{j,k}}}{h_y}; \\ \varepsilon_0 \varepsilon_{j,k} \frac{E^{m+1}_{x_{j,k}} - E^m_{x_{j,k}}}{h_t} &= \frac{H^{m+0,5}_{z_{j+0,5,k}} - H^{m+0,5}_{z_{j-0,5,k}}}{h_y} - \frac{H^{m+0,5}_{y_{j,k+0,5}} - H^{m+0,5}_{y_{j,k-0,5}}}{h_z}. \end{aligned} \qquad (2)$$

Задавая граничные условия, положим

$E^m_{x_{0,k}} = 0$ и $E^m_{x_{J,k}} = 0$ при $0 \le \mathrm{m} \le M$ *и* $0 \le \mathrm{k} \le K$;

$E^m_{x_{j,0}} = 0$ и $E^m_{x_{j,K}} = 0$ при $0 \le \mathrm{m} \le M$ *и* $0 \le \mathrm{j} \le J$,

что соответствует электрической стенке.

Начальное условие записывается как

$E^0_{x_{j,k}} = \Phi_2(jh_y, kh_z) \quad (1 \le j \le J-1, \quad 1 \le k \le K-1).$

2. Векторизация алгоритма

Достоинством решения по схеме (2) является возможность его векторизации. Вычислительные процедуры по классическим схемам Yee основываются на чередовании временных слоев: сеточные функции на следующем временном слое выражаются исключительно через функции на предыдущих двух слоях (как в (2)).

При реализации вычислений принято хранить по одному временному слою для каждой проекции. Тогда имеет смысл следующий векторный алгоритм [3], записанный в традиционной нотации Дж. Голуба [13]:

$$
\begin{bmatrix}
\textit{for} \quad m = 1:M \quad \%\textit{проход по всем временным слоям} \\
\%\textit{вычисление проекции } E_x \textit{ на данном временном слое} \\
E_x(2:J-1,2:K-1) = E_x(2:J-1,2:K-1) + \\
+\left(\frac{h_t}{h_y\varepsilon_0}\left(H_z(2:J-1,2:K-1) - H_z(1:J-2,2:K-1)\right) - \frac{h_t}{h_z\varepsilon_0}\left(H_y(2:J-1,2:K-1) - H_y(2:J-1,1:K-2)\right)\right)./\varepsilon(2:J-1,2:K-1); \\
\%\textit{вычисление проекции } H_y \textit{ на данном временном слое} \\
H_y(2:J-1,1:K-1) = H_y(2:J-1,1:K-1) - \frac{h_t}{h_z\mu_0}\left(E_x(2:J-1,2:K) - E_x(2:J-1,1:K-1)\right); \\
\%\textit{вычисление проекции } H_z \textit{ на данном временном слое} \\
H_z(1:J-1,2:K-1) = H_z(1:J-1,2:K-1) + \frac{h_t}{h_y\mu_0}\left(E_x(2:J,2:K-1) - E_x(1:J-1,2:K-1)\right);
\end{bmatrix}
$$

Каждая строка алгоритма может быть развернута в две циклические конструкции, соответствующие проходу по строкам и столбцам области Ω_h.

На любом временном слое сеточные функции выражаются через операции сложения векторов и saxpy.

3. Реализация по технологии CUDA

Вычислительные эксперименты по определению ускорения проводились на видеокарте NVIDIA GeForce GT 240 (Таблица 1) и процессоре Intel Core Duo E6500 (Таблица 2).

Таблица 1. Основные характеристики GPU NVIDIA GeForce GT 240

Характеристика	Значение
Количество мультипроцессоров, шт.	12
Размер видеопамяти, Мб	512
Максимальное число потоков в блоке, шт.	512
Максимальная размерность блока потоков (x, y, z), шт.	512×512×64
Максимальная размерность сетки блоков, шт.	65535×65535×1
Тактовая частота ядра, МГц	550
Тактовая частота памяти, МГц	1000

Таблица 2. Основные характеристики CPU Intel Core Duo E6500

Характеристика	Значение
Тактовая частота ядра, ГГц	2,93
Тактовая частота шины CPU, МГц	1066

Кеш L1, Кб	64×2
Кеш L2, Кб	2048
Пропускная способность шины процессор-чипсет, ГБ/с	8,528
Ширина шины L2 кеша, бит	512

Не задаваясь целью провести исследование какой-либо физической модели, автор сосредоточила свое внимание на изучении ускорения вычислений при распространении плоской однородной монохроматической волны в вакууме. Область вычислительного эксперимента характеризовалась линейными размерами L_y=L_z=1500 мкм и дискретизацией J=K от 100 до 4000 узлов сеточной области. Длина волны принималась равной $\lambda = 0.3 мкм$. Дискретизация по времени T=400 узлов.

Исследование велось в операционной системе Microsoft Windows XP, 32-bit (Service Pack 3) с установленным драйвером NVIDIA CUDA 2.3 driver. Скалярный и векторный алгоритмы написаны с использованием программного пакета Microsoft Visual Studio 2005.

Стремясь сравнить ускорение вычислений на GPU относительно CPU, автор реализовала алгоритм несколькими способами: скалярно на CPU, на GPU с использованием матричных и векторных функций библиотеки CUBLAS, а также без использования функций библиотеки с применением CUDA runtime API.

Ниже приведено ядро (kernel в терминах [9]) для GPU, выполняющее основные операции описанного выше векторного алгоритма без использования функций библиотеки:

```
// расчет матриц Ex, Hy, Hz
__global__ void matrix(float* Ex, float* Hy, float* Hz, int Je, int Ke, float Ca, float Cb, float Db, float Wt, int Js, int Ks)
{// номер элемента в блоке
  int idx = blockIdx.x*blockDim.x + threadIdx.x;
  int idy = blockIdx.y*blockDim.y + threadIdx.y;
  Jb = Je + 1;// размерность матрицы Ex
  if ((idx > 0) && (idx < Je) && (idy > 0) && (idy < Ke)) // расчет элементов матрицы Ex
    Ex[idy*Jb + idx] = Ca*Ex[idy*Jb + idx] + Cb*(Hz[idy*Je + idx] - Hz[idy*Je + idx - 1] + Hy[(idy - 1)*Jb + idx] - Hy[idy*Jb + idx]);
  __syncthreads();
  if ((idx == Js) && (idy == Ks)) //Задание источника
    Ex[idy*Jb + idx] = sin(Wt);
  __syncthreads();
  if ((idx > 0) && (idx < Je) && (idy < Ke)) // расчет элементов матрицы Hy
    Hy[idy*Jb + idx] = Hy[idy*Jb + idx] + Db*(Ex[idy*Jb + idx] - Ex[(idy + 1)*Jb + idx]);
  __syncthreads();
  if ((idx < Je) && (idy > 0) && (idy < Ke)) // расчет элементов матрицы Hz
    Hz[idy*Je + idx] = Hz[idy*Je + idx] + Db*(Ex[idy*Jb + idx + 1] - Ex[idy*Jb + idx]);
  __syncthreads();
}
```

Матрицы Ex, Hy, Hz задают электрическое и магнитное поля, Je, Ke – количество узлов по осям y, z. Размер блока равен 100×5 потоков (thread в терминах [9]), размер сетки (grid в терминах [9]) выбирался равным Je/100×Ke/5. На каждом мультипроцессоре может выполняться совокупность потоков до 24 (24 warp в терминах [9]), размер warp=32 потока, за счет чего и достигается ускорение.

Этот же алгоритм может быть реализован с использованием как векторных (cublasSscal, cublasSaxpy), так и матричных (cublasSgemm) функций библиотеки CUBLAS [12], что позволяет не вычислять номер thread и не выполнять их синхронизацию, что делает код более понятным пользователю, незнакомому с особенностями CUDA. При этом использовалась современная версия библиотеки CUBLAS 2.3.

На рисунке 1 приведена зависимость времени выполнения алгоритма от числа узлов сеточной области, что иллюстрирует эффект от применения векторных вычислений:

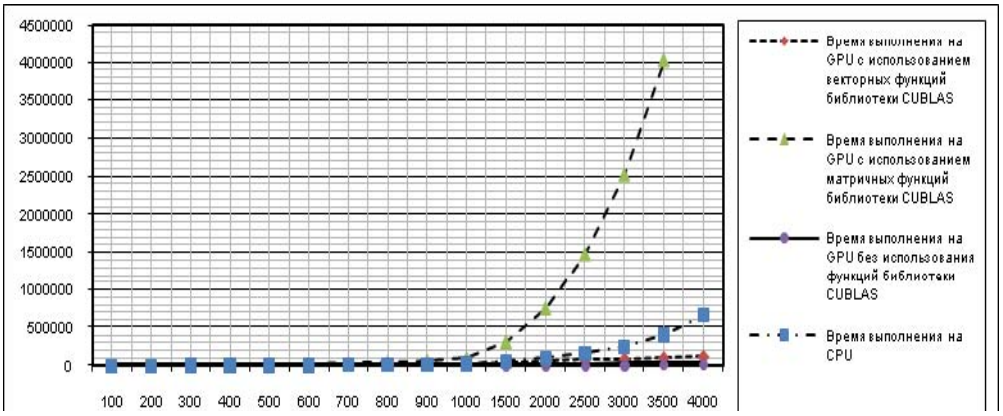


Рис. 1. – Время выполнения алгоритма в зависимости от числа узлов сеточной области

Как видно из рисунка 1, зависимость времени выполнения от размерности задачи для векторных алгоритмов на GPU линейная, а при использовании матричных функций библиотеки CUBLAS и для скалярного алгоритма на CPU – параболическая. Параболическая зависимость времени выполнения от размерности задачи для CPU объясняется тем, что с ростом размерности задачи число скалярных операций увеличивается пропорционально квадрату размерности сеточной области, а число векторных операций растет линейно. Для матриц небольшой размерности существенной разницы нет, что связано с возможностью центрального процессора быстро обрабатывать небольшое количество данных, которое может быть помещено в кеш (cache). С ростом размерности задачи большое количество времени уходит на пересылку данных из оперативной памяти в кеш.

При большой размерности матриц использование алгоритмов FDTD, реализованных на CUDA, дает существенное ускорение относительно CPU: для 1500 отсчетов в 2 раза при реализации через векторные функции библиотеки CUBLAS и в 25 раз без их использования, а для 4000 отсчетов – в 5,5 и 42,9 раза соответственно. Этот результат следует объяснить особенностями архитектуры GPU: на CPU большую часть ядра занимает кеш и сложные арифметико-логические устройства (АЛУ), на графическом ядре размещено большое количество упрощенных АЛУ, которые имеют общую память на кристалле.

Использование функций библиотеки CUBLAS заметно увеличивает время вычислений. Однако с ростом размерности задачи выигрыш во времени при реализации с применением CUDA runtime API относительно реализации с помощью векторных функций библиотеки CUBLAS снижается с 30 раз при J=1000 до 7,8 при J=4000, что свидетельствует о наличии оптимизации библиотеки для больших задач. Использование же матричных функций библиотеки CUBLAS не рационально, т. к. не только не дает выигрыша во времени относительно CPU, но и приводит к дополнительным затратам памяти (при использовании функции cublasSgemm для вычисления матриц E_x, H_y, H_z необходимо также хранение двух единичных матриц размера (K-2)×(K-2), (K-1)×(K-1)).

Заключение

В настоящей работе векторный FDTD-алгоритм реализован с использованием технологии CUDA как с помощью функций библиотеки CUBLAS, так и с применением CUDA runtime API. Анализ времени выполнения задач показывает, что наиболее эффективной является реализация без использования функций библиотеки (увеличение

производительности по сравнению со скалярным алгоритмом в 42,9 раза для размерности задачи 4000×4000). Применение векторных функций библиотеки CUBLAS целесообразно лишь для задач большой размерности. Таким образом, предпочтительнее запись векторных алгоритмов с использованием технологии CUDA.

Литература

1. Методы компьютерной оптики (Издание второе, исправленное) / Под ред. В.А. Сойфера – М: Физматлит, 2003. - 688 с.
2. Yee, K.S. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media / K.S. Yee // IEEE Trans. Antennas Propag., 1966.- AP-14.- p. 302-307.
3. Taflove, A. Computational Electrodynamics: The Finite-Difference Time-Domain Method: 2nd. ed. / A. Taflove, S. Hagness – Boston:Arthech House Publishers, 2000.- 852 p.
4. Fidel, B. Hybrid ray-FDTD moving window approach to pulse propagation / B. Fidel, E. Heyman, R. Kastner and R.W. Zioklowski // Journal of Computational Physics, 1997.- vol. 138, Issue 2.- p. 480-500.
5. Perlik, A.T. Predicting scattering of electromagnetic fields using FD-TD on a connection machine / A.T. Perlik, A. Taflove, T. Opsahl // IEEE Transactions on magnetics, 1989.-vol.25, № 4.- p. 2910-2912.
6. Yu, W. Parallel Finite-Difference Time-Domain Method / W. Yu, R. Mittra, T. Su, Y. Liu, H. Yang – Boston:Arthech House Publishers, 2006.- 272 p.
7. Price, D.K. GPU-based accelerated 2D and 3d FDTD solvers / D.K. Price, J.R. Humphrey and E.J. Kelmelis – Phisics and Simulation of Optoelectronic Devices XV, vol. 6468 of Proceedings of SPIE, San Jose, Calif, USA, January 2007.
8. Adams, S. Finite Difference Time Domain (FDTD Simulations Using Graphics Processors / S. Adams, J. Payne, R. Boppana // Proceedings of the 2007 DoD High Performance Computing Modernization Program Users Group Conference, 2007 – p. 334-338
9. NVIDIA: Nvidia CUDA Compute Unified Device Architecture. Programming Guide version 2.3. (2007)
10. Valcarce, A. A GPU approach to FDTD for Radio Coverage Prediction / A. Valcarce, G. De La Roche, J. Zhang // IEEE 11th International Conference on Communication Systems, November 2008.
11. Valcarce, A. Applying FDTD to the Coverage Prediction of WiMAX Femtocells / A. Valcarce, G. De La Roche, J. Zhang //EURASIP Journal on Wireless Communications and Networking, February 2009
12. Борн, М. Основы оптики/ М. Борн, Э. Вольф, пер. с англ.- М.:Наука, 1973.- 720 с.
13. Голуб, Дж. Матричные вычисления / Дж. Голуб, Ч. Ван Лоун – М.: Мир, 1999. – 548 с
14. NVIDIA: CUBLAS Library version 2.3. (2007)

Секция 3

ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И МАТЕМАТИЧЕСКАЯ ФИЗИКА

Расчет параметров ракетных двигателей с учетом вероятностных отклонений начальных данных

Д.М. Борисов, А.М. Камай
Московский физико-технический институт (государственный университет) факультет аэрофизики и космических исследований, кафедра тепловых процессов, НИЦ «Исследовательский центр имени М. В. Келдыша», Москва

Введение

Вопрос о влиянии вероятностных отклонений физических величин, а именно исходных данных на производные параметры (те физические величины, которые получаются уже по известным формулам) становиться актуальным, если без эксперимента можно рассчитать степень надежности летательного аппарата либо любой другой конструкции. С последующими уточнениями, производимыми после эксперимента, и хорошо с ним согласующимися.

Актуальность данного вопроса состоит в том, что учет заводской ошибки (отклонений хотя бы геометрических параметров) позволит внести коррективы в значения функций надежности.

В данный момент исследования и расчеты проводятся в критических условиях на предмет возможности или невозможности отказа. Мы же хотим получить зависимости во всем диапазоне реализуемых режимов роботы оборудования, это как и нормальный рабочий режим, так и реальная область критических значений. А именно, можно выделить два направления современных экспериментов по надежности ракетных двигателей:

1) многочисленные запуски оборудования для исследования на предмет достижения максимума при тех или иных условиях (дорого и долго).

2) эксперименты проводятся при значениях параметров, гораздо выше максимально ожидаемых. После чего заменяется оборудование, и граница ищется в меньшем диапазоне (дорого).

Реализация данного исследования позволит сэкономить огромные средства, затрачиваемые на экспериментальные образцы, оптимизировать работу экспериментаторов по поиску критических областей, и уменьшить количество экспериментов, носящих статистический характер.

В данный момент нами ведется работа анализа, применимости и обобщения, ранее полученных подходов к расчетам подобного плана, для дальнейшего построения математического аппарата (алгебры многомерных функций распределения случайных величин).

Целью работы является количественная оценка показателей надежности

агрегатов, а именно ракетных двигателей на этапе проектирования (расчет предельных отклонений производных величин); априорное определение надежности, как вероятности достижения критического режима; создание максимально универсального метода для исследования вероятности выхода из строя различных агрегатов; построение и исследование законов композиции случайных величин (функций распределения, пространства функций распределения).

В работе все физические величины рассматриваются как случайные (случайные многомерные вектора) и характеризуются номинальным законом распределения. Выбор распределения основан на следующих математических реализациях процессов, а именно:

1) внезапные отказы, носящие случайный характер, обычно довольно хорошо описываются экспоненциальным законом;

2) отказы, возникающие в результате износа, в результате необратимых физико-химических изменений физических параметров элемента, носящие название постепенных, во многих случаях довольно хорошо описываются нормальным законом.

Вероятность того, что случайная величина (случайный многомерный вектор) А меньше определенного значения В определяется

$$P = \int_{a\min}^{a\max} K_a \cdot f(x, m_a, \sigma_a) \left(\int_{\max(x, b\min)}^{b\max} K_b \cdot f(x, m_b, \sigma_b) dy \right) dx$$

где К – коэффициент масштабирования, $f(x, m, \sigma) = \exp\left(-\frac{(x-m)^2}{2 \cdot \sigma^2}\right) \Big/ \sigma \cdot \sqrt{2\pi}$

На первых этапах исследования данную работу можно представить как следующую модель предположений, с последующим усовершенствованием:



Наиболее простым примером иллюстрации данной работы является задание отклонений геометрических размеров. На данном этапе исследований рассматриваем эти характеристики как независимые. Для этих величин исследуются способы построения алгебры функций распределения.

После расчета выходные параметры сопоставляются с ранее произведенными экспериментами. Верификация полученных данных, а именно оценка достоверности надежности, что позволит уменьшить интервал вероятности отказа.

Литература

1. *Арнольд В.И.* Жесткие и мягкие математические модели. – М. МЦНМО, 2000

2. *Барлоу Р. Прошан Ф.* Математическая теория надежности. – М. «Советское радио», 1969

3. *Болотин В.В*. Прогнозирование ресурса машин и конструкций. – М. «Машиностроение», 1984

4. *Вентцель Е.С., Овчаров Л.А***.** Теория вероятностей и ее инженерные приложения. – М. «Наука», 1988

5. *Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д.* Математические методы в теории надежности. – М. «Наука», 1965

6. *Хеннан Э*. Представление теории групп и прикладная теория вероятностей. – М. «Мир», 1970

Численное моделирование противотока в разделительной газовой центрифуге

И.В. Ериклинцев, С.А. Козлов, О.В.Трошкин
Московский физико-технический институт (государственный университет), г. Долгопрудный, Московская область

Обзор работы

Основным свойством газового потока, возникающего во вращающемся со сверхзвуковой скоростью цилиндре, является возникновение в нем при определенных условиях вторичного движения циркуляционного типа (более медленного по сравнению с основным) или противотока. Причиной его появления может служить механическое возбуждение – различие в угловых скоростях вращения боковой поверхности и торцов цилиндра; при тепловом возбуждении циркуляция обеспечивается градиентами температуры на поверхности цилиндра и его торцах. Возможно внешнее возбуждение течения за счет втекающих в систему потоков, а также различные комбинации указанных способов [1].

Данный эффект активно применяется в устройстве для получения высокообогащенных стабильных изотопов различных химических элементов, используемых в науке, технике и медицине – газовой разделительной центрифуге. Одним из наиболее важных применений этого устройства является разделение изотопов урана в форме гексафторида урана UF_6. С учетом ничтожности содержания в разделяемой смеси урана–234 (0,0054%), она обычно рассматривается как двухкомпонентная смесь изотопов урана–235 (0,7114%) и урана–238 [2].

Современная конструкция противоточной газовой центрифуги была разработана и освоена на промышленном уровне впервые в нашей стране [2]. Центрифуга представляет собой вращающийся ротор (цилиндр). Что касается главного эффекта разделения изотопов – скорости получения необходимой концентрации обогащаемой легкой фракции, то, как показали выкладки Ю.Б.Харитона [3], только одно вращение смеси вокруг неподвижной оси не дает никаких преимуществ центрифугирования перед, например, газодиффузионным методом (известным по первым опытам разделения изотопов неона с 1910 г.). Эффект разделения бинарной смеси по радиусу за счет вращения может быть значительно умножен путем организации осевого противотока. Последний переносит газ, обогащенный тяжелым компонентом, по линиям тока от боковой стенки ротора к одному из концов, а газ, находящийся в центральной зоне ротора и обогащенный легким компонентом, – к другому его концу.

Важной и актуальной задачей является создание газовых центрифуг, эффективно работающих в различных диапазонах масс рабочих газов. В общем случае эта задача, являющаяся оптимизационной и многопараметрической, требует детального анализа процессов переноса внутри центрифуги. Сложность состоит в том, что для современных газовых центрифуг при линейных скоростях вращения ротора около 600 м/с проведение прямых экспериментальных исследований ее газодинамических характеристик невозможно. Попытки экспериментального измерения поля скоростей в роторе газовой центрифуге до настоящего времени носят качественный характер [4]. Практически единственным методом исследования рабочих характеристик газовой центрифуги, позволяющим использовать полученные результаты для оптимизации поля течения и процесса диффузии, является теоретический подход, основанный на численном моделировании.

В данной работе решение поставленной задачи совершается в два этапа. На первом этапе в рамках линейной модели, описывающей однокомпонентный газ,

производится исследование противоточного вихря промышленной центрифуги, возникающего в основной области разделения компонентов бинарной смеси - газодинамическом ядре. Рассчитываются поля скорости, температуры, плотности, давления, а также зависимости положения условного центра циркуляции (точки инверсии) от некоторых параметров. Рассматриваются различные способы возбуждения противотока: торможением верхнего торца вращающегося газового цилиндра, а также с помощью градиента температуры на боковой поверхности. Данный этап, является необходимым, для исследования самого процесса разделения изотопов, так как именно на нем рассчитываются газодинамические характеристики течения, без которых невозможны расчеты пространственного распределения компонент разделяемой изотопной смеси.

На втором этапе в рамках модели “пассивной примеси” для бинарной смеси, описывающей двухкомпонентный газ, проводится исследование распределений концентраций разделяемых фракций в найденном на первом этапе установившемся режиме течения сверхзвукового вращающегося газового потока в разделительной центрифуге.

В данной работе построена линейная модель механики и термодинамики двухкомпонентного газа в разделительной центрифуге. В основе рассмотренной модели лежит подход, предложенный в работе [5]: линеаризация уравнений газовой динамики с учетом твердотельного приближения вращающегося газового цилиндра. Сделанные упрощения при расчете структуры установившихся полей газодинамических характеристик не стали препятствием для получения результатов, согласующихся с уже известными. Недостаток линейного приближения заключается в изотермическом приближении твердотельного вращения, но при расчете задач в диапазоне небольших температурных градиентов, модель хорошо описывает физический процесс.

Полученные результаты позволяют сделать некоторые выводы относительно созданной модели вращающегося со сверхзвуковой скоростью газового цилиндра.

При постановке численных экспериментов, для граничных условий не сильно отличающихся от режима “твердотельного” вращения газа, линейная модель дает весьма хорошее приближение; произведенные расчеты газодинамических характеристик разделяемой смеси находятся в согласии с результатами как физических, так и численных экспериментов, использующих модели, учитывающие нелинейные члены.

Полученные оценки для времени выхода течения на стационар и времени установления поля концентраций, подтверждают приемлемость используемого подхода разделения физического процесса на два: движения газа в изотопном приближении (пренебрежение разностью молекулярных масс двух компонент), и установление поля концентраций при постоянном, установившемся поле других газодинамических параметров.

Показана независимость установившегося противотока при различных значениях выбора левой границы расчетной области (вакуумной стенки). Диапазон приемлемых значений при этом может достигать $0.94R$, что сильно сужает расчетную область и дает возможность, разрешать основную область разделения компонентов бинарной смеси - газодинамическое ядро с большой точностью, без необходимости при этом увеличения количества узлов расчетной сетки.

Произведенные расчеты так же подтверждают, возникновение дополнительных слоев (более удаленных от вращающейся поверхности) противотоков – “обертоков”, которые по сравнению с основным противотоком являются слабыми и практически не участвуют во вращении основной массе смеси, но влияют на структуру главного

противотока. Проведенное сравнение двух способов возбуждения противотока, подтверждает преимущество термического способа перед механическим.

Литература

1. *Борисевич В.Д., Левин Е.В., Наумочкин В.В.* Численное исследование вторичных течений вязкого газа во вращающемся цилиндре при наличии источников и стоков // Изв. АН СССР. Механика жидкости и газа. 1989. № 4, С. 39–44.

2. Обогащение урана. Сб. статей под ред. С. Виллани. Пер. с англ. под ред. И.К. Кикоина. – М.: Энергоатомиздат, 1983.

3. Разработка и создание газоцентрифужного метода разделения изотопов в СССР (России). Сб. статей. – С.–Петербург: ЛНПП «Облик», 2002.

4. *Айсен Э.М., Борисевич В.Д., Левин Е.В.* Моделирование течения и диффузии в газовой центрифуге для разделения многокомпонентных изотопных смесей // Математическое моделирование. 1997. Т. 9. № 4. С. 27–38.

5. Cloutman L. D. A Novel Approach to the Computation of Axisymmetric Countercurrent Patterns in Gas Centrifues. Los Alamos Scientific Laboratory report LA-8258-MS. 1980.

Ударные волны и автоволновые структуры в газовой среде с релаксационным процессом и источником энергии

Р.Н. Галимов, В.Г. Макарян, Н.Е. Молевич
Самарский филиал Учреждения Российской академии наук Физического института им. П.Н. Лебедева РАН, Самара
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В последнее время огромное количество экспериментальных работ свидетельствуют об аномальном поведении ударных волн в неравновесных средах. Например, в слабоионизированном газе наблюдается усиление ударной волны, распад фронта ударной волны и появление предвестника [1-3]. Усиление и модификация слабых ударных волн может быть вызвана новыми дисперсионно-вязкостными свойствами неравновесных сред. В [2-4] подробно рассмотрены принципиальные различия между акустикой равновесных сред и акустикой таких неравновесных сред как колебательно-возбужденный газ, неизотермическая плазма, химически активная смесь, среды с неравновесными фазами и др. В таких средах коэффициент второй (объемной) вязкости ξ и дисперсия звука могут быть отрицательными: $\xi < 0$ и $c_0 > c_\infty$. Здесь, c_0, c_∞ равновесная (низкочастотная) и замороженная (высокочастотная) скорости звука соответственно. Среды, имеющие отрицательную вязкость могут быть акустически активными. Кроме того, низкочастотный коэффициент газодинамической нелинейности Ψ_0 является сложной функцией от степени неравновесности. Только замороженный коэффициент газодинамической нелинейности имеет обычную форму $\Psi_\infty = (\gamma_\infty + 1)/2$. Существуют области степени неравновесности, в которых $\Psi_0 < 0$. В настоящей работе исследуется влияние новых акустических свойств стационарно неравновесных сред с релаксационным процессом и источником энергии на структуру ударных волн в таких средах.

Структура ударных волн в неравновесной среде. Бифуркационная диаграмма.

Исходная система газодинамических уравнений для исследуемой модели среды имеет следующий вид:

$$\frac{d\rho}{dt} + \rho\frac{\partial v}{\partial x} = 0,$$
$$\rho\frac{dv}{dt} = -\frac{\partial P}{\partial x},$$
$$C_{V\infty}\frac{dT}{dt} + \frac{dE_v}{dt} - \frac{T}{\rho}\frac{d\rho}{dt} = Q - I,$$
$$\frac{dE_v}{dt} = \frac{E_e - E_v}{\tau_v(T,\rho)} + Q \qquad (1)$$

В уравнении (1), E_v энергия колебательных степеней свободы молекул, E_e её равновесное значение, τ_v время колебательной релаксации и Q мощность внешнего источника колебательной энергии (в частности, электрическая накачка в разряде, химическая или оптическая накачка), который поддерживает степень неравновесности

$S=(E_{v0}-E_{e0})/T_0=Q\tau_v/T_0$, v, T, ρ, P соответственно скорость, температуры, плотность и давление. $I=Q$ теплоотвод и $d/dt=\partial/\partial t+v\partial/\partial x$.

Газ со стационарной неравновесностью и зависимостью времени релаксации от температуры и плотности в форме Ландау-Теллера имеет пять областей степени неравновесности S с качественно различными свойствами:

Область 1: $S<S_{thr}=C_v/(C_{V\infty}-\tau_T)$, где $C_v=dE_{e0}/dT_0$, $\tau_T=\partial\ln\tau_{v0}/\partial\ln T_0$. В этой области вязкость положительна $\xi>0$, дисперсия положительна $c_0<c_\infty$ и коэффициент нелинейности положительный $\Psi_0\approx(\gamma_0+1)/2$ подобно равновесным средам.

Область 2: $S_{thr}<S<S_n$. Дисперсия и вторая вязкость отрицательные ($\xi<0, c_0>c_\infty$). Низкочастотный коэффициент нелинейности $\Psi_0>0$. Здесь S_n определяется из уравнения $\Psi_0(S_n)=0$, где

$$\Psi_0=[\frac{S_0\tau_T(1+S_0)}{C_{P0}C_{V0}}+\frac{1+2C_{V0}}{2C_{V0}}-\frac{S_0(1+S_0)^2}{2C_{P0}C_{V0}^{\ 2}}\tau_{TT}],\ \tau_{TT}=\frac{T_0^{\ 2}}{\tau_{v0}}\frac{\partial^2\tau_{v0}}{\partial T^2_{\ 0}},$$

$C_{V0}=C_{V\infty}+C_v+S_0\tau_T$, $C_{P0}=C_{P\infty}+C_v+S_0(\tau_T+1)$ низкочастотные теплоёмкости при постоянном объеме и давлении в колебательно-возбужденном газе.

Область 3: $S_n<S<S_V=-(C_{V\infty}+C_v)/\tau_T$. Здесь $\xi<0$, $c_0>c_\infty$, $\widetilde{\Psi}_0=\gamma_0\Psi_0<0$.

Область 4: $S_V<S<S_P$, $S_P=-(C_{P\infty}+C_v)/(\tau_T+1)$. Здесь $\xi<0$, $\gamma_0<0$, $C_{V0}<0$, $C_{P0}>0$.

Область 5: $S>S_P$. Здесь $\xi<0$, $c_0<c_\infty$, $\widetilde{\Psi}_0>0$, $C_{V0}<0$, $C_{P0}<0$.

Структура ударной волны за скачком фронта $\rho_d=(\gamma_\infty+1)D^2\rho_0/[(\gamma_\infty-1)D^2+2c_\infty^{\ 2}]$ была получена с помощью численного решения уравнения [5]:

$$\frac{d\rho}{dz}=-\frac{\rho\{[E_e(\rho)-E_v(\rho)]/\tau_v(\rho)+Q\}}{\rho_0 D(dE_e/d\rho)}\equiv\frac{A(\rho)}{B(\rho)},\qquad(2)$$

$$z=x-Dt$$

$$E_v(\rho)=E_{v0}+M[C_{P\infty}\frac{P_0}{\rho_0}+\frac{D^2}{2}-\frac{C_{P\infty}}{\rho}(P_0+\rho_0D^2(1-\frac{\rho_0}{\rho}))-\frac{1}{2}(\frac{\rho_0D}{\rho})^2]$$

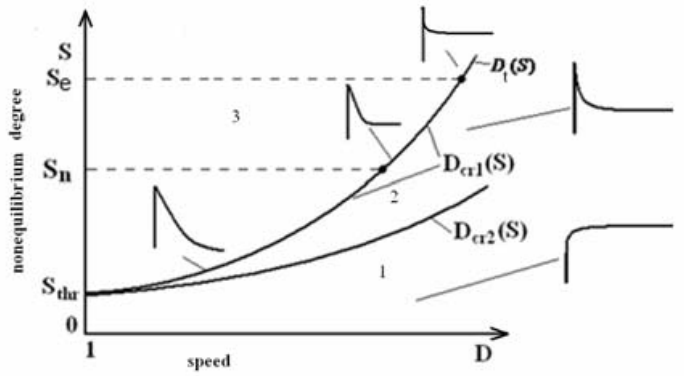


Рисунок 1. Бифуркационная диаграмма.

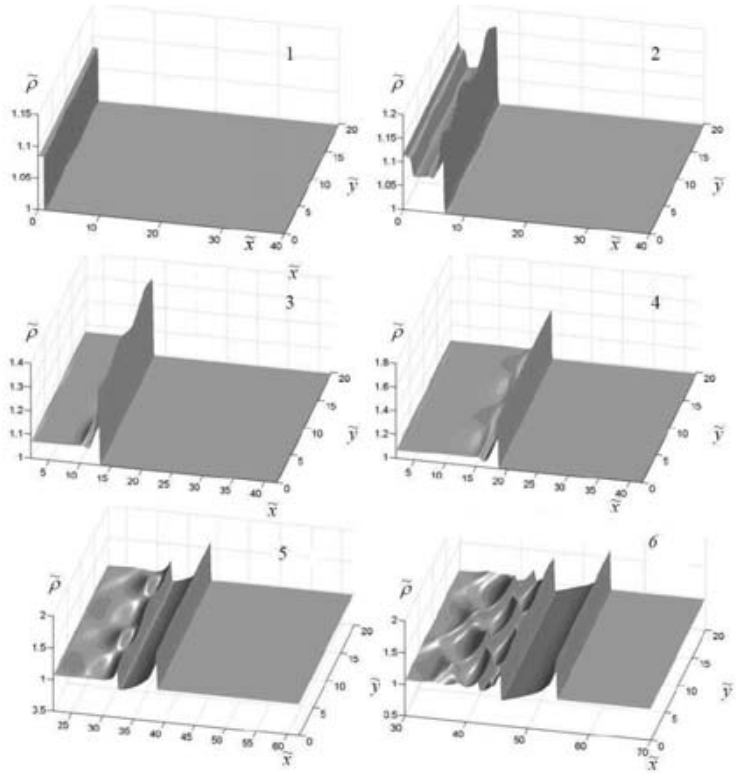


Рисунок 2. Устойчивость плоского фронта автоволновой структуры для двухмерной геометрии. Начальное поперечное синусоидальное возмущение на меняет самоподдерживающуюся структуру.

Интегральные кривые и возможные стационарные волновые решения уравнения (2) были получены в [6]. Все результаты могут быть представлены на бифуркационной диаграмме (рисунок 1). Границы бифуркационной диаграммы $D_{cr1}(S), D_{cr2}(S), D_t(S)$ получены в аналитическом виде и имеют сложные громоздкие вырежиня.

Для малых степеней неравновесности получаем

$$\frac{D_{cr1}}{c_\infty} = 1 - \frac{\mu\Psi_\infty}{(2\Psi_\infty - \gamma_0\Psi_0/\gamma_\infty)},$$

$$\frac{D_{cr2}}{c_\infty} = 1 - \frac{\mu\Psi_\infty}{(2\Psi_\infty - \gamma_0\Psi_0/\gamma_\infty)},$$

$$\mu = \frac{\gamma_0 - \gamma_\infty}{\gamma_\infty} << 1.$$

В области 1 бифуркационной диаграммы для скорости ударной волны $D > D_{cr2}$, быстрое сжатие продолжается медленным сжатием к конечному состоянию. Для $D_{cr1} < D < D_{cr2}$ (область 2) за быстрым сжатием следуем медленное разряжение к конечному состоянию. При $D = D_{cr2}(S)$ ударная волна имеет профиль ступени. В третьей области решений в виде стационарной ударной волны не существует.

Как известно, ударная волна становится неустойчивой если скорость распространения звука за фронтом волны меньше чем скорость самой ударной волны. Для исследуемой модели среды это условие совпадает с условием $D < D_{cr1}$ при $S_{thr} < S < S_e$ и $D < D_t$ при $S > S_e$ (область 3 рисунок 1). Скорость D_t является

аналогом скорости Жуге в теории детонации. Для $D = D_{cr1}(S)$ или $D = D_t$ стационарная волна имеет форму импульса с амплитудой ρ_d или волны с ненулевой асимптотой. Для малой степени неравновесности, этот автоволновый импульс имеет ударноволновой фронт и экспоненциальный «хвост» $\rho(z) = \rho_d \exp\{(x - x_0)C_{V0}\Psi_0/(2\Psi_\infty C_{V\infty} c_\infty \tau_{v0})\}$ [5].

Нестационарная эволюция волн при $D < D_{cr1}, D_t$

Для исследования эволюции ударных волн в стационарно неравновесной среде с релаксационным процессом и источником энергии в настоящей работе численно промоделирована эволюция возмущения в виде ступеньки. Ударные волны в областях 1 и 2 бифуркационной диаграммы эволюционно устойчивы. Для области 3 неустойчивых ударных волн $S - D$ диаграммы, были получены как эволюции две автоволновые структуры (импульс или волна с ненулевой асимптотой). Их форма, амплитуда и скорость не зависят от амплитуды начального возмущения.

Для $S_{thr} < S < S_n$ нестабильные волны при $D < D_{cr1}$ распадаются на последовательность автоволновых импульсов одининаковой амплитуды, распространяющихся со скоростью $D = D_{cr1}(S)$. Для $S_n < S < S_e$ нестабильные волны распадаются на последовательность автоволн с ненулевой асимптотой и подобной скоростью $D = D_{cr1}(S)$ (рисунок 2). Формирование последовательности автоволн при $S_{thr} < S < S_e$ связано с акустической активностью, сохраняющихся за фронтом волны. Напротив, при $S > S_e$ нестабильные волны трансформируются в одну автоволну, распространяющуюся со скоростью $D = D_t(S)$.

Полученные автоволновые структуры устойчивы по отношению к поперечным возмущениям (рисунок 2).

Благодарности

Работа была частично поддержана исследовательской программой ГР 01200805605, аналитической целевой программой <Развитие научного потенциала высшей школы> (2009 - 2010 гг.) , проект 2.1.1/309, Федеральной целевой программы <Научные и научно-педагогические кадры инновационной России> на 2009 - 2013 годы (проект НК- 410П/10 (контракт П2315), НК-422П/22 (контракт П2450)) министерства образования и науки РФ и фондом «Династия».

Литература

1. Климов А.И., Битюрин В., Серов Ю., 2001, Неизотермическое приближение в плазменной эародинамике, Paper AIAA, 0348

2. Молевич Н.Е., 2004, Акустические свойства неравновесных сред, Paper AIAA,020.

3. Молевич Н.Е., Климов А.И. и Макарян В.Г., 2005, Влияние термодинамической неравновесности на акустические свойства газов, Международный журнал аэроакустики, Вып. 4, №3&4, 345-355.

4. Макарян В.Г. и Молевич Н.Е., 2007, Стационарные ударные волны в неравновесной сред, *Plasma Sources Sci. Thechnol.,* Том. **16**, 124-131.

5. Макарян В.Г. и Молевич Н.Е., 2004, Структура газодинамических возмущений в термодинамически неравновесной среде со степенной зависимостью времени релаксации, Механика жидкости и газ, том 39, №5, 836-845.

6. Галимов Р.Н. и Молевич Н.Е., 2009, Структура и буфуркации плоских ударных волн в колебательно-возбужденном газе с внешним источником энергии, Механика жидкости и газа, том 44, №1, 158-169.

Численное моделирование структуры вихревого потока в присутствии разряда

И.П. Завершинский, В.Г. Макарян, Н.Е. Молевич, Д.П. Порфирьев
[1] *Самарский филиал Физического института им. П.Н. Лебедева РАН, Самара*
[3] *Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара*

В последнее время в связи с развитием технологий создания быстропроточных электроразрядных лазеров [1, 2], формирования характеристик потока плазмы в плазмотронах [3, 4] и т.д. актуальной становится задача изучения особенностей разрядов, существующих стационарно в потоке газа, а также сопутствующих проблем плазмохимической кинетики и процессов стимулированной релаксации в разряде [5, 6]. По целому ряду свойств (механизм протекания тока, особенности развития неустойчивостей и др.) разряд в вихревом потоке существенно отличается от обычного разряда в трубках. Актуальной проблемой для работы данных устройств является поиск конфигураций газового потока, обеспечивающих устойчивое существование разряда при высоких давлениях и энерговкладах, а также способов управления характеристиками потока. В этой связи представляет интерес использование для поставленных целей сильнозакрученных газовых потоков.

В работах [5, 6] исследован поперечный газовый разряд постоянного тока в высокоскоростном закрученном потоке в воздухе и азоте. В результате этого эксперимента были зафиксированы спиральная структура разрядной области вблизи электродов и вытянутая вдоль потока светящаяся коническая область (рис. 1а). Причем, эта область росла по потоку только при относительно малых расходах газа. С ростом расхода газа возникал протяженный горячий конус ($T\sim3000\ K$) длиной $\sim 0.1 \div 1\ м$, растущий уже против потока (рис 1 б).

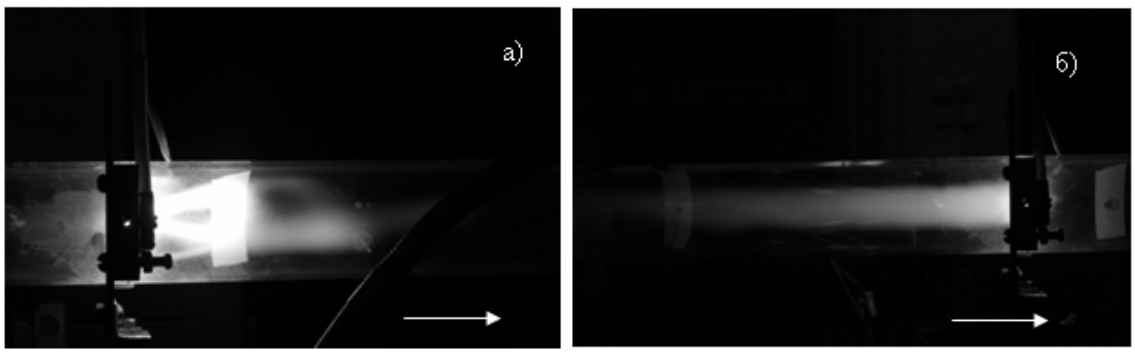


Рис. 1. Протяженное светящееся образование – экспериментальные данные [6]. Давление P =1 атм., энерговклад – 1.2 кВт, а) массовый расход 2 г/с, б) массовый расход 9 г/с. Стрелкой обозначено направление потока.

Поток воздуха или азота атмосферного давления создавался в вихревых камерах диаметром *2-6 см* с торцевым вдувом. Геометрия камеры позволяла создавать дозвуковой поток с осевым и тангенциальным числом Маха $M_z,\ M_\varphi \le 0.3$. Угол закрутки потока составлял 45^0 . Параметры разряда постоянного тока соответствовали напряжению $U \le 10$ кВ, току $I \le 2$ А, мощности $P \le 3$ кВт. Условия разряда были близки к равновесным с концентрацией электронов $n_e \sim 10^{14}$ см$^{-3}$, что соответствует температуре, оцениваемой по формуле Саха, $T \sim 5000K$ [1].

Вначале отметим, что в рассматриваемых условиях выделяются два существенно разных временных масштаба. Время формирования разрядной области $\tau_r = (k_r n_e)^{-1} \sim 10^{-7}\ c$, определяется гибелью заряженных частиц, где $k_r \sim 10^{-7}$ см$^3\cdot$с$^{-1}$

– константа скорости диссоциативной рекомбинации [1]. При используемых осевых скоростях в рассматриваемой вихревой камере продольный размер разрядной области фактически совпадает с диаметром электродов. Ионы, вследствие малой скорости их дрейфа, увлекаются закрученным потоком, формируя спиральную структуру в плоскости *(r,φ)*. В разрядной области присутствуют также продукты эрозии электродов и плазмохимии, нагреваемые до ~ *3000-5000 K*, которые увлекаются потоком и существуют на продольных масштабах ~ $v_z\tau$, формируя фиксируемый в [5, 6] непрерывный спектр теплового излучения и полосатый спектр колебательно-возбуждённых молекул. Здесь τ - характерное время остывания этих частиц. В условиях [5, 6] оно существенно больше времени рекомбинации $\tau >> \tau_r$, поэтому светящиеся области могут протянуться далеко от разрядного промежутка, как это и наблюдалось. Гораздо более интересным является изменение направления прорастания светящегося конуса на противоположное с ростом расхода газа.

Известно, что структура закрученных течений заметно отличается от структуры течений без закрутки. Обзоры аналитических моделей, описывающих структуру таких течений в отсутствие объемных источников нагрева, приведены в [7-13]. Согласно этим моделям, условия образования рециркуляционной (противоточной) зоны определяются, прежде всего, углом закрутки потока, а не расходом газа. Аналитических решений, описывающих структуру закрученного потока с объемным подводом энергии к настоящему времени не существует. Численный анализ эволюции и структуры незакрученного потока в присутствии однородного источника нагрева в плоской геометрии проводился в работах [14, 15]. Стационарная структура закрученного течения в дозвуковом радиально неограниченном потоке с приосевым объемным подводом энергии численно исследовалась в работах [16-18]. При этом использовались приближения, не допускающие образования возвратных течений.

Цель работы – на основе численного расчёта структуры закрученного потока в трубе с источником объемного тепловыделения дать качественное объяснение причин свечения от разрядной области, наблюдаемого в зависимости от расхода газа по направлению или против потока. Для описания структуры областей люминесценции вначале рассмотрим структуру дозвукового закрученного течения, формирующегося в цилиндрической вихревой трубке с торцевым вдувом с учетом нагрева газа в разрядной области. В стационарном случае структура осесимметричного закрученного потока описывается уравнениями

$$P = \rho T / M \ ,$$

$$\frac{\partial(\rho v_z)}{\partial z} + \frac{\partial(\rho v_r)}{\partial r} + \frac{\rho v_r}{r} = 0 \ ,$$

$$c_V \rho\left(v_r \frac{\partial T}{\partial r} + v_Z \frac{\partial T}{\partial z} \right) - T\left(v_r \frac{\partial \rho}{\partial r} + v_Z \frac{\partial \rho}{\partial z} \right) = \frac{1}{r}\frac{\partial}{\partial r}\left(r\kappa \frac{\partial T}{\partial r} \right) + \frac{\partial}{\partial z}\left(\kappa \frac{\partial T}{\partial z} \right) +$$

$$+ \mu \frac{\partial v_i}{\partial x_k}\left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} - \frac{2}{3}\delta_{ik}\frac{\partial v_i}{\partial x_k} \right) + Q(r) \ ,$$

$$\frac{1}{r}\frac{\partial\left(r\rho v_z v_r\right)}{\partial z} + \frac{1}{r}\frac{\partial\left(r\rho v_r^2\right)}{\partial r} = -\frac{\partial P}{\partial r} + \frac{1}{r}\frac{\partial}{\partial z}\left[r\mu\left(\frac{\partial v_z}{\partial r} + \frac{\partial v_r}{\partial z} \right) \right] + \frac{1}{r}\frac{\partial}{\partial r}\left[r\mu\left(2\frac{\partial v_r}{\partial r} - \frac{2}{3}\nabla\vec{v} \right) \right] -$$

$$- 2\mu\frac{v_r}{r^2} + \frac{2}{3}\frac{\mu}{r}\nabla\vec{v} + \rho\frac{v_\varphi^2}{r} \ ,$$

$$\frac{1}{r}\frac{\partial\left(r\rho v_z v_\varphi\right)}{\partial z}+\frac{1}{r}\frac{\partial\left(r\rho v_r v_\varphi\right)}{\partial r}=\frac{1}{r}\frac{\partial}{\partial z}\left[r\mu\frac{\partial v_\varphi}{\partial z}\right]+\frac{1}{r^2}\frac{\partial}{\partial r}\left[r^3\mu\frac{\partial}{\partial r}\left(\frac{v_\varphi}{r}\right)\right]-\frac{\rho v_r v_\varphi}{r},$$

$$\frac{1}{r}\frac{\partial\left(r\rho v_z^2\right)}{\partial z}+\frac{1}{r}\frac{\partial\left(r\rho v_z v_r\right)}{\partial r}=-\frac{\partial P}{\partial z}+\frac{1}{r}\frac{\partial}{\partial z}\left[r\mu\left(2\frac{\partial v_z}{\partial z}-\frac{2}{3}\nabla\vec{v}\right)\right]+\frac{1}{r}\frac{\partial}{\partial r}\left[r\mu\left(\frac{\partial v_z}{\partial r}+\frac{\partial v_r}{\partial z}\right)\right],$$

$$\nabla\vec{v}=\frac{\partial v_z}{\partial z}+\frac{\partial v_r}{\partial r}+\frac{v_r}{r}$$

с граничными условиями на стенке

$$v_r(R,z)=v_\varphi(R,z)=v_z(R,z)=0\,,\ T(R,z)=T_0\,,\ \rho(R,z)=\rho_0$$

и условиями на входе в трубу

$$v_r(r,0)=0,\ v_\varphi(r,0)=f_1(r),\ v_z(r,0)=f_2(r)\,,T(r,0)=T_0\,,\ \rho(r,0)=\rho_0\,,$$

где v_r, v_φ, v_z – радиальная, тангенциальная и продольная компоненты скорости течения, соответственно; P – давление; T – температура; ρ - плотность; μ - коэффициент вязкости; κ - коэффициент теплопроводности; c_V - удельная теплоёмкость при постоянном объёме, зависящая от температуры; Q – мощность объемного источника нагрева. Предполагалось, что переменные не зависят от угловой координаты [7-11]. Невозмущённое потоком давление полагалось равным атмосферному.

Моделирование структуры закрученного течения в дозвуковом потоке с объемным подводом энергии в трубе проводилось с использованием программного комплекса FLUENT 6.3 в режиме двумерного аксиально-симметричного закрученного потока. Расчетная сетка содержала 83331 элемент и сгущалась вблизи стенок трубы и в окрестности источника. Радиус трубы $R = 3$ *см*, длина трубы - *0.5 м*. В начальном сечении задавалась температура $T_0 = 300$ К. Радиальные распределения тангенциальной и осевой скоростей $f_1(r), f_2(r)$ выбирались близкими к экспериментально измеренным в торцевом завихрителе с углом закрутки 45^0 и соответствовали степени закрутки потока, близкой к единице [11]. Радиальная зависимость объёмного источника нагрева (моделирующего нагрев в разряде) соответствовала преимущественному выделению тепла (*~70%*) в кольцевом цилиндре с радиусами $0.5R < r < R$, длиной $L = 1$ *см*. Положение центра источника $z = 0.4$ *м*. Полная мощность разряда считалась равной *1.2 кВт*.

На открытом конце граничное условие соответствовало условию сохранения массового расхода. Для проверки ряд расчётов проведено также при постановке граничного условия в виде постоянного давления в дальней зоне. Качественно приведённые ниже результаты при этом не менялись, но подобный расчёт требовал гораздо больше расчётного времени.

Согласно полученным результатам численного моделирования при используемом в [5, 6] угле закрутки потока возникает противоточная зона засасываемого в центральную область атмосферного газа с вогнутым профилем тангенциальной скорости. Вогнутый профиль тангенциальной скорости в центральной области объясняется тем, что засасываемый в трубу атмосферный воздух первоначально не имеет вращательной составляющей скорости и только позднее, растекаясь от центра, периферийными слоями увлекается во вращение [9]. Этот эффект наиболее сильно выражен для частично закрученных потоков с малой степенью закрутки. Длина и ширина этой зоны растёт с ростом параметра крутки. Осевая

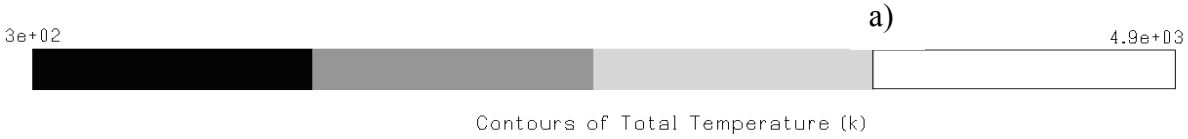
скорость внутри этой области мала, а на границе с основным потоком вообще достигает нуля.

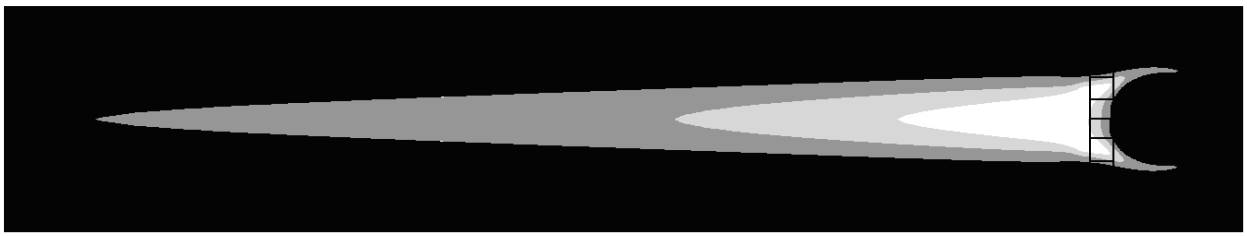
При наличии локализованного источника тепловыделения эта приосевая зона пониженных скоростей сильно перегревается и плотность в ней резко падает. В результате на границе с этой областью также резко растёт радиальная составляющая прямого потока, направленная к оси трубы по градиенту температуры $v_{r1}/v_{r0} = \rho_0/\rho_1 \approx T_1/T_0$, где ρ_0, T_0 и ρ_1, T_1 – плотности и температуры газа в холодной и нагретой областях, соответственно. Следствием этого является преломление линий тока к оси трубки. На аналогичный эффект разрядной области в закрученном потоке ранее было указано в [1].

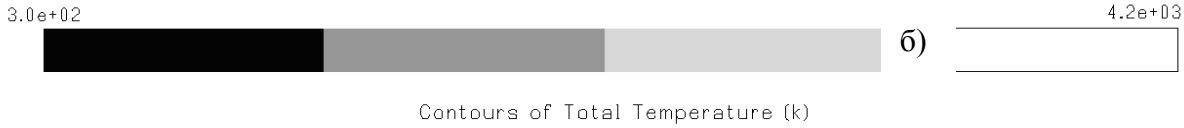
В результате проникновения закрученных периферийных областей потока в осевую область тангенциальная скорость вблизи оси трубы увеличивается по сравнению со случаем без источника. Осевая скорость также растёт. Эффективность конвективного охлаждения растёт с ростом расхода газа. При малых расходах газа (или больших энерговкладах) приосевая зона нагревается сильнее. В результате преломление линий тока происходит под большим углом, и основной поток протекает через локальную зону источника к выходному отверстию, пробивая и разрушая зону противотока. За источником по потоку образуется менее плотная горячая область (рис. 2а). При увеличении расхода газа, то есть более эффективном охлаждении, преломление линий тока оказывается не столь значительным и разрушение зоны противотока происходит только вблизи источника или не происходит вовсе. В результате область наиболее горячего разреженного газа выдувается против потока газа (рис. 2б), причём в зоне источника и перед источником тангенциальная скорость в осевой зоне гораздо больше, чем за источником.



а)







б)

Рис. 2. Изолинии температуры закрученного потока в воздухе, массовый расход: а) 2 г/с, б) 9 г/с. Поток распространяется слева направо. Положение источника указано решеткой.

Полученные закономерности изменения структуры потока с локальным источником нагрева в зависимости от расхода газа позволяют качественно объяснить наблюдаемые особенности светимости поперечного разряда в закрученном потоке. Очевидно, что интенсивность излучения будет больше в более нагретой области. При малых расходах газа сильно нагретые и возбуждённые в разрядной области продукты эрозии электродов и плазмохимических реакций уносятся в прямом направлении от разрядной области, образуя светящийся след по потоку. С увеличением расхода длина этого следа уменьшается. Дальнейшее увеличение расхода газа приводит к выносу светящихся продуктов плазмохимии осевым противотоком, образуя наблюдаемый светящийся конус, растущий в направлении завихрителя.

Работа частично поддержана АВЦП «Развитие научного потенциала высшей школы» (2009 – 2010 гг.), проект 2.1.1/309, ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 гг.

Литература

1. Райзер Ю.П. Основы современной физики газоразрядных процессов. М.: Наука, 1980. 416 с.

2. Адясов С.Ф., Воронов А.И., Катулин В.А., Михеев П.***А***., Николаев В.Д., Петров ***А.***Л., Шепеленко ***А.А***. Электроразрядный CO_2 – лазер с вихревым потоком газа // Квантовая электроника. 1990. Т.17. № 5. С. 537.

3. Лелевкин В.М., Семенов В.Ф. Влияние диафрагмы на вихревую термализацию дуги в канале // Письма в ЖТФ. 2002. Т.28. № 17. С. 31.

4. Михайлов Б.И. Влияние геометрии дуговой камеры, давления, расхода и рода плазмообразующего газа на осевую стабилизацию дуги в газовихревых плазмотронах // Теплофизика и аэромеханика. 2001. Т. 8. № 1. С. 133.

5. Klimov A. et al. Longitudinal Vortex Plasmoid Created by Capacity HF Discharge // Paper AIAA 2008-1386.

6. Klimov A. et al. Study of a Longitudinal Plasmoid Created by Capacity Coupled HF Discharge in Vortex Airflow // Paper AIAA 2009-1046.

7. Гупта А., Лилли Д., Сайред Н. Закрученные потоки. М.: Мир, 1987. 588 с.

8. Гуцол А.Ф. Эффект Ранка // УФН. 1997. Т. 167. № 6. С. 665.

9. Смульский И.И. Аэродинамика и процессы в вихревых камерах. Новосибирск. Наука, 1992. 301 с.

10. Гольдштик М.А. Вихревые потоки. Новосибирск: Наука, 1981. 368 с.

11. Щукин В.К., Халатов А.А. Теплообмен, массообмен и гидродинамика закрученных потоков в осесимметричных каналах. М.: Машиностроение, 1982. 200 с.

12. Казанцева О.В., Пиралишвили Ш.А., Фузеева А.А. Численное моделирование закрученных течений в вихревых трубах // ТВТ. 2005. Т.43. №4. С. 606.

13. Митрофанова О.В. Гидродинамика и теплообмен закрученных потоков в каналах с завихрителями // ТВТ. 2003. Т.41. №4. С.587.

14. Суржиков С.Т. Бифуркация дозвукового газового потока при обтекании локализованного объема низкотемпературной плазмы // ТВТ. 2002. Т.40. № 4. С.591.

15. Железнякова А.Л., Суржиков С.Т. Расчет дозвукового обтекания локальной области тепловыделения // Физико-химическая кинетика в газовой динамике / www.chemphys.edu.ru/pdf/2008-09-01-034.pdf. С.1.

16. Казаков А.В. Влияние объемного подвода энергии на закрученные течения в спутном дозвуковом потоке // Механика жидкости и газа. 1998. № 6. С.47.

17. Казаков А.В. Влияние объемного подвода энергии на устойчивость закрученного дозвукового потока // Механика жидкости и газа. 2003. № 4. С.56.

18. Асмолов Е.С., Казаков А.В., Киселёв А.Ф., Русьянов Д.А. К расчёту закрученных турбулентных многофазных течений вязкого теплопроводного газа с объёмным тепловыделением // ТВТ. 2005. Т.43. №4. С. 594.

Расчет стеклопластиковой панели с учетом ползучести при одностороннем высокотемпературном нагреве

А.А. Абдрахманова, В.П. Павлов
Уфимский государственный авиационный технический университет

Представлены математическая модель и результаты расчета на прочность и жесткость при высоких температурах защемленной по контуру панели, выделенной из прямой круговой конической оболочки.

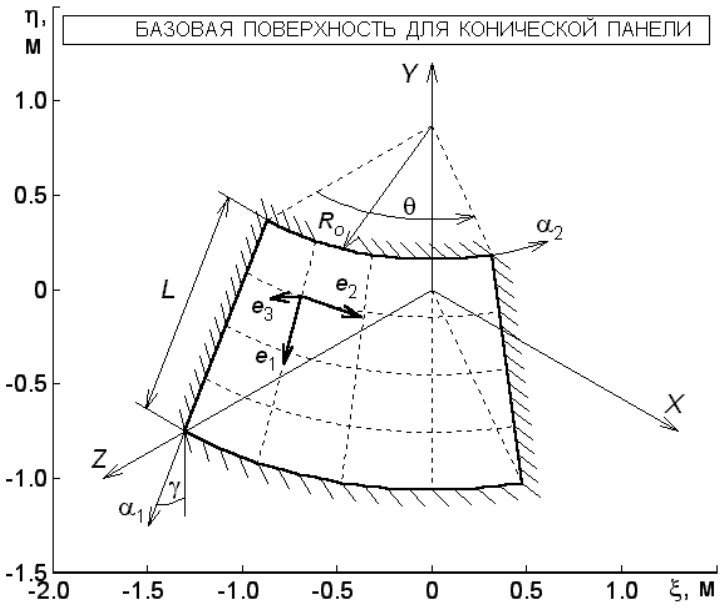


Рисунок 1

Материалом панели является стеклопластик КТ-11-К-Ф, толщина панели 7 мм, температура $T = T(\alpha_3, t)$ в панели изменяется во времени t и по координате α_3 (вдоль орта $\boldsymbol{e}_3$), оставаясь одинаковой вдоль координатных направлений α_1 и α_2. Внешняя распределенная нагрузка равна нулю: $q_1 = 0,\ q_2 = 0,\ q_3 = 0$. Базовой поверхностью оболочки является часть прямой круговой конической поверхности с радиусом верхнего основании радиуса $R_o = 1\text{м}$, длиной образующей $L = 1\text{м}$, углом конусности $\gamma = \pi/6$ и двухгранным углом $\theta = \pi/3$.

Вдоль образующей базовой конической поверхности направлена координатная линия α_1: $0 \le \alpha_1 \le L$ (рис.1). Координатные линии α_1 являются образующими конической поверхности. Вдоль них направлен орт $\boldsymbol{e}_1$.

Вторая криволинейная координата α_2 является двухгранным углом с областью изменения $0 \le \alpha_2 \le \theta$ (рис.1). Координатные линии α_2 на базовой поверхности являются дугами окружностей, получаемыми при пересечении конической базовой поверхности плоскостями, перпендикулярными к оси оболочки. Вдоль координатной линии α_2 направлен орт $\boldsymbol{e}_2$.

Орт нормали к базовой поверхности $\boldsymbol{e}_3$ определяется как векторное произведение $\boldsymbol{e}_3 = \boldsymbol{e}_1 \times \boldsymbol{e}_2$.

Считается справедливой гипотеза прямых нормалей, при которой деформирование оболочки описывается функциями перемещения точек базовой поверхности $u_i = u_i(\alpha_1, \alpha_2),\ i = 1,2,3$.

Панель защемлена по контуру, и граничные условия для ее функций перемещения принимают вид:

$$\begin{cases} u_1 = 0 \quad \text{при} \quad \alpha_1 = 0 \quad \text{и} \quad \alpha_1 = L, \\ u_2 = 0 \quad \text{при} \quad \alpha_2 = 0 \quad \text{и} \quad \alpha_2 = \theta, \\ u_3 = \partial u_3 / \partial \alpha_1 = 0, \quad \text{при} \quad \alpha_1 = 0 \quad \text{и} \quad \alpha_1 = L, \\ u_3 = \partial u_3 / \partial \alpha_2 = 0, \quad \text{при} \quad \alpha_2 = 0 \quad \text{и} \quad \alpha_2 = \pi / 6. \end{cases}$$

При расчете панели применялся пошаговый метод, т. е. решения находились в последовательные моменты времени $t_i, i = 1, 2, \ldots$.

Температурное поле описано в работе [2] и задано на рис. 2.

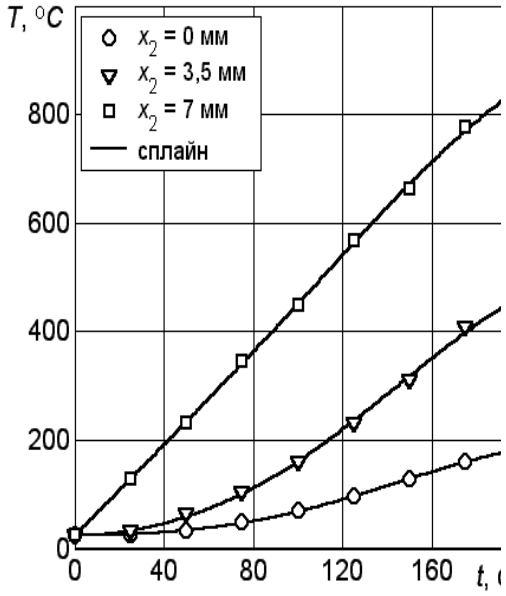


Рис.2. Изменение температуры T во времени t в слоях с координатами α_3 при одностороннем нагреве стержня из КТ-11-К-Ф, закрепленного по концам в опорах заданной жесткости

Уравнения равновесия оболочки произвольного очертания [3] приняты в форме системы из шести уравнений

$$\begin{cases} \dfrac{1}{A_1 A_2}\left[\dfrac{\partial(N_1 A_2)}{\partial \alpha_1} + \dfrac{\partial(S_2 A_1)}{\partial \alpha_2} + S_1 \dfrac{\partial A_1}{\partial \alpha_2} - N_2 \dfrac{\partial A_2}{\partial \alpha_1}\right] + k_1 G_1 = -q_1, \\ \dfrac{1}{A_1 A_2}\left[\dfrac{\partial(S_1 A_2)}{\partial \alpha_1} + \dfrac{\partial(N_2 A_1)}{\partial \alpha_2} + S_2 \dfrac{\partial A_2}{\partial \alpha_1} - N_1 \dfrac{\partial A_1}{\partial \alpha_2}\right] + k_2 G_2 = -q_2, \\ \dfrac{1}{A_1 A_2}\left[\dfrac{\partial(G_1 A_2)}{\partial \alpha_1} + \dfrac{\partial(G_2 A_1)}{\partial \alpha_2}\right] - k_1 N_1 - k_2 N_2 = -q_3, \\ \dfrac{1}{A_1 A_2}\left[\dfrac{\partial(M_1 A_2)}{\partial \alpha_1} + \dfrac{\partial(H_2 A_1)}{\partial \alpha_2} + H_1 \dfrac{\partial A_1}{\partial \alpha_2} - M_2 \dfrac{\partial A_2}{\partial \alpha_1}\right] - G_1 = 0, \\ \dfrac{1}{A_1 A_2}\left[\dfrac{\partial(M_2 A_1)}{\partial \alpha_2} + \dfrac{\partial(H_1 A_2)}{\partial \alpha_1} + H_2 \dfrac{\partial A_2}{\partial \alpha_1} - M_1 \dfrac{\partial A_1}{\partial \alpha_2}\right] - G_2 = 0, \\ S_1 - S_2 + k_1 H_1 - k_2 H_2 = 0, \end{cases} \tag{1}$$

где α_1, α_2 - криволинейные координаты, A_1, A_2 - параметры Ляме, k_1, k_2 - кривизны по направлениям α_1, α_2, N_1, N_2 - продольные силы, G_1, G_2 - поперечные силы, S_1, S_2 - сдвигающие силы, M_1, M_2 - изгибающие моменты, H_1, H_2 - изгибающие моменты, q_1, q_2, q_3 - составляющие внешней распределенной нагрузки.

Было учтено, что в пластинах из композиционных материалов при переменных во времени повышенных температурах возникают значительные тепловые деформации,

приводящие к появлению существенных напряжений, следствием которых являются приведенные к базовой плоскости пластины продольные N_1, N_2 и сдвигающие S_1, S_2 силы. Эти силы, даже при малых прогибах пластины, создают значительные изгибающие моменты, которые соответственно приводят к изменению изгибной деформации пластины.

По методике, изложенной в [2], построена дискретная модель ползучести стеклопластика, где для каждого расчетного момента времени $t_i, i = 2,...,n$ составляется система из трех дифференциальных уравнений равновесия в частных производных по функциям перемещений $u_1(\alpha_1,\alpha_2)$, $u_2(\alpha_1,\alpha_2)$, $u_3(\alpha_1,\alpha_2)$.

В данной работе предлагается решить полученную систему методом двумерных сплайн функций пятой степени. Двумерные сплайны пятого порядка лежат в основе построения численных решений двумерных задач механики, описываемых дифференциальными уравнениями в частных производных до четвертого порядка включительно.

Область задания сплайна ограничивается прямоугольником $\Omega = [a, b]\times[c, d]$, с прямоугольной сеткой линий на нем $\Delta = \Delta_x \times \Delta_y$, где

$$\begin{cases} \Delta_x: \quad a = x_1 < x_2 < ... < x_N = b, \\ \Delta_y: \quad c = x_1 < x_2 < ... < x_M = d. \end{cases}$$

Сетка Δ расчленяет область Ω на прямоугольные ячейки

$$\begin{cases} \Omega_{i,j} = [x_i, x_{i+1}]\times[y_j, y_{j+1}], \\ \quad i = 1,..., N-1, \quad j = 1,..., M-1. \end{cases}$$

Через $C^{4,4}[\Omega]$ обозначено множество непрерывных на Ω функций $f(x,y)$, имеющих непрерывные частные и смешанные производные $D^{r,s}f(x,y)$, $r \le 4, s \le 4$.

Строится функция $W_{5,5,1,1}(x,y)$ - сплайн двух переменных степени 5 дефекта 1 по x и степени 5 дефекта 1 по y с линиями склейки на сетке Δ, удовлетворяющая двум условиям:

Условие 1. В каждой ячейке $\Omega_{i,j}$ функция $W_{5,5,1,1}(x,y)$ является многочленом степени 5 по x и степени 5 по y, то есть

$$\begin{cases} W_{5,5,1,1}(x,y) = \sum\limits_{\alpha=0}^{5}\sum\limits_{\beta=0}^{5} a_{\alpha,\beta}^{(i,j)}(x-x_i)^{\alpha}(y-y_j)^{\beta}, \\ \qquad\qquad i = 1,..., N-1, \quad j = 1,..., M-1; \end{cases}$$

Условие 2. $W_{5,5,1,1}(x,y) \in C^{4,4}[\Omega]$

Считаем, что пространство сплайнов двух переменных $W_{5,5,1,1}(\Delta)$ совпадает с тензорным произведением двух пространств сплайнов одной переменной $W_{5,1}(\Delta_x)$ и $W_{5,1}(\Delta_y)$

$$W_{5,5,1,1}(\Delta) = W_{5,1}(\Delta_x) \otimes W_{5,1}(\Delta_y)$$

В нашем случае пространства одномерных сплайнов $W_{5,1}(\Delta_x)$ и $W_{5,1}(\Delta_y)$ имеют размерности $N+4$ и $M+4$ соответственно [1], а, значит, размерность пространства $W_{5,5,1,1}(\Delta)$ будет равна $(N+4)\times(M+4)$.

Строящийся сплайн задается прямоугольной матрицей $\boldsymbol{Q}$ обобщенных параметров сплайна, имеющей размерность $(N+4)\times(M+4)$, элементами которой являются значения сплайна во всех узловых точках сетки и некоторые из частных и смешанных производных в контурных точках

$$\begin{cases} W_{i,j}^{(\alpha,\beta)} = \dfrac{\partial^{(\alpha+\beta)} W_{5,5,1,1}(x_i, y_j)}{\partial x^{\alpha} \partial y^{\beta}}, \\ i = 1,..., N, \qquad j = 1,..., M, \\ \alpha = 0,...,5, \qquad \beta = 0,...,5, \end{cases}$$

где i и j - нижние индексы, определяющие координаты рассматриваемой узловой точки (x_i, y_j);

α и β - верхние индексы, задающие порядок производных от сплайна соответственно по координатам x и y.

В результате решения данной задачи для каждого расчетного момента времени $t_i, i = 1,2,...$ были определены функции перемещения $u_1 = u_1(\alpha_1,\alpha_2), u_2 = u_2(\alpha_1,\alpha_2), u_3 = u_3(\alpha_1,\alpha_2)$, внутренние силовые факторы $N_1 = N_1(\alpha_1,\alpha_2), N_2 = N_2(\alpha_1,\alpha_2), S_1 = S_1(\alpha_1,\alpha_2), S_2 = S_2(\alpha_1,\alpha_2), M_1 = M_1(\alpha_1,\alpha_2), M_2 = M_2(\alpha_1,\alpha_2), H_1 = H_1(\alpha_1,\alpha_2), H_2 = H_2(\alpha_1,\alpha_2)$ и напряжения $\sigma_{ij} = \sigma_{ij}(\alpha_1,\alpha_2,\alpha_3), i, j = 1, 2$.

На основе описанной методики выполнены расчеты для N_1 и σ_{11}, результаты которых представлены на рис. 3 и рис. 4.

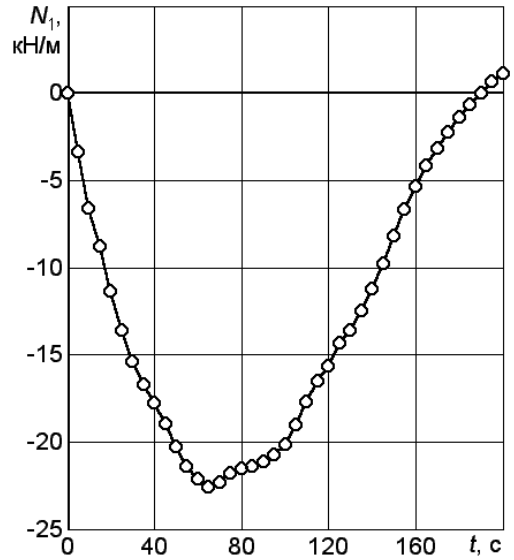


Рис.3. Зависимость погонной продольной силы N_1 от времени t в центральной точке с координатами $\alpha_1 = L/2$ и $\alpha_2 = \theta/2$

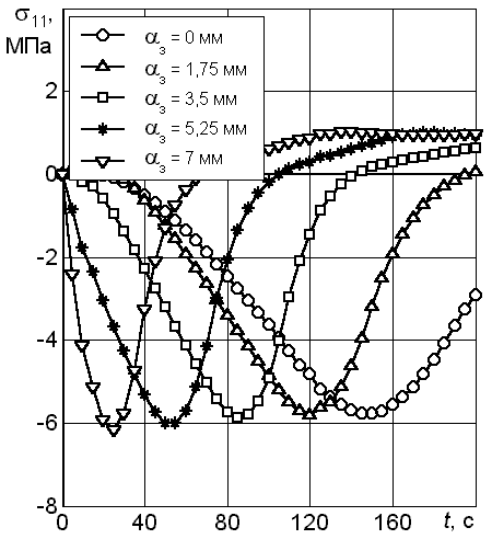


Рис.4. Изменение нормального напряжения σ_{11} от времени t в слоях с координатами α_3 для точки с координатами $\alpha_1 = L/2$ и $\alpha_2 = \theta/2$

На рис. 5 показаны зависимости коэффициента запаса прочности $K = \sigma_{пч}/\sigma_{11}$ от времени в различных слоях для центральной точки на панели при $\alpha_1 = L/2$ и $\alpha_2 = \theta/2$.

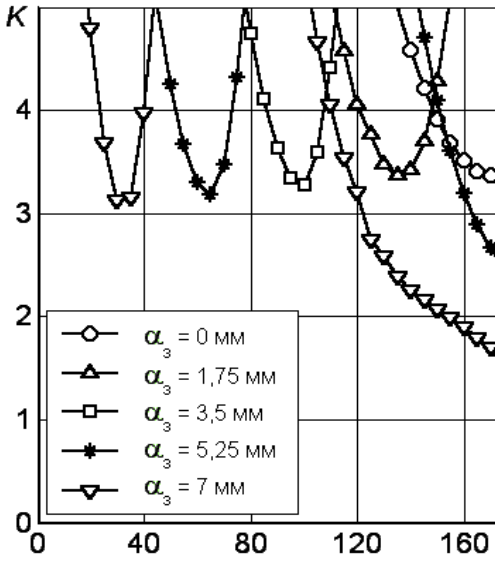


Рис. 5. Изменение коэффициентов запаса прочности K от времени t в слоях с координатами α_3 для точки с координатами $\alpha_1 = L/2,\ \alpha_2 = \theta/2$

Видно (рис. 5), что на наружной поверхности панели $K \to 1$, т. е. напряженное состояние близко к предельному состоянию, что может привести конструкцию к разрушению.

Аналогичным образом можно рассмотреть все остальные точки панели и сделать заключение об ее прочности.

Литература

1. Абдрахманова А.А. Вариант метода сплайна для расчета изгиба балок. /«Вестник Самарского государственного университета» - Самара. 2007. № 2(52) С. 5-15

2. Павлов В.П. Тепловая деформация, прочность и термовязкоупругость стеклопластиков при высокой переменной во времени температуре в условиях термодеструкции. Экспериментальные исследования и математическое моделирование/ Уфа: Уфимск. гос. авиац. техн. ун-т. Уфа, 2004. 218 с.

3. Филин А. П. Элементы теории оболочек. - Л.: Стройиздат, 1975. - 256 с.

Численное моделирование пространственно-временной динамики широкоапертурного лазера с отстройкой частоты генерации

Д.А. Анчиков, А.А. Кренц, Н.Е. Молевич

Самарский государственный аэрокосмический университет (национальный исследовательский университет), Самарский филиал Учреждения Российской академии наук Физического института им. П.Н. Лебедева РАН, г. Самара

Введение

В настоящее время широко исследуются процессы, происходящие в поперечном сечении световой волны, распространяющейся в широкоапертурных лазерных и пассивных оптических системах. Было показано, что с увеличением числа Френеля в широкоапертурном резонаторе Nd-YAG лазера происходит переход от стационарных поперечных картин к периодическим, квазипериодическим и хаотическим картинам. Аналогичная картина получена для электроразрядного CO_2 лазера, работающего на одной продольной моде. В широкоапертурном лазере (при числах Френеля больших 30) почти постоянная интенсивность становится промодулирована в пространстве периодической структурой и слабо промодулирована по времени с частотой ~150 кГц, которая зависела от величины отстройки моды от центра линии перехода. С увеличением числа Френеля, отстройки или накачки картина становилась сильно нерегулярной с полной потерей пространственной и временной корреляции. Настоящая работа посвящена численному расчету пространственно-временной динамики поперечного профиля оптического поля в широкоапертурном лазере с отстройкой частоты генерации.

Основные уравнения. Линейный анализ устойчивости

В качестве исходной рассмотрим следующую систему уравнений

$$\frac{\partial E}{\partial t} - i\frac{\partial^2 E}{\partial x^2} = \frac{\nu}{2}E\left(\frac{N}{1+\Delta^2{}_0} - 1\right) + \frac{i\nu E}{2}\left(2\Delta_{cav} - \frac{\Delta_0}{1+\Delta^2{}_0}N\right), \qquad (1)$$

$$\frac{\partial N}{\partial t} = N_{un} - N\left(1 + \frac{J}{1+\Delta_0{}^2}\right).$$

Система уравнений (1) описывает пространственно-временную динамику лазера в предположении, что поляризация мгновенно следует за изменениями оптического поля. Такая модель в литературе часто называется стандартным адиабатическим исключением поляризации. Предполагается также, что генерация происходит на одой продольной моде резонатора Фабри-Перо. Здесь E - медленно меняющаяся амплитуда поля в широкоапертурном лазере в одномерном (планарном) приближении, обезразмеренная на величину E_s, E_s — амплитуда поля насыщения в активной среде, $N = g/g_t$; g, g_t – коэффициенты усиления активной среды на центральной частоте лазерного перехода и потери, усредненные по длине резонатора ; безразмерное время t и координата x связаны с размерными величинами t_d и x_d как $t = t_d/T_i$, $x = x_d\left(2k/T_i c\right)^{1/2}$, k - волновое число; c – скорость света; T_i – время релаксации населенности уровней активной среды; $\nu = cT_i g_t$ – коэффициент, определяющий отношение времени релаксации населенности активной среды ко времени жизни фотонов в резонаторе ; $\Delta_{cav} = \left(\omega - \omega_c\right)\Big/ cg_t$ – отстройка частоты генерации от моды пустого резонатора, обезразмеренная на ширину линии резонатора ; $\Delta_0 = \left(\omega_0 - \omega\right)T_P$ –

отстройка частоты генерации от центра линии усиления АС, обезразмеренная на полуширину линии усиления ; T_p – время релаксации поляризации; $J = |E|^2$, $N_{un} = g_{un} / g_t$, g_{un} – ненасыщенный коэффициент усиления на частоте ω_0.

Будем искать решение системы уравнений (1) в автомодельном виде в виде бегущих с постоянной скоростью волн. Для этого используем автомодельную замену переменных $\xi = t - \beta x$, где $w = 1/\beta -$ скорость волны, распространяющейся в поперечном к оси резонатора направлении. В этой модели конечность апертуры резонатора не учитывается.

Линейный анализ показал, что если частотная отстройка $\Delta_0 < 0$, то при

$$\beta > \beta_{cr} = \left[\frac{-(1+I_{st})\Delta_0}{(1+I_{st})^2 + \Delta_0{}^2 \nu I_{st}} \right]^{\frac{1}{2}}, \; I_{st} = \frac{J_{st}}{1+\Delta_0{}^2}.$$

режим однородной стационарной генерации оказывается неустойчивым. При $\beta = \beta_{cr}$ происходит бифуркация Андронова-Хопфа [1], в результате чего рождается семейство периодических волн интенсивности малой амплитуды, параметризованных при помощи β.

Моделирование точечной автомодельной системы уравнений

Нелинейную стадию развития возмущений режима однородной стационарной генерации при $\Delta_0 < 0$ исследуем методами численного моделирования системы уравнений (1), записанной в автомодельном виде (с использованием замены координат $\xi = t - \beta x$). Для этого комплексную величину поля представим в виде $E = E' + iE''$. Таким образом, получим систему обыкновенных дифференциальных уравнений размерности $n = 5$. Начальные условия соответствуют малым отклонениям E и N от их стационарных значений.

При малых значениях управляющего параметра $\varepsilon = (\beta - \beta_{cr})/\beta_{cr}$ любые малые отклонения от положения равновесия притягиваются к аттрактору, который является двумерным эргодическим тором [2-4]. На рисунке 1 построена фазопараметрическая диаграмма, иллюстрирующая переход к хаосу через каскад бифуркаций удвоения. Как видно из рисунка 2, при увеличении управляющего параметра ε наблюдаются бифуркации удвоения тора, приводящие в конечном итоге к хаотическому режиму колебаний.

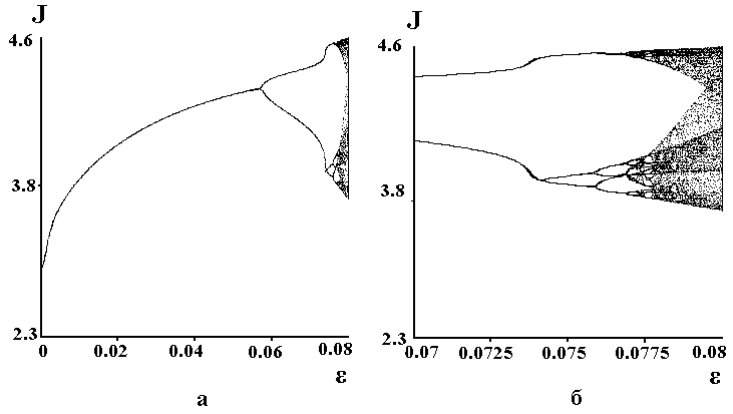


Рисунок 1. Фазопараметрическая диаграмма режимов (а) и её увеличенный фрагмент (б)

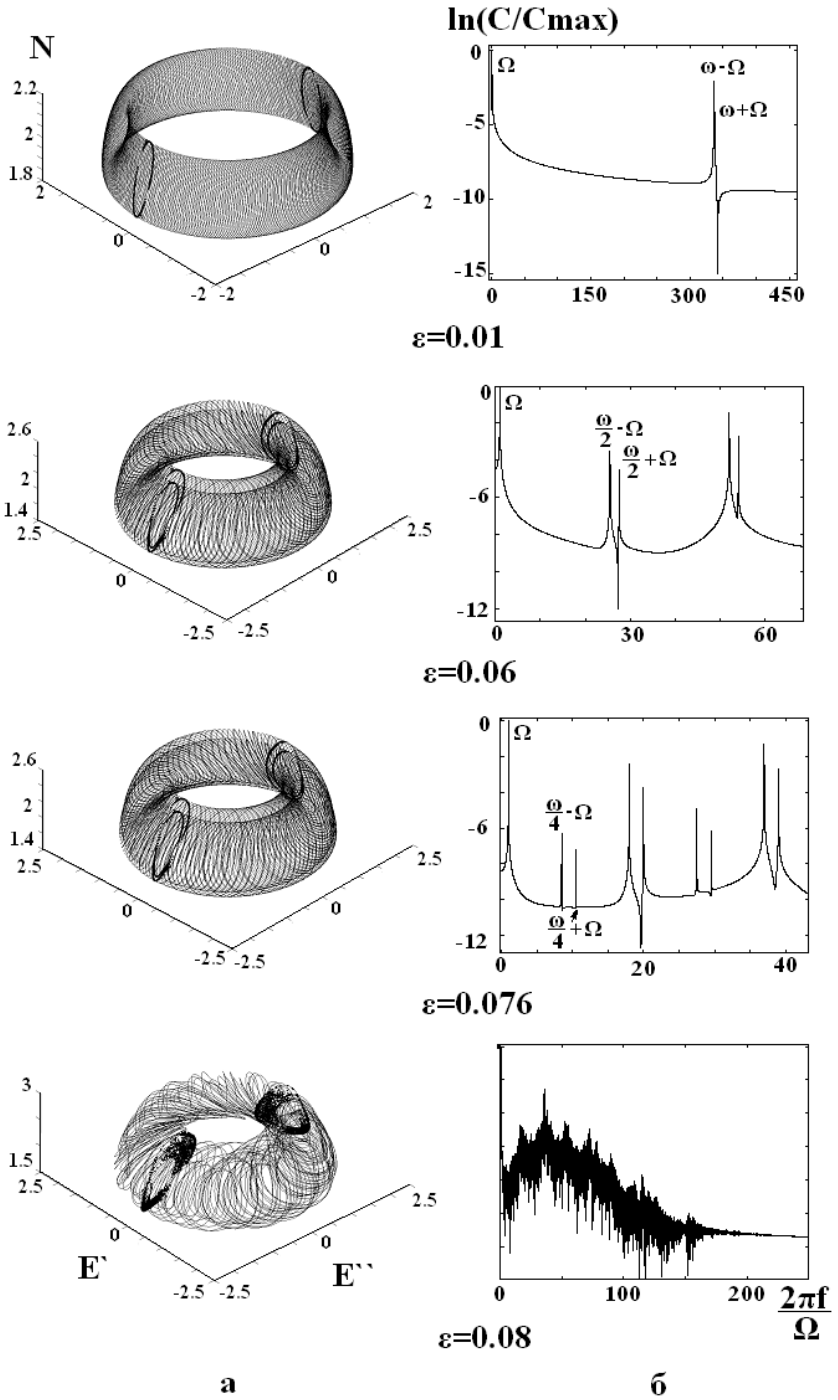


Рисунок 2. Бифуркации удвоения тора

Моделирование распределенной системы уравнений

С помощью численных методов решалась полная (не автомодельная) система уравнений в частных производных. Свойства лазерного резонатора моделировались заданием граничных условий и длины активной среды. В качестве управляющего параметра выбрано время жизни фотона в резонаторе v. Показано, что в распределенной системе при малых значениях управляющего параметра наблюдается режим стационарной генерации (рис. 3).

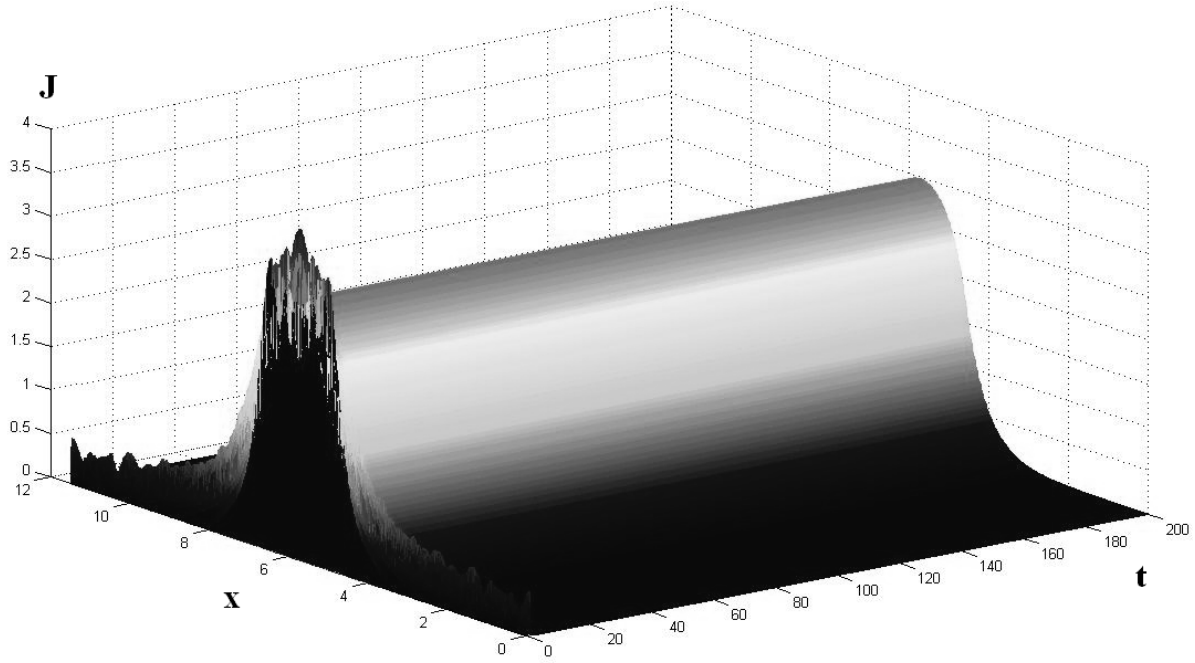
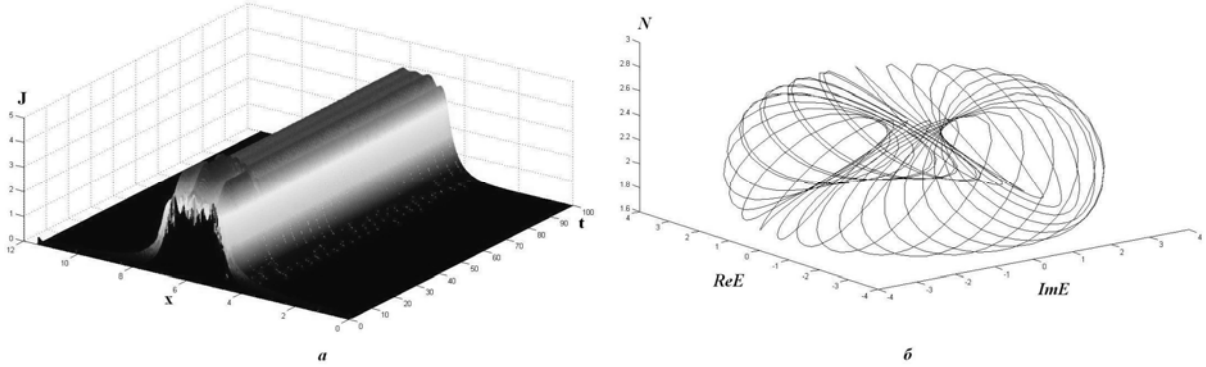


Рисунок 3. Стационарный профиль интенсивности при $\nu = 7$

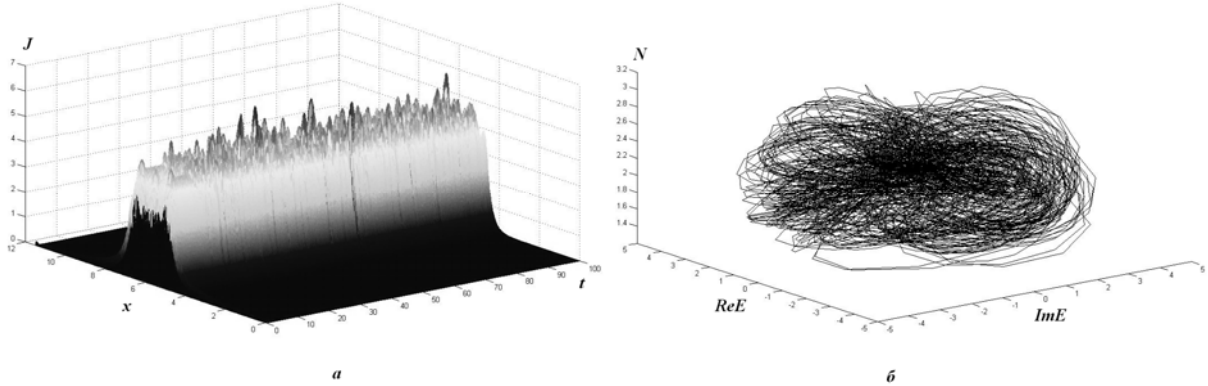
При увеличении параметра рождаются волны интенсивности, бегущие поперек апертуры (рис. 4а). Если снимать значения переменных в точке, то в фазовом объеме рождается устойчивый двумерный тор (рис. 4б).



а *б*

Рисунок 4. а) Волны интенсивности при $\nu = 30$ б) Динамика поля и коэффициента усиления в центре резонатора

При дальнейшем увеличении управляющего параметра, происходит каскад бифуркаций удвоения тора, приводящий к разрушению тора и переходу лазера к хаотическому режиму колебаний (рис. 5).



а *б*

Рисунок 5. а) Пространственно-временной хаос при $\nu = 60$ б) Динамика поля и коэффициента усиления в центре резонатора

Заключение

Таким образом, показано, что в лазерной модели Максвелла-Блоха с адиабатически исключенной поляризацией, как в автомодельной так и в распределенной, наблюдается переход к режиму динамического хаоса через каскад бифуркаций удвоения тора. Найдены условия существования стационарного профиля интенсивности, условия рождения автоволн интенсивности, бегущих поперек лазерной апертуры. Найдены значения лазерных параметров, при которых система переходит к режиму динамического хаоса.

Литература

1. Заикин А.П., Кургузкин А.А., Молевич Н.Е. Периодические автоволновые структуры в широкоапертурном лазере с отстройкой частоты. 1. Бифуркационный анализ // Квантовая электроника. - 1999. - Т. 27, №3. - С. 246 –248.

2. Кренц, А.А. Каскад бифуркаций удвоения тора в лазере с отстройкой частоты [текст]/А.А. Кренц, Н.Е. Молевич//Квантовая электроника. – 2009. - №8. – С. 751-756.

3. Кренц, А.А. Исследование сценария перехода к хаосу в динамической системе с особой окружностью на фазовой плоскости [текст]/А.А. Кренц, Н.Е. Молевич//Известия Самарского научного центра Российской академии наук. – 2010. - №4. – С.108-112

4. Кренц, А.А. Рождение устойчивого тора из замкнутой особой кривой и его бифуркации в лазерной системе с отстройкой частоты [текст]/А.А. Кренц, Н.Е. Молевич // Известия вузов. Прикладная нелинейная динамика. 2010. - №5

Об адекватности некоторых моделей самоподобного трафика

А.В. Благов, А.Ю. Привалов
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

В последние двадцать лет сетевой телекоммуникационный трафик является объектом исследования многих ученых [1]. Было установлено, что сетевой трафик обладает так называемым свойством самоподобия [2]. Несмотря на популярность исследования самоподобного трафика до сих пор нет его единой общепризнанной модели. Некоторые из них не дают желаемых результатов, другие имеют очень высокую вычислительную сложность [3].

Практически актуальными представляются задачи построения имитационной модели самоподобного сетевого трафика для использования её в программах имитационного моделирования при исследованиях характеристик моделей реальных сетей, проводимых на стадии проектирования. Кроме того, хорошая модель может служить не только для имитации реального трафика, но и выявлять его структуру, параметры порождающих трафик источников, ситуацию в сети, по которой трафик распространялся и т.д.

Исследователями было предложено большое количество различных моделей самоподобного сетевого трафика [4, 5, 6]. Как правило проверкой работоспособности модели предлагается соответствие различных статистических характеристик, таких как параметр Херста (описывающий степень самоподобия), математической ожидание, корреляционная функция, одномерное распределение трафика. Реже встречаются оценки средней и максимальной длины очереди [5].

В данной работе авторами предлагается проверка моделей на вход в очередь, а если быть более точным – такая качественная оценка, как двумерное распределение очереди. Моделями для исследования взяты: “Input M/G/∞” и “On-Off Sources”.

Краткое описание моделей

В модель “Input M/G/∞” [4] сеть связи рассматривается, как система с дискретным временем $t = 0, 1, 2...$. Промежуток времени $[t, t+1)$ называется временным окном t. По предположению модели новые источники возникают в системе в каждом временном окне, образуя пуассоновский поток с параметром λ:, т.е., в частности

$$\Pr\{A_t = k\} = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in Z_+, \tag{1}$$

где A_t - это число новых источников трафика, появившихся в системе во временном окне t. Также каждый источник имеет конечное время работы в системе, после истечения которого он исчезает из системы. «Время жизни» источников является случайной величиной τ, имеющей распределение с «тяжёлым хвостом»

$$\Pr\{\tau = k\} \xrightarrow[k \to \infty]{} \frac{C}{k^{2+\beta}}, \tag{2}$$

где C – некоторая константа, а $0 < \beta < 1$. Кроме того, каждый источник имеет определенную скорость генерации пакетов. Модель “Input M/G/∞” хорошо работает для генерации глобальных (TCP трафика) трасс [7]

Модель “On-Off Sources” представлена конечным числом источников. Каждый из них имеет два состояния: состояние активности (On), в котором источником генерируются данные и состояние молчания (Off) - когда источник молчит [7]. Эти

периоды попеременно сменяют друг друга. В каждом On периоде источник генерирует информацию с определённой интенсивностью (скоростью), равной случайной величине S_i. Длительность On периодов всех источников является случайной величиной τ, имеющей распределение « с тяжёлым хвостом». Длительность Off периодов всех источников является независимой случайной величиной, имеющей геометрическое распределение. Как и в модели "Input M/G/∞" каждый источник имеет скорость генерации пакетов, определенную для каждого ON периода [4]. Модель "On-Off Sources" неплохо работает для генерации локальных сетей (Ethernet трассы).

Краткое описание исследуемого материала

В качестве исследуемого материала используем известный архив трасс трафика [8]. Исследуем как трассы трафика локальных сетей (Ethernet трассы), так и глобальных (трассы TCP трафика). Глобальные TCP трассы: dec-pkt-1.tcp, dec-pkt-2.tcp, LBL-PKT-4.TCP [9] – представлены трассами часовой записи трафика между Digital Equipment Corporation и остальным Интернетом (первые две) и трассами часовой записи трафика между Lawrence Berkeley Laboratory и остальным Интернетом (последняя). Трассы локальных Ethernet сетей представлены трассами часовой записи трафика из коллекции корпорации Bellcore: BC-pAug89.TL, BC-pOct89.TL [10], а именно трассы трафика в опорной магистрали корпоративной локальной сети Bellcore.

Помимо указанных выше широко известных трасс, возьмем для исследования еще одну Ethernet трассу часовой записи в магистрали локальной сети Metromax-Самара, снятую в ноябре 2007 г. в качестве более современного материала.

Каждая трасса представляется процессом с дискретным временем, значение которого в определенный момент времени - сумма размеров всех пакетов, прибывших в данном временном окне (шаге дискретизации) [4].

Двумерное распределение очереди

Исследуем поведение системы массового обслуживания дискретного времени со входным трафиком, и найдём двумерное распределение очереди в такой СМО.

Пусть входным потоком заявок будет сам трафик, т.е. данные, поступающие в виде пакетов в каждом временном интервале. Интенсивность обслуживания примем равным среднему количеству поступающих требований, в нашем случае математическому ожиданию, помноженному на коэффициент ρ, который будем менять в пределах от 0 до 1 в различных испытаниях. При обозначенных параметрах входного и выходного потоков проведем проверки на вход в очередь реальных трафиковых трасс и их смоделированных аналогов.

Построим двумерное распределение очереди Q_{t_i}:

$\Pr\{Q_{t1} = a,\ \ Q_{t2} = b\} = \Pr_{t1,t2}(a,b) = \Pr_{t1-t2}(a, b)$, где a и b – меняются от 0 до максимального количества длины очереди $Q_{\max}$. При этом двумерное распределение представляется в виде матрицы размером $Q_{\max}$ x $Q_{\max}$ со значением каждого элемента с номером (a, b) равным $\Pr_{t1-t2}(a, b)$, а $\Delta t = t1 - t2$ выбирается в зависимости от длины самой трассы.

По результатам проведенных экспериментов со всеми трассами, с различными значениями Δt (в нашем случае мы брали равным 10, 100 и 1000), и вариациями коэффициента ρ от 0,3 до 0,8, мы получили относительно схожие результаты (матрицы) как глобальных (при моделировании по обеим моделям), так и локальных (при моделировании по On-Off) трасс (рисунок 1-6).

0.859290 0.003613 0.002860 0.002822 0.002438
0.003662 0.000014 0.000007 0.000007 0.000000
0.002752 0.000007 0.000017 0.000010 0.000014
0.002804 0.000010 0.000010 0.000017 0.000007
0.002372 0.000021 0.000003 0.000007 0.000010
0.002324 0.000028 0.000007 0.000014 0.000007
0.002538 0.000010 0.000014 0.000007 0.000003
0.005937 0.000021 0.000021 0.000024 0.000010
0.004101 0.000010 0.000010 0.000031 0.000003
0.002396 0.000010 0.000003 0.000003 0.000010
0.002192 0.000007 0.000007 0.000000 0.000003
0.001988 0.000007 0.000000 0.000000 0.000007

Рисунок 1. Двумерное распределение очереди по реальной трассе dec1

0.871238 0.003064 0.002481 0.002581 0.002401
0.003060 0.000010 0.000017 0.000003 0.000010
0.002432 0.000003 0.000007 0.000017 0.000007
0.002602 0.000017 0.000010 0.000010 0.000010
0.002353 0.000007 0.000010 0.000014 0.000000
0.001849 0.000003 0.000000 0.000000 0.000007
0.002633 0.000010 0.000021 0.000007 0.000000
0.005348 0.000021 0.000014 0.000007 0.000010
0.004113 0.000035 0.000010 0.000017 0.000028
0.002032 0.000010 0.000000 0.000007 0.000010
0.001884 0.000003 0.000007 0.000007 0.000003
0.001698 0.000010 0.000003 0.000010 0.000017

Рисунок 2. Двумерное распределение очереди по «Input M/G/∞» смоделированной трассе dec1

0.748615 0.002872 0.002247 0.002064 0.001634
0.002832 0.000017 0.000000 0.000000 0.000006
0.002282 0.000006 0.000006 0.000000 0.000029
0.001984 0.000011 0.000000 0.000011 0.000006
0.001600 0.000017 0.000023 0.000000 0.000000
0.001428 0.000006 0.000000 0.000006 0.000011
0.001433 0.000006 0.000017 0.000023 0.000000
0.001336 0.000017 0.000006 0.000006 0.000000
0.001158 0.000000 0.000000 0.000000 0.000000
0.000808 0.000006 0.000000 0.000006 0.000000
0.000768 0.000000 0.000000 0.000006 0.000000

Рисунок 3. Двумерное распределение очереди по реальной трассе Oct.

0.734463 0.001475 0.001564 0.001285 0.001291
0.001498 0.000006 0.000006 0.000000 0.000000
0.001635 0.000006 0.000000 0.000006 0.000006
0.001297 0.000000 0.000000 0.000000 0.000000
0.001267 0.000006 0.000000 0.000000 0.000000
0.000983 0.000000 0.000000 0.000000 0.000000
0.000705 0.000000 0.000000 0.000000 0.000000
0.000776 0.000000 0.000000 0.000000 0.000000
0.000734 0.000000 0.000000 0.000000 0.000000
0.000841 0.000000 0.000000 0.000006 0.000000
0.000503 0.000000 0.000000 0.000000 0.000000

Рисунок 4. Двумерное распределение очереди по «On-Off Sources» смоделированной трассе Oct

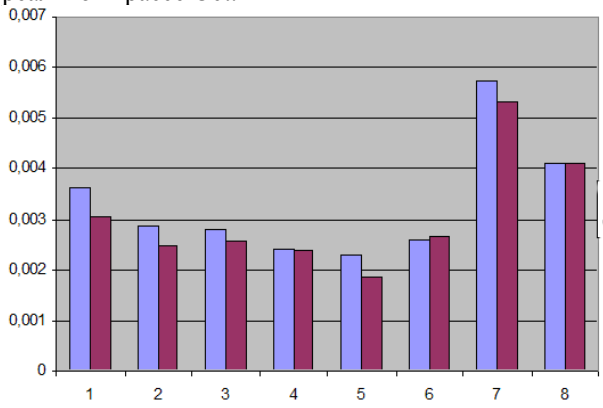


Рисунок 5. Графическое сравнение элементов первого ряда двумерное распределения очереди по реальной и смоделированной трассе dec1

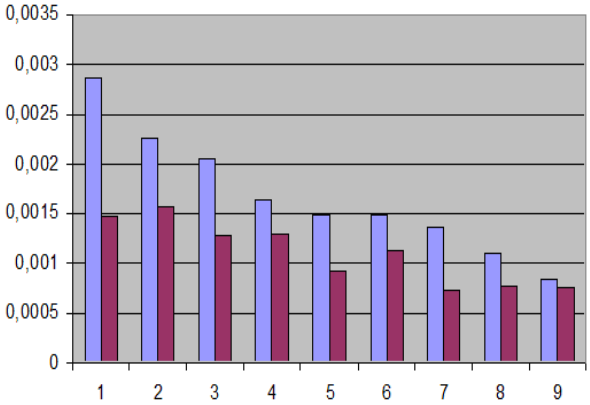


Рисунок 6. Графическое сравнение элементов первого ряда двумерное распределения очереди по реальной и смоделированной трассе Oct.

Для более строгой оценки наших моделей введем такой параметр как СКО по первым наиболее значимым 100 элементам (квадрат 10x10) двумерных распределений (σ). Причем сравнивать полученные результаты будем с результатами, полученными при моделировании без аппроксимации по корреляционной функции (для «входной M/G/∞») и без оптимизации по одномерному распределению – распределению скоростей (для обеих моделей).

Таблица 1. Сравнительные оценки по двумерному распределению очереди

Трасса, нагрузка СМО и интервал времени в двумерном распределении			σ без приближения одномер. распред. и корр. функции	σ с приближением одномер. распред.	σ с приближением распред. и корр. ф.
Dec1	**ρ=0.4**	**Δ=1000**	0.001499	$7.1\cdot10^{-4}$	$3.8\cdot10^{-4}$
		Δ=100	0.001588	$8.3\cdot10^{-4}$	$4.2\cdot10^{-4}$
	ρ=0.8	**Δ=1000**	0.236777	$8.7\cdot10^{-3}$	$6.9\cdot10^{-3}$
		Δ=100	0.239779	$9.9\cdot10^{-3}$	$8.3\cdot10^{-3}$
Dec2	**ρ=0.4**	**Δ=1000**	0.001789	$4.5\cdot10^{-6}$	$3.6\cdot10^{-6}$
		Δ=100	0.001918	$4.3\cdot10^{-6}$	$3.3\cdot10^{-6}$

	ρ=0.8	**Δ=1000**	0.978629	$5.3\cdot10^{-3}$	$2.6\cdot10^{-3}$
		Δ=100	0.981528	$5.5\cdot10^{-3}$	$2.95\cdot10^{-3}$
lbl-pkt4	**ρ=0.4**	**Δ=1000**	0.134501	$1.2\cdot10^{-3}$	$1.1\cdot10^{-3}$
		Δ=100	0.129529	$1.2\cdot10^{-3}$	$1.1\cdot10^{-3}$
	ρ=0.8	**Δ=1000**	0.284557	0.042554	0.013577
		Δ=100	0.284493	0.09155	0.042686
lbl-pkt5	**ρ=0.4**	**Δ=1000**	0.106974	$1.2\cdot10^{-3}$	$1.1\cdot10^{-4}$
		Δ=100	0.106003	$1.7\cdot10^{-3}$	$4.9\cdot10^{-5}$
	ρ=0.8	**Δ=1000**	0.387547	0.050898	0.075888
		Δ=100	0.378847	0.060657	0.042004
Oct	**ρ=0.4**	**Δ=1000**	0.560480	$2.1\cdot10^{-4}$	$4.7\cdot10^{-3}$
		Δ=100	0.591422	$2.8\cdot10^{-4}$	$1.2\cdot10^{-4}$
	ρ=0.8	**Δ=1000**	0.817393	$2.4\cdot10^{-4}$	$7.4\cdot10^{-4}$
		Δ=100	0.822923	$2.4\cdot10^{-4}$	$2.8\cdot10^{-4}$
Aug	**ρ=0.4**	**Δ=1000**	0.931578	$1.1\cdot10^{-4}$	$1.3\cdot10^{-4}$
		Δ=100	0.934616	$1.3\cdot10^{-4}$	$1.0\cdot10^{-4}$
	ρ=0.8	**Δ=1000**	0.915703	$6.9\cdot10^{-3}$	$8.6\cdot10^{-3}$
		Δ=100	0.916172	$7.8\cdot10^{-3}$	$6.9\cdot10^{-3}$
Met-ro-max	**ρ=0.4**	**Δ=1000**	0.772726	$1.3\cdot10^{-4}$	$1.5\cdot10^{-3}$
		Δ=100	0.99	$2.0\cdot10^{-4}$	$8.6\cdot10^{-4}$
	ρ=0.8	**Δ=1000**	0.959726	$9.7\cdot10^{-4}$	$1.5\cdot10^{-3}$
		Δ=100	0.983493	0.050447	$1.1\cdot10^{-4}$

Выводы. Безусловно к вопросу адекватности модели сетевого трафика необходимо подходить комплексно. А предложенный авторами метод может служить для демонстрации практической необходимости имитации трафика, являясь при этом достаточно простым в реализации и логически подходящим к такому объекту как телекоммуникационный самоподобный трафик.

Литература

1. Шелухин О.И. Фрактальные процессы в телекоммуникациях./ О.И. Шелухин, А.М. Тенякшев, А.В. Осин. - М.: Радиотехника, 2003.- 480 с.
2. Willinger, W. A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks/ W. Willinger, M.S.Taqqu, A. Erramilli. – Clarendon press, Oxford, 1996. Pp 339-366.
3. Melamed, B. An overview of TES process and modeling metodology. In L.Donatiello and R.Nelson, editors, «Models and techniques for Performance Evaluation of Computer and Communications Systems». – Springer Verlag, New-York, 1993. -Pp.359-393.
4. Благов А.В. Модификации моделей типа «входная M/G/∞» и «On-Off источники» для имитационного моделирования самоподобного телекоммуникационного трафика./ А.В. Благов, А.Ю. Привалов. – М.: Труды МАИ, 2010.
5. Ильницкий, С.В. Самоподобный (self-similar) входной поток в системах M/M/1 и G/M/1 /С.В. Ильницкий/Computer Science/Рижский тех. ун-т. -Рига, 2004.
6. Громов, Ю.Ю. Фрактальный анализ и процессы в компьютерных сетях/ Ю.Ю Громов,. Н.А. Земской, О.Г. Иванова, А.В. Лагутин - Тамбов: Изд-во технич. ун-та, 2007. - 65с.
7. Privalov, A.Yu. Some Models Parameters Calculation for Simulation of Network Traffic Marginal Distribution and Self-similarity/ A.Yu. Privalov, A.V. Blagov – 23 European conference on modeling and simulation. – Madrid, 2009.
8. Internet Traffic Archive. DOI= http://ita.ee.lbl.gov.
9. V. Paxson and S. Floyd. Wide-Area Traffic: The Failure of Poisson Modeling, *IEEE/ACM Transactions on Networking*, 3(3), pp. 226-244, June 1995.

Приближенное решение гиперсингулярных интегральных уравнений

И. В.Бойков, А.И. Бойкова
Пензенский государственный университет, г. Пенза

Предложены и обоснованы сплайн–коллокационные вычислительные схемы нулевого порядка для приближенного решения гиперсингулярных интегральных уравнений, полигиперсингулярных интегральных уравнений, многомерных гиперсингулярных интегральных уравнений. В частности, получены условия однозначной разрешимости и скорости сходимости приближенного решения уравнения Прандтля крыла конечного размера.

Введение

Хорошо известен спектр применения сингулярных интегральных уравнений в физике, механике и технике: теория упругости, термоупругости, аэродинамика, электродинамика. При изучении более сложных процессов в этих же областях происходит естественный переход от сингулярных к гиперсингулярным интегральным уравнениям. Однако, как точное вычисление сингулярных и гиперсингулярных интегралов, так и точное решение сингулярных и гиперсингулярных интегральных уравнений возможно лишь в исключительных случаях.

Поэтому основным аппаратом при решении прикладных задач, описываемых сингулярными и гиперсингулярными интегральными уравнениями, являются численные методы. Подробное изложение приближенных методов вычисления сингулярных и гиперсингулярных интегралов дано в книгах [1], [2], а приближенные методы решения сингулярных и гиперсингулярных интегральных уравнений и подробная библиография содержатся в книгах [3], [4].

Напомним определения гиперсингулярных интегралов.

Определение 1 [5]. Интеграл вида $\int_a^b \frac{\varphi(\tau)\,d\tau}{(b-\tau)^{p+\alpha}}$ при целом p и $0<\alpha<1$ определяет величину («конечную часть») рассматриваемого интеграла, как предел при $t \to b$ суммы

$$\int_a^b \frac{\varphi(\tau)\,d\tau}{(b-\tau)^{p+\alpha}} = \lim_{t\to b}\left[\int_a^t \frac{\varphi(\tau)\,d\tau}{(b-\tau)^{p+\alpha}} + \frac{B(t)}{(b-t)^{p+\alpha-1}}\right],$$

если предположить, что $\varphi(t)$ имеет p производных в окрестности точки b. Здесь $B(t)$ – любая функция, на которую налагаются два условия:

а) рассматриваемый предел существует,

б) $B(t)$ имеет по крайней мере p производных в окрестности точки $t=b$.

Определение 2 [6]. Пусть функция $\varphi(t)$ имеет непрерывные производные до $p-1$ порядка в окрестности точки с $(a<c<b)$. Интегралом $\int_a^b \frac{\varphi(\tau)\,d\tau}{(\tau-c)^p}$ в смысле главного значения Коши-Адамара называется следующий предел

$$\int_a^b \frac{\varphi(\tau)\,d\tau}{(\tau-c)^p} = \lim_{v\to 0}\left[\int_a^{c-v} \frac{\varphi(\tau)\,d\tau}{(\tau-c)^p} + \int_{c+v}^b \frac{\varphi(\tau)\,d\tau}{(\tau-c)^p} + \frac{B(v)}{v^{p-1}}\right],$$

где $B(v)$ –некоторая функция, выбранная так, чтобы указанный предел существовал.

Определение 3. Пусть функция $\varphi(t)$ и ее производные до $p-1$ порядка имеют разрыв первого рода в точке $t=c$, причем как для самой функции, так и для ее

производных существуют пределы слева и справа. Интегралом $\int_a^b \frac{\varphi(\tau)d\tau}{(\tau-c)^p}$ называется следующий предел

$$\int_a^b \frac{\varphi(\tau)d\tau}{(\tau-c)^p} = \lim_{v\to 0}\left[\int_a^{c-v} \frac{\varphi(\tau)d\tau}{(\tau-c)^p} + \frac{B_1(v)}{v^{p-1}} + B_2(v)\ln v + \int_{c+v}^b \frac{\varphi(\tau)d\tau}{(\tau-c)^p} + \frac{D_1(v)}{v^{p-1}} + D_2(v)\ln v\right],$$

где функции $B_1(v)$ и $D_1(v)$ имеют непрерывные левые и правые производные в окрестности нуля, функции $B_2(v)$ и $D_2(v)$ удовлетворяют условиям Дини–Липшица в окрестности нуля. Функции $B_1(v)$, $B_2(v)$, $D_1(v)$, $D_2(v)$ подбираются так, чтобы предел существовал.

Рассмотрим интеграл

$$H\varphi = \iint_G \frac{\varphi(\tau_1,\tau_2)d\tau_1 d\tau_2}{((\tau_1-t_1)^2+(\tau_2-t_2)^2)^{p/2}},$$

где p –вещественное число, $p>2$. Обозначим через $R(t,r)$, $t=(t_1,t_2)$ круг с центром в точке t радиуса r.

Определение 4 [2]. Регуляризацией интеграла $H\varphi$ называется предел

$$H\varphi = \lim_{\varepsilon\to 0}\left(\iint_{G\setminus R(t,\varepsilon)} \frac{\varphi(\tau_1,\tau_2)d\tau_1 d\tau_2}{((\tau_1-t_1)^2+(\tau_2-t_2)^2)^{p/2}} - \frac{B(\varepsilon)}{\varepsilon^{p-1}}\right),$$

где $B(\varepsilon)$ –любая функция, на которую налагаются два условия:

а) рассматриваемый предел существует,

б) функция $B(\varepsilon)$ имеет, по крайней мере, $p-1$ непрерывную производную в окрестности нуля.

2. Приближенное решение гиперсингулярных интегральных уравнений

Рассмотрим линейное гиперсингулярное интегральное уравнение

$$Kx \equiv a(t)x(t) + b(t)\int_{-1}^{1} \frac{x(\tau)d\tau}{(\tau-t)^p} + \int_{-1}^{1} h(t,\tau)x(\tau)d\tau = f(t),\ p=2,3,\dots. \qquad (2.1)$$

На коэффициенты и правую часть уравнения (2.1) наложим следующие условия: 1) функции $a(t)$, $b(t)$, $f(t)$, $h(t,\tau)$ непрерывны; 2) $b(t)\neq 0$ при $t\in[-1,1]$; 3) уравнение (2.1) однозначно разрешимо.

Вначале рассмотрим приближенный метод решения уравнения (2.1) в предположении, что p –четное число.

Введем узлы $t_k = -1+2k/N$, $k=0,1,\dots,N$ и $\overline{t}_k = t_k + 1/N$, $k=0,1,\dots,N-1$. Пусть $\Delta_k = [t_k, t_{k+1})$, $k=0,1,\dots,N-2$, $\Delta_{N-1}=[t_{N-1},1]$.

Приближенное решение уравнения (2.1) будем искать в виде кусочно–постоянной функции

$$x_N(t) = \sum_{k=0}^{N-1} \alpha_k \psi_k(t), \qquad (2.2)$$

где $\psi_k(t)=1$ при $t\in\Delta_k$, $\psi_k(t)=0$ при $t\in[-1,1]\setminus\Delta_k$.

Значения $\{\alpha_k\}$, $k=0,1,\dots,N-1$, определяются из системы линейных алгебраических уравнений

$$a(\overline{t}_k)\alpha_k + \sum_{l=0}^{N-1}{}' \alpha_l \int_{\Delta_l} \frac{d\tau}{(\tau - \overline{t}_k)^p} + \sum_{l=0}^{N-1} \alpha_l h(\overline{t}_k, \overline{t}_l)\frac{2}{N} = f(\overline{t}_k),\ k = 0,1,...,N-1. \qquad (2.3)$$

Здесь $\sum'$ означает суммирование по $l \neq k-v, k-v+1, ..., k-1, k+1, ..., k+v-1$. Целочисленное значение v зависит от абсолютных величин коэффициентов $a(t)$, $b(t)$ и от значения N.

Теорема 2.1. [7]. Пусть уравнение (2.1) имеет единственное решение $x^*(t)$ и пусть v – минимальное значение при котором выполняется условие

$$B_* \frac{2}{p-1} N^{p-1} - A^* - \frac{2}{N} H^* > B^* \frac{N^{p-1}}{p-1}\left(\frac{1}{(2v-1)^{p-1}} - \frac{1}{(N-1)^{p-1}}\right) + 2H^*, \text{ где}$$

$$A_* = \min_{-1\le t\le 1} |a(t)|,\ A^* = \max_{-1\le t\le 1} |a(t)|,\ B_* = \min_{-1\le t\le 1} |b(t)|,\ B^* = \max_{-1\le t\le 1} |b(t)|,\ H_* = \min_{-1\le t_1 t_2\le 1} |h(t_1,t_2)|,$$

$$H^* = \max_{-1\le t_1 t_2\le 1} |h(t_1,t_2)|.$$

Тогда при данном v система уравнений (2.3) имеет единственное решение $x_N^*(t)$.

Пусть p – нечетное целое число. В этом случае приближенное решение уравнения (2.1) ищется в виде кусочно–постоянной функции (2.2), коэффициенты $\{\alpha_k\}$ которой определяются из системы линейных алгебраических уравнений

$$a(t_k)\alpha_k + d(t_k)\sum_{l=0}^{N-1}{}' \alpha_k \int_{\Delta_l} \frac{d\tau}{(\tau - t_k)^p} + \frac{2}{N}\sum_{l=0}^{N-1} h(t_k,t_l)\alpha_l = f(t_k),\ k = 0,1,...,N-1, \qquad (2.4)$$

где $\sum_{l=0}^{N-1}{}'$ означает суммирование по $l \neq k-2, k-1, k+1$.

Теорема 2.2. Пусть уравнение (2.1) имеет единственное решение $x^*(t)$ и выполнены условия: $|b(t)| \neq 0$,

$$|b(t_i)|\frac{N^{p-1}}{2^{p-1}(p-1)} - |a(t_i)| - \frac{2}{N}|h(t_i,t_i)| > |b(t_i)|\left(\left(\frac{N}{4}\right)^{p-1}\frac{1}{p-1} - \frac{1}{2^{p-1}(p-1)}\right) + 2H^*,\ i = 0,1,N-2,N-1,$$

$$|b(t_k)|\frac{N^{p-1}}{2^{p-1}(p-1)} - |a(t_k)| - \frac{2}{N}|h(t_k,t_k)| >$$

$$> |b(t_k)|\left(\frac{2}{p-1}\left(\frac{N}{4}\right)^{p-1} - \frac{1}{p-1}\left(\frac{N}{2N-2k}\right)^{p-1} - \frac{1}{p-1}\left(\frac{N}{2k}\right)^{p-1}\right) + 2H^*,$$

$$k = 2,...,N-3.$$

Тогда система уравнений (2.4) имеет единственное решение x_N^*.

Рассмотрим бигиперсингулярное интегральное уравнение

$$a(t_1,t_2)x(t_1,t_2) + b(t_1,t_2)\int_{-1}^{1}\frac{x(\tau_1,t_2)}{(\tau_1-t_1)^p}d\tau_1 + c(t_1,t_2)\int_{-1}^{1}\frac{x(t_1,\tau_2)}{(\tau_2-t_2)^p}d\tau_2 +$$

$$+ d(t_1,t_2)\int_{-1}^{1}\int_{-1}^{1}\frac{x(\tau_1,\tau_2)d\tau_1 d\tau_2}{(\tau_1-t_1)^p(\tau_2-t_2)^p} + \int_{-1}^{1}\int_{-1}^{1} h(t_1,t_2,\tau_1,\tau_2)x(\tau_1,\tau_2)d\tau_1 d\tau_2 = f(t_1,t_2). \qquad (2.5)$$

Распространение полученных ниже утверждений на полигиперсингулярные интегральные уравнения не представляет затруднений. Для простоты обозначений ниже в уравнении (2.5) опускается компактный оператор.

Пусть N –целое число, $v_k = -1 + 2k/N, \quad k = 0,1,...,N,$ $\overline{v}_k = v_k + k/N,\ k = 0,1,\cdots,N-1, \quad \Delta_k = [v_k, v_{k+1}), \quad k = 0,1,...,N-2, \quad \Delta_k = [v_{N-1},1],$ $\Delta_{k,l} = [v_k, v_{k+1}) \times [v_l, v_{l+1}), \quad k,l = 0,1,...,N-2, \quad \Delta_{N-1,l} = [v_{N-1} v_N] \times [v_l, v_{l+1}), \quad l = 0,1,...,N-2,$ $\Delta_{k,N-1} = [v_k, v_{k+1}) \times [v_{N-1}, v_N], \quad k = 0,1,...,N-1, \ \Delta_{N-1,N-1} = [v_{N-1}, v_N] \times [v_{N-1}, v_N].$

Приближенное решение уравнение (2.5) будем искать в виде кусочно–постоянной функции

$$x_{N,N}(t_1,t_2) = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} \alpha_{k_1 k_2} \psi_{k_1 k_2}(t_1,t_2), \tag{2.6}$$

где $\psi_{k_1 k_2}(t_1,t_2) = 1,$ если $(t_1,t_2) \in \Delta_{k_1 k_2},$ $\psi_{k_1 k_2}(t_1,t_2) = 0,$ если $(t_1,t_2) \in [-1,1]^2 \setminus \Delta_{k_1 k_2}.$

Пусть p – четное число.

Коэффициенты $\{\alpha_{k_1,k_2}\}$ находятся из системы уравнений

$$\begin{aligned} &a(\overline{v}_i,\overline{v}_j) a_{ij} + b(\overline{v}_i,\overline{v}_j) {\sum_{k=0}^{N-1}}' a_{k,j} \int_{\Delta_k} \frac{d\tau_1}{(\tau_1 - \overline{v}_i)^p} + c(\overline{v}_i,\overline{v}_j) {\sum_{k=0}^{N-1}}' a_{ij} \int_{\Delta_i} \frac{d\tau_2}{(\tau_2 - \overline{v}_i)^p} + \\ &+ d(\overline{v}_i,\overline{v}_j) {\sum_{k=0}^{N-1}}' {\sum_{k=0}^{N-1}}' a_{k,j} \int\int_{\Delta_{ki}} \frac{d\tau_1 d\tau_2}{(\tau_1 - \overline{v}_i)^p (\tau_2 - \overline{v}_i)^p} = f(\overline{v}_i,\overline{v}_j), \ i,j = 0,1,...,N-1, \end{aligned} \tag{2.7}$$

где $\sum_k' \left(\sum_i' \right)$ означает суммирование по $k \neq i-1,\ i+1 (l \neq j-1, j+1).$

Теорема 2.3 [7]. Пусть выполнены условия:

1) уравнение (2.5) имеет единственное решение $x^*(t_1,t_2),\ 2) \ |d(t_1,t_2)| > 0$

при $(t_1,t_2) \in [-1,1]^2.$ Тогда при достаточно больших N система уравнений (2.7) имеет единственное решение $x_N^*(t_1,t_2).$

Рассмотрим теперь случай, когда p – нечетное число.

Приближенное решение будем искать в виде функции (2.6), коэффициенты которой определяются из системы уравнений

$$\begin{aligned} &a(v_{k_1},v_{k_2}) \alpha_{k_1 k_2} + b(v_{k_1},v_{k_2}) {\sum_{l_1=0}^{N-1}}' \alpha_{l_1 k_2} \int_{\Delta_{l_1}} \frac{d\tau_1}{(\tau_1 - v_{k_1})^p} + c(v_{k_1},v_{k_2}) {\sum_{l_2=0}^{N-1}}'' \alpha_{k_1 l_2} \int_{\Delta_{l_2}} \frac{d\tau_2}{(\tau_2 - v_{k_2})^p} + \\ &+ d(v_{k_1},v_{k_2}) {\sum_{l_1=0}^{N-1}}^* {\sum_{l_2=0}^{N-1}}^* \iint_{\Delta_{l_1 l_2}} \frac{d\tau_1 d\tau_2}{(\tau_1 - v_{k_1})^p (\tau_2 - v_{k_2})^p} = f(v_{k_1},v_{k_2}),\ k_1,k_2 = 0,1,...,N-1. \end{aligned} \tag{2.8}$$

Здесь ${\sum_{l_1=0}^{N-1}}'$ означает суммирование по $l_1 \neq k_1 - 1,$ ${\sum_{l_2=0}^{N-1}}''$ означает суммирование по $l_2 \neq k_2 - 1,$ ${\sum_{l_1=0}^{N-1}}^* {\sum_{l_2=0}^{N-1}}^*$ означает суммирование по $l_1 \neq k_1 - 2, k_1 - 1, k_1 + 1, l_2 \neq k_2 - 2, k_2 - 1, k_2 + 1.$

Теорема 2.4. Пусть выполнены условия:

1) уравнение (2.5) имеет единственное решение $x^*(t_1,t_2),\ 2) \ |d(t_1,t_2)| > 0$

при $(t_1,t_2) \in [-1,1]^2.$ Тогда при достаточно больших N система уравнений (2.8) имеет единственное решение $x_{NN}^*(t_1,t_2).$

Рассмотрим многомерное гиперсингулярное интегральное уравнение

$$a(t_1,t_2)\,x(t_1,t_2)+b(t_1,t_2)\iint\limits_{\Omega}\frac{x(\tau_1,\tau_2)d\tau_1 d\tau_2}{\left((\tau_1-t_1)^2+(\tau_2-t_2)^2\right)^{p/2}}+$$
$$+\quad\iint\limits_{\Omega}h(t_1,t_2,\tau_1,\tau_2)\,x(\tau_1,\tau_2)d\tau_1 d\tau_2=f(t_1,t_2). \qquad (2.9)$$

Приближенное решение уравнения (2.9) будем искать в виде функции (2.6), коэффициенты $\{\alpha_{kl}\}$ которой определяются из системы уравнений

$$a\left(\overline{v}_k,\overline{v}_l\right)\alpha_{kl}+b\left(\overline{v}_k,\overline{v}_l\right)\sum_{k_1=0}^{N-1}\sum_{l_1=0}^{N-1}{}'\iint\limits_{\Delta_{k_1l_1}}\frac{x_N(\tau_1,\tau_2)d\tau_1 d\tau_2}{((\tau_1-v_{k_1})^2+(\tau_2-v_{k_2})^2)^{p/2}}+$$
$$+\frac{4}{N^2}\sum_{k_1=0}^{N-1}\sum_{l_1=0}^{N-1}h(\overline{v}_k,\overline{v}_l,\overline{v}_{k_1},\overline{v}_{l_1})=f(\overline{v}_k,\overline{v}_l),\ k,l=0,1,...,N-1. \qquad (2.10)$$

Здесь $\sum\limits_{k_1}\sum\limits_{l_1}{}'$ означает суммирование по всем квадратам $\Delta_{k_1l_1}$, не соприкасающимся с квадратом Δ_{kl}. При этом интеграл по Δ_{kl} входит в сумму.

Теорема 2.5 [7]. Пусть выполнены следующие условия: 1) уравнение (2.9) имеет единственное решение $x^*(t_1,t_2)$, 2) справедливо неравенство $\min\limits_{t_1,t_2}|b(t_1,t_2)|\geq b>0$. Тогда при достаточно больших значениях N система уравнений (2.10) имеет единственное решение $x^*_{NN}(t_1,t_2)$.

Замечание. Для уравнения Прандтля $b(t_1,t_2)\iint\limits_{\Omega}\frac{x(\tau_1,\tau_2)d\tau_1 d\tau_2}{\left((\tau_1-t_1)^2+(\tau_2-t_2)^2\right)^{3/2}}=f(t_1,t_2)$ получены оценки быстроты сходимости и погрешности.

Литература

1. Бойков И.В. Приближенные методы вычисления сингулярных и
2. гиперсингулярных интегралов. Часть первая. Сингулярные интегралы. Пенза: Издательство Пензенского государственного университета. 2005. 360 с.
3. Бойков И.В. Приближенные методы вычисления сингулярных и
4. гиперсингулярных интегралов. Часть вторая. Гиперсингулярные интегралы. Пенза: Издательство Пензенского государственного университета. 2009. 252 с.
5. Вайникко Г.М., Лифанов И.К., Полтавский Л.Н. Численные методы в гиперсингулярных интегральных уравнениях и их приложения. М.: Янус-К. 2001. 508 с.
6. И.В. Бойков Приближенное решение сингулярных интегральных уравнений.Пенза: Издательство Пензенского государственного университета. 2004. 316 с.
7. Адамар Ж. Задача Коши для линейных уравнений с частными производными гиперболического типа. – М.: Наука, 1978. – 352 с.
8. Чикин Л. А. Особые случаи краевой задачи Римана и сингулярные интегральные уравнения // Ученые записки Казанского гос. ун-та. – 1953. – Т. 113. – № 10. – С. 53–105.
9. Boykov I.V., Ventsel E.S., Boykova A.I. An approximate solution of hypersingular integral equations// Applied Numerical Mathematics. 60 (2010) P. 607-628.

Применение современных моделей турбулентности к расчету аэродинамических характеристик профиля при умеренных числах рейнольдса

И.Н. Варюхина
Московский физико-технический институт (ГУ), г. Жуковский

1. Введение. В настоящее время все более широкое применение находят наблюдательные комплексы, построенные на основе беспилотных летательных аппаратов (БЛА). БЛА, входящие в состав таких комплексов, имеют небольшие размеры (хорда крыла 0,2–0,5 м) и летают со скоростями 15-50 м/с. Соответствующие этим режимам числа Рейнольдса лежат в диапазоне 2×10^5 – $1{,}5\times10^6$. Обтекание при числе Рейнольдса до 1 млн характеризуется заметным влиянием ламинарно-турбулентного перехода. Такие режимы не могут рассматриваться как обтекание развитым турбулентным потоком летального аппарата и требуют привлечения моделей ламинарно-турбулентного перехода (ЛТП).

Целью работы является изучение возможностей и верификация модели турбулентности SST, дополненной моделью ЛТП Лангтри-Ментера, реализованной в комплексе ANSYS CFX 11, в задаче расчета аэродинамических характеристик (АДХ) крыльевого профиля E387.

2. Постановка задачи. Рассматривается обтекание типичного планерного профиля E387, изображенного на рис. 1, при условиях, соответствующих условиям экспериментов, описанных в работе [4]. Число Рейнольдса Re = 300000, уровень турбулентности набегающего потока T_u < 0,2%, хорда профиля c = 0,3 м, скорость набегающего потока V = 14,95 м/с. Основной результат решения задачи заключается в определении коэффициентов сил и моментов и структуры поля течения (границ «ламинарного пузыря») в сравнении с экспериментальными данными.

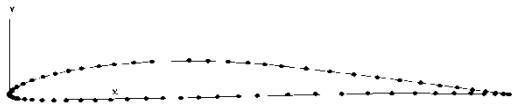


Рис.1. Форма профиля E387.

3. Теоретическая часть. В настоящее время создано большое количество разнообразных моделей для расчёта турбулентных течений. В данном случае выбрана модель турбулентности SST (shear-stress transport), так как она более устойчива во внутренней части пограничного слоя, лучше учитывает влияние сжимаемости и неблагоприятного градиента давления (отрыв). В своей структуре она содержит два уравнения переноса для *K* (кинетической энергии турбулентности)и ω(характерной частоты пульсаций).

Модель для описания ЛТП строится на двух уравнениях переноса для перемежаемости (intermittency) γ и числа Рейнольдса по толщине потери импульса Re_θ. Эта модель была описана в работе [3] и называется γ-Re_θ моделью.

Модель ЛТП модифицирует модель турбулентности SST только для уравнение переноса *K,* для ω уравнение переноса остается каким же.

4. Параметры расчета. Расчетное исследование проводилось с помощью комплекса ANSYS CFX v.11 (12). Расчетная сетка строилась в комплексе Pointwise Gridgen v. 15.13. Для сравнения результатов использовалась программа XFoil (версия XFLR5_v4.17). Форма профиля представлена 80 точками. Расчетная сетка C-типа должна быть достаточно подробной как по нормали к стенке, так и вдоль стенки. Это

связано с тем, что должно быть обеспечено хорошее разрешение ламинарного пограничного слоя от носка профиля, а также границ «ламинарного пузыря» на поверхности профиля.

5. Результаты расчета. Полученные в расчете интегральные аэродинамические характеристики в сравнении с экспериментальными данными представлены на рис. 2 а,б

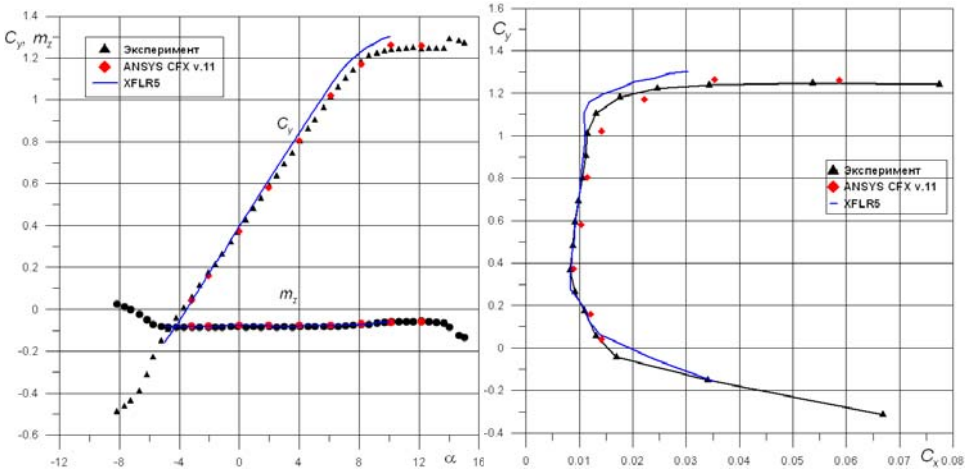


Рис.2. АДХ профиля E387

XFoil дает несколько завышенное значение производной $dC_y/d\alpha$, что приводит к расхождению расчетной и экспериментальной поляр при углах атаки $\alpha > 6°$.

Получено поле скорости и поле давления около профиля при $\alpha = 4°$.

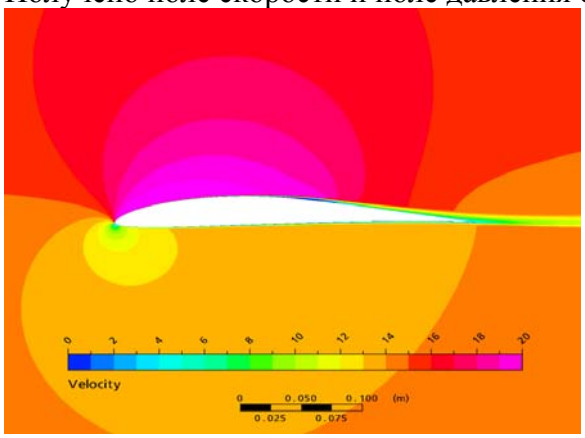


Рис.3.поле скорости при $\alpha = 4°$.

На рис. 4 представлены зависимость положения границ отрывной зоны («ламинарного пузыря») от угла атаки при $\alpha = 0°$. Границы отрывной зоны определялись по поведению напряжения трения на стенке.

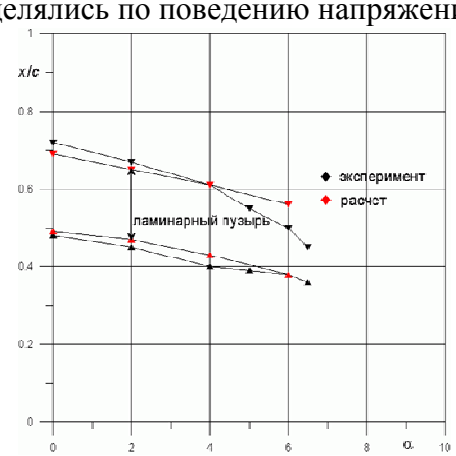


Рис.4. Положение границ отрывной зоны

6. Выводы

Модель ЛТП Лангтри-Ментера корректно воспроизводит структуру течения около профиля E387 при умеренных числах Рейнольдса и малых (крейсерских) углах атаки. При увеличении угла атаки точность полуэмипирической модели и точность численного расчета падают. Анализ методики экспериментального исследования течений с «ламинарным пузырем» в аэродинамических трубах показывает, что она также нуждается в уточнении при малых размерах пузыря.

Литература

1. Павлушенко М.И., Евстафьев Г.М., Макаренко И.К. *Беспилотные летательные аппараты: история, применение, угроза распространения и перспективы развития* // Научные записки ПИР-центра: национальная и глобальная безопасность. 2004. № 2.
2. Selig M.S., Gopalarathnam A., Giguere P., Lyon C.A. *Systematic Airfoil Design Studies at Low Reynolds Numbers* // In book: "Fixed and Flapping Wing Aerodynamics for Micro Air Vehicle Applications", T.J. Mueller (Editor). AIAA Progress in Astronautics and Aeronautics, Vol. 195. Reston, VA, 2001. P. 143-167.
3. Langtry R.B., Menter F.R.*Correlation-Based Transition Modeling for Unstructured Parallelized Computational Fluid Dynamics Codes* // AIAA Journal. 2009. V. 47. N 12. P. 2894 – 2906.
4. Selig M.S., McGranahan B.D. *Wind Tunnel Aerodynamic Tests of Six Airfoils for Use on Small Wind Turbines* // NREL Report. 2004. SR-500-34515.
5. Blazek J. *Computational Fluid Dynamics: Principles and Applications*. – Elsevier, 2001.
6. Cebeci T. *Analysis of Turbulent Flows*. – Elsevier, 2004.

Конечно-элементный подход к расчету ресурсной прочности огневой стенки камеры двигателя 14Д23

С.А. Владимиров, С.И. Трефилов
ЦНИИМАШ, Королёв
Московский физико-технический институт (ГУ), Долгопрудный

Объектом исследования является камера двигателя 14Д23 (рисунок 1). Цель работы – прочностная оценка мероприятий по повышению ресурсной прочности паяного соединения между внутренней оболочкой и наружной стенкой сопла верхнего. Для анализа степени нагружённости создана осесимметричная модель, по которой выполнены два типа расчетов: определение температурных полей и оценка напряжений и деформаций. Для расчета напряженного состояния используются два типа элементов: осесимметричные конечные элементы для описания поведения осесимметричных деталей камеры и элементы плоско-напряжённого состояния для ребер рубашки охлаждения. Главными факторами, определяющими ресурсную прочность огневой стенки, являются тяга камеры, межрубашечное давление охладителя, температурные поля. Перед расчетом накопления пластических деформаций от пуска к пуску выполнено моделирование стационарного температурного поля в деталям камеры. При решении стационарной задачи теплопроводности используются только температурные граничные условия. Результаты решения задачи теплопроводности используются далее в качестве периодической , повторяющейся от пуска к пуску, тепловой нагрузки при решении прочностной задачи. Механические характеристики материалов предполагаются упруго-пластическими и имеют зависимость от температуры. Накопление деформаций моделируется кинематической теорией пластичности. При расчёте малоцикловой прочности кинематическая теория имеет преимущество перед изотропной, поскольку позволяет рассчитывать малоцикловую прочность конструкции в запас.

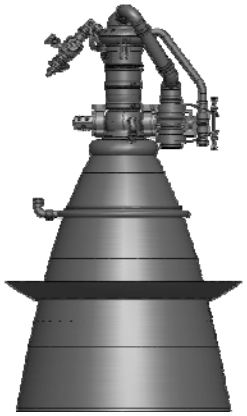


Рисунок 1

Тестирование модели выполнено на задаче однократного нагружения. Результаты хорошо совпадают с результатами, полученными на более простых моделях.

С использованием разработанной математической модели камеры двигателя 14Д23 решена термопрочностная задача и определены запасы малоцикловой прочности огневой стенки для штатной и усиленной конструкции. Расчёты показали снижение уровня напряжений в опасных местах усиленной конструкции.

Определение условий неустойчивости типа жесткой моды и анализ уравнения для параметров порядка в модели морфогенеза Гирера-Майнхардта во внешнем случайном поле

К.А. Головкин

Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Теоретическое исследование морфогенеза относится к наиболее важным аспектам изучения процессов индивидуального развития. Современное моделирование морфогенеза основано на идее о том, что благодаря диффузии и реакции отдельных химических веществ образуется прообраз структуры, так называемое «морфогенное поле». Оно влияет на гены, вызывая дифференциацию клеток. Подобные модели частично подтверждаются прямыми наблюдениями определенных химических веществ, например, фактора роста нейронов.

Модели морфогенеза Тьюринга [1] и Гирера – Майнхардта [2] показали, что нелинейная автокаталитическая диссипативная система, «дальнодействующая» сигнализация в которой обеспечивается диффузией, оказывается способной к пространственной и временной самоорганизации, то есть к образованию диссипативных структур (ДС). Эти модели не утратили своего значения и в настоящее время, так как форма учета дальнодействия членами диффузионного типа является достаточно общей. Она охватывает не только диффузию, но и другие процессы, в частности, взаимодействие за счет упругих сил [3].

В работе [3] отмечается, что за счет стохастического характера ферментативных реакций уровень флуктуаций «выхода» реакции может быть достаточно большим. Флуктуации вызываются различными причинами. В частности, концентрация молекул может изменяться из-за тепловой и турбулентной диффузии через виртуальные или вполне реальные стенки элементарного объема (например, оболочки клеток, внутриклеточные мембраны). Необходимо также учитывать флуктуации за счет стохастического характера взаимодействия между молекулами [3]. В этом случае динамические переменные модели являются случайными функциями.

В данной работе исследуется влияние аддитивного однородного изотропного поля гауссовых флуктуаций динамических переменных на образование ДС при мягком режиме их возбуждения в модели Гирера - Майнхардта.

Исследование появления неустойчивости типа жесткой моды

Запишем модель Гирера-Майнхардта в следующей форме:

$$\begin{aligned} \frac{\partial a}{\partial t} &= \rho + k\frac{a^2}{h} - \mu a + D_a \Delta a + F_1(\vec{r},t); \\ \frac{\partial h}{\partial t} &= ca^2 - \nu h + D_h \Delta h + F_2(\vec{r},t). \end{aligned} \qquad (1)$$

Здесь ρ – скорость образования автокаталитической переменной, μ – ее постоянная распада, D_a – коэффициент диффузии активатора; ν – постоянная скорости распада демпфирующей переменной, член ca^2 задает скорость образования ингибитора, D_h – коэффициент диффузии ингибитора. Случайные функции $F_i(\vec{r},t)$ определяют пространственные и временные гауссовы флуктуации скорости изменения концентраций активатора и ингибитора.

В данной работе с использованием концепции параметра порядка аналитически исследовано поведение системы (2).

Вводя новые переменные: $\vec{r}' = \vec{r}\sqrt{\nu / D_a}$; $t' = \nu t$; $a' = ca / k$; $h' = \nu ch / k^2$, получим

$$\begin{aligned} \frac{\partial a'}{\partial t} &= \rho' + \frac{a'^2}{h'} - \mu' a' + \Delta a' + F_1'(\vec{r}', t'); \\ \frac{\partial h'}{\partial t} &= a'^2 - h' + D' \Delta h' + F_2'(\vec{r}', t'), \end{aligned} \tag{2}$$

где $\rho' = \rho c / \nu k$; $\mu' = \mu / \nu$; $D' = D_h / D_a$.

В дальнейшем штрихи будем опускать. Чтобы учесть влияние случайного воздействия, перепишем (4) в виде:

$$\left(\frac{\partial}{\partial t} - K(\Delta) \right) \vec{q} = \vec{g} + \vec{F}, \tag{3}$$

где

$$K(\Delta) = \begin{pmatrix} \mu\left(\frac{2}{\rho + 1} - 1 \right) & \frac{-\mu^2}{(\rho+1)^2} \\ \frac{2}{\mu}(\rho + 1) & -1 + D\Delta \end{pmatrix};$$

$$\vec{q} = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}; \qquad q_1 = a - a_0, \quad q_2 = h - h_0.$$

q_1, q_2 описывают малые отклонения концентраций относительно стационарных состояний $h_0 = (\rho+1)^2 / \mu^2; a_0 = (\rho+1)/\mu$, а вектор $\vec{g}$ содержит квадратичные и кубические нелинейности, полученные разложением в ряд правой части (3) и определяется так:

$$\vec{g} = \left((q_1^2 / h_0 - \left(q_1^2 q_2 + 2a_0 q_1 q_2 \right) / h_0^2 + \left(2a_0 q_1 q_2^2 + a_0^2 q_2^2 \right) / h_0^3 - a_0^2 q_2^3 / h_0^4), q_1^2 \right).$$

Показано, что неустойчивость типа мягкой моды возникает при двух критических значениях волновых чисел $k_{кр} = \pm \sqrt[4]{\mu / D}$.

Представим вектор $\vec{q}$ в виде

$$q(x,t) = \exp(ikx + \lambda t). \tag{4}$$

Решение соответствующего характеристического уравнения дает

$$\lambda_{1,2} = \frac{\alpha(k)}{2} \pm \sqrt{\frac{\alpha^2(k)}{4} - \beta(k)}, \tag{5}$$

где $\alpha(k) = -(D+1)k^2 + 2\mu/(\rho+1) - \mu - 1$; $\beta(k) = (k^2 + \mu)(Dk^2 + 1) - 2\mu Dk^2/(\rho+1)$.

Условие появления неустойчивости типа жесткой моды имеет вид

$$\mathrm{Re}\{\lambda_{1,2}\} \geq 0, \ \mathrm{Im}\{\lambda_{1,2}\} \neq 0. \tag{6}$$

Таким образом можно получить следующее:

$$\begin{aligned} \rho &\leq \frac{2\mu}{(\mu+1) + (D+1)k^2} - 1; \\ \rho &> \frac{2\mu D k^2}{(1 + Dk^2)(\mu + k^2)} - 1. \end{aligned} \tag{7}$$

Чтобы сначала возникла неустойчивость типа жесткой моды, необходимо потребовать, чтобы

$D > 2\mu + 1 + 2\sqrt{\mu + 1}$

Рассмотрим уравнение (3) и представим вектор $\vec{q}$ в виде суперпозиции

$$\vec{q}(\vec{r},t)=\sum_{\vec{k},j}\vec{O}^{(j)}\xi_{\vec{k}}^{(j)}(t)e^{i\vec{k}\vec{r}}, \tag{8}$$

где $\vec{O}^{(j)}$ - собственные векторы линейного оператора *K(Δ);* $\vec{O}^{*(j)}$ - собственные вектора оператора, сопряженного к $K(\nabla^2)$.

$$\vec{O}^{(j)}=\begin{pmatrix}1\\ \dfrac{2(\rho+1)}{\mu\left(1+Dk^2+\lambda_j\right)}\end{pmatrix},\quad \vec{O}^{*(j)}=\begin{pmatrix}(-1)^j\dfrac{1+Dk^2+\lambda_j}{\lambda_2-\lambda_1}\\ (-1)^j\dfrac{\mu\left(1+Dk^2+\lambda_1\right)\left(1+Dk^2+\lambda_2\right)}{2(\rho+1)(\lambda_1-\lambda_2)}\end{pmatrix}.$$

Неустойчивые моды заключены в узкой полосе значений волновых векторов, что дает возможность построения волновых пакетов путем суммирования по волновым векторам, заключенным в малых интервалах [5]. Таким образом, получаем несущие моды с дискретными значениями волновых векторов и медленно меняющимися амплитудами $\xi_{\vec{k}}^{(j)}(t)$.

После преобразований система уравнений (3) для амплитуд мод $\xi_{\vec{k}}^{(j)}(t)$ принимает вид:

$$\frac{d\xi_{\vec{k}}^{(j)}}{d\tau}-\lambda_j(\vec{k})\xi_{\vec{k}}^{(j)}=\sum_{jj'',\vec{k}'\vec{k}''}\sigma_{jj''}^{(j)}(\vec{k},\vec{k}',\vec{k}'')\xi_{\vec{k}'}^{(j')}\xi_{\vec{k}''}^{(j'')}\delta(\vec{k}'+\vec{k}''-\vec{k})+$$

$$+\sum_{j'j''j''',\vec{k}'\vec{k}''\vec{k}'''}\sigma_{j'j''j'''}^{(j)}(\vec{k},\vec{k}',\vec{k}'',\vec{k}''')\xi_{\vec{k}'}^{(j')}\xi_{\vec{k}''}^{(j'')}\xi_{\vec{k}'''}^{(j''')}\delta(\vec{k}'+\vec{k}''+\vec{k}'''-\vec{k})+\sum_{\varepsilon=1}^{2}O_{\varepsilon}^{*(j)}(\vec{k})z_{\varepsilon,\vec{k}}(\tau).$$

где

$$\sigma_{jj''}^{(j)}(\vec{k},\vec{k}',\vec{k}'')=\frac{\mu^2\overline{O}_{1,k}^{(j)}}{(1+\rho)^2}\left\{\sum_{\vec{k}',\vec{k}''}\left(O_{1,k'}^{(1)}\xi_{\vec{k}'}^{(1)}+O_{1,k'}^{(2)}\xi_{\vec{k}'}^{(2)}\right)\left(O_{1,k''}^{(1)}\xi_{\vec{k}''}^{(1)}+O_{1,k''}^{(2)}\xi_{\vec{k}''}^{(2)}\right)+\right.$$

$$\left.+\frac{\mu^2}{(1+\rho)^2}\sum_{\vec{k}',\vec{k}''}\left(O_{2,k'}^{(1)}+O_{2,k'}^{(2)}\right)\left(O_{2,k''}^{(1)}+O_{2,k''}^{(2)}\right)-\frac{2\mu}{(1+\rho)}\sum_{\vec{k}',\vec{k}''}\left(O_{1,k'}^{(1)}+O_{1,k'}^{(2)}\right)\left(O_{2,k''}^{(1)}+O_{2,k''}^{(2)}\right)\right. \tag{9}$$

$$\sigma_{j'j''j'''}^{(j)}(\vec{k},\vec{k}',\vec{k}'',\vec{k}''')=-\frac{\mu^2}{(1+\rho)^2}\sum_{\vec{k}',\vec{k}'',\vec{k}'''}\left(O_{1,k'}^{(1)}+O_{1,k'}^{(2)}\right)\left(O_{1,k''}^{(1)}+O_{1,k''}^{(2)}\right)\times$$

$$\times\left(O_{2,k'''}^{(1)}+O_{2,k'''}^{(2)}\right)+\frac{2\mu^3}{(1+\rho)^3}\sum_{\vec{k}',\vec{k}'',\vec{k}'''}\left(O_{2,k'}^{(1)}+O_{2,k'}^{(2)}\right)\left(O_{2,k''}^{(1)}+O_{2,k''}^{(2)}\right)\left(O_{1,k'''}^{(1)}+O_{1,k'''}^{(2)}\right)-$$

$$-\frac{\mu^4}{(1+\rho)^4}\sum_{\vec{k}',\vec{k}'',\vec{k}'''}\left(O_{2,k'}^{(1)}+O_{2,k'}^{(2)}\right)\left(O_{2,k''}^{(1)}+O_{2,k''}^{(2)}\right)\left(O_{2,k'''}^{(1)}+O_{2,k'''}^{(2)}\right)$$

$$z_{\varepsilon,\vec{k}}(\tau)=\int F_{\varepsilon}(\vec{r},\tau)e^{-i\vec{k}\vec{r}}d\vec{r}\,.$$

Функции $\delta(\vec{k}'+\vec{k}''-\vec{k})$ и $\delta(\vec{k}'+\vec{k}''+\vec{k}'''-\vec{k})$ дают «правила отбора», которые важны для определения возникающих пространственных структур.

Система (20) содержит как устойчивые, так и неустойчивые моды. Необходимо исключить нежелательные или ненужные затухающие моды.

Проведя процедуру адиабатического исключения и усреднив выражение, окончательно получим

$$\frac{\partial\left\langle \xi_{\vec{k}u}^{(j)}\right\rangle}{\partial\tau}-\lambda_j(\vec{k}_u)\left\langle \xi_{\vec{k}u}^{(j)}\right\rangle=\frac{1}{2}\sum_{\varepsilon\varepsilon'\varepsilon''\varepsilon''',\vec{k}'u} g_{\varepsilon'\varepsilon'}\left(\left|\frac{\vec{k}_u-\vec{k}'_u}{2}\right|\right)\rho_{\varepsilon\varepsilon'}^{(j)}(\vec{k}_u,\frac{\vec{k}_u+\vec{k}'_u}{2},\frac{\vec{k}_u-\vec{k}'_u}{2})\times$$

$$\times\rho_{\varepsilon''\varepsilon'''}^{(\varepsilon)}(\frac{\vec{k}_u+\vec{k}'_u}{2},\vec{k}'_u,\frac{\vec{k}_u-\vec{k}'_u}{2})\left\langle \xi_{\vec{k}''u}^{(\varepsilon'')}\right\rangle+$$

$$+\sum_{\varepsilon\varepsilon',\vec{k}'u\vec{k}''u}\sigma_{\varepsilon\varepsilon'}^{(j)}(\vec{k}_u,\vec{k}'_u,\vec{k}''_u)\left\langle \xi_{\vec{k}'u}^{(\varepsilon)}\xi_{\vec{k}''u}^{(\varepsilon')}\right\rangle\delta(\vec{k}'_u+\vec{k}''_u-\vec{k}_u)+ \tag{10}$$

$$+\sum_{\varepsilon\varepsilon',\vec{k}'u\vec{k}''u\vec{k}'''u}[\sigma_{\varepsilon\varepsilon\varepsilon}^{(j)}(\vec{k}_u,\vec{k}'_u,\vec{k}''_u,\vec{k}'''_u)\delta_{\varepsilon,\varepsilon'}+\omega_{\varepsilon\varepsilon'}^{(j)}(\vec{k}_u,\vec{k}'_u,\vec{k}''_u,\vec{k}'''_u,\vec{k}_s)]\times$$

$$\times\left\langle \xi_{\vec{k}'u}^{(\varepsilon)}\xi_{\vec{k}''u}^{(\varepsilon')}\xi_{\vec{k}'''u}^{(\varepsilon')}\right\rangle\delta(\vec{k}'_u+\vec{k}''_u+\vec{k}'''_u-\vec{k}_u),$$

где

$$\lambda_j=\lambda_j(\vec{k}_u)+\frac{1}{2}\sum_{\varepsilon\varepsilon'\varepsilon''\varepsilon'''}g_{\varepsilon'\varepsilon'}(0)\rho_{\varepsilon\varepsilon'}^{(j)}(\vec{k}_u,\vec{k}_u,0)\rho_{\varepsilon''\varepsilon'''}^{(\varepsilon)}(\vec{k}_u,\vec{k}_u,0)\,. \tag{11}$$

Из выражения (11) можно показать, что для собственных чисел λ действительная часть пропорциональна интенсивности флуктуаций.

Таким образом, в модели Гирера - Майнхардта, вблизи маргинального состояния достаточно сильные флуктуации динамических переменных могут привести к увеличению области неустойчивых мод. Показано, что действительная часть собственных чисел неустойчивых мод пропорциональна интенсивности флуктуаций и некоторой степени радиуса корреляции, зависящей от размерности пространства реакционно-диффузионной системы.

Приведенный анализ будет качественно справедлив и для других распределенных моделей типа реакция - диффузия.

Заключение

Была рассмотрена модель Морфогенеза Гирера-Майнхардта. Рассмотренная модель имеет весьма широкую область применения. В результате работы были получены следующие основные результаты:

1. Исследованы условия возникновения неустойчивости типа жесткой моды в детерминированной системе Гирера-Майнхардта.

2. Получены уравнения для параметров порядка в поле аддитивных флуктуаций.

3. Исследовано влияние внешнего случайного поля на инкременты неустойчивых мод. Показано, что при наличии внешних случайных сил инкременты неустойчивых мод увеличивается.

Литература

1. Meinhardt H., Gierer A. Generation and regeneration of sequences of structures during morphogenesis. J. Theor. Biol., 1980.
2. Keener I.P. Activaters and ingibitors in pattern formation. Studies and Applied Mathematics, 1978.
3. Turing A.M. The chemical basis of the morphogenesis. Proc. Roq. Soc., B, 1952.
4. Романовский Ю.М., Степанова Н.В., Чернавский Д.С. Математическое моделирование в биофизике (Введение в теоретическую биофизику). М.: ИКИ, 2004. 472 с.
5. Хакен Г. Синергетика. М.: Мир, 1980. 406 с.

Исследование погрешности численного моделирования процесса прохождения электромагнитного излучения в волноводе с распределенным коэффициентом преломления

А.А. Дегтярев, Е.С. Козлова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Рассмотрим обобщенное волновое уравнение для напряженности электрического поля $\vec{E}$ для изотропной, неоднородной среды в предположении отсутствия токов и распределенных по объему зарядов [1]:

$$-\nabla\left[\ln\mu\right]\times\left[\nabla\times\vec{E}\right]-\nabla\left[\nabla\left[\ln\varepsilon\right]\cdot\vec{E}\right]-\Delta\vec{E}+\varepsilon\mu\frac{\partial^2\vec{E}}{\partial t^2}=0. \quad (1)$$

где $\vec{E}$ – напряженность электрического поля, μ – магнитная проницаемость, ε – диэлектрическая проницаемость, t – время.

Используем обозначения:

$$\mu_0=\frac{\mu}{\mu_0},\quad \varepsilon_0=\frac{\varepsilon}{\varepsilon_0}, \quad (2)$$

где μ_0 – относительная магнитная проницаемость, μ_0 – магнитная постоянная, ε_0 – относительная диэлектрическая проницаемость, ε_0 – электрическая постоянная.

Считая среду не поглощающей, представим диэлектрическую проницаемость в виде:

$$\sqrt{\varepsilon_0}=n(x,y)=\overline{n}+h(x,y) \quad (3)$$

где n – коэффициент преломления, $\overline{n}$ – среднее значение показателя преломления, а $h(x,y,z)$ – уклонение от среднего значения.

Введем обозначения:

$$k=\frac{2\pi n}{\lambda},\quad \overline{k}=\frac{2\pi\overline{n}}{\lambda}, \quad (4)$$

где λ – длина волны в среде.

Рассмотрим монохроматическую, гармоническую волну вида:

$$\vec{E}(x,y,z,t)=\vec{\mathrm{E}}(x,y,z)e^{i(\omega t+\overline{k}z)} \quad (5)$$

где ω – угловая частота.

Подставив выражение (5) в уравнение (1), получим следующее уравнение:

$$-\omega^2 n^2\varepsilon_0\mu_0\vec{\mathrm{E}}-\frac{\partial^2\vec{\mathrm{E}}}{\partial z^2}-2i\overline{k}\frac{\partial\vec{\mathrm{E}}}{\partial z}+\overline{k}^2\vec{\mathrm{E}}-\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\vec{\mathrm{E}}-\nabla\left[\nabla\left[\ln n^2\right]\cdot\vec{\mathrm{E}}\right]=0. \quad (6)$$

Предположим, что вдоль оси z амплитуда меняется медленнее, нежели в перпендикулярном этой оси сечении:

$$\left|\frac{\partial^2\vec{\mathrm{E}}}{\partial z^2}\right|=2\overline{k}\left|\frac{\partial\vec{\mathrm{E}}}{\partial z}\right|. \quad (7)$$

Тогда с учетом вышеприведенного предположения (7) и соотношений (2) – (4) мы получим однонаправленное уравнение Гельмгольца:

$$\frac{\partial \vec{\mathrm{E}}}{\partial z} = \frac{\imath}{2\bar{k}}\left\{\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\vec{\mathrm{E}} + \left(k^2 - \bar{k}^2\right)\vec{\mathrm{E}} + \nabla\left[\nabla\left[\ln \varepsilon\right]\cdot \vec{\mathrm{E}}\right]\right\}. \quad (8)$$

В случае ТЕ-поляризации оно примет вид:

$$\frac{\partial \mathrm{E}_z}{\partial z} = \frac{\imath}{2\bar{k}}\left[\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\mathrm{E}_z + \left(k^2 - \bar{k}^2\right)\mathrm{E}_z\right]. \quad (9)$$

Для удобства записи заменим E_z на E. Теперь, добавив к данному уравнению начальные и граничные условия, мы получим математическую модель описания процесса распространения электромагнитного излучения в волноводе прямоугольного сечения [2,3]:

$$\begin{cases} \dfrac{\partial \mathrm{E}}{\partial z} = \dfrac{\imath}{2\bar{k}}\left[\Delta_{xy}\mathrm{E} + \left(k^2 - \bar{k}^2\right)\mathrm{E}\right], \\ x \in [0, l_x], \quad y \in [0, l_y], \quad z \in [0, L]; \\ \mathrm{E}\big|_{z=0} = \psi(x, y), \qquad x \in [0, l_x], \quad y \in [0, l_y]; \\ \mathrm{E}\big|_{x=0} = \mathrm{E}\big|_{x=l_x} = 0, \quad y \in (0, l_y), \quad z \in (0, L]; \\ \mathrm{E}\big|_{y=0} = \mathrm{E}\big|_{y=l_y} = 0, \quad x \in (0, l_x), \quad z \in (0, L], \end{cases} \quad (10)$$

где $\Delta_{xy} = \dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}$ – оператор Лапласа, l_x, l_y, L – ширина, толщина и длина волновода соответственно, $\psi(x, y)$ – распределение амплитуды на входе в волновод. Для удобства будем считать, что грани волновода покрыты тонкой оболочкой, идеально проводящей излучение.

Рассмотрим частный случай, когда профиль показателя преломления волновода является параболическим:

$$n(x) = \bar{n} + \frac{h - \bar{n}}{\bar{n}} \cdot \frac{x^2}{a^2},$$

где $\bar{n}$, h – показатели преломления среды на оси волновода и на гранях волновода соответственно; a - ширина области с неоднородным распределением показателя преломления.

Из теории известно, что в данном случае уравнение решается аналитически, а решением являются моды Гаусса-Эрмита [3]. Таким образом, есть возможность тестирования работы программы на конкретном примере.

Для решения задачи (10) предлагается использовать схему Писмена-Рэкфорда:

$$\begin{cases} \dfrac{\mathrm{E}_{ij}^{s+1/2} - \mathrm{E}_{ij}^{s}}{0,5h_z} = \dfrac{\imath}{2\bar{k}}\left[\Lambda_{h_x}\mathrm{E}_{ij}^{s+1/2} + \Lambda_{h_y}\mathrm{E}_{ij}^{s} + \left(k_{ij}^2 - \bar{k}^2\right)\mathrm{E}_{ij}^{s+1/2}\right], \quad i = \overline{1, I-1}; \quad j = \overline{1, J-1}; \\ \dfrac{\mathrm{E}_{ij}^{s+1} - \mathrm{E}_{ij}^{s+1/2}}{0,5h_z} = \dfrac{\imath}{2\bar{k}}\left[\Lambda_{h_x}\mathrm{E}_{ij}^{s+1/2} + \Lambda_{h_y}\mathrm{E}_{ij}^{s+1} + \left(k_{ij}^2 - \bar{k}^2\right)\mathrm{E}_{ij}^{s+1/2}\right], \quad j = \overline{1, J-1}; \quad i = \overline{1, I-1}; \\ s = \overline{0, S-1}; \\ \mathrm{E}_{ij}^{s}\big|_{s=0} = \psi_{ij}, \qquad i = \overline{0, I}, \quad j = \overline{0, J}; \\ \mathrm{E}_{ij}^{s}\big|_{i=0} = \mathrm{E}_{ij}^{s}\big|_{i=I} = 0, \qquad s = \overline{1, S}, \quad j = \overline{1, J-1}; \\ \mathrm{E}_{ij}^{s}\big|_{j=0} = \mathrm{E}_{ij}^{s}\big|_{j=J} = 0, \qquad s = \overline{1, S}, \quad i = \overline{1, I-1}, \end{cases} \quad (11)$$

где $\Lambda_{h_x} \mathrm{E}_{ij}^s = \dfrac{\mathrm{E}_{i+1j}^s - 2\mathrm{E}_{ij}^s + \mathrm{E}_{i-1j}^s}{h_x^2}$, $\Lambda_{h_y} \mathrm{E}_{ij}^s = \dfrac{\mathrm{E}_{ij+1}^s - 2\mathrm{E}_{ij}^s + \mathrm{E}_{ij-1}^s}{h_y^2}$ – разностные операторы Лапласа по переменным x и y соответственно.

Конечно-разностная схема Писмена-Рэкфорда (11) при любых шагах является устойчивой.

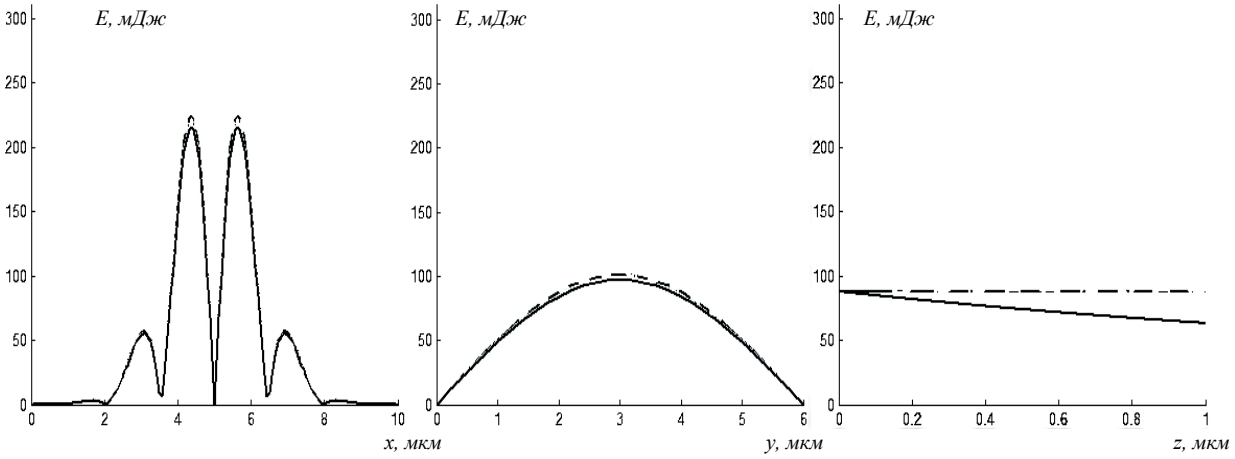
Для исследования погрешности численного решения используем тестовый пример с начальным условием вида:

$$\psi(x, y) = H_9(x) \sin\left(\frac{\pi}{l_y} y\right), \qquad (12)$$

где – $H_9(x)$ мода Гаусса – Эрмита девятого порядка. Данная функция является собственной для оператора дифференциального уравнения системы (10). Таким образом, при прохождении по волноводу, она должна сохранять свой вид.

На Рисунке 1 приведены графики изменения интенсивности комплекснозначной амплитуды вдоль осей, полученные с помощью разработанного комплекса. Пунктирной линией обозначено аналитическое, а непрерывной численное. Расчеты были произведены при следующих значениях параметров:

$\lambda = 0.63$мкм; $n_1 = 1,2$; $n_2 = 1,192$; $l_x = 10$мкм; $l_y = 6$ мкм; $L = 1$ мкм; $I = 160$; $J = 30$; $K = 1000$.



а) **б)** **в)**

Рисунок 1. Графики интенсивности вдоль осей: а) *Ox , б) Oy, в) Oz.*

Заметим что, численное и аналитическое решение близки. Ошибка численного моделирования составляет величину порядка 6-8%.

На Рисунке 2 приведен график зависимости рассогласования численного и аналитического решений от шага дискретизации, которое вычислялось в равномерной (верхний график) и среднеквадратической (нижний график) нормах:

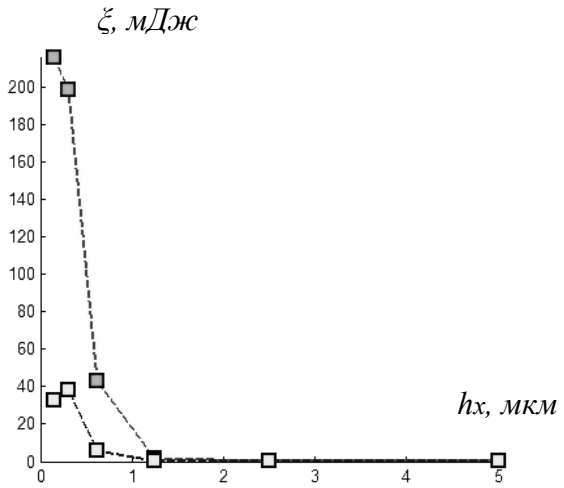


Рисунок 2 График погрешности

Так же значения величины рассогласования решений приведены в Таблице 3.

Таблица 1 Значения погрешности ξ

ξ_p, мДж	ξ_c, мДж	h_x, h_y, мкм	h_z, мкм
220	35	0,156	0,003
200	40	0,313	0,006
46	10	0,625	0,012
5	4	1,25	0,024
1	0,7	2,5	0,048
0,05	0,02	5	0,096

Возникающие погрешности можно объяснить воздействием на численное решение схемы (11) граничных условий. Дело в том, что решение задачи (10) в виде мод Гаусса-Эрмита было получено в предположении неограниченности среды распространения. Следовательно, при наложении граничных условий мы будем получать решение, несколько отличающееся от первоначального распределения.

На Рисунке 3 представлена величина рассогласования численного и аналитического решений, связанная с наложением граничных условий в схеме (11), где a – расстояние от центра до границы волновода:

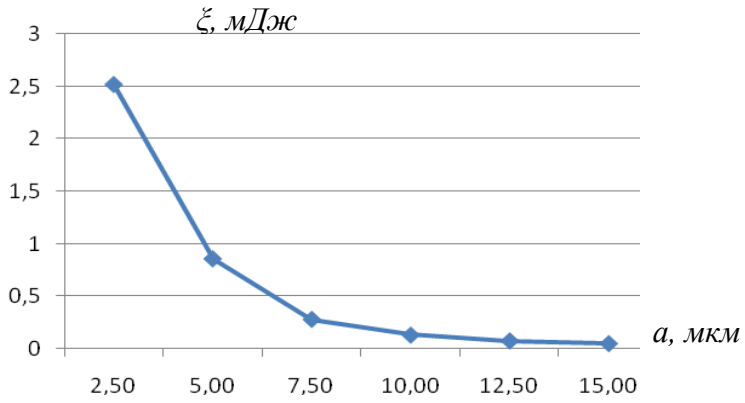


Рисунок 3 Зависимость среднеквадратической погрешности от параметров волновода

Из графика видно, что удаляя границы волновода, мы уменьшаем их воздействие на численное решение. И таким образом, численное решение задачи (10) будет все ближе подходить к решению системы (10), полученному для задачи без учета граничных условий. Различия между численным и аналитическим решением задачи (10) с удалением границ становится ничтожно малым. Так для $a = 15$ мкм рассогласование ξ можно характеризовать величиной не больше $0{,}1$ мДж, что представляет собой 0,1%.

Литература

1. Неганов В.А., Раевский С.Б., Яровой Г.П. Линейная макроскопическая электродинамика. т1. // М: Радио и связь, 2000. – 509с.
2. Виноградова М.Б., Руденко О.В., Сухоруков А.П. Теория волн // М: Наука, 1979.
3. Самарский А.А. Численные методы // М: Наука, 1958.
4. Адамас М., Кривошлыков С.Г., Черных В.А., Сисакян И.С. Введение в теорию оптических волноводов // М: Мир, 1984. – 512 с.
5. Белоусов А.А., Гаврилов А.В., Дегтярев А.А. Анализ погрешностей численного решения задачи о распространении электромагнитного излучения в радиально-симметричном волноводе // Компьютерная оптика 2003. Вып. 25. С 29-35.
6. Самарский А.А., Тихонов А.Н. Уравнения математической физики // М: Наука, 1966. – 724с.

Разработка и исследование алгоритма поиска множества решений нелинейного дифференциального уравнения эллиптического типа

А.А. Дегтярев, Г.М. Комендантян
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Предположим, что плоскопараллельный термически тонкий прозрачный оптический элемент, имеющий форму диска толщиной l и радиусом R (l<<R), облучается лазерным пучком света, который обладает круговой симметрией и падает нормально и соосно на рабочую поверхность оптического элемента.

Из-за поглощения части энергии излучения элемент нагревается, причем процесс теплопоглощения сопровождается интерференционными эффектами [1], обуславливающими его нелинейность.

Процесс теплового взаимодействия элемента с лазерным излучением описывается следующей математической моделью [2]:

$$\begin{cases} c\dfrac{\partial u}{\partial t} = k\left(\dfrac{\partial^2 u}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial u}{\partial r}\right) - \dfrac{\alpha}{l}(u) + \dfrac{A(u)}{l} I(r), \quad 0 \le t \le T, \quad 0 < r < R; \\ u\big|_{t=0} = 0, \quad 0 \le r \le R; \\ \left.\dfrac{\partial u}{\partial r}\right|_{r=0} = 0, \quad 0 \le t \le T; \\ \left.\dfrac{\partial u}{\partial r}\right|_{r=R} = 0, \quad 0 \le t \le T, \end{cases} \tag{1}$$

где $u = u(r,t)$ - превышение температуры оптического элемента над температурой окружающей среды; $r \in [0; R]$ - радиальная координата полярной системы; $t \in [0; T]$ - координата времени; c и k - коэффициенты объёмной теплоёмкости и теплопроводности материала, из которого выполнен элемент, соответственно; α - коэффициент теплоотвода с граней оптического элемента во внешнюю среду; $A(u)$ - функция поглощательной способности элемента, определяющая нелинейное поглощение элементом лазерного излучения и имеющая следующий вид [1]:

$$A(u) = \frac{b_1}{b_2 + b_3 \sin^2(\theta + \omega u)}, \tag{2}$$

где b_1, b_2, b_3, θ и ω – некоторые константы.

Таким образом, термодинамический процесс в элементе описывается нелинейным дифференциальным уравнением параболического типа. На рисунке 1 приведен график зависимости температуры от времени, полученный при помощи численного решения задачи (1) при следующих значениях параметров:

$$\begin{aligned} &k = 0.011,\ c = 1.6,\ l = 0.5, R = 4.0, \\ &b_1 = 0.1, b_2 = 1.0,\ b_3 = 3.0,\ \theta = 0.0, \\ &\omega = 3.0,\ \alpha = 0.03, \\ &I(r) = \begin{cases} \dfrac{P}{\pi a^2}, 0 \le r \le a, \\ 0, r > a, \end{cases},\ P = 10,\ a = 0.8. \end{aligned} \tag{3}$$

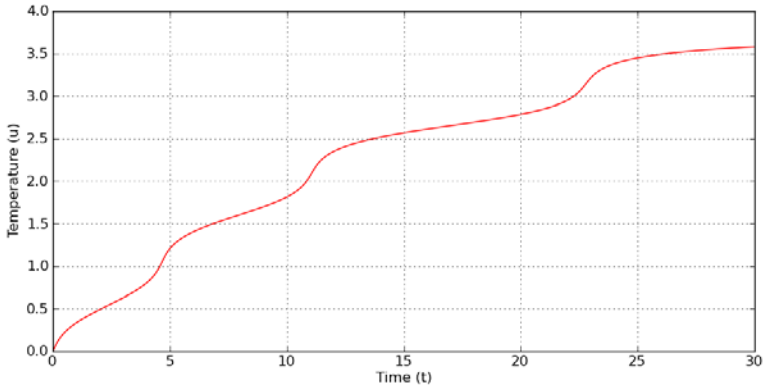


Рисунок 1 Зависимость температуры оптического элемента от времени при $r = 0$

Логично предположить, что при заданной мощности лазера тепловое поле с течением времени стремится к некоему установившемуся состоянию. Установившееся распределение температуры в системе не изменяется со временем, то есть, справедливо равенство:

$$\frac{\partial u}{\partial t} = 0, \quad 0 \le r \le R. \tag{4}$$

Следовательно, установившееся тепловое поле будет описываться следующей математической моделью:

$$\begin{cases} k\left(\dfrac{\partial^2 u}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial u}{\partial r}\right) - \dfrac{\alpha}{l}(u) + \dfrac{A(u)}{l} I(r) = 0, 0 < r < R; \\ \left.\dfrac{\partial u}{\partial r}\right|_{r=0} = 0; \quad \left.\dfrac{\partial u}{\partial r}\right|_{r=R} = 0. \end{cases} \tag{5}$$

Наиболее интересным фактом, выявленным при исследовании, является множественность решений задачи (5).

Объяснить факт неединственности решения задачи (5) можно следующим образом. Анализируя уравнение теплопроводности в системе (1) нетрудно заметить, что за теплоотвод отвечает член $Q_1 = -\dfrac{\alpha}{l}u$, а за подвод тепла к системе от лазера – $Q_2 = \dfrac{A(u)}{l} I(r)$. Функция $A(u)$ – нелинейная функция температуры, отражающая процесс нелинейности теплопоглощения, вызванного интерференцией. Очевидно, что тепловое равновесие в элементе возникает в случае равенства теплоотвода и теплопоглощения. На рисунке 2 приведены зависимости теплоотвода (линейная) и теплопоглощения (нелинейная периодическая) от температуры.

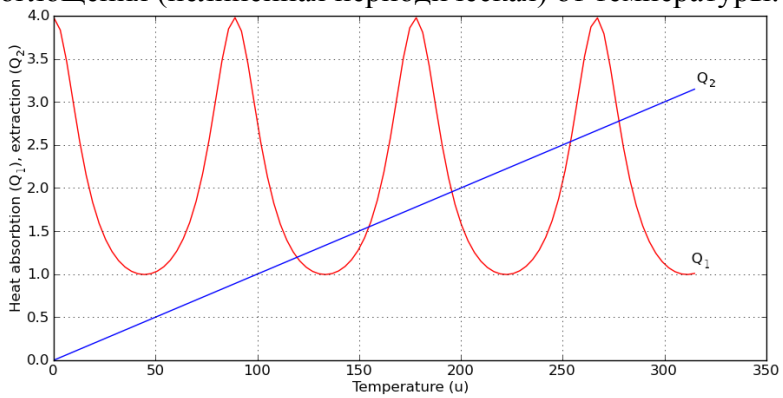


Рисунок 2. Схематичные графики теплоотвода Q_1 и теплопоглощения Q_2

Видно, что графики пересекаются в нескольких точках, которые определяют состояния равновесия теплового процесса. Каждая из этих точек определяет установившееся состояние. Устойчивыми из них являются лишь те, для которых повышение температуры понижает теплопоглощение, а понижение температуры – повышает теплопоглощение.

Для численного решения задачи (5) используем следующую итерационную разностную схему [4].

$$\begin{cases} C\dfrac{u_i^{k+1}-u_i^k}{\tau} = p\left(\dfrac{u_{i+1}^{k+1}-2u_i^{k+1}+u_{i-1}^{k+1}}{h_r^2}+\dfrac{1}{r_i}\cdot\dfrac{u_{i+1}^{k+1}-u_{i-1}^{k+1}}{2h_r}\right)-\dfrac{\alpha}{l}u_i^{k+1}+ \\ +\dfrac{A\left(u_i^k\right)}{l}I\left(r_i\right),\ i=\overline{1,I-1},\ \ k=\overline{0,K-1}; \\ C\dfrac{u_0^{k+1}-u_0^k}{\tau} = 2p\left(\dfrac{u_1^{k+1}-2u_0^{k+1}+u_{-1}^{k+1}}{h_r^2}\right)-\dfrac{\alpha}{l}u_0^{k+1}+\dfrac{A\left(u_0^k\right)}{l}\cdot I\left(r_0\right), \\ k=\overline{0,K-1}; \\ u_i^0=0,\ \ i=\overline{0,I}; \\ \dfrac{u_I^{k+1}-u_{I-1}^{k+1}}{h_r}=0,\ k=\overline{0,K-1}, \end{cases} \qquad (6)$$

u_i^k – значение температуры в i-м узле сетки $r_i = ih_r$, k – номер итерации, а τ – параметр итерационного процесса.

Фактически, схема (6) аппроксимирует задачу (1) при достаточно малых τ. При этом τ можно рассматривать как шаг по времени.

Для решения данной системы использовался метод прогонки.

На рисунке 3 при $r=0$ показан переход в установившееся состояние. Видно, что по прошествии 50-60 условных временных единиц динамический тепловой процесс устанавливается.

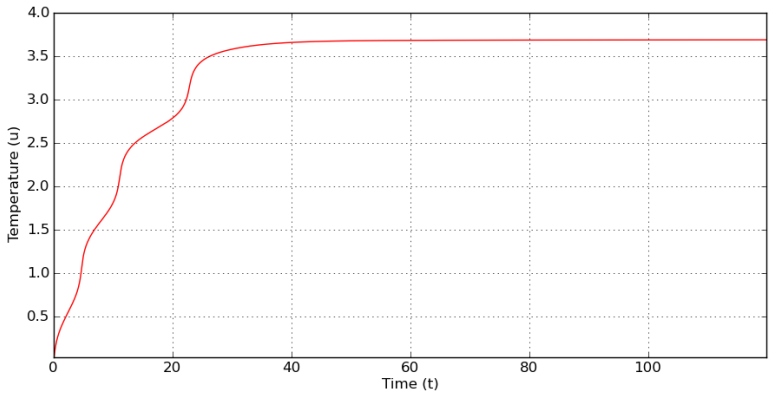


Рисунок 3. Расчет установившегося решения

Для расчета альтернативных решений задачи (5), предпримем следующие действия. Значительно (для нашего эксперимента – в 10 раз) увеличим мощность лазера P. После выхода на установившийся режим, вернем величину мощности к прежнему значению. Результат этого эксперимента приведен на рисунке 4. Видно, что после возвращению мощности к первоначальному значению решение задачи (1) устанавливается на другом уровне (установившееся решение не совпадает с предыдущим). Ясно, что оба решения задачи (1) являются решениями задачи (5).

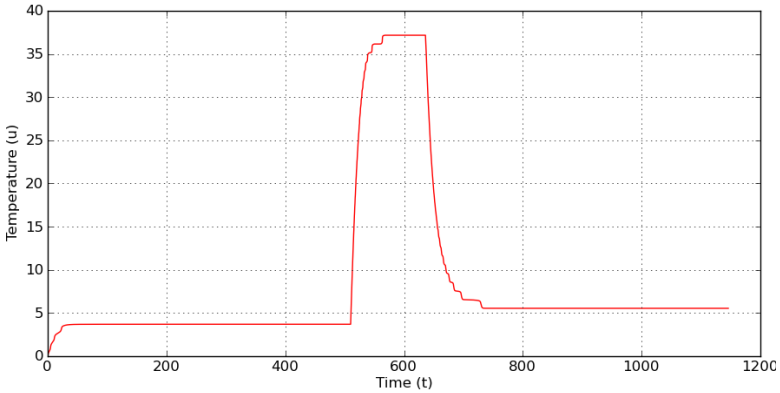


Рисунок 4. Поиск альтернативных установившихся решений

Теперь выясним, есть ли установившиеся состояния между двумя полученными. При помощи варьирования увеличения мощности лазера P (с последующим возвратом к прежнему значению аналогично предыдущему случаю), было найдено еще одно решение задачи (5) при вышеуказанных параметрах.

Полученные решения задачи (5) приведены на рисунке 5.

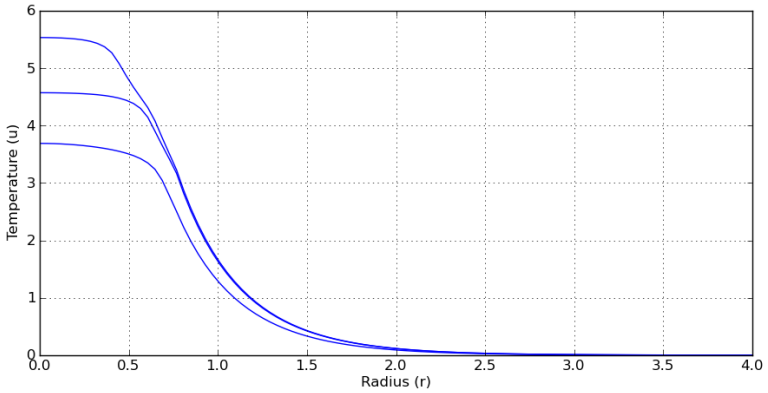


Рисунок 5. Решения задачи (5)

Каждая из фунций, график которых приведен на рисунке 5, получена с помощью итерационного процесса (6) с точностью порядка 10^{-4}.

Литература

1. Карлов Н.В., Кузьмин Г.П., Сисакян Е.В. Интерференционные эффекты при измерении коэффициентов поглощения прозрачных материалов. Квантовая электроника, 1977, т.4, №8, с. 1816-1818.

2. Дегтярев А.А., Шамаров П.А. Комбинированный алгоритм вычисления оптимальных оценок параметров оптических элементов. Перспективные информационные технологии в научно-исследовательском проектировании и обучении. Самара, СГАУ, 1995.

3. Дегтярев А.А., Сисакян Е.В., Сойфер В.А. Модифицированный лазерный калориметрический метод автоматизированных измерений малых оптических потерь. Препринт ИОФАН СССР, 1985, №51, с. 46.

4. Самарский А. А. Теория разностных схем. М.: Наука, 1977. 656 с.

Дисперсия стохастического процесса распределения ресурса

А.В. Докучаев
Самарский государственный технический университет, Самара

Введение

При моделировании физических (в обработке сигналов), в бортовых вычислительных устройствах систем управления оснащением летательных аппаратов, в авиации (для расчёта устойчивости движения и при проектировании элементов конструкций) – часто приходится решать различные задачи динамического программирования.

Оптимизационные задачи в стохастической постановке предполагают учет случайного характера процессов, протекающих в моделируемых системах. Применение того или иного метода оптимизации для решения задачи в стохастической постановке обусловлено спецификой прикладной задачи. В данной статье речь пойдет о принципиальной возможности учета случайного характера моделируемых процессов в рамках динамического программирования в дискретном времени и о конкретном классе задач, весьма перспективном для применения моделей динамического программирования в стохастической постановке – распределении ресурсов [1-2].

1. Постановка задачи о стохастическом процессе распределения ресурса

Распределяется некоторый ресурс (например, сигнал и т.п.)

$X = \sum x_i$, $n \in N = \{1, 2, ..\}$, $0 \le x_i \le X$, $i \in \overline{1,n}$, $x \in R^+$,

по стохастическим функциям освоения (откликам) $f_i(x_i)$, $f_i : R^+ \to R^+$ с заданным равномерным шагом дискретизации ресурса $h = X/(j-1)$, j – целое, $1 < j \le x+1$, $h \in R^+$, с целью максимизации (минимизации) математического ожидания суммарного эффекта *M[Z]:*

$$\begin{cases} M[Z] = \sum\limits_g M_g[Z^*_{p,k}(f(y_j))], \\ M_g = \sum\limits_k Z^*_{p,k}(f(y_j)) P\left(Z^*_{p,k}(f(y_j))\right), \\ Z^*_{p,k}(f(y_j)) = opt\{f(x_1^*, x_2^*, ..., x_n^*)_g\}, \\ \sum\limits_k P\left(Z^*_{p,k}(f(y_j))\right) = 1, \\ y_j = jh, \\ \sum y_j = X, \end{cases}$$

где M_g – математическое ожидание *g*-ой точки спектра суммарного эффекта *Z*; $Z^*_{p,k}$ – суммарный эффект *g*-ой детерминированной подзадачи (получаемой при разбиении исходной стохастической задачи), найденный с помощью динамического программирования на основе рекуррентных соотношений Беллмана; *k* – количество спектральных величин ряда распределения *p* случайной точки задачи (случайной точкой называем значение $f(y_j)$, если оно представляет собой дискретную случайную величину с заданным рядом распределения), y_j – распределяемая на *j*-ом шаге порция ресурса *X*; $P\left(Z^*_{p,k}\right)$ – вероятность осуществления суммарного эффекта детерминированной подзадачи, соответствующей индексам *p* и *k*; $(x_1^*, x_2^*, ..., x_n^*)_g$ - вектор распределения ресурсов *g*-й подзадачи (находимый динамическим программированием,

g_1 соответствует подзадача для первой спектральной величины первой случайной точки первой функции освоения – если таковая имеется, g_2 – второй спектральной точки и так далее, потом осуществляется переход к следующей случайной точке первой функции освоения, потом переход к спектральным точкам следующей функции освоения и далее по аналогии), где x_i^* – ресурс, при распределении которого функция $f_i(x_i)$ обеспечивает оптимум; $f(y_j)$ – значения функций эффектов в *j*-й точке дискретизации ресурса; через *opt* обозначен оптимум целевой функции.

Таблица.1. Исходные данные задачи

Порция ресурса	Функции освоения					
	f_1	f_2	…	f_i	…	f_n
y_0	f_{01}	f_{02}	…	f_{0i}	…	f_{0n}
y_1	f_{11}	f_{12}	…	f_{1i}	…	f_{1n}
…	…	…	…	…	…	…
y_j	f_{j1}	f_{j2}	…	f_{ji}	…	f_{jn}
…	…	…	…	…	…	…
y_n	f_{n1}	f_{ni}	…	f_{ni}	…	f_{nn}

Таблицу значений $f_i(y_j)$ (см. табл. 1) назовем таблицей исходных данных задачи. Количество получаемых при разбиении исходной стохастической задачи подзадач будет определяться как

$$g = \prod_p k(p)k(p-1)..k(1),$$

где *k(p)* – мощность спектра *p*-й случайной точки. Вероятность достижения суммарного эффекта *g*-й спектральной точки определяется по формуле полной вероятности.

Для определённости будем рассматривать задачу максимизации. Функции откликов f_i в общем случае зависят от распределяемого ресурса *X* нелинейно.

2. Исследование дисперсии стохастического процесса распределения ресурса. Пусть функции освоения заданы числами $f_i(y_j)$ для конкретного шага дискретизации *h* (таблица 2).

Таблица.2. Пример задания исходных данных

y_j	$f_1(y_j)$	$f_2(y_j)$	$f_3(y_j)$	$f_4(y_j)$
0	0	0	0	0
1	3	2	2,5	2,2
2	5	2	4	4,5
3	8	3	6	7
4	9	4	6	8
5	10	4	8	8

Оптимальный суммарный эффект, найденный методом динамического программирования, $Z^* = Z\left(X^*\right) = 12,7$.

Пусть теперь значения функций освоения $f_3(1)$ и $f_4(1)$ представляют собой дискретные случайные величины, связанные матрицей вероятностей:

$$P = \left(\begin{array}{ccc|c} 0 & 2 & 5 & f_4(1)/f_3(1) \\ \hline 0,1 & 0 & 0,2 & 1 \\ 0 & 0,3 & 0 & 2 \\ 0,1 & 0,3 & 0 & 4 \end{array}\right).$$

Суммируя вероятности, получим ряды распределения случайных величин $f_3(1)$ и $f_4(1)$ (пример 4 в табл. 3). Вычислим математическое ожидание $M(Z^*(X^*))$ и дисперсию $D(Z^*(X^*))$ оптимального значения суммарного эффекта, а также обобщенную дисперсию случайных величин – $D_{обобщ}$.

Обобщенная дисперсия двух случайных точек задачи находится как определитель ковариационной матрицы:

$$D(\xi,\eta)=\begin{Vmatrix} \sigma_\xi^2 & \operatorname{cov}(\eta,\xi) \\ \operatorname{cov}(\xi,\eta) & \sigma_\eta^2 \end{Vmatrix},$$

где ξ и η – случайные точки задачи со среднеквадратическими отклонениями $\sigma_\xi^2, \sigma_\eta^2$ и ковариацией $\operatorname{cov}(\xi,\eta)$:

$$\operatorname{cov}(\xi,\eta)=\operatorname{cov}(\eta,\xi)=M(\xi\eta)-m_\xi m_\eta .$$

Найдем $M(f_3(1), f_4(1))=4,6$, $\operatorname{cov}(f_3(1), f_4(1))=-0,9$. То есть случайные точки $f_3(1)$ и $f_4(1)$ коррелированны. Их обобщенная дисперсия $D_{обобщ}(f_3(1), f_4(1))=4,69$. Результаты расчетов приведем в табл.3 (пример 4), где $f_i(X_j)$ – случайная точка задачи, $f_{ik}(X_j)$ – дискретный спектр её значений с распределением $P(f_{ik}(X_j))$ и дисперсией $D(f_{ik}(X_j))$, $Z_{p,k}{}^*(X^*)$ – оптимальный суммарный эффект g-й детерминированной подзадачи (в таблице обозначены как f_{ik}, P_k, D_k, Z_k^* соответственно).

Остальные примеры постановок задачи (1-3, 5,6 в табл. 3) приведены для некоррелированных случайных величин.

Утверждение 1. Если стохастичность случайных точек не влияет на оптимальный суммарный эффект получаемых при разбиении стохастической задачи детерминированных подзадач (по сравнению с оптимальным суммарным эффектом в модели с усреднением случайных параметров), то дисперсия оптимального суммарного эффекта стохастической задачи равна нулю (пример 1 в табл. 3).

Таблица 3. Примеры постановок задачи с исходными данными в таблице 2.

№	$f_i(X_j)$	f_{ik}	P_k	D_k	$D_{обобщ}$	$D(Z^*(X^*))$	Z_k^*	$M(Z^*(X^*$
1	$f_1(4)$	8,9	1/2	0,01	0,0001	0	12,7; 12,7; 12,7; 12,7.	12,7
		9,1	1/2					
	$f_3(4)$	5,9	1/2	0,01				
		6,1	1/2					
2	$f_1(3)$	7,5	1/6	0,05	0,0005	0,0125	12,5; 12,5; 12,8; 12,8.	12,75
		8,1	5/6					
	$f_1(4)$	8,9	1/2	0,01				
		9,1	1/2					
3	$f_1(3)$	7,5	1/6	0,05	0,96	0,32	12,5; 12,5; 14; 12,8; 12,8; 14.	13,1
		8,1	5/6					
	$f_2(4)$	1	13/25	19,2				
		2	5/25					
		11	7/25					
4	$f_3(1)$	1	3/10	1,65			12,5;	13,6

		2	3/10		3,41	0,94	15; 12,5; 14; 14.	
		4	4/10					
	$f_4(1)$	0	2/10	2,56				
		2	6/10					
		5	2/10					
5	$f_1(3)$	4	3/7	12	230,4	2,16	12,5; 12,5; 14; 5,7; 15,7; 5,7.	14,31
		11	4/7					
	$f_2(4)$	1	13/25	19,2				
		2	5/25					
		11	7/25					
6	$f_1(3)$	4	1/2	16	307,2	3,8	12,5; 12,5; 14; 16,7; 16,7; 16,7.	14,81
		12	1/2					
	$f_2(4)$	1	13/25	19,2				
		2	5/25					
		11	7/25					

Утверждение 2. Рост дисперсии в случайной точке (для нескольких случайных точек в исходных данных задачи – суммы их дисперсий) при неизменной мощности спектра вызывает рост дисперсии суммарного эффекта и может вызывать рост математического ожидания суммарного эффекта (видно из сравнения результатов расчета примеров 1 и 2, а также примеров 3,5 с 6 в табл. 3).

Проводя расчет различных постановок задачи и сравнивая результаты расчета для двух некоррелированных случайных точек в исходных данных задачи (примеры 1-3 и 5-6 в табл.3) с решением для двух коррелированных случайных точек (пример 4 в табл. 3) построим графики (рис 1.).

Утверждение 3. Коррелированность случайных точек задачи не влияет на приращение математического ожидания суммарного эффекта и его дисперсию (график на рис.1 слева).

Обобщенная дисперсия случайных точек в коррелированной постановке задачи меньше обобщенной дисперсии случайных точек в постановке задачи с некоррелированными параметрами (см. рис. 1 справа).

3. Влияние мощности спектра случайных точек задачи на математическое ожидание суммарного эффекта

В математических экспериментах по исследованию приращения математического ожидания суммарного эффекта в зависимости от изменения количества точек спектра и дисперсии, проводимых с использованием современных информационных технологий, было установлено, что не каждое увеличение дисперсии случайных точек влияет на математическое ожиданием суммарного эффекта, но каждому изменению математического ожидания соответствует изменение суммы дисперсий случайных точек.

Итак, если при усреднении точек спектра происходит уменьшение дисперсии случайной точки, то получаемый суммарный эффект усредненной задачи в ряде случаев будет локально-оптимальным и заниженным. Если происходит увеличение дисперсии, то в ряде случаев суммарный эффект задачи, не учитывающей стохастические особенности процесса распределения ресурса, будет завышенным.

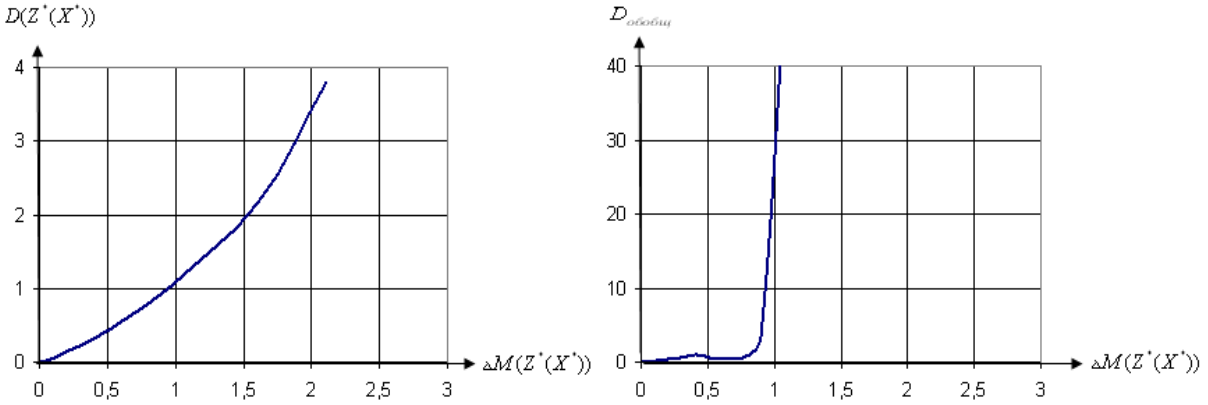


Рис. 1. Зависимость дисперсии суммарного эффекта и обобщенной дисперсии случайных точек от приращения математического ожидания суммарного эффекта. $\Delta M(Z^*(X^*)) = M(Z^*(X^*)) - M_0(Z^*(X^*))$, где за $M_0(Z^*(X^*))$ обозначено значение суммарного эффекта, получаемое при усреднении параметров задачи.

Как таковое количество точек спектра не влияет на математическое ожидание оптимального суммарного эффекта – если при этом дисперсия суммарного эффекта (для нескольких случайных точек в исходных данных задачи, для одной же случайной точки – дисперсия случайной точки) не изменяется.

Утверждение 4. Для одной случайной точки в исходных данных задачи – усреднение спектральных величин случайной точки не влияет на математическое ожидание суммарного эффекта исходной стохастической задачи распределения ресурса, если при этом не изменяется дисперсия случайной точки. Для нескольких случайных точек – усреднение спектральных величин не влияет на математическое ожидание суммарного эффекта, если при этом не изменяется дисперсия суммарного эффекта. Но если дисперсия случайной точки (суммарного эффекта) изменяется – то это не всегда ведет к изменению математического ожидания суммарного эффекта.

Литература

1. Докучаев А.В. Использование байесовского подхода в стохастической постановке задачи динамического программирования для оценки влияния погрешностей исходных данных. // Тр. Международной молодежной научной конференции "XXXIV Гагаринские чтения". – М., 2008. – С. 54–55.

2. Докучаев А.В., Котенко А.П. Алгоритмы решения стохастических задач динамического программирования большой размерности. // Вестник Самарского государственного технического университета. Серия физ.-мат. науки. – 2008. – № 2(17). – С. 203–210.

Вычисление параметра регуляризации методом перекрёстной значимости на основе разложения Холесского

В.В. Долишний, А.И. Жданов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Постановка задачи

Рассмотри систему линейных алгебраических уравнений (СЛАУ)

$$\mathbf{A}x = \widetilde{\mathbf{b}},\ \mathbf{A} \in \mathbf{R}^{m\times n},\ \widetilde{\mathbf{b}} = \mathbf{b} + \boldsymbol{\xi} \in \mathbf{R}^{m},\ m \geq n, \qquad (1)$$

в которой точное значение вектора правой части $\mathbf{b}$ неизвестно, а известно его значение с погрешностью $\widetilde{\mathbf{b}} = \mathbf{b} + \boldsymbol{\xi}$.

Подобные СЛАУ часто встречаются в различных прикладных задачах. Обычно, исходя из их происхождения и физического смысла, при отсутствии погрешностей в исходных данных они являются совместными и имеют единственное решение $\bar{x}$. Однако наличие погрешностей, в частности, погрешности $\boldsymbol{\xi}$ вектора правой части делает такие системы несовместными.

Задача состоит в нахождении вектора $\widetilde{x}$, наиболее близкого по норме к вектору $\bar{x}$.

Трудностями отыскания вектора $\widetilde{x}$ в реальных задачах, как правило, являются плохая обусловленность матрица системы $\mathbf{A}$ (число обусловленности $\kappa_2(\mathbf{A}) \approx 1/\varepsilon_{mash}$, где ε_{mash} - машинное эпсилон) и её большая размерность.

Для получения устойчивого решения плохообусловленных систем применяют различные регуляризирующие алгоритмы, наиболее известным и часто применяемым из которых является регуляризация Тихонова.

Регуляризованная система нормальных уравнений для системы (1) имеет вид:

$$\left(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I}\right)x = \mathbf{A}^T\widetilde{\mathbf{b}}, \qquad (2)$$

где $\alpha \in \mathbf{R}$, $\alpha > 0$ - параметр регуляризации.

Эффективность регуляризации Тихонова определяется параметром α. Для выбора оптимального параметра регуляризации α_{opt} существуют различные методы. Некоторые из них требуют использование априорной информации о погрешности ξ, которая обычно отсутствует. Другие требуют многократного решения системы (2) при различных α, что влечёт недопустимо большие вычислительные затраты, если размерность матрицы системы велика.

Одним из эффективных методов выбора квазиоптимального параметра регуляризации при отсутствии априорной информации о погрешности является метод перекрёстной значимости (cross validation method) [1]. Согласно этому методу параметр регуляризации α следует выбирать как значение, обеспечивающее минимум функции

$$G_{\mathbf{A},\tilde{b}}(\alpha) = \frac{n\left\|\left(\mathbf{I} - \mathbf{A}\mathbf{A}_\alpha^*\right)\widetilde{\mathbf{b}}\right\|^2}{\left(\operatorname{trace}\left(\mathbf{I} - \mathbf{A}\mathbf{A}_\alpha^*\right)\right)^2}, \qquad (3)$$

где $\mathbf{A}_\alpha^* = \left(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I}\right)^{-1}\mathbf{A}^T$.

Существует эффективный способ вычисления квазиоптимального параметра регуляризации методом перекрёстной значимости на основе сингулярного разложения матрицы системы $\mathbf{A}$ [2]. Однако вычисление сингулярного разложения требует значительного объёма оперативной памяти компьютера и невозможно для матриц

больших размерностей (порядка $10^3 \sim 10^4$ в зависимости от параметров конкретной вычислительной машины).

В работе предлагается альтернативных способ вычисления параметра регуляризации методом перекрёстной значимости, позволяющий эффективно использовать возможности параллельных и векторных вычислительных машин.

Вычисление параметра регуляризации

Согласно методу перекрёстной значимости, задача нахождения квазиоптимального параметра регуляризации $\widetilde{\alpha}_{opt}$ сводится к задаче отыскания минимума функции (3), что при использовании численной минимизации в свою очередь требует способа эффективного вычисления значений $G_{\mathbf{A},\tilde{b}}(\alpha)$ при различных α.

Рассмотрим следующую организацию процесса вычислений значений функции $G_{\mathbf{A},\tilde{b}}(\alpha)$.

Перепишем $G_{\mathbf{A},\tilde{b}}(\alpha)$ в виде:

$$G_{\mathbf{A},\tilde{b}}(\alpha) = \frac{n\left\|\left(\widetilde{\mathbf{b}} - \mathbf{A}\left(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I}\right)^{-1}\mathbf{A}^T\widetilde{\mathbf{b}}\right)\right\|^2}{\left(\sum_{i=1}^{m}\left(e_i - \mathbf{A}\left(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I}\right)^{-1}\mathbf{A}^T e_i\right)_i\right)^2} =$$

$$= \frac{n\left\|\left(\widetilde{\mathbf{b}} - \mathbf{A}\bar{x}_{\alpha,0}\right)\right\|^2}{\left(\sum_{i=1}^{m}\left(e_i - \mathbf{A}\bar{x}_{\alpha,i}\right)_i\right)^2} =$$

$$= \frac{n\left\|\left(\widetilde{\mathbf{b}} - \mathbf{A}\bar{x}_{\alpha,0}\right)\right\|^2}{\left(m - \sum_{i=1}^{m}\mathbf{a}_{i\cdot}\bar{x}_{\alpha,i}\right)^2}, \qquad (4)$$

где $\bar{x}_{\alpha,0}$ - решение системы

$$\left(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I}\right)x = \mathbf{A}^T\widetilde{\mathbf{b}}, \qquad (5)$$

векторы $\bar{x}_{\alpha,i}$, $i = \overline{1,m}$ удовлетворяют системам

$$\left(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I}\right)x = \mathbf{a}_{i\cdot}^T,\ i = \overline{1,m}, \qquad (6)$$

$\mathbf{a}_{i\cdot}$ - i-тая строка матрицы $\mathbf{A}$.

Обозначим

$$\widetilde{\mathbf{A}}_\alpha = \left(\mathbf{A}^T\mathbf{A} + \alpha\mathbf{I}\right)$$

и перепишем системы (5), (6) в виде:

$$\widetilde{\mathbf{A}}_\alpha x = \mathbf{d}_i,\ i = \overline{0,m}, \qquad (7)$$

где $\mathbf{d}_i = \begin{cases}\mathbf{A}^T\widetilde{\mathbf{b}}, i = 0,\\ \mathbf{a}_{i\cdot}^T, i = \overline{1,m.}\end{cases}$, $\mathbf{d}_i = \mathbf{a}_{i\cdot}^T$

Таким образом процесс вычисления функции $G_{\mathbf{A},\tilde{b}}(\alpha)$ сводится к решению $(m+1)$-ой СЛАУ (7) с одной и той же симметричной положительноопределённой матрицей $\widetilde{\mathbf{A}}_\alpha$ ($\alpha > 0$) и различными правыми частями.

Для решения систем (7) эффективно подходит использование как параллельного, так и векторного компьютера.

В случае использования векторной вычислительной машины перепишем (7) в виде

$$\widetilde{\mathbf{A}}_\alpha x = \mathbf{D},$$

где матрица $\mathbf{D}$ составлена из столбцов $\mathbf{d}_i$ $i = \overline{0,m}$.

Если использовать разложение Холесского матрицы $\widetilde{\mathbf{A}}_\alpha$:

$$\widetilde{\mathbf{A}}_\alpha = \widetilde{\mathbf{L}}_\alpha \widetilde{\mathbf{L}}_\alpha^T, \tag{8}$$

то решение системы (7) со многими правыми частями сводится к векторным операциям над строками матрицы $\mathbf{D}$. Отметим, что вычисление матрицы $\mathbf{A}^T\mathbf{A}$ также допускает использование векторных вычислений.

Рассмотрим организацию процесса вычислений для параллельного компьютера.

Для решения систем (7) будем также используя разложения Холесского (8).

Вычисления организуем следующим образом. Вычислим матрицу $\mathbf{A}^T\mathbf{A}$, используя параллельные алгоритмы матричного умножения. Найдём разложение (8) для стартового параметра α. Далее, параллельно будем решать системы (7) на основе найденного разложения (8) и одновременно искать разложение (8) для двух возможных следующих значений α.

Оценим вычислительные затраты предложенного алгоритма.

Вычисление матрицы $\mathbf{A}^T\mathbf{A}$, учитывая её симметричность, и вектора $\mathbf{A}^T\widetilde{\mathbf{b}}$ требует $O\left(mn^2\right)$ флопов. Получение разложения (8) требует $O\left(n^3/3\right)$ флопов. Для решения всех систем (7) и получения $\bar{x}_{\alpha,i}$, $i = \overline{0,m}$, необходимо затратить $O\left(2mn^2\right)$. Расчет функции $G_{\mathbf{A},\tilde{b}}(\alpha)$ по формуле (4) требует $O\left(n^2\right)$ флопов, и им мы пренебрежём.

Итого, при последовательном выполнении всех вычислений требуется в общей сложности $mn^2 + \left(n^3/3 + 2mn^2\right)\tau + O\left(n^2\right)$, где τ - необходимое число вычисления функции $G_{\mathbf{A},\tilde{b}}(\alpha)$, операций над числами с плавающей запятой.

Пусть вычисления выполняются на машине с p процессорами. Тогда для вычисления $\mathbf{A}^T\mathbf{A}$ последовательно потребуется выполнить mn^2/p флопов. Вычисление разложения (8) для двух возможных следующих значений параметра α можно выполнять на одном или двух отдельных процессорах одновременно с решением систем (7). В частности, если выполняется условие $p \le 3m/n$, то для вычисления разложения (8) достаточно одного процессора, Если же $3m/n \le p \le 6m/n$, то для отсутствия простаивания процессоров, выполняющих решение систем (7), разложение (8) лучше проводить на двух процессорах для каждого α отдельно. Используя описанную организацию вычислений, последовательно потребуется выполнить $n^3/3 + mn^2\left(2\tau + 1\right)/p + O\left(n^2\right)$ флопов.

Оценим требуемый алгоритмом объём оперативной памяти.

Пусть исходная матрица размером $m \times n$ уже хранится в оперативной памяти. Для хранения матриц $\widetilde{\mathbf{A}}_\alpha$, $\widetilde{\mathbf{L}}_\alpha$ и главной диагонали матрицы $\mathbf{A}^T\mathbf{A}$ потребуется дополнительно $n\left(n+2\right)$ ячеек памяти. Также требуется по одному массиву ячеек длиной n для каждого процессора системы для хранения текущего вычисляемого вектора $\bar{x}_{\alpha,i}$. Итого, помимо памяти для хранения исходной матрицы системы потребуется ещё $n\left(n+2\right)+np$ ячеек памяти, где p - число процессоров в системе.

Выводы и анализ алгоритма

Описанный алгоритм позволяет эффективно использовать возможности параллельных и векторных вычислительных машин. При этом скорость вычислений по описанному алгоритму приближается к скорости SVD-алгоритма при выполнении условий $p/3 \le m/n$ и $\tau/p \approx 1$, то есть при условии, что число уравнений много больше числа неизвестных и вычислительная система имеет достаточно большое число процессоров в случае использования параллельных вычислений или большую длину векторных регистров в случае применения векторных.

Литература

1. *Воскобойников Ю. Е., Литвинов Л.А.* Выбор момента останова в итерационных алгоритмах восстановления сигналов и изображений // Автометрия, 2004. - Т. 40, №4. - С. 3-10.

2. *Chung J., Nagy J.G., O'Leary D.P.* A weighted-GCV method for Lanczos-hybrid regularization // Electronic Transaction on Numerical Analysis, 2008. - Volume 28, pp. 149-167.

Исследование механизмов ценообразования опционов в моделях обобщенного случайного блуждания в непрерывном времени с негауссовскими распределениями скачков логарифмов цен

А.Н. Евдокимов, А.П. Зубарев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Одной из центральных задач стохастической финансовой математики является определение справедливой цены производных финансовых активов. Классическое решение этой задачи было предложено Блэком, Шоулсом и Мертоном в их знаменитой модели, за что авторы были удостоены нобелевской премии. Стохастическое моделирование динамики цен в этом подходе основывалось на модели геометрического броуновского движения, представляющего непрерывный диффузионный процесс:

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t},$$

где S_t – текущая цена; S_0 – начальная цена; μ – коэффициент сноса; σ – коэффициент волатильности; W_t – стандартный винеровский процесс (броуновское движение), заданный на вероятностном пространстве (Ω, F, P); T – дата исполнения.

Тем не менее, изменение цен финансовых активов, наблюдаемое на рынках, особенно для низколиквидных активов, не считается непрерывным. Более того, статистические характеристики изменения цен финансовых активов во времени, полученные на основе анализа экспериментальных данных на реальных рынках, имеют закономерности, которые не описываются моделью геометрического броуновского движения.

В связи с этим актуальной задачей становится исследование моделей ценообразования производных финансовых активов, основанных на случайных процессах с разрывными траекториями, что может принципиально повысить качество прогнозирования рынка финансовых инструментов. К таким типам случайных процессов относится обобщенное непрерывное во времени случайное блуждание (continious time random walk или CTRW).

Случайный ценовой процесс $X(t) = Z(t) - \langle Z(t) \rangle$, где $Z(t) = \ln\left(\frac{S(t)}{S(t_0)}\right)$ с торгуемой стоимостью актива S(t) и начальным временем t_0 (причем Z(t) не зависит от t_0), удовлетворяет требованиям CTRW и состоит из серии функций-скачков. X(t) меняется в случайные, непрерывно распределенные моменты времени t_0, t_1, t_2,... t_n,… Временные интервалы между двумя совершившимися переходами (времена паузы или ожидания) $T(t_n) = t_n - t_{n-1}$ есть независимые, одинаково распределенные случайные переменные с функцией плотности распределения $\psi(t)$:

$$\psi(t)dt = \mathrm{Prob}\{t < T_n < t + dt\}.$$

В отличие от броуновского движения в CTRW число шагов n, сделанных блуждающей частицей в интервале времени (0,T), является случайной величиной.

Процесс X(t) эквивалентен набору случайных переходов $\Delta X_n = X(t_n) - X(t_{n-1})$, функция плотности распределения которого:

$$h(x)dx = \mathrm{Prob}\{x < \Delta X_n < x + dx\}.$$

Одним из возможных направлений исследования, представляющим интерес, можно рассматривать создание метода расчета европейского и американского опциона

при рэлеевском и равномерном вероятностном распределении логарифмов ценовых изменений и показательном вероятностном распределении времен ожидания ценового изменения, базирующегося на принципах физической модели CTRW.

Решение заявленной задачи исследования включает в себя следующие этапы:

- вывод формул расчета европейского опциона колл (на покупку) и опциона пут (на продажу);
- разработка методики получения справедливой стоимости американского опциона;
- проведение вычислительных экспериментов и анализ полученных результатов.

Выбор опциона в качестве объекта изучения функционирования модели CTRW в области прогнозирования поведения рынка финансовых производных инструментов не случаен, т.к. именно финансовый инструмент такого типа обладает низкой ликвидностью. Низкая ликвидность подразумевает то обстоятельство, что открытие, изменение и закрытие позиций по опциону производится в некоторые случайные дискретные моменты времени (раз в час, раз в сутки и т.д.).

При этом опцион обладает свойством волатильности, т.е. его стоимость может изменяться даже при условии, что цена базового актива, лежащего в его основе (акция, товар, другая финансовая производная бумага и т.д.), остается постоянной. Таким образом, можно заключить, что опцион является нелинейным финансовым инструментом, и для его описания нельзя использовать традиционные в экономике методы дисконтирования.

Основополагающим механизмом опционов, помимо спекулятивной функции, было и остается хеджирование − редуцирование рисков возможных потерь при неблагоприятном изменении цен на базовый актив за счет внесения первоначальной сравнительно небольшой инвестиции. Знание справедливой стоимости позволяет построить грамотную хеджирующую стратегию. Достижение именно такой цели может быть осуществлено посредством рассматриваемой в данной работе модели CTRW.

При определении справедливой стоимости опциона подразумевается, что рынок эффективен, т.е. на нем отсутствуют арбитражные возможности в виде внешних вмешательств при расчете прибыли. Таким образом, безрисковый рынок не дает прибыли, и все биржевые участники остаются «при своих». На математическом языке это значит, что дисконтирующий ценовой процесс удовлетворяет условию мартингальности $E^P[S(t')e^{-rt'}|F(t)]=S(t)e^{-rt}$, поэтому исследуется класс марковских процессов. Вероятностная мера Q описания рыночного процесса с неопределенным риском отличается от физической вероятностной меры P, но является ей эквивалентной. Поэтому описывает класс процессов подобного характера.

Единственный случай марковской модели CTRW наблюдается, только если функция плотности распределения времени ожидания ценового скачка имеет убывающий экспоненциальный вид с характерным параметром интенсивности λ, зависящим от типа используемого вероятностного распределения ценовых скачков (r – безрисковая процентная ставка):

$$\psi(t) = \lambda e^{-\lambda t}, \quad \lambda = \frac{r}{\tilde{h}(-i) - 1}.$$

Было обнаружено, что вероятностное распределение Рэлея $h(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ и постоянное на отрезке числовой оси [-a;a] вероятностное распределение $h(x) = \frac{1}{b-a}$

подходят для описания такого класса ситуаций, но лишь в определенных областях параметра среднеквадратичного отклонения σ (для рэлеевского) и ширины интервала a (для постоянного), удовлетворяющих условию положительности параметра интенсивности:

$0 < \lambda < \infty,\ 1 < \tilde{h}(-i) < \infty.$

Формула стоимости при рэлеевском распределении ценовых скачков выводится посредством решения восстанавливающего уравнения, предложенного теоретиками модели CTRW:

$$C(x,\bar{t}) = \Phi(x)e^{-(\lambda+r)\bar{t}} + \lambda\int_0^{\bar{t}} e^{-(\lambda+r)(\bar{t}-\bar{t}')}d\bar{t}'\int_{-\infty}^{+\infty} h(y-x)C(y,\bar{t}')dy$$

После определения вида платежной функции (например, ванильного опциона колл $\Phi(x) = \begin{cases} 0, & x < \ln K \\ e^x - K, & x \geq \ln K \end{cases}$, где K – страйковая цена, по которой можно купить или продать базовый актив в будущем) были рассмотрены два возможных способа решения проблемы и на их основании получены следующие расчетные формулы нахождения стоимости опциона для рэлеевского распределения ценовых скачков:

$$C(x,t) = \left(e^x - K\right)e^{-(\lambda+r)(T-t)} +$$

$$+ \sum_{n=1}^{N} \frac{(-1)^n\pi^n}{n!}\left(\frac{1}{\sqrt{n}}\right)^n \frac{1}{\sqrt{2\pi n\sigma^2}}(T-t)^n\lambda^n e^{-(\lambda+r)(T-t)}\int_{\ln K}^{+\infty} e^{-\frac{(x-y)^2}{2n\sigma^2}} H_n\left(-\frac{y-x}{\sqrt{2n\sigma^2}}\right)\left(e^y - K\right)dy$$

И для постоянного распределения:

$$C(x,t) = \sum_{n=1}^{N} \frac{e^{-(\lambda+r)(T-t)}\lambda^{n-1}(T-t)^{n-1}}{(n-1)!}\frac{1}{2a}\int_{-a}^{a}\left(\frac{\sin\left(\frac{a(y-x)}{n-1}\right)}{a(y-x)}\right)^{n-1}\left(K - e^y\right)dy.$$

Полученные формулы представляют собой бесконечно сходящийся ряд и при взятии достаточного количества членов ряда вполне пригодны для машинного вычисления стоимости опциона.

Американский опцион, в отличие от европейского, может быть исполнен в любой момент времени от начала до даты экспирации включительно. При рассмотрении проблемы нахождения справедливой стоимости американского опциона необходимо решать вспомогательную задачу об определении функции времени исполнения опциона Z(τ). Математическим аналогом такой задачи является нахождение $S_{[a,b]}(t\text{-}t_0, x_0)$ – вероятности выживания физической частицы на некотором отрезке [a;b]. Функция плотности вероятности выживания частицы в свою очередь удовлетворяет интегральному уравнению с разностным ядром на конечном промежутке числовой оси:

$$\tilde{S}_{[a,b]}(s,x_0) = \tilde{\psi}(s) + \tilde{\psi}(s)\int_a^b h(x-x_0)\tilde{S}_{[a,b]}(s,x_0)dx,$$

где $\tilde{\psi}(s,x_0) = \frac{1}{s}\tilde{\psi}(s)\int_{-\infty}^{+\infty} h(x-x_0)dx$ – Лаплас-образ функции вероятности того, что следующее временное пребывание будет вне заданной области после ценового скачка $x - x_0$.

Дальнейшим направлением исследования следует считать поиск метода решения подобного вида уравнений с учетом вида разностного интегрального ядра, зависящего от функции плотности распределения ценовых изменений, в целях получения удобных аналитических результатов, которые будут использоваться для подстановки в выражение для расчета стоимости американского опциона.

При проверке численным моделированием результатов работы полученных выражений для вычисления стоимости европейского опциона отмечается их согласованность, но не повторяемость по отношению к результатам классического подхода Блэка-Шоулса, что служит доказательством принципиальной возможности и актуальности подхода, основанного на принципах CTRW, и подтверждением справедливости выведенных формул (см. рисунок).

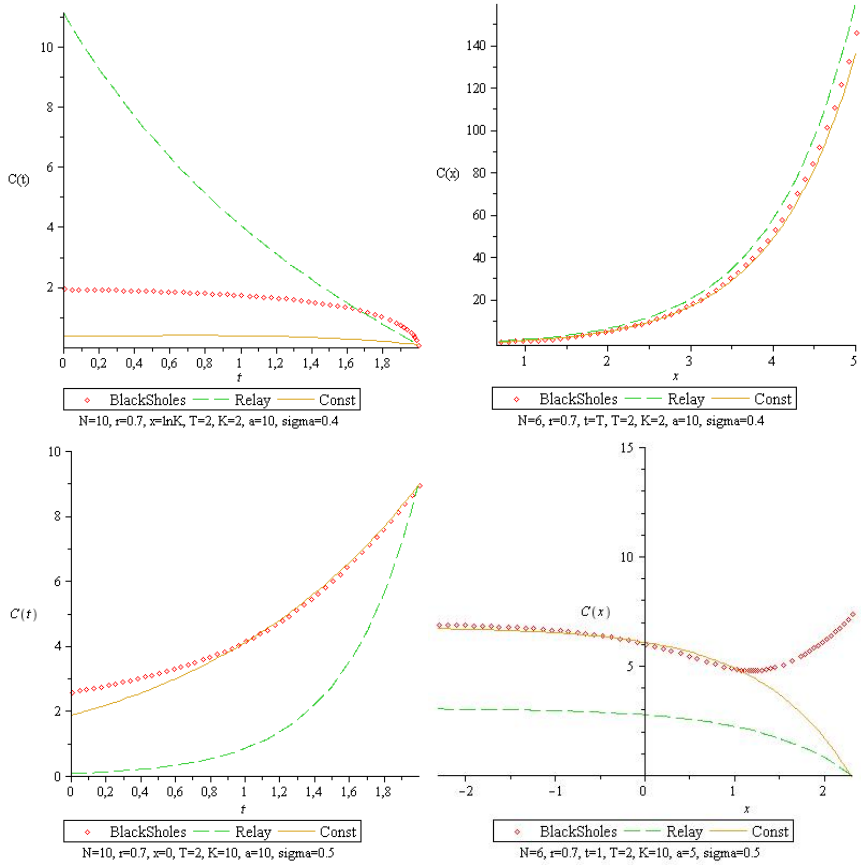


Рисунок – Сравнительные графики формулы Блэка-Шоулса и подхода CTRW при определении стоимости европейского ванильного опциона колл (верхние) и опциона пут (нижние) при фиксированной ценовом скачке x и при фиксированном моменте времени t

Исследование фотодиссоциирующих областей в межзведной среде

Д.И. Завершинский, Н.Е. Молевич
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара
[2]Самарский филиал Учреждения Российской академии наук Физического института им. П.Н. Лебедева РАН

Аннотация
В работе впервые получено нелинейное акустическое уравнение для среды с обобщенной функцией тепловыделения. Определены коэффициенты нелинейности, дисперсии, усиления для современной модели фотодиссоциирующей области[1,2]. На основе аналитического решения этого уравнения с учетом полученных коэффициентов определены стационарные структуры акустических возмущений и слабых ударных волн. Нелинейная эволюция акустических возмущений исследована как с помощью акустического уравнения, так и на основе системы газодинамических уравнений в форме Лагранжа. Показано, что в межзвездных облаках с изоэнтропической неустойчивостью возможно возникновение генерации последовательности автоволновых импульсов, что подтверждает гипотезу Краснобаева[3].

Межзвездная среда тесно связана со звездами и межгалактической средой. Из межзвездного газа образуются звезды, которые на поздних стадиях эволюции снова отдают часть своего вещества межзвездной среде. В результате этого обмена межзвездная среда обогащается создаваемыми в недрах звезд тяжелыми элементами. Звезды поставляют также в межзвездную среду электромагнитное излучение и космические лучи (особенно при взрывах сверхновых звезд). Часть вещества из межзвездной среды выбрасывается в межгалактическое пространство, а горячий межгалактический газ своим излучением и давлением может оказывать влияние на ионизацию межзвездной среды и ее динамику.

В настоящее время наиболее важные наблюдения межзвездной среды сделаны в радио и ИК-области электромагнитного излучения. Многие виды межзвездного вещества (такие как твердые тела с диаметром более одного миллиметра) почти невозможно обнаружить на основе их излучения или поглощения. Масса этих видов материи может быть даже больше, чем наблюдаемая масса всех других видов вместе взятых. Однако верхний предел полной массы межзвездной материи может быть выведен на основе ее гравитационных эффектов.

Весь нейтральный атомарный водород и большая часть молекулярного газа в Млечном Пути и иных галактиках находится в фотодисоциирующих областях (ФДО), и ФДО источник большинства незвездного (ИК) и миллиметрового CO излучения из галактики. На поверхности межзвездных облаках, что соответствует малым величинам визуальной экстинкции, происходит поглощение коротковолнового ультрафиолетового излучения частицами пыли и излучение в линиях CII, OI, H_2 связанных с вращательными и вибрационными переходами, а также полициклических ароматических гидрокарбонов (Polycyclic Aromatic Hydrocarbon).

Исторически изучение ФДО были связаны с наблюдениями за массивными звездообразующими областями Орион А и M17 в линиях CII 158 μm и OI 63 μm. Эти наблюдения показали наличие нейтральных областей светящихся в инфракрасном диапазоне, находящихся вне области ионизированного водорода H II.

Один из наиболее интересных эффектов, который может наблюдаться в таких средах это межзвездные ударные волны. Межзвездные ударные волны часто связаны с

рождением и смертью звезд, они могут существовать на протяжении всей жизни звезд высокой массы; звездные ветра, расширяющиеся области ионизированного водорода, и остатки суперновой, могут являться источником ударных волн. Также источником ударных волн могут являться различные рода неустойчивости в межзвездной среде: тепловые неустойчивости, неустойчивости связанные с ион-нейтральными потоками, неустойчивости посредством космических лучей, и неустойчивости связанные с взрывной волной [4].

В [5] было показано, что условие изоэнтропической неустойчивости совпадает с условием наличия в тепловыделяющей среде отрицательной второй (объёмной) вязкости. Описание стационарного состояния подобной однородной среды плотностью ρ_0 и температурой T_0 было проведено с помощью обобщённой функцией тепловыделения $\Im = L - \Gamma$, где L, Γ это мощности охлаждения и нагрева в эрг/(г с). Тогда уравнение энергии (в пренебрежении теплопроводностью) имеет вид

$$\Im(\rho_0, T_0) = 0$$

В тепловыделяющем газе возможны три типа тепловой неустойчивости: изохорическая, изобарическая и изоэнтропическая (акустическая). Для межзвёздной наиболее важны изобарическая неустойчивость, приводящая к тепловым конденсациям с образованием плотных холодных облаков и горячего разреженного газа, и изоэнтропическая неустойчивость, приводящая к усилению акустических возмущений.

При изохорических условиях стационарная среда неустойчива, если

$$\Im_T < 0,$$

Изобарическая неустойчивость реализуется при условии

$$\Im_T - \frac{\rho_0}{T_0}\Im_\rho < 0,$$

Условие акустической неустойчивости соответствует

$$\Im_T + \frac{\rho_0}{(\gamma - 1)T_0}\Im_\rho < 0,$$

где $\Im_T$, $\Im_\rho$ частные производные по температуре и плотности соответственно, γ - показатель адиабаты (отношение удельных теплоёмкостей).). Для межзвёздной наиболее важны изобарическая неустойчивость, приводящая к тепловым конденсациям с образованием плотных холодных облаков и горячего разреженного газа, и изоэнтропическая неустойчивость, приводящая к усилению акустических возмущений.

Стоит сказать о модели среды, которая нами была использована. Модель описывает ФДО, а именно ту ее часть, которая соответствует малым значения визуальной экстинкции и находится вблизи к так называемому поверхностному слою или по другому ионизационному фронту. Нагрев среды происходит за счет фотоэлектрической эмиссии от частиц пыли и полициклических ароматических гидрокарбонов, а охлаждение происходит за счёт перехода между электронными уровнями иона углерода.

Для простоты нами была рассмотрена одномерная задача, и мы пренебрегли сдвиговой вязкостью и температуропроводностью. Ось x направлена вдоль линии распространения ударной волны, то есть перпендикулярно к её фронту. Поскольку задача одномерная, все величины зависят только от координаты x и времени t. Компоненту скорости частиц вдоль оси x обозначена через u. Воспользовавшись, описанными выше условиями, исходная система уравнений выглядит как:

$$\frac{d\rho}{dt} + \rho\frac{\partial \mathrm{u}}{\partial x} = 0,\ \rho\frac{d\mathrm{u}}{dt} = -\frac{\partial P}{\partial x},\ c_{V\infty}\frac{dT}{dt} - \frac{P}{\rho^2}\cdot\frac{d\rho}{dt} = -\Im,\ P = \frac{k_B\rho T}{m},$$

где $d/dt = \partial/\partial t + u\partial/\partial x$ - полная производная по времени, ρ - плотность среды, T - температура среды, P - давление в среде, u - скорость, $\Im$ - обобщенная функция теплопроводности, $c_{V\infty}$ - высокочастотная теплоёмкость при постоянном объёме, k_B - постоянная Больцмана.

Значение m зависит от области межзвездной среды (например, в *H I* областях т.е. там где преобладает атомарный водород, $m = m_H$, в межзвездных молекулярных облаках $m = 2m_H$, m_H это масса атомарного водорода).

Необходимо упомянуть об условиях применимости нелинейного акустического уравнения. Первое состоит в малости возмущения это связано с тем, что нелинейное акустическое уравнение получается с помощью теории возмущений.

А второе состоит в слабой дисперсии:

$$|\beta| = \left|\frac{c_0^2 - c_\infty^2}{c_\infty^2}\right| << 1 ,$$

где c_∞, c_0 – скорости высокочастотного и низкочастотного звуков.

Линеаризовав исходную систему до величин второго порядка малости, получим во втором порядке малости теории возмущения нелинейное акустическое уравнение вида:

$$\tau_0 c_{V\infty}\left[\frac{\partial^2\rho'}{\partial t^2} - c_\infty^2\frac{\partial^2\rho'}{\partial x^2} - \frac{\Psi_\infty c_\infty^2}{\rho_0}\frac{\partial^2\rho'^2}{\partial x^2}\right]_t + c_{V0}\left[\frac{\partial^2\rho'}{\partial t^2} - c_0^2\frac{\partial^2\rho'}{\partial x^2} - \frac{\Psi_0 c_0^2}{\rho_0}\frac{\partial^2\rho'^2}{\partial x^2}\right] = 0,$$

$$\Psi_\infty = \frac{\gamma_\infty + 1}{2}, \qquad \Psi_0 = \frac{2\gamma_0 - 1}{\gamma_0} - \frac{1}{T_0\Im_{0T}}\left[\frac{T_0^2\Im_{0TT}(\gamma_0 - 1)^2}{2\gamma_0} + \frac{\rho_0^2\Im_{0\rho\rho}}{2\gamma_0} + \frac{T_0\rho_0\Im_{0\rho T}(\gamma_0 - 1)}{\gamma_0}\right],$$

$$\Im_{0TT} = \left(\frac{\partial^2\Im}{\partial T^2}\right)_{\rho=\rho_0, T=T_0}, \quad \Im_{0\rho T} = \left(\frac{\partial^2\Im}{\partial\rho\partial T}\right)_{\rho=\rho_0, T=T_0}, \Im_{0\rho\rho} = \left(\frac{\partial^2\Im}{\partial\rho^2}\right)_{\rho=\rho_0, T=T_0},$$

где величины Ψ_∞, Ψ_0 - высокочастотный и низкочастотный показатель нелинейности, соответственно, ρ' – возмущение плотности, γ_∞, γ_0 – высокочастотный и низкочастотный показатели адиабаты, C_{V0}, C_{P0}, $C_{V\infty}$, $C_{P\infty}$ – низкочастотные и высокочастотные теплоемкости в колебательно-возбужденном газе при постоянном объёме и давлении, T_0, ρ_0, τ_0 – невозмущенные значения температуры, плотности и времени релаксации.

Нелинейное акустическое уравнение имеет три качественно различных решения представленных на рисунках 1-3, причем все параметры ударноволновых фронтов можно находить аналитически.

Также в работе было проведено численное моделирование одномерной системы уравнений газодинамики с учетом нагрева и охлаждения межзвёздной среды по модели [2,3], и и показан распад неустойчивой ударной волны на последовательность автоволновых импульсов с разрывными фронтами. Тот же результат получен при численном решении нелинейного акустического уравнения. Эволюция локализованного импульса также приводила к формированию импульсов автоволновой формы.

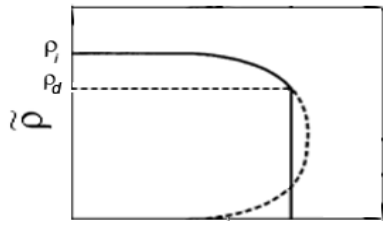


Рисунок 1

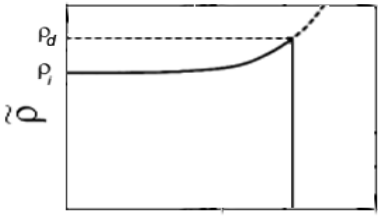


Рисунок 2

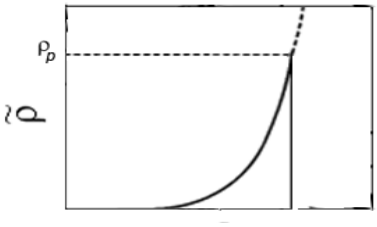


Рисунок 3

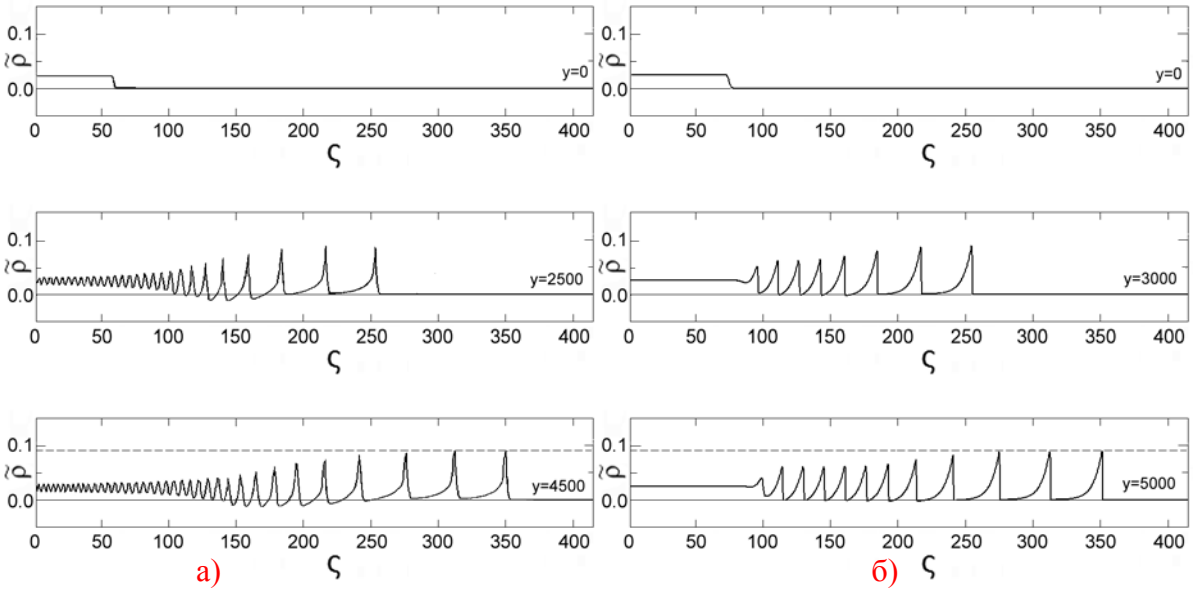


Рисунок 4. Результаты численного моделирования с помощью исходной системы уравнений (а) и нелинейного акустического уравнения (б). Показан распад начального возмущения в виде «ступеньки» на серию автоволновых импульсов.

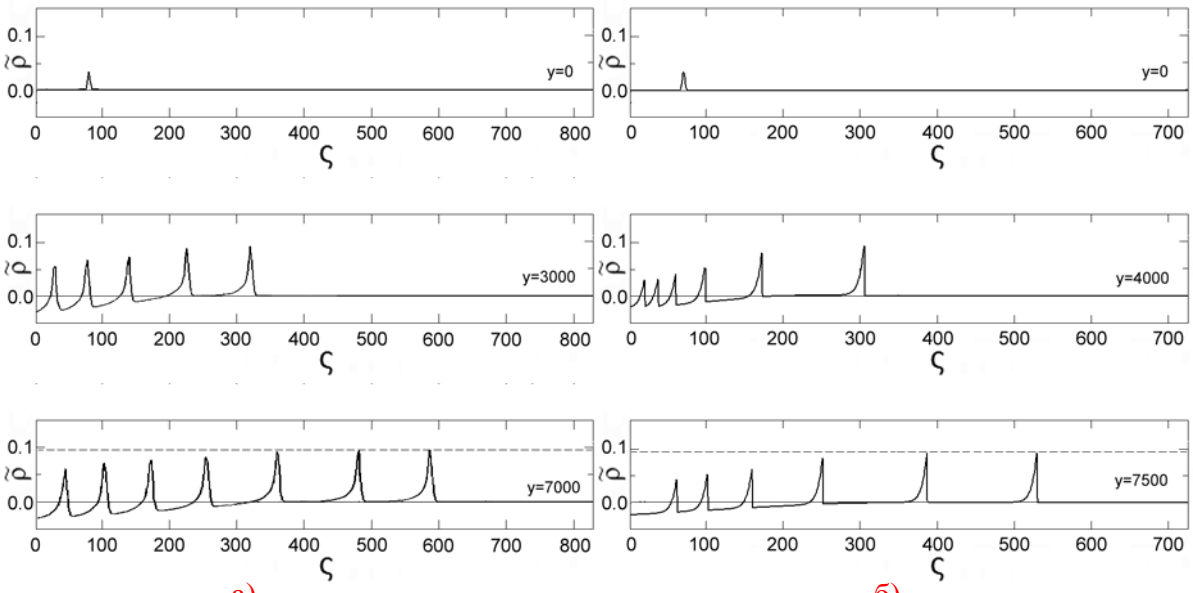


Рисунок 5. Результаты численного моделирования с помощью исходной системы уравнений (а) и нелинейного акустического уравнения (б). Показан распад начального возмущения в виде локализованного гауссова импульса на серию автоволновых импульсов.

Как можно видеть численное моделирование подтверждает гипотезу Краснобаева[3] о самопроизвольном распаде начального возмущения на серию автоволновых импульсов в средах с изоэнтропической неустойчивостью. Также стоит отметить, что нелинейное акустическое уравнение, хотя, и получено при некоторых ограничениях с хорошей точностью описывает эволюцию возмущения в среде.

Работа частично поддержана аналитической целевой программой «Развитие научного потенциала высшей школы» (2009 – 2010 гг.) , проект 2.1.1/309, Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009 – 2013 годы, НК-410 (38) и НК -422 (59)

Литература

1. Hollenbach D., Takahashi T., Tielens, Astrophys. J., 1991 Part 1, V. 377, p. 192-209
2. Wolfire M.G., McKee C. F., Hollenbach D., Tielens, Astrophys. J., 2003 V.587, I.1, p. 278-311
3. Краснобаев К.В., Сысоев Н.Е., Тарев В.Ю. - в сб. Ядерная физика, физика космических излучений и астрономия. М.: МГУ-1993. 243 с.
4. Draine, B. T.; McKee, C. F. Annual review of astronomy and astrophysics 1993 Vol. 31, P. 373-432.
5. Молевич Н.Е., Ораевский А.Н. ЖЭТФ. 1988. Т. 94. №3. С. 128-132.

Метод L-кривой для итерационных алгоритмов решения плохообусловленных задач

А.А. Иванов

Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

Основная трудность применения метода регуляризации Тихонова, лежащего в основе теории решения некорректных задач, состоит в формулировке алгоритмических принципов выбора параметра регуляризации [1]. К современным популярным алгоритмам определения регуляризирующего параметра задачи, не требующим априорной информации о решении, относят метод перекрестной значимости и метод, так называемой, L-кривой.

Существует два подхода к определению параметра регуляризации на основании L-кривой. Суть обоих подходов [2,3,4] состоит в определении точки максимальной кривизны («угла») L-кривой, с тем отличием, что во втором подходе [2,3] кривизна определяется в логарифмических координатах, и L-кривая имеет сходство с английской буквой L. Здесь рассматриваются алгоритмы выбора параметра регуляризации, в которых L-кривая строится в обычных – не логарифмических координатах.

В работе указывается алгоритм выбора параметра регуляризации для задачи регрессионного анализа, основанный на комбинации стратегии метода L-кривой [4] и получения устойчивых решений с несмещенным квадратом длины [5].

Предлагается использовать полученные теоретические результаты для оценки оптимального номера останова итераций при решении возмущенной СЛАУ специальными итерационными методами. Для модифицированного алгоритма наискорейшего спуска в задаче решения интегрального уравнения типа свертки экспериментально показывается, что номер итерации является бифуркационным параметром.

Работа выполнена при поддержке РФФИ (проект № 10-01-00723-а).

Постановка задачи

Рассмотрим систему линейных алгебраических уравнений (СЛАУ)

$$A\,u = \tilde{f},\ A \in R^{M\times N},\ u \in R^{N},\ \tilde{f} \in R^{M},\ M > N, \qquad (1)$$

также положим, что матрица коэффициентов СЛАУ (1) имеет полный столбцовый ранг, т.е. $rank(A) = N$ и вектор $\tilde{f}$ состоит из детерминированной и стохастической компоненты: $\tilde{f} = f + \Delta f$, где Δf - реализация вектора случайных величин с дисперсией $D(\Delta y_i) = M\left(f_i - \tilde{f}_i\right)^2 = d^2 < \infty$ и математическим ожиданием $E(\Delta f_i) = 0$. Запишем регуляризированную систему нормальных уравнений для СЛАУ (1)

$$\left(A^T A + a I\right)u_a = A^T \tilde{f}, \qquad (2)$$

где A^T - суть транспонированная матрица A, а I - единичная матрица размерности $N \times N$.

О методе L-кривой (LC)

L-кривой называют параметрически заданную кривую от параметра a, определяемую соотношениями

$$x(a) = \|u_a\|^2, \quad y(a) = \|A u_a - f\|^2, \tag{3}$$

где x - квадрат нормы регуляризованного решения, а y - соответствующая невязка. В [1] показано, что функции x и y непрерывны по a. При этом $y(a)$ строго монотонно возрастает, а $x(a)$ строго монотонно убывает. Параметр регуляризации a выбирается из условия $a_L = \arg\max_a K(a)$, где $K(a)$ - функция кривизны для (3). С использованием сингулярного (SVD) разложения $A = U S V^T$ и соответствующего представления функций x и y в виде конечных сумм можно получить аналитическое выражение для функции кривизны

$$K(a) = \left(\frac{dy}{da}\right)\left(\frac{d^2x}{da^2}\right) - \frac{d^2y}{da^2}\left(\frac{dx}{da}\right)\left(\left(\frac{dy}{da}\right)^2 + \left(\frac{dx}{da}\right)^2\right)^{-3/2} = -\left(1 + a^2\right)^{-3/2}\left(\frac{dx}{da}\right)^{-1}. \tag{4}$$

В (4) производная может быть вычислена, например, так

$$\frac{dx}{da} = -2\sum_{i=1}^{N} s_i^2 \left(s_i^2 + a\right)^{-3} g_i^2, \tag{5}$$

где $g = U^T f$ и $s_i = S_{i,i}$ - сингулярные числа матрицы A. Стратегию выбора параметра регуляризации будем называть **сходящейся**, если

$$\lim_{d \to 0} \sup_{\|f - f_\delta\| \le \delta} \left\| A_a^+ f_\delta - u^+ \right\| = 0 \tag{6}$$

где $A_a^+ = \left(A^T A + a I\right)^{-1} A^T$, а $u^+ = \left(A^T A\right)^{-1} A^T f$ - псевдорешение точной СЛАУ. Можно показать, что стратегия LC не является сходящей и функция кривизны $K(a)$ строго монотонно возрастает при $0 < a \le \|A\|^2$. Таким образом, всегда $a_L \le \|A\|^2$ независимо от уровня возмущений d.

Сходящаяся стратегия выбора параметра регуляризации

Пусть возмущения задачи удовлетворяют условию

$$\|f - f_\delta\| \le \delta < \|f_\delta\|. \tag{7}$$

В [4] предлагается следующая модифицированная стратегия выбора параметра регуляризации на основании L-кривой:

Алгоритм 1: строится L-кривая (3), строится **вертикальная** линия $y = x^2$, определяется точка пересечения L-кривой и построенной вертикали, строится касательная к L-кривой в точке пересечения, вычисляется тангенс угла наклона (далее наклон) m касательной, параметр принимается $a = -m$.

Предлагаемая стратегия обладает [4] свойством сходимости (6). С использованием сингулярного (SVD) разложения, легко показать, что $\frac{dy}{da} = -a\frac{dx}{da}$, тогда наклон касательной к кривой (3) $\frac{dy}{dx} = \frac{dy}{da}\left(\frac{dx}{da}\right)^{-1} = -a$.

Изложенный алгоритм основан на использовании **принципа невязки** [1], а именно, выбор параметра регуляризации может осуществляться из решения (например, итерационным методом касательных Ньютона) уравнения $r(a) = \left\| A u_a - f \right\|^2 = x^2$.

Выбор параметра регуляризации для оценки решений с несмещенным квадратом длины

В [5] показано, что использование метода наименьших квадратов (МНК) приводит к «дефекту» средней величины суммы квадратов невязок. Для устранения этого дефекта там же предлагается способ вычисления параметра регуляризации из уравнения

$$\left\| A u_a - f \right\|^2 = Q_{\min} + g,\ g > 0, \tag{8}$$

где $Q_{\min} = \left\| (A A^+ - I) f \right\|^2$ - невязка МНК, а A^+ - псевдообратная матрица Мура-Пенроуза. Параметр g просто определить следующим образом $g = N (M - N)^{-1} Q_{\min}$.

Регуляризованные оценки, определяемые из совместного решения уравнений

$$\left\| A u_a - f \right\|^2 = M (M - N)^{-1} Q_{\min} \tag{9}$$

и (2) существуют, единственны и удовлетворяют условию $E\left\| u_a \right\|^2 = \left\| u \right\|^2$, где $u = A^+ f$ - решение точной СЛАУ [5]. Далее нам понадобится следующее очевидное неравенство

$$E\left\| f \right\|^2 = \sum_{i=0}^{M-1} E\left(f_i + \Delta f_i\right)^2 = \sum_{i=0}^{M-1} f_i^2 + \sum_{i=0}^{M-1} E(\Delta f_i)^2 = \sum_{i=0}^{M-1} f_i^2 + M d^2 > M d^2. \tag{10}$$

Воспользуемся соотношением

$$E(Q_{\min}) = (M - N) M^{-1} d^2 \Rightarrow d^2 = M (M - N)^{-1} E(Q_{\min}). \tag{11}$$

Тогда учитывая (10), (11) и (7)

$$E\left\| f - f \right\|^2 = M d^2 = (M - N)^{-1} E(Q_{\min}) \Rightarrow x^2 < E\left\| f \right\|^2. \tag{12}$$

Таким образом, показано, что использование алгоритма 1 для выбора параметра регуляризации при $x = (M - N)^{-1} E(Q_{\min})$ позволяет получить устойчивое решение, в том смысле, что $M\left\| u_a \right\|^2 = \left\| u \right\|^2$ и при этом стратегия выбора параметра регуляризации сходится.

О критерии останова итерационного алгоритма

Уравнение невязки разрешимо относительно параметра a в том случае, если x^2 выбрано из множества допустимых значений. Проведем аналогию между параметром регуляризации a и номером останова некоторого итерационного алгоритма $u_{k+1} = F(x_k, h)$, где h - параметр определяющий малость шага между итерациями. При этом будем считать, что последовательность приближений сходится к псевдорешению СЛАУ (1), а именно

$$\lim_{k \to \infty} \left\| u_k \right\| = \left\| A^+ f \right\|. \tag{13}$$

Обозначим k^{stop} - номер останова итераций и $u_{k_{stop}}$ - полученное приближение. Можно показать, что результат итераций с большим значением k^{stop} соответствует решению СЛАУ (2) при малых a и наоборот. Рассмотрим последовательности $x_k = \|u_k\|_2^2$ и $y_k = \|A u_k - f\|_2^2$. Будем рассматривать такие итерационные методы, в которых при увеличении итераций последовательность x_k не убывает, а y_k не возрастает.

Рассмотрим **модифицированный** алгоритм наискорейшего градиентного [6] спуска

$$u_{k+1} = u_k - b_k w_k,\ w_{k+1} = w_k - b_k B w_k,\ b_k = h \|u_k\|^2 (B w_k, w_k)^{-1}, \qquad (14)$$

где матрица $B = A^T A$ - очевидно, симметричная и положительно определенная. Суть модификации в том, что на каждой итерации требуется лишь одно умножение вектора на матрицу, а не два, как в классическом алгоритме. Итерационный процесс (14) удовлетворяет условию сходимости (13), но при этом обладает сложным характером накопления ошибок, что экспериментально показывается ниже.

Используя (4), предлагается новый критерий останова итерационного алгоритма, основанный на наблюдении за поведением последовательности

$$K(k) = \left(1 + w(k)^2\right)^{3/2} (x_k - x_{k-1})^{-1}, \qquad (15)$$

где $w(k)$ - некоторая возрастающая последовательность, зависящая от номера итерационного процесса. Содержательный смысл критерия останова (15) раскрывается в докладе.

Уравнение невязки (9) также позволяет предложить критерий останова итераций, а именно

$$k^{stop} = \min\left\{k : \|A u_k - f\|^2 \le M (M - N)^{-1} Q_{\min}\right\}, \qquad (16)$$

где k^{stop} - номер останова итераций. В докладе приводятся результаты применения (15) и (16) при решении различных задач обработки изображений с использованием (14).

Выше отмечалось, что итерационный метод (14) обладает сложным характером накопления ошибок. При этом номер итерации является бифуркационным параметром, начиная с некоторого k^{bif}, итерационные приближения становятся неустойчивыми.

Номер останова как бифуркационный параметр алгоритма

В знаменателе (15) находится разность $z(x_k, x_{k-1}) = x_k - x_{k-1}$ (рис.1), наблюдая за которой можно выделить устойчивый и неустойчивый этапы сходимости. Аналогичная картина наблюдается и для последовательности $z(y_k, y_{k-1}) = y_k - y_{k-1}$. С увеличением возмущения правой части СЛАУ (1) неустойчивый этап начинает преобладать, и наоборот. Даже при отсутствии шума, когда СЛАУ (1) совместна, имеет место вычислительная неустойчивость, связанная с ограниченностью разрядной сетки для представления чисел в ЭВМ.

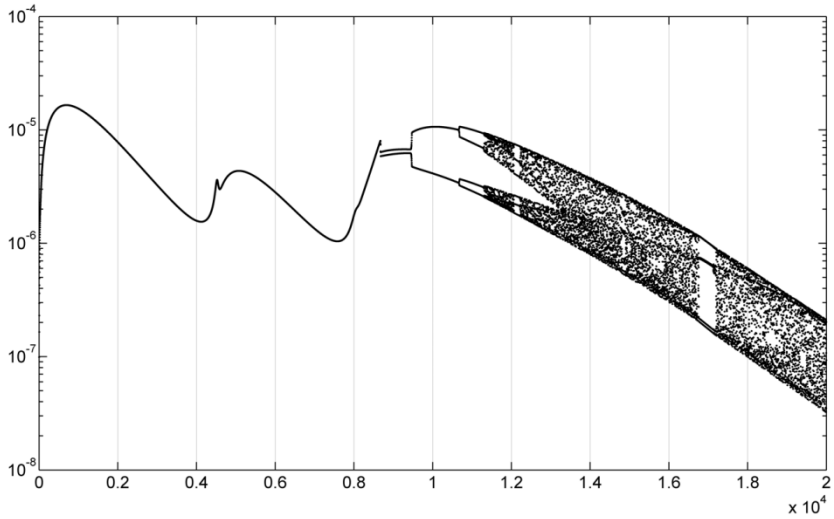


Рис.1. Значения последовательности x_{k+1} - x_k для модифицированного метода наискорейшего спуска для СЛАУ с 1% возмущением вектора правой части.

Литература

1. Морозов В.А. Регулярные методы решения некорректно поставленных задач. – М.: Наука. 1987. – 240 с.

2. Hansen, P.C. The L-Curve and its Use in the Numerical Treatment of Inverse Problems / P.C. Hansen // Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering, vol. 4, WIT Press, Southampton, 2000, pp. 119–142.

3. Vogel, C.R. Non-convergence of the L-curve regularization parameter selection method / C.R. Vogel // Inverse Problems. – 1996. – N 12. – P. 535-547.

4. Engl, H.W. Using the L-curve for determining optimal regularization parameters / H.W. Engl, W. Grever // Numer. Math. – 1994. – N 69. – P. 25-31.

5. Жданов, А.И. Оптимальная регуляризация решений приближенных стохастических систем линейных алгебраических уравнений // ЖВМ и МФ. – 1990. – Т. 29, № 10. – С. 1588-1593.

6. Бахвалов, Н.С. Численные методы / Н.С. Бахвалов, Н.П. Жидков, Г.М. Кобельков. – 4-е изд. – М.: БИНОМ. Лаборатория знаний, 2006. – 636 с.

Теоретический анализ динамики систем реакция-диффузия вблизи точки бифуркации в поле внешних флуктуаций

С.Е. Курушина
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Базовые математические модели самоорганизации едины для широкого класса физических, химических, биологических и экономических систем. К таким моделям относятся обобщенные уравнения Гинзбурга-Ландау, решения которых хорошо изучены. Линейные члены уравнений Гинзбурга-Ландау определяют положение точки бифуркации в пространстве параметров задачи. Нелинейные слагаемые определяют правила отбора взаимодействующих мод и, следовательно, вид возникающих структур, а также ограничивают секулярный рост амплитуд неустойчивых мод. Поэтому разработка новых или модификация существующих аналитических методов сведения исходной системы к базовым моделям представляется важной.

Метод получения обобщенных уравнений Гинзбурга-Ландау для систем, в правой части которых содержатся флуктуирующие во времени, но не в пространстве, дельта - коррелированные силы, обусловленные взаимодействием с внешними резервуарами и диссипацией внутри системы, описан в работе [1]. В [1] не учтены пространственно-распределенные флуктуации и не исследован случай мультипликативных шумов.

Здесь разработан метод получения обобщенных уравнений Гинзбурга-Ландау для систем реакция-диффузия, находящихся в поле внешних флуктуаций:

$$\frac{\partial x_1}{\partial \tau} = P(x_1, x_2, \chi_0, \chi_{10} \ldots \chi_n) + f_1(\vec{r}, \tau) P_2(x_1, x_2, \chi_{k+1}, \ldots \chi_n) + F_1(\vec{r}, \tau) + \nabla^2 x_1,$$

$$\frac{\partial x_2}{\partial \tau} = Q(x_1, x_2, \eta_0, \eta_{10} \ldots \eta_n) + f_2(\vec{r}, \tau) Q_2(x_1, x_2, \eta_{k+1}, \ldots \eta_n) + F_2(\vec{r}, \tau) + D\nabla^2 x_2, \qquad (1)$$

где x_1, x_2 – динамические переменные, $\chi_0 \ldots \chi_n$, $\eta_0 \ldots \eta_n$ – параметры задачи, D, – отношение коэффициентов диффузии, P, Q, P_2 и Q_2 – нелинейные функции, описывающие взаимодействие подсистем, вид которых определяется конкретной задачей. Случайные поля $f_i(\vec{r}, t)$ определяют пространственно-временные гауссовы флуктуации параметров χ_1, η_1 с корреляционной функцией:

$$\left\langle f_i(\vec{r}', t') f_j(\vec{r}'', \tau) \right\rangle = \Phi_i(|\vec{r}' - \vec{r}''|) \exp(-k_{ti} |t' - \tau|) \delta_{ij} \quad (i,j=1,2) \qquad (2)$$

и нулевыми средними значениями. Функция $\Phi_i(|\vec{r}' - \vec{r}''|)$ задает пространственную зависимость корреляций случайного поля. Время корреляции $r_t = k_t^{-1}$ - характерный временной масштаб внешних флуктуаций. Случайные поля $F_i(\vec{r}', \tau)$ с функцией корреляции вида (2) и нулевыми средними значениями учитывают воздействие случайных сил.

Вид корреляционной функции (2) выбран в соответствии с результатами работы [2], где показано, что для большинства реальных систем, имеющих немарковское поведение, хорошим приближением для моделирования флуктуирующей окружающей среды является стационарное однородное изотропное гауссово поле с экспоненциальной функцией корреляции.

Система (1) представляется в операторном виде:

$$\frac{\partial \vec{q}}{\partial \tau} - K(\nabla^2)\vec{q} = \vec{g} + \vec{F}(\vec{r}, \tau) . \qquad (3)$$

Здесь вектор $\vec{q}$ описывает отклонения динамических переменных от устойчивого положения равновесия x_{10} и x_{20}, возникающего при некоторых значениях параметра, выбранного в качестве бифуркационного. Линейный оператор $K(\nabla^2)$ и нелинейная часть $\vec{g}$ получены разложением в ряд детерминированной части (1), $\vec{F}(\vec{r},\tau)$ - случайная компонента уравнений, получающаяся в результате разложения в ряд функций $P_2(x_1,x_2,\chi_{k+1},...\chi_n)$ и $Q_2(x_1,x_2,\eta_{k+1},...\eta_n)$.

$$\vec{q}=\begin{pmatrix} x_1-x_{10} \\ x_2-x_{20} \end{pmatrix},\ K(\nabla^2)=\begin{pmatrix} a_{11}+\nabla^2 & a_{12} \\ a_{21} & a_{22}+D\nabla^2 \end{pmatrix}, a_{1j}=\left.\frac{\partial P}{\partial x_j}\right|_{x_{10},x_{20}}, a_{2j}=\left.\frac{\partial Q}{\partial x_j}\right|_{x_{10},x_{20}},$$

$$\vec{g}=\begin{pmatrix} \sum\limits_{\mu,\nu=1}^{2} g^{(2)}_{1,\mu\nu} q_\mu q_\nu + \sum\limits_{\mu,\nu,\kappa=1}^{2} g^{(3)}_{1,\mu\nu\kappa} q_\mu q_\nu q_\kappa \\ \sum\limits_{\mu,\nu=1}^{2} g^{(2)}_{2,\mu\nu} q_\mu q_\nu + \sum\limits_{\mu,\nu,\kappa=1}^{2} g^{(3)}_{2,\mu\nu\kappa} q_\mu q_\nu q_\kappa \end{pmatrix},$$

$$\vec{F}=\begin{pmatrix} F_1(\vec{r},\tau)+f_1(\vec{r},t)\left[p^{(0)}_1+\sum\limits_{\mu=1}^{2} p^{(1)}_{1,\mu} q_\mu + \sum\limits_{\mu,\nu=1}^{2} p^{(2)}_{1,\mu\nu} q_\mu q_\nu \right] \\ F_2(\vec{r},\tau)+f_2(\vec{r},t)\left[p^{(0)}_2+\sum\limits_{\mu=1}^{2} p^{(1)}_{2,\mu} q_\mu + \sum\limits_{\mu,\nu=1}^{2} p^{(2)}_{2,\mu\nu} q_\mu q_\nu \right] \end{pmatrix}.$$

Здесь $g^{(n)}, p^{(m)}$ - коэффициенты разложения нелинейных функций (1) в ряд Тейлора.

Используя процедуру построения волновых пакетов и представляя решение задачи (3) в виде суперпозиции $\vec{q}(\vec{r},\tau)=\sum\limits_{\vec{k},j}\vec{O}^{(j)}(\vec{k})\xi^{(j)}_{\vec{k}}(\tau)e^{i\vec{k}\vec{r}}$, где $\vec{O}^{(j)}$ - собственные векторы линейного оператора $K(\nabla^2)$, получаются уравнения для амплитуд собственных мод:

$$\frac{d\xi^{(j)}_{\vec{k}}}{d\tau}-\lambda_j(\vec{k})\xi^{(j)}_{\vec{k}}=\sum_{\substack{lm \\ \vec{k}'\vec{k}''}}\sigma^{(j)}_{lm}\xi^{(l)}_{\vec{k}'}\xi^{(m)}_{\vec{k}''}+\sum_{\substack{lmn \\ \vec{k}'\vec{k}''\vec{k}'''}}\sigma^{(j)}_{lmn}\xi^{(l)}_{\vec{k}'}\xi^{(m)}_{\vec{k}''}\xi^{(n)}_{\vec{k}'''}+F_{\vec{k}}\left(\xi,z\right), \qquad (4)$$

где $\sigma^{(j)}_{km}=\sigma^{(j)}_{lm}(\vec{k},\vec{k}',\vec{k}'')\delta(\vec{k}'+\vec{k}''-\vec{k})$, $\sigma^{(j)}_{lmn}=\sigma^{(j)}_{lmn}(\vec{k},\vec{k}',\vec{k}'',\vec{k}''')\delta(\vec{k}'+\vec{k}''+\vec{k}'''-\vec{k})$. Показано, что для аддитивных шумов функция $F_{\vec{k}}\left(\xi,z\right)$ совпадает с полученной в работе [1] и имеет вид $F_{\vec{k}}\left(\xi,z\right)=\sum\limits_{\varepsilon=1}^{2}O^{*(j)}_{\varepsilon}(\vec{k})z_{\varepsilon,\vec{k}}$. При наличии мультипликативного шума функция $F_{\vec{k}}(\xi,z)$ получена впервые и имеет вид

$$F_{\vec{k}}\left(\xi,z\right)=\sum_{\varphi=1}^{2}O^{*(j)}_{\varphi}p^{(0)}_{\varphi}z_{\varphi,\vec{k}}+\sum_{\varphi=1}^{2}\sum_{l,\vec{k}'}\varepsilon^{(j)}_{\varphi,l}(\vec{k},\vec{k}')\xi^{(l)}_{\vec{k}'}z_{\varphi,\vec{k}-\vec{k}'}+\sum_{\varphi=1}^{2}\sum_{\substack{lm, \\ \vec{k}'\vec{k}''}}\varepsilon^{(j)}_{\varphi,lm}(\vec{k},\vec{k}',\vec{k}'')\xi^{(l)}_{\vec{k}'}\xi^{(m)}_{\vec{k}''}z_{\varphi,\vec{k}-\vec{k}'-\vec{k}''}.$$

Здесь $\lambda_j(\vec{k})$ - корни характеристического уравнения линеаризованной детерминированной системы. Функции $\sigma^{(j)}_{lm}(\vec{k},\vec{k}',\vec{k}'')$, $\sigma^{(j)}_{lmn}(\vec{k},\vec{k}',\vec{k}'',\vec{k}''')$, $\varepsilon^{(j)}_{\varphi,l}(\vec{k},\vec{k}')$, $\varepsilon^{(j)}_{\varphi,lm}(\vec{k},\vec{k}',\vec{k}'')$ определяются собственными векторами оператора $K(\nabla^2)$, коэффициентами $g^{(n)}, p^{(m)}$ и $O^{*(j)}_{\varphi}=O^{*(j)}_{\varphi}(\vec{k})$ - компонентами собственных векторов оператора, сопряженного к $K(\nabla^2)$. $z_{j,\vec{k}}=z_{j,\vec{k}}(t)=\int e^{-i\vec{k}\vec{r}}f_j(\vec{r},t)d\vec{r}$ - компоненты случайного векторного поля $\vec{z}(t)$. В предположении, что времена корреляций

случайных полей гораздо меньше всех характерных времен задачи (1), корреляционные функции для компонент поля $\vec{z}(t)$ будут иметь вид: $\left\langle z_{j,\vec{k}}(t) z_{l,\vec{k}'}(\tau) \right\rangle = g_{jl}(|\vec{k}|)\, \delta(\vec{k}-\vec{k}')\delta(t-\tau)\delta_{jl}$.

Собственные моды с использованием дисперсионного соотношения детерминированной системы (1) разделяются на устойчивые υ_K и неустойчивые $\zeta_{\vec{k}}$. Считая, что изменения амплитуд происходят согласовано $\upsilon \sim \zeta^2$ и, полагая, что $|\upsilon| << |\zeta|$, проводится процедура адиабатического исключения устойчивых мод, в результате чего получаются обобщенные стохастические уравнения Гинзбурга-Ландау

$$\frac{d\zeta_{\vec{k}}}{d\tau} - \lambda_1(\vec{k})\zeta_{\vec{k}} = \Lambda^{(0)}_{\vec{k}} + \sum_{\vec{k}'} \Lambda^{(1)}_{\vec{k}'}\zeta_{\vec{k}'} + \sum_{\vec{k}'\vec{k}''} \Lambda^{(2)}_{\vec{k}'\vec{k}''}\zeta_{\vec{k}'}\zeta_{\vec{k}''} + \sum_{\vec{k}'\vec{k}''\vec{k}'''} \Lambda^{(3)}_{\vec{k}'\vec{k}''\vec{k}'''}\zeta_{\vec{k}'}\zeta_{\vec{k}''}\zeta_{\vec{k}'''} + \ldots \qquad (5)$$

В частном случае аддитивных шумов без учета пространственной корреляции (5) совпадает с системой, полученной в [1].

Ниже приводятся выражения для операторов $\Lambda^{(n)}$ для случая мультипликативных шумов при наличии в исходной системе неустойчивости типа мягкой моды ($\mathrm{Im}(\lambda_{1,2}) = 0$ и $\mathrm{Re}(\lambda_1) \geq 0$). Оператор

$$\Lambda^{(0)}_{\vec{k}} = \sum_{\varphi=1}^{2} O^{*}_{\varphi}(\vec{k}) p^{(0)}_{\varphi} z_{\varphi,\vec{k}} - \sum_{\psi,\varphi,\varphi'=1}^{2} \sum_{\vec{K}} \varepsilon^{(1)}_{\varphi,\psi}(\vec{k},\vec{K}) \frac{O^{*(\psi)}_{\varphi'}(\vec{K})}{\lambda_{\psi}(\vec{K})} p^{(0)}_{\varphi'} z_{\varphi,\vec{k}-\vec{K}} z_{\varphi',K}$$

приводит к смещению статистически стационарных состояний мод в поле мультипликативных флуктуаций, оператор

$$\Lambda^{(1)}_{\vec{k}'} = \sum_{\varphi=1}^{2} \varepsilon^{(1)}_{\varphi,1}(\vec{k},\vec{k}') z_{\varphi,\vec{k}-\vec{k}'} - \sum_{\psi,\varphi=1}^{2} \sum_{\vec{K}} (\sigma^{(1)}_{1\psi}(\vec{k},\vec{k}',\vec{K}) + \sigma^{(1)}_{\psi 1}(\vec{k},\vec{K},\vec{k}')) \frac{O^{*(\psi)}_{\varphi}(\vec{K})}{\lambda_{\psi}(\vec{K})} p^{(0)}_{\varphi} \delta(\vec{k}'+\vec{K}-\vec{k}) z_{\varphi,K} -$$

$$- \sum_{\psi,\varphi,\varphi'=1}^{2} \sum_{\vec{K},} \varepsilon^{(1)}_{\varphi,\psi}(\vec{k},\vec{K}) \frac{\varepsilon^{(\psi)}_{\varphi',1}(\vec{K},\vec{k}')}{\lambda_{\psi}(\vec{K})} z_{\varphi',\vec{K}-\vec{k}'}(\tau) z_{\varphi,\vec{k}-\vec{K}} - \sum_{\psi,\varphi,\varphi'=1}^{2} \sum_{\vec{K}} (\varepsilon^{(1)}_{\varphi,1\psi}(\vec{k},\vec{k}',\vec{K}) + \varepsilon^{(1)}_{\varphi,\psi 1}(\vec{k},\vec{K},\vec{k}')) \frac{O^{*(\psi)}_{\varphi'}(\vec{K})}{\lambda_{\psi}(\vec{K})} p^{(0)}_{\varphi'} z_{\varphi,\vec{k}-\vec{k}'-\vec{K}} z_{\varphi',\vec{K}},$$

определяет линейную по интенсивностям флуктуаций поправку к инкременту моды $\zeta_{\vec{k}}$, оператор

$$\Lambda^{(2)}_{\vec{k}'\vec{k}''} = \sigma^{(1)}_{11}(\vec{k},\vec{k}',\vec{k}'')\delta(\vec{k}'+\vec{k}''-\vec{k}) + \sum_{\varphi=1}^{2} \varepsilon^{(1)}_{\varphi,11}(\vec{k},\vec{k}',\vec{k}'') z_{\varphi,\vec{k}-\vec{k}'-\vec{k}''} -$$

$$- \sum_{\psi,\varphi=1}^{2} \sum_{\vec{K}} (\sigma^{(1)}_{1\psi}(\vec{k},\vec{k}',\vec{K}) + \sigma^{(1)}_{\psi 1}(\vec{k},\vec{K},\vec{k}')) \frac{\varepsilon^{(\psi)}_{\varphi,1}(\vec{K},\vec{k}'')}{\lambda_{\psi}(\vec{K})} \delta(\vec{k}'+\vec{K}-\vec{k}) z_{\varphi,\vec{K}-\vec{k}''} -$$

$$- \sum_{\psi,\varphi=1}^{2} \sum_{\vec{K}} \varepsilon^{(1)}_{\varphi,\psi}(\vec{k},\vec{K}) \frac{\sigma^{(\psi)}_{11}(\vec{k},\vec{k}',\vec{k}'')}{\lambda_{\psi}(\vec{K})} \delta(\vec{k}'+\vec{k}''-\vec{K}) z_{\varphi,\vec{k}-\vec{K}}$$

является оператором трехмодового взаимодействия неустойчивых мод (первое и второе слагаемые) и устойчивых мод с неустойчивыми (третье и четвертое слагаемые) и также определяет линейную поправку к инкременту моды $\zeta_{\vec{k}}$, оператор $\Lambda^{(3)}_{\vec{k}'\vec{k}''\vec{k}'''} = \delta(\vec{k}'+\vec{k}''+\vec{k}'''-\vec{k})\omega(\vec{k},\vec{k}',\vec{k}'',\vec{k}''',\vec{K})$ является оператором четырехмодового взаимодействия.

Уравнения (5) усредняются по ансамблю реализаций с учетом их формальных решений и правил раскрытия моментов и расщепления корреляторов.

В результате впервые получена система уравнений для моментов параметров порядка, структура которой имеет вид

$$\frac{\partial\left\langle\zeta_{\vec{k}}\right\rangle}{\partial t}-\lambda_1(\vec{k})\left\langle\zeta_{\vec{k}}\right\rangle=L_{\vec{k}}^{(0)}+L_{\vec{k}\vec{k}',\vec{k}'',\vec{K}}^{(1)}\left\langle\zeta_{\vec{k}''}\right\rangle+L_{\vec{k},\vec{k}',\vec{k}'',\vec{k}''',\vec{K}}^{(2)}\left\langle\zeta_{\vec{k}''}\zeta_{\vec{k}'''}\right\rangle+L_{\vec{k},\vec{k}',\vec{k}'',\vec{k}''',\vec{k}^{IV},\vec{K},\vec{K}'}^{(3)}\left\langle\zeta_{\vec{k}''}\zeta_{\vec{k}'''}\zeta_{\vec{k}^{IV}}\right\rangle. \quad (6)$$

Выражения для операторов $L_{\vec{k}}^{(0)}$, $L_{\vec{k}\vec{k}',\vec{k}'',\vec{K}}^{(1)}$, $L_{\vec{k},\vec{k}',\vec{k}'',\vec{k}''',\vec{K}}^{(2)}$, $L_{\vec{k},\vec{k}',\vec{k}'',\vec{k}''',\vec{k}^{IV},\vec{K},\vec{K}'}^{(3)}$ здесь не представлены ввиду их громоздкости.

Полагая в (2) $\Phi_j(|\vec{r}'-\vec{r}''|)=\theta_j\varphi_j(|\vec{r}'-\vec{r}''|)$, при малых значениях интенсивностей флуктуаций θ_j, когда можно ограничиться членами, линейными по интенсивностям шума, из (6) следуют выражения для инкрементов незатухающих мод, которые получены для частных случаев неустойчивостей типа мягкой моды ($\mathrm{Im}(\lambda_{1,2})=0$ и $\mathrm{Re}(\lambda_1)\geq 0$), типа жесткой моды ($\mathrm{Im}(\lambda_{1,2})\neq 0$ и $\mathrm{Re}(\lambda_{1,2})\geq 0$) при аддитивном шуме и неустойчивости типа мягкой моды ($\mathrm{Im}(\lambda_{1,2})=0$ и $\mathrm{Re}(\lambda_1)\geq 0$) при мультипликативном шуме.

Анализ системы (6) позволил выявить следующие особенности поведения системы типа «реакция-диффузия» во флуктуирующих полях вблизи точки бифуркации Тьюринга.

После усреднения возникают дополнительные детерминированные слагаемые, входящие в оператор $L_{\vec{k}}^{(0)}$ и не зависящие от амплитуд параметров порядка. Наличие таких слагаемых в уравнениях типа (6) приводит к смещению статистически стационарного состояния каждой моды.

При малой интенсивности мультипликативных флуктуаций в процессе образования диссипативных структур изменяются условия резонансного взаимодействия неустойчивых мод: возникает дополнительное условие $2\vec{k}'-\vec{k}''=\vec{k}$. При увеличении интенсивности шума взаимодействие становится многомодовым: условия отбора включают пять и более мод.

Из выражений для инкрементов неустойчивых мод следует, что наличие внешнего шума приводит к изменению инкремента нарастания каждой моды и изменению области неустойчивости. Это обуславливает зависимость скорости образования пространственных диссипативных структур от параметров случайного поля и смещение точки бифуркации Тьюринга (шумоиндуцированное параметрическое возбуждение системы).

Литература

1. Хакен Г. Синергетика. М.: Мир, 1980. 406 с.

2. Хорстхемке В., Лефевр Р. Индуцированные шумом переходы*:* теория и применение в физике, химии и биологии. М.: Мир, 1987. 400 с.

Исследование процессов формо- и фазообразования в неравновесных малых металлических частицах

Н.И. Лиманова, Е.А. Талалова, М.М. Ворона, Ю.Р. Хайруллова, П.А. Шишкин[1]
Тольяттинский государственный университет, г. Тольятти
[1]*Московский физико-технический институт (государственный ун-т), г. Москва*

Введение

Математическое моделирование процессов формо- и фазообразования в малых металлических частицах (ММЧ), получаемых путем осаждения из газовой фазы или на начальных этапах процесса электрокристаллизации, представляет собой актуальную задачу. Подобные процессы получения нано- и микрочастиц являются существенно неравновесными. Существенно неравновесной и нестабильной является и структура частиц, образующихся в таких процессах. Эволюция частиц сопровождается структурной релаксацией, изменением их формы и порообразованием. Однако требования, предъявляемые в настоящее время к технологиям получения нано- и микрообъектов, предполагают стабильность (метастабильность) таких объектов и их свойств при определенных условиях в течение заданных промежутков времени. Моделирование структурообразования ММЧ является важным инструментом в исследовании данных процессов. Исследование свойств металлических наноразмерных образований актуально как с фундаментальной, так и с прикладной точек зрения. Свойства малых частиц весьма чувствительны к малым изменениям их структуры, формы, размеров. В данной работе процесс эволюции неравновесных ММЧ моделируется на основе метода клеточных автоматов. В рамках данного метода применяется модель ячеечной перколяции, уточненная авторами на основе задания параметра связности. Объединение метода клеточных автоматов и перколяционной модели позволяет описать процессы формо- фазообразования в ММЧ.

Диффузионно-деформационная модель

ММЧ, получаемые при осаждении из газовой фазы или на начальных этапах процесса электрокристаллизации, являются неравновесными напряженными структурами. В процессе эволюции состояние частицы претерпевает значительные изменения, характер которых напрямую зависит от начального состояния частицы и условий релаксации. В данной работе начальное состояние частицы моделируется как аморфное с дефектом дисклинационного типа. Обычно диффузионные процессы направлены таким образом, что частица, первоначально находящаяся в неравновесном состоянии, релаксирует к равновесному или метастабильному состоянию. В ММЧ наблюдаются некоторые особенности протекания релаксационных процессов, что связано с тем, что частица обладает развернутыми границами и неоднородными напряжениями. Релаксационные процессы в частице и формирующаяся при этом структура существенным образом регулируются ее размерами, формой, состоянием и внешними условиями. Структурную релаксацию и релаксацию напряжений можно свести к процессу диффузии. Диффузионными процессами обычно называют процессы перераспределения вещества в среде путем диссипации или его направленного переноса под действием градиента химического потенциала. При этом внешние и внутренние границы являются активными стоками, а межфазные границы и поверхности обладают существенно более высоким коэффициентом диффузии по сравнению с объемным значением в кристаллической фазе, наличие напряжений в объеме частицы и на поверхности приводят к появлению восходящих диффузионных потоков. В качестве основного уравнения для моделирования было взято уравнение

диффузии для изотермических потоков, полученное на основе термодинамики неравновесных процессов:

$$\vec{j_v} = -D_v\left(\nabla C_v - \frac{C_v \nabla \sigma \Omega}{kT}\right),$$

где $\vec{j_v}$ — поток свободного объема (вакансий), D_v — коэффициент диффузии вакансий, C_v — концентрация вакансий (в рамках разработанной модели рассматривалась диффузия вакансий, моделирующая свободный объем), σ — напряжения, Ω — удельный объем. В изотропном веществе тензор напряжений можно заменить функцией. Уравнение диффузии становится неоднородным, когда в объеме, где происходит чистая диффузия, имеются источники-стоки диффундирующего компонента. Выбор граничных и начальных условий определялся физикой процесса и морфологией частиц.

Граничные условия задавались в зависимости от состояния поверхности частицы в начальный момент времени — ее температуры, структуры и свойств граничащей с ней фазы. Простейший случай статических граничных условий — граничные условия 1-ого рода — внешняя поверхность - поверхность стока вакансий $C_v(\vec{r}_b, t) = 0$. Здесь $\vec{r}_b$ задается уравнением поверхности частицы.

Метод клеточных автоматов позволяет реализовать динамические граничные условия. В этом случае массив клеток поверхности пересчитывается, при этом и форма, и свойства поверхности частицы в процессе диффузии изменяются. Граничные условия третьего рода позволяют отслеживать стоки на границах и являются более подходящими в условиях нестационарной диффузии: $\left(-D_v \cdot \nabla C_v(\vec{r}_b, t)\big|_{\vec{r}_b} \cdot \vec{n}\right) = \alpha \cdot C_v(\vec{r}_b, t).$

Здесь α характеризует скорость вакансий на границе. Граничные условия третьего рода задаются при условии протекания процесса электроосаждения. Мы использовали комбинацию динамических граничных условий на поверхности частицы и потоковых на внутреннем дисклинационном стоке.

Компьютерная реализация модели

Процесс эволюции кристалла, моделируемый методом клеточных автоматов, сопровождается изменением количества атомов в каждой ячейке с течением времени, соответственно, изменяется свободный объем и пространственная связность с соседними клетками автомата. Важным моментом является то, что перколяция на конечных ячейках с разной вероятностью наблюдается в некотором диапазоне вероятностей заполнения. Причем ширина этой области зависит от размеров ячейки. С физической точки зрения эта область соответствует метастабильной области фазового перехода, которая тем шире, чем меньше размер рассматриваемой системы. При этом особую роль играет зависимость вероятности перколяции P_p от вероятности случайного заполнения $P_з$, характеризующая поведение системы в области фазового перехода. Таким образом, одной из задач, решаемых в данной работе, является расчет зависимостей вероятностей перколяции от вероятности случайного заполнения для случаев связности по одному, по двум и по трем направлениям для решеток разных типов и размеров. В компьютерной реализации модели для идентификации кластеров и определения порога перколяции был использован алгоритм Хошена-Копельмана [1].

Авторами предложены перколяционные модели, уточненные на основе задания параметра связности и реализованные на плоских квадратной, треугольной, гексагональной и объемной кубической решетках. Модели позволяют анализировать наличие связности по одному, двум и по трем направлениям. Разработан комплекс программ, реализующих перколяционные модели для простой квадратной, треугольной, гексагональной и простой кубической решеток. Программы созданы в

среде разработки Visul C++, визуализация результатов выполнена с использованием графической библиотеки Open GL. Программное обеспечение предоставляет пользователю возможность задания размеров области моделирования (максимум 1000x1000 ячеек для плоских решеток) и вероятности перколяции. Для проверки правильности кластеризации маркировка кластеров для всех моделей выполняется различными, не повторяющимися цветами. Результатом вычислений является количественная информация о наличии горизонтальной или вертикальной и одновременно горизонтальной и вертикальной перколяции, а также визуальное отображение перколяционных кластеров для каждого вычислительного эксперимента.

Результаты

Ниже представлены зависимости вероятностей перколяции от вероятности случайного заполнения, полученные на основе вычислительного эксперимента. Значения вероятностей перколяции рассчитывались с помощью разработанного комплекса программ для перколяционных моделей на плоских квадратной, треугольной, гексагональной и объемной кубической решетках. В рамках вычислительного эксперимента каждое значение вероятности перколяции определялось на основе 1000 запусков программы для соответствующей вероятности случайного заполнения на решетках с различными условными линейными размерами для случаев связности по одному и по двум направлениям для плоских решеток, а также по одному, по двум и по трем направлениям для кубической решетки. На рис. 1 приведены указанные зависимости для случаев связности по одному и по двум направлениям на основе модели на плоской квадратной решетке для полей размером 10 x 10, 50 x 50, 100 x 100, 500 x 500 и 1000 x 1000, на рис. 2 — зависимости, полученные на основе модели на плоской треугольной решетке и на рис. 3 — зависимости, полученные на основе плоской гексагональной решетки. Размер поля моделирования выбирался в зависимости от реализуемой модели и имел соответствие с пространственным масштабом ячейки клеточного автомата. С размера 10 x 10 начинается нанометровая область, а размер 1000 x 1000 соответствует микродиапазону.

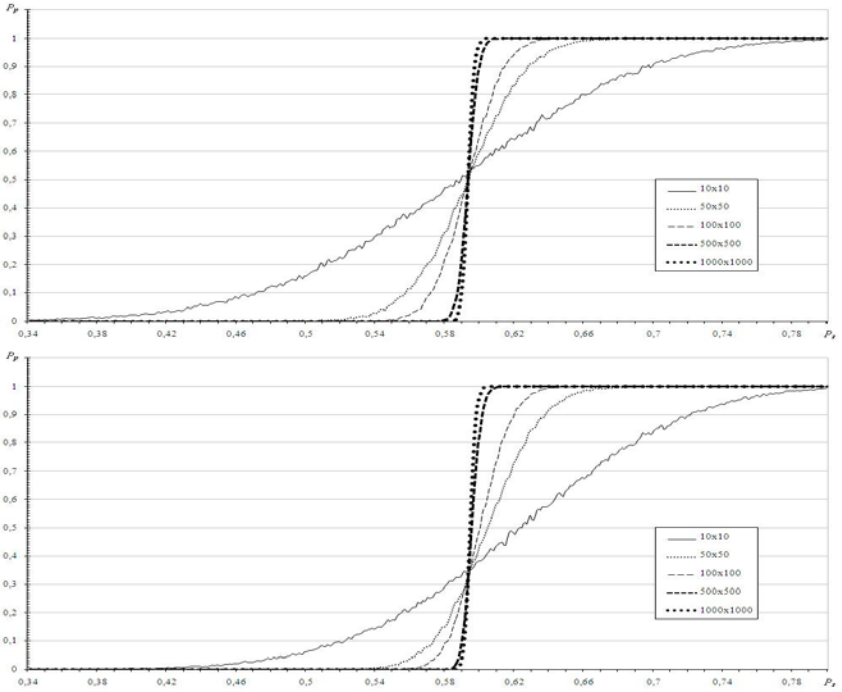


Рис. 1. Зависимости вероятностей перколяции от вероятности случайного заполнения для случаев связности по одному направлению (сверху) и одновременно по двум направлениям (снизу) для полей размером 10 x 10, 50 x 50, 100 x 100, 500 x 500 и 1000 x 1000 для квадратной решетки

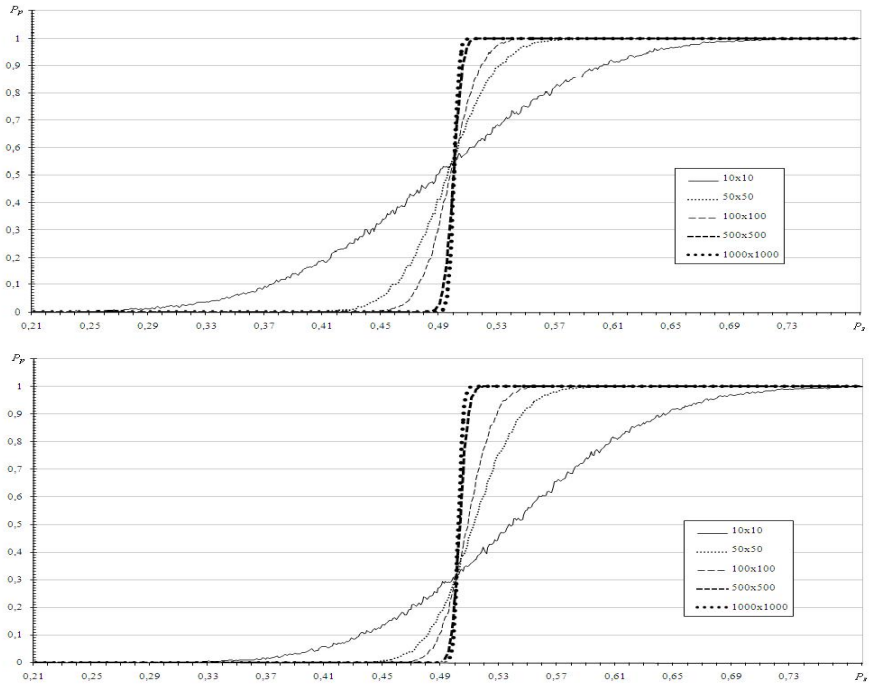


Рис. 2. Зависимости вероятностей перколяции от вероятности случайного заполнения для случаев связности по одному направлению (сверху) и одновременно по двум направлениям (снизу) для полей размером 10 x 10, 50 x 50, 100 x 100, 500 x 500 и 1000 x 1000 для треугольной решетки

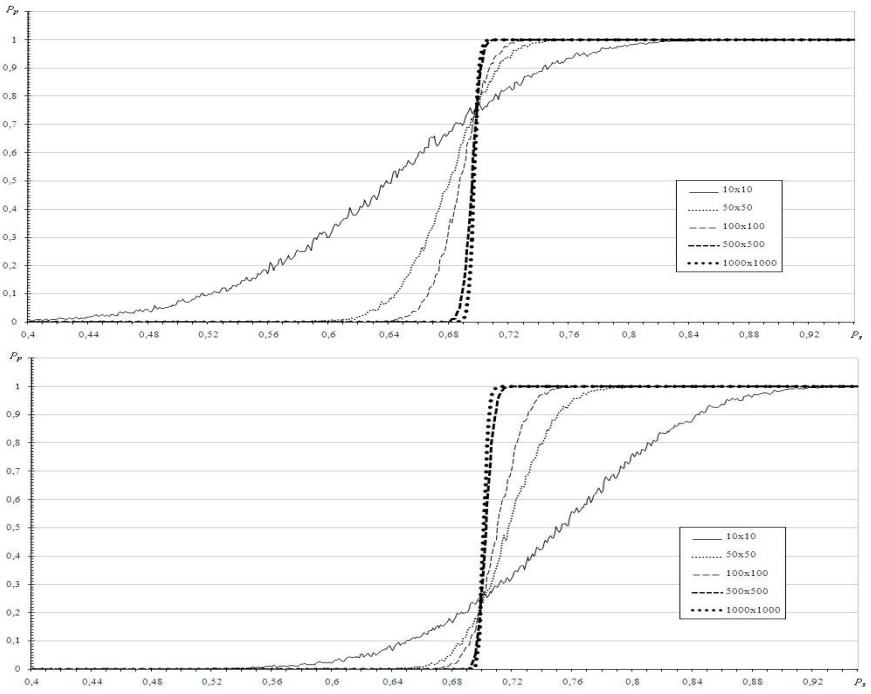


Рис. 3. Зависимости вероятностей перколяции от вероятности случайного заполнения для случаев связности по одному направлению (сверху) и одновременно по двум направлениям (снизу) для полей размером 10 x 10, 50 x 50, 100 x 100, 500 x 500 и 1000 x 1000 для гексагональной решетки

Аналогичные зависимости были получены для перколяционных моделей на кубической решетке. Графики этих зависимостей представлены на рис. 4.

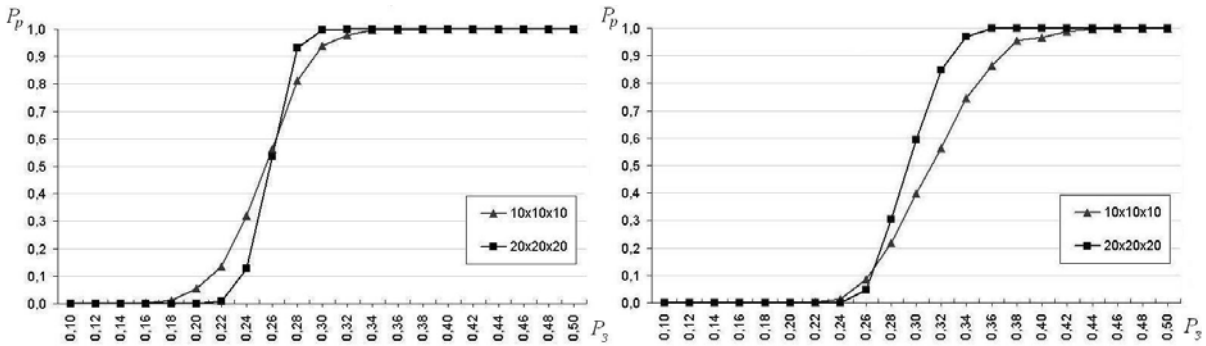


Рис. 4. Зависимости вероятностей перколяции от вероятности случайного заполнения для случаев связности по одному направлению (слева) и одновременно по трем направлениям (справа) для областей моделирования 10x10x10 и 20x20x20 для кубической решетки

На графиках зависимостей вероятностей перколяции для конечных решеток точка пересечения графиков соответствует пределу перколяции для бесконечной решетки соответствующего типа. Полученные таким образом значения пределов перколяции хорошо согласуются с теоретическими данными [2,3]. При этом предел перколяции для бесконечной решетки обеспечивает возникновение соединяющего кластера на конечной решетке только с некоторой вероятностью, значение которой уменьшается с увеличением размерности связности. При этом метастабильная область теряет симметричность относительно предела перколяции для бесконечной решетки.

Размер поля моделирования имеет соответствие с пространственным масштабом ячейки клеточного автомата.

Заключение

Результаты моделирования позволили в явном виде получить зависимости вероятностей перколяции от вероятности случайного заполнения на квадратных, треугольных, гексагональных и кубических решетках с различными условными линейными размерами. При этом было обнаружено, что зависимости существенно различаются для случаев одномерной, двумерной и трехмерной перколяции. Результаты двумерного моделирования могут быть использованы при исследованиях процессов фазовых превращений в тонких пленках и жидких кристаллах. Трехмерные модели могут применяться для изучения механизмов структурообразования в объемных материалах.

Работа выполнена по заданию МО в рамках тематического плана Тольяттинского государственного ун-та № 01.2.00 951120.

Литература

1. *Hoshen J., Kopelman R.* Percolation and cluster distribution // Phys. Rev. B. — 1976. — Vol. 14, N. 8. — P. 3438–3445.
2. *Stauffer D., Aharony A*. Introduction to Percolation Theory. — London: Taylor & Francis, 1992.
3. *Grimmet G.* Percolation. — Berlin: Springer-Verlag, 1999.

Моделирование напряженно-деформированного состояния многослойных гетерогенных структур

А.Н. Литвинов, М.А. Литвинов[1]
Пензенский государственный университет, г. Пенза
[1]*ФГУП ФНПЦ "ПО "Старт" им. М.В. Проценко", г. Заречный*

Гетерогенные структуры широко применяются в конструкциях технических систем различного назначения, работающих в сложных условиях эксплуатации при действии статических и динамических нагрузок [1]. К таким структурам относятся микросборки приборов, микросхемы, упругие элементы датчиков для измерения механических величин, многослойные вибродемпфирующие покрытия различного конструктивного исполнения для радиоэлектронной и специальной аппаратуры военного применения.

Одним из способов исследования напряженно-деформированного состояния (НДС) в гетерогенных структурах является создание моделей с приведенными физико-механическими характеристиками. В этом случае основной подход состоит в том, что гетерогенная структура заменяется эквивалентной ей однородной упругой системой, свойства которой характеризуются приведенными физико-механическими характеристиками и в ней имеет место НДС, соответствующее исходной гетерогенной системы. Наиболее сложным является вопрос о выборе критерия эквивалентности между этими системами. Этот способ широко применяется для слоистых и волокнистых материалов, для неоднородных материалов, состоящих из различных фракций, для композиционных материалов различной структуры. Первые работы в этом направлении принадлежат Фойхту и Рейссу, которые предложили проводить осреднение физико-механических свойств по объему, которые описываются прямыми или обратными тензорами второго, третьего или четвертого рангов [2].

Например, для макроскопического изотропного композита, представляющего собой смесь изотропных компонентов с модулями E_j, приведенный модуль упругости по Фойхту (E_ϕ) и Рейссу (E_p)определяются по формулам

$$E_\phi = \sum_{j=1}^{n} \psi_j \cdot E_j, \qquad E_p = \left(\sum_{j=1}^{n} \psi_j \cdot E_j^{-1} \right)^{-1} \tag{1}$$

где ψ_j – относительное объемное содержание; E_j – модуль упругости материала j-го компонента в гетерогенной системе; j =1, 2,…, n; n – число компонентов в системе. Метод Фойхта соответствует предположению об однородности поля обобщенных перемещений, а метод Рейсса – предположению об однородности поля обобщенных сил.

Существенной особенностью этих подходов является то, что метод Фойхта дает оценку сверху, а метод Рейса – оценку снизу для приведенных характеристик гетерогенной системы. При этом расхождение между этими оценками тем значительнее, чем больше отличаются упругие характеристики (например, E_j) компонентов системы. Таким образом, выполняются условия

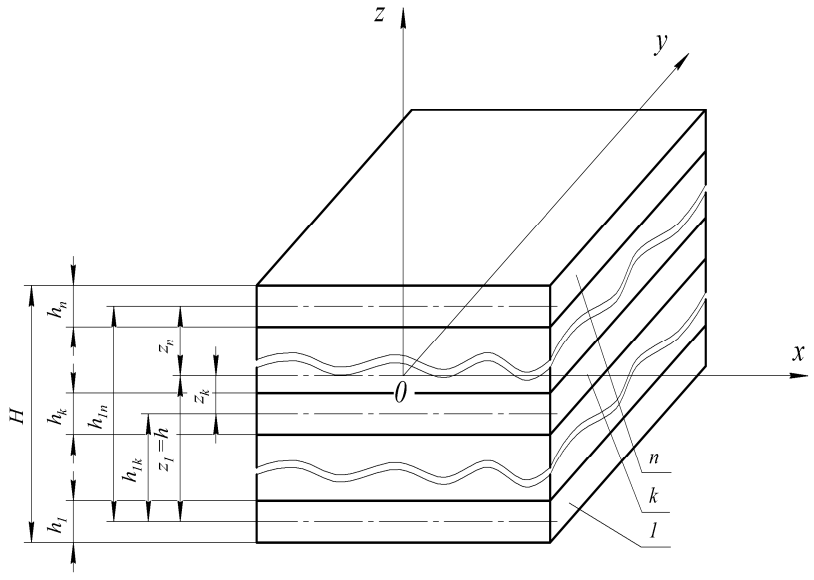
$$E_p < E < E_\phi \ , \tag{2}$$

где E – приведенный модуль гетерогенной системы. Аналогичные оценки можно получить и для других приведенных физико-механических характеристик гетерогенной системы.

Различные модели и критерии для определения приведенных характеристик гетерогенных структур предложены в работах [3, 4], где в отличие от подходов Фойхта и Рейса, осреднение производиться в сочетании с принятием дополнительных гипотез о полях напряжений, деформаций и перемещений, а также конструктивными особенностями рассматриваемой гетерогенной системы. Существенной особенностью различных моделей, в соответствии с которыми определяются приведенные характеристики, является то, что значения этих характеристик укладываются в интервал, определяемый оценками (2).

Модели, основанные на определении приведенных физико-механических характеристик эквивалентной упругой системы, являются весьма перспективными для выполнения практических расчетов на прочность и анализ НДС сложных гетерогенных структур, т.к. в ряде случаев допускают аналитические решения для систем, подверженных различным внешним воздействиям.

Для построения обобщенной модели рассмотрим плоскую многослойную структуру в координатах x, y, z, состоящую из n слоев (рисунок 1). Считаем, что все слои являются упругими, изотропными и характеризуются толщинами h_k, модулями упругости E_k, коэффициентами Пуассона ν_k, где k = 1, 2, …, n – номер слоя. В соответствии с классификацией, принятой в [5], все слои считаем жесткими, т. е. они работают на растяжение (сжатие) и изгиб. Трансверсальной податливостью слоев в направлении оси z пренебрегаем. Заменим плоскую многослойную гетерогенную систему эквивалентной ей в смысле жесткости однослойной конструкцией, которая характеризуется приведенными модулем упругости E, толщиной H и коэффициентом Пуассона ν. В качестве приведенной толщины принимаем суммарную толщину пакета $H = \sum_{k=1}^{n} h_k$, а в качестве приведенного коэффициента Пуассона его усредненное значение – $\nu = \frac{1}{n}\sum_{k=1}^{n} \nu_k$.



__ . __. __ – срединная поверхность k-ого слоя

Рисунок 1. Схема гетерогенной структуры

Цилиндрические жесткости эквивалентной обобщенной модели (D_{np}) и многослойной системы (D) вычисляется как

$$D_{np} = \frac{EH^3}{12\left(1-\nu^2\right)}, \qquad D = \sum_{k=1}^{n}\left(D_k + A_k z_k^2\right) \tag{3}$$

где $D_k = \frac{E_k h_k^3}{12\left(1-\nu_k^2\right)}$ – цилиндрическая жесткость; $A_k = \frac{E_k h_k}{1-\nu_k}$ – жесткость на растяжение k-ого слоя структуры; z_k – расстояние от срединной поверхности k-ого слоя до нейтральной поверхности всего пакета, которая совпадает с плоскостью xoy, в которой расположено начало координат (см. рисунок 1).

Положение нейтральной поверхности определяем выражением

$$\sum_{k=1}^{n} A_k z_k = 0\,. \tag{4}$$

и параметром $h = z_1$, который отсчитывается от срединной поверхности первого слоя. Приведенную жесткость обобщенной модели представим в виде

$$D_{np} = \beta D_1\,, \tag{5}$$

где β – безразмерный коэффициент приведения, учитывающий переход от многослойной системы к эквивалентной однослойной; D_1 – цилиндрическая жесткость первого слоя (k = 1). С учетом (4) и (5) получим следующую зависимость для коэффициента β:

$$\beta = 1 + \sum_{k=2}^{n} \overline{E}_k \overline{h}_k + 12\left(\overline{z}_1^2 + \sum_{k=2}^{n} \overline{E}_k \overline{h}_k \overline{z}_k^2\right). \tag{6}$$

Здесь $\overline{h}_k = \frac{h_k}{H}$; $\overline{E}_k = \frac{E_k\left(1-\nu_1^2\right)}{E_1\left(1-\nu_k^2\right)}$; $\overline{z}_k = \frac{z_k}{H}$ – безразмерные параметры системы.

Параметр, определяющий положение нейтральной поверхности с учетом (4), определяется выражением:

$$\overline{h} = \overline{z}_1 = \left[\sum_{k=2}^{n} \overline{E}_k \overline{h}_k \left(0{,}5\left(1+\overline{h}_n\right) + \sum_{k=2}^{n-1} \overline{h}_k\right)\right] \cdot \left(1 + \sum_{k=2}^{n} \overline{E}_k \overline{h}_k\right)^{-1}. \tag{7}$$

В соответствии с (3) выражение для приведенного модуля упругости принимает вид

$$E = E_1 \frac{\left(1-\nu^2\right)}{\left(1-\nu_1^2\right)} \left(\frac{h_1}{H}\right)^3 \beta\,. \tag{8}$$

Применяя предложенную модель и заменяя многослойную гетерогенную систему однослойной с приведенными характеристиками E, ν, H, известными методами теории упругости можно решать задачи на воздействие различных видов внешних нагрузок на гетерогенную систему. В том числе достаточно просто рассматриваются и контактные задачи. Анализ напряженно-деформированного состояния плоской гетерогенной структуры при статическом нагружении следует проводить в следующей последовательности:

1.Используя предложенную модель, заменить слоистую гетерогенную структуру эквивалентной ей системой с приведенными характеристиками E, ν, H. Остальные геометрические размеры конструкции не изменяются.

2.Решить задачу по определению поля перемещений $w(x, y)$, $u(x, y)$, $v(x, y)$ по осям z, x, y соответственно и, используя соотношения Коши, определить поле относительных деформаций $\varepsilon_x^{(k)}$, $\varepsilon_y^{(k)}$, $\varepsilon_z^{(k)}$ в каждом слое при заданной внешней нагрузке.

3.Используя обобщенный закон Гука определить поле напряжений в любом слое гетерогенной структуры с учетом характеристик соответствующего слоя (E_k, ν_k), т. е. осуществить обратный переход от эквивалентной системы с приведенными характеристиками к исследуемой слоистой гетерогенной системе:

$$\sigma_x^{(k)} = \frac{E_k}{1-\nu_k^2}\left(\varepsilon_x^{(k)} + \nu_k \varepsilon_y^{(k)}\right); \qquad \sigma_y^{(k)} = \frac{E_k}{1-\nu_k^2}\left(\varepsilon_y^{(k)} + \nu_k \varepsilon_x^{(k)}\right), \tag{9}$$

где k – номер слоя, в котором определяются напряжения.

Предложенная модель позволяет исследовать НДС дискретно непрерывных слоистых гетерогенных структур при любом числе слоев и допускает следующие упрощения и предельные переходы:

– если коэффициенты Пуассона для материалов слоев близки между собой, то, полагая $\nu_k = \nu$, получим выражение для приведенного модуля упругости $E = E_1\left(\frac{h_1}{H}\right)^3 \beta$;

– если некоторые слои являются мягкими и работают только на сдвиг, что характерно для вибродемпфирующих покрытий, то полагая в (3) $E_k \equiv 0$ следует перейти к модели, предложенной в [5];

– если между соседними слоями нет жесткой связи и имеет место проскальзывание, то в (3) следует положить $A_k = 0$ и в этом случае выражение для приведенного модуля упругости принимает вид $E = \left(\sum_{k=1}^{n} E_k h_k^3\right) \cdot \left(\sum_{k=1}^{n} h_k\right)^{-3}$;

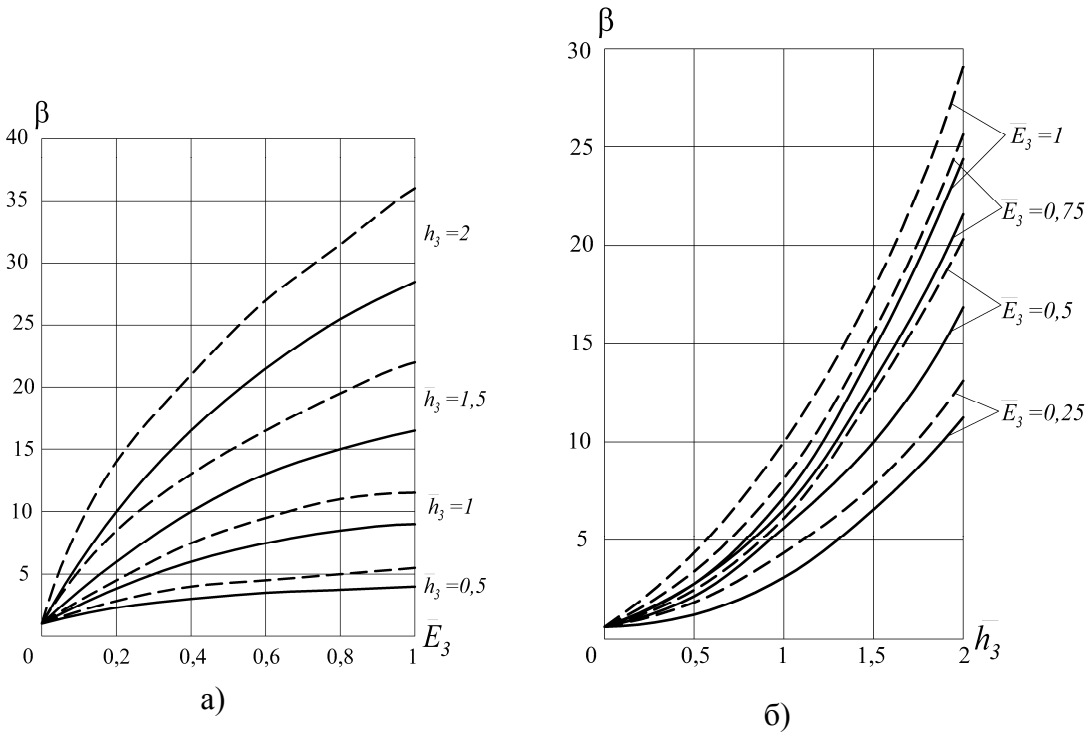
– если в гетерогенной структуре $h_1 >> h_k$, при $k = 2, 3, \ldots, n$, то в модели можно принять $H = h_1$ и всю структуру привести к нижнему слою ($k = 1$) аналогично [5]. В этом случае, заменяя H на h_1, получим $E = E_1\beta$;

– если в конструкции имеется несущий слой, обладающий большой жесткостью (например слой с номером $k = m$), то именно его следует взять в качестве слоя приведения полагая $H = h_m$, т. к. $h_m >> h_j$ при $j \neq m$.

В зависимости от конструктивных особенностей конкретной гетерогенной структуры возможны и другие варианты выбора толщины H эквивалентной структуры в рамках предложенной модели.

На рисунке 2 показаны результаты численных расчетов коэффициента β для трехслойной гетерогенной системы ($n = 3$) при различных значениях параметров, соответствующих реальным микросборкам приборов. При расчете принято, что $\nu_k = \nu = 0{,}3$.

Приведенные зависимости показывают влияние характеристик слоев гетерогенной структуры на коэффициент приведенной жесткости и соответственно на величину приведенного модуля упругости.



$----\bar{h}_2 = 0,1$; —— $\bar{h}_2 = 0,05$ $\bar{E}_2 = 0,1$

Рис.2. Зависимость коэффициента приведенной жесткости βот физико-механических и геометрических характеристик слоев

Предложенная модель позволяет выполнять дальнейшие обобщения при переходе к вязкоупругим гетерогенным структурам, а также использовать ее при учете ползучести и пластичности слоев гетерогенных структур различного назначения и конструктивного исполнения.

Литература

1. Динамика гетерогенных структур. В 12 т. Т. 5. Моделирование гетерогенных структур преобразователей информации ракетно-космических систем/ В.А. Васильев, Н.И. Волчихина, В.В. Смогунов; под общ. ред. В.В. Смогунова. – Пенза: Изд-во Пенз. гос. ун-та, 2004. – 468 с.

2. Сендецки, Дж. Механика композиционных материалов/ Дж.Сендецки. – М.: Мир. – 1978. – 567с.

3. Фокин, А.Г. Эффективные модули упругости композита, составленного из анизотропных слоев./ А.Г.Фокин, Т.Д.Шермергор// Механика полимеров, 1975, №3. – С.408-413.

4. Малмейстер, А.К. Сопротивление жестких полимерных материалов./ А.К.Малмейстер, В.П.Тамуж, Г.А.Тетерс.// Рига: Зинатне. – 1972 . – 498с.

5. Литвинов, А.Н. Прикладные модели механики гетерогенных структур изделий приборостроения: монография/ А.Н. Литвинов, М.А. Литвинов, В.В. Смогунов; под ред. В.В. Смогунова. – Пенза: Изд-во Пенз. гос. ун-та. – 2009. – 320с.

Программный комплекс для исследования пространственно-временной динамики систем реакция-диффузия в поле мультипликативных флуктуаций параметров

В.В. Максимов

Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Разработан программный комплекс (ПК) [1], позволяющий провести анализ дисперсионных уравнений для усредненных по ансамблю реализаций амплитуд незатухающих мод, моделирование эволюции и визуализацию решений двухкомпонентных систем реакция-диффузия, находящихся во внешней флуктуирующей среде, имеющих вид:

$$\frac{\partial x_1}{\partial t} = P(x_1, x_2, \chi_{10} \ldots \chi_{k0} \ldots \chi_n) + \sum_{i=1}^{k} f_i(\vec{r}, t) P_i'(x_1, x_2, \chi_{k+1}, \ldots \chi_m) + D_1 \nabla^2 x_1,$$

$$\frac{\partial x_2}{\partial t} = Q(x_1, x_2, \eta_{10} \ldots \eta_{l0} \ldots \eta_j) + \sum_{i=1}^{l} \phi_i(\vec{r}, t) Q_i'(x_1, x_2, \eta_{l+1}, \ldots \eta_s) + D_2 \nabla^2 x_2 . \quad (1)$$

Здесь x_1, x_2 – динамические переменные, $\chi_1 \ldots \chi_n$, $\eta_1 \ldots \eta_j$ – параметры задачи, $\chi_{10} \ldots \chi_{k0}$, $\eta_{10} \ldots \eta_{l0}$ – пространственно временные средние соответствующих параметров, D_1, D_2, коэффициенты диффузии компонент, P, Q, P' и Q' – нелинейные функции, описывающие взаимодействие подсистем. Однородные изотропные гауссовы поля $f_i(\vec{r}, t)$ и $\phi_i(\vec{r}, t)$ с нулевыми средними и функциями корреляции вида $\left\langle f_i(\vec{r}, t) f_j(\vec{r}', t') \right\rangle = \Phi_i(|\vec{r} - \vec{r}'|) \exp(-k_{1ti} |t - t'|) \delta_{ij}$ моделируют воздействие внешнего флуктуирующего окружения на систему (1). Функция $\Phi_i(|\vec{r} - \vec{r}'|)$ задает пространственную зависимость корреляций случайного поля. Время корреляции $r_t = k_t^{-1}$ - характерный временной масштаб внешних флуктуаций.

Разработанный комплекс программ осуществляет решение смешанной задачи Коши для систем Гирера-Майнхардта [2] и Шеффера [3], в которых учитывается влияние флуктуирующего окружения и являющихся частными случаями (1), с граничными условиями Дирихле или Неймана. Однако он может быть легко модифицирован для других конкретных систем рассматриваемого типа.

Структурная схема данного программного комплекса (ПК) показана на рис. 1.

Подсистема ввода предназначена ввода параметров, необходимых для моделирования эволюции системы. Здесь же осуществляется выбор граничных и начальных условий.

Подсистема анализа дисперсионного соотношения системы Гирера – Майнхардта позволяет провести анализ действительной и мнимой частей собственных значений детерминированной модели Гирера – Майнхардта и исследовать зависимость инкрементов усредненных по ансамблю реализаций амплитуд неустойчивых мод в соответствующей стохастической модели от волновых чисел и параметров случайного поля. Данная подсистема позволяет определить максимальное значение инкремента и соответствующее ему значение волнового числа. Здесь же осуществляется вывод полученных результатов расчетов на экран в виде соответствующих графиков или окон со значениями переменных.

Подсистема проверки и контроля данных. В данной подсистеме осуществляется контроль формата данных, диапазона вхождения переменной, а также вспомогательные функции, необходимые для функционирования программы.

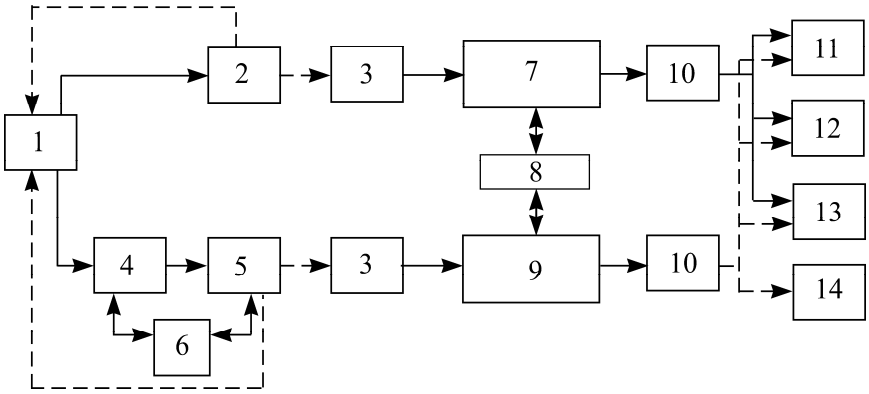


Рис. 1. Структурная схема комплекса программ, разработанного для исследования пространственно-временной динамики моделей Гирера – Майнхардта и Шеффера в поле мультипликативных флуктуаций параметров. 1. Подсистема ввода. 2. Подсистема анализа дисперсионного соотношения системы Гирера – Майнхардта. 3. Подсистема проверки и контроля данных. 4. Подсистема генерации решений детерминированной локальной системы Шеффера и построение фазового портрета. 5. Подсистема анализа дисперсионного соотношения системы Шеффера. 6. Подсистема вычисления особых точек и определения их устойчивости. 7. Подсистема генерации решений в стохастической модели Гирера – Майнхардта. 8. Подсистема моделирования случайного поля. 9. Подсистема генерации решений в стохастической модели Шеффера. 10. Подсистема вывода в текстовые файлы. 11. Подсистема вычисления усредненных по поверхности слоя флуктуаций динамических переменных. 12. Подсистема вычисления норм матриц разностей распределений динамических переменных. 13. Подсистема генерации видеоизображений и графической визуализации. 14. Подсистема определения положения точки обрыва волнового фронта спиральной волны.

Подсистема генерации решений детерминированной локальной системы Шеффера и построение фазового портрета. Данная система предназначена для предварительного анализа модели [3]. В ней осуществляется построение решений детерминированной локальной системы Шеффера при выбранных параметрах моделирования, построение фазового портрета системы для некоторой области начальных условий, определяемых пользователем в подсистеме ввода, построение главных изоклин. Интегрирование уравнений модели осуществляется при помощи метода Рунге-Кутта четвертого порядка точности. Полученные результаты расчетов выводятся на экран в графическом виде.

Подсистема анализа дисперсионного соотношения системы Шеффера позволяет провести анализ действительной и мнимой частей собственных значений, а также расчет функций определяющих тип возникающей неустойчивости, для детерминированной модели Шеффера и исследовать зависимость инкрементов усредненных по ансамблю реализаций амплитуд неустойчивых мод соответствующей стохастической модели от волновых чисел и параметров случайного поля. Данная подсистема позволяет определить максимальное значение инкремента и соответствующее ему значение волнового числа. Здесь же осуществляется вывод полученных результатов расчетов на экран в виде соответствующих графиков или окон со значениями переменных.

Подсистема вычисления особых точек и определения их устойчивости предназначена для нахождения особых точек детерминированной локальной системы Шеффера, их классификации и определения устойчивости. Особые точки определяются как точки пересечения главных изоклин системы. Осуществляется вывод значений особых точек в текстовом представлении на экран в виде окон.

Подсистема генерации решений в стохастической модели Гирера – Майнхардта является одной из основных в данном программном комплексе. Она предназначена для моделирования эволюции этой системы.

Подсистема моделирования случайного поля осуществляет интеграцию работ подсистем 7 и 9 с математическим пакетом MATLAB 7.0 и производит генерацию случайного поля с заданной функцией корреляции стандартными средствами MATLAB.

Подсистема генерации решений в стохастической модели Шеффера является второй из основных в данном программном комплексе. Она предназначена для моделирования эволюции этой системы.

Подсистема вывода в текстовые файлы предназначена для вывода полученных результатов моделирования в текстовом представлении на диск в рабочую папку в виде текстовых файлов определенного формата.

Подсистема вычисления усредненных по поверхности слоя флуктуаций динамических переменных. В данной подсистеме происходит загрузка из рабочей папки текстовых файлов, полученных из подсистемы вывода, в MATLAB, преобразование их в матрицу распределения соответствующей динамической переменной и расчет с помощью стандартной процедуры (std(X,0)) MATLAB усредненных по поверхности слоя флуктуаций этой динамической переменной в зависимости от времени. Здесь же осуществляется вывод полученных результатов расчетов на экран в виде соответствующих графиков.

Подсистема вычисления норм матриц разностей распределений динамических переменных. В данной подсистеме происходит загрузка из рабочей папки текстовых файлов, полученных из подсистемы вывода, в MATLAB, преобразование их в матрицы распределений соответствующих динамических переменных и расчет с помощью стандартной процедуры (normest(A,inf)) MATLAB норм матриц разностей распределений динамических переменных в зависимости от времени. Здесь же осуществляется вывод полученных результатов расчетов на экран в виде соответствующих графиков.

Подсистема генерации видеоизображений и графической визуализации предназначена для построения графических изображений и создания видеоизображений для демонстрации эволюции систем. Подсистема использует стандартные возможности MATLAB.

Подсистема определения положения точки обрыва волнового фронта спиральной волны. Положение кончика спиральной волны, определяется как точка пересечения изолиний с определенными уровнями динамических переменных. В данной подсистеме происходит загрузка из рабочей папки текстовых файлов, полученных из подсистемы вывода, в MATLAB, преобразование их в матрицы распределений соответствующих динамических переменных, построение с помощью процедуры contour(Z,n) изолиний определенного уровня для соответствующего распределения и определение пересечения построенных изолиний с помощью дополнительного модуля расчета точки пересечения изолиний (*http://www.mathworks.com/matlabcentral/fileexchange/ 8997*). Здесь же осуществляется вывод полученных результатов расчетов на экран в виде соответствующих графиков.

Рис. 2. демонстрирует возможности представленного выше комплекса. На нем изображены эволюция спиральных волн, возникающих в модели Шеффера в поле внешних флуктуаций, а также траектории дрейфа точки обрыва волнового фронта спиралей.

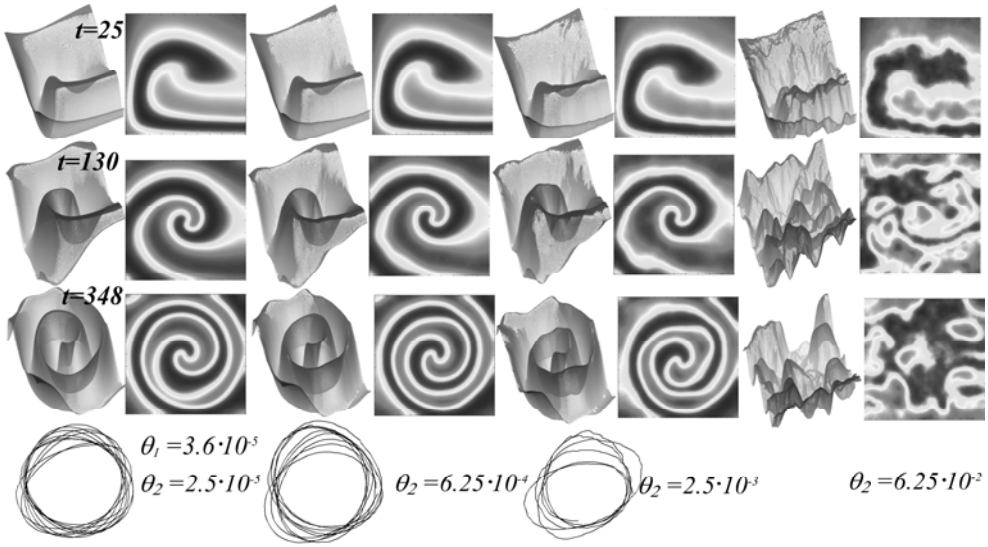


Рис.2. Эволюция спиральных популяционных волн в модели Шеффера при увеличении интенсивности флуктуаций θ_2. θ_1= 3,6·10^{-5}(1%). $k_{f1} = k_{f2} = 1$.

Литература

1. Курушина С.Е., Иванов А.А., Максимов В.В. Моделирование пространственных и пространственно-временных диссипативных структур в системах типа «реакция-диффузия» в поле мультипликативных флуктуаций // Свидетельство о государственной регистрации программы для ЭВМ № 2010613948 от 18.06.2010г. 50 с.

2. Meinhardt H., Gierer A. Applications of a theory of biological pattern formation based on lateral inhibition // Journ. Cell. Sci. 1974. V.15. P.321.

3. Malchow H. Motional instabilities in prey-predator systems // J. Theor. Biol. 2000. V/204. P.639-647.

Статистическая модель и программа оценки вероятностных характеристик вращательного движения капсулы на тросе

Ю.М. Заболотнов, О.Н. Наумов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Работа посвящена разработке программного комплекса статистического анализа вращательного движения капсулы при управляемом развертывании орбитальной тросовой системы с космического аппарата находящегося на круговой орбите. Программный закон развертывания троса реализуется с помощью системы регулирования, работающей по принципу обратной связи. Для организации обратной связи измеряются длина троса и его скорость, и производится их сравнение с заданными программными значениями. Для моделирования упругих свойств троса применяется односторонний закон Гука. Для моделирования вращательного движения капсулы используются классические уравнения Эйлера, записанные в традиционной форме. Статистический анализ проводится на основе численного моделирования при действии различных внешних возмущений и возмущений, возникающих при работе системы управления. В основу статистического моделирования положен классический метод статистического моделирования – метод Монте-Карло. Обработка результатов статистического моделирования осуществляется при помощи метода статистической линеаризации и метода наименьших квадратов. Для реализации данной модели использовалась среда программирования Delphi.

1. Математическая модель

Для описания движения спускаемой капсулы (СК) относительно центра масс, соединенной тросом с космическим аппаратом (КА), используется следующая совокупность систем координат: $OXYZ$, $OX_oY_oZ_o$, $Cx_oy_oz_o$, $Cx_ty_tz_t$, $cxyz$. Здесь $OXYZ$ и $OX_oY_oZ_o$ - орбитальные геоцентрические системы координат, $Cx_oy_oz_o$ - орбитальная система координат, связанная с центром масс КА; $cx_ty_tz_t$ - тросовая система координат, $cxyz$ - связанная система координат, c - центр масс СК. Оси OX и OX_o направлены в точку восходящего узла орбиты и по радиусу вектору КА, оси OZ и OZ_o совпадают и параллельны вектору кинетического момента движения КА, оси систем координат $OX_oY_oZ_o$ и $Cx_oy_oz_o$ параллельны, ось cx направлена по линии, соединяющей центр масс СК и точку крепления троса. Все применяемые системы координат являются правыми. Система координат $OXYZ$ считается неподвижной. По сравнению с общепринятыми системами координат, здесь введена так называемая тросовая система координат $cx_ty_tz_t$, для которой ось cx_t параллельна вектору силы натяжения троса T, приложенной к СК; плоскость cx_ty_t параллельна плоскости, проходящей через радиус вектор точки крепления троса к КА и натянутый трос.

Уравнения движения СК относительно ее центра масс имеют вид [1]

$$
\begin{aligned}
I_x \cdot \frac{d\omega_x}{dt} + \omega_y \cdot \omega_z \cdot (I_z - I_y) &= \sum M_x, \\
I_y \cdot \frac{d\omega_y}{dt} + \omega_x \cdot \omega_z \cdot (I_x - I_z) &= \sum M_y, \\
I_z \cdot \frac{d\omega_z}{dt} + \omega_x \cdot \omega_z \cdot (I_y - I_x) &= \sum M_z,
\end{aligned}
\tag{1.1}
$$

$$
\begin{aligned}
\frac{d\psi}{dt} &= \frac{\omega_x \cdot \sin\varphi + \omega_y \cdot \cos\varphi}{\sin\theta}, \\
\frac{d\varphi}{dt} &= \omega_z - \frac{\left(\omega_x \cdot \sin\varphi + \omega_y \cdot \cos\varphi\right) \cdot \cos\theta}{\sin\theta}, \\
\frac{d\theta}{dt} &= \omega_x \cdot \cos\varphi - \omega_y \cdot \sin\varphi.
\end{aligned}
\tag{1.2}
$$

где I_X, I_Y, I_Z - моменты инерции СК в главных связанных осях; ω_i и $\sum M_i$ $(i = x, y, z)$ - проекции угловых скоростей вращения груза и суммы действующих на него моментов на оси главной связанной системы координат; ψ, φ, θ - углы Эйлера, определенные относительно геоцентрической неподвижной системы координат; $M_i = M_{grav_i} + M_{aer_i} + M_{elast_i}$. M_{grav} - гравитационный момент; M_{aer} - аэродинамический момент; M_{elast} -момент силы упругости троса.

Динамику работы управляющего механизма опишем приближенными уравнениями

$$
m_u \frac{dV_l}{dt} = T - F_c, \quad \frac{dl}{dt} = V_l, \tag{1.3}
$$

где l и V_l - длина и скорость выпуска троса, m_u - коэффициент, характеризующий инерционность механизма управления; $F_c = F_{cp}(t) + p_l\left[l - l_p(t)\right] + p_V\left[V_l - V_{lp}(t)\right]$ - сила сопротивления выпуску троса в механизме управления (управляющая сила), где $F_{cp}(t)$, $l_p(t)$ и $V_{lp}(t)$ - программные (номинальные) зависимости силы, длины и скорости троса от времени; p_l, p_V - коэффициенты обратной связи.

Для вычисления силы натяжения троса T используется односторонний закон Гука

$$
T = \begin{cases} ES(1-\gamma) & при\ \gamma < 1 \\ 0 & при\ \gamma \geq 1 \end{cases}, \tag{1.4}
$$

где E и S - модуль Юнга и площадь поперечного сечения троса, $1-\gamma$ - относительное удлинение троса, $\gamma = l / AB$, l - длина нерастянутого троса, выпущенного из механизма управления развертыванием ОТС; AB - расстояние между точками крепления троса.

Уравнения (1.1, 1.2, 1.3) необходимо дополнить уравнениями движения центров масс спускаемой капсулы и КА, в данной работе данные уравнения записывались в

неподвижной геоцентрической системе координат с использованием второго закона Ньютона.

2. Случайные возмущения и статистическая обработка результатов моделирования

Случайные факторы вызывают в процессе выпуска троса отклонения параметров движения СК от их заданных значений. Для большинства возмущающих воздействий статистические характеристики либо неизвестны, либо известны недостаточно полно, поэтому необходимо делать предположения о законах распределения случайных возмущений. Для случайных факторов, был принят ряд предположений о законах распределения и их параметрах:

1. Начальные угловые скорости $\omega_{xo}, \omega_{yo}, \omega_{zo}$, отделения в связанных осях распределены по нормальному закону с нулевым математическим ожиданием m_ω и со стандартным отклонением $\sigma_\omega = 0.01\ c^{-1}$

2. Начальные угловые скорости $\omega_{xo}, \omega_{yo}, \omega_{zo}$, КА в связанных осях распределены по нормальному закону с нулевым математическим ожиданием m_ω и со стандартным отклонением $\sigma_\omega = 0.002\ c^{-1}$

3. Начальные значения углов прецессии и собственного вращения распределены равномерно на отрезке $[0, 2\pi]$. Начальное значение угла нутации распределено равномерно на отрезке $[0, 10\,град]$.

4. Моменты инерции концевого груза сферической формы $I = I_x = I_y = I_z$ распределены нормально с математическим ожиданием равным моменту инерции сферы и со стандартным отклонением $\sigma_I = 0.01I$.

5. Две координаты точки крепления троса к СК (относительно номинального направления, соединяющего центр масс тела и точку крепления) распределены нормально с нулевым математическим ожиданием и со стандартным отклонением равным $\sigma_B = 0.01OB$.

где OB - номинальное расстояние между точкой крепления и центром масс СК.

6. Ошибки измерения длины троса и скорости распределены нормально с нулевыми математическими ожиданиями и со стандартными отклонениями равными

$$\sigma_V = 2\ см/с,\ \sigma_L = 2\ см$$

7. Ошибки работы механизма управления распределены нормально с нулевым математическим ожиданием и со стандартным отклонением равным $\sigma_F = 0.01F_c$.

Для оценки действия различных возмущений на величину дисперсии характеристик движения использовались следующие методы:

1. Статистической линеаризации (СЛ) [2]:

$$y = b_0 + b_1 \bar{x}_1 + ... + b_n \bar{x}_n \qquad (2.1)$$

где $b_0 = M(y)$, $b_i = \dfrac{M\{y \cdot \bar{x}_i\}}{D_{xi}}$ - коэффициенты, которые необходимо определить по результатам статистических испытаний; $\bar{x}_i$ - центрированные входные случайные

величины. $M(y)$, $M\{y \cdot \bar{x}_i\}$ - оценки математического ожидания некоторая характеристика движения СК y и корреляционного момента по результатам статистических испытаний, $M(\bar{x}_i), D_{\bar{x}_i}$ - математическое ожидание и дисперсия входных величин.

2. Наименьших квадратов (НК) [3]:

$$y = B_0 + \sum_{i=1}^{n} B_i \cdot x_i \qquad (2.2)$$

где B_0 - свободный коэффициент; B_i - вектор весовых коэффициентов (определяемые по известной методике метода НК); x_i - исходные случайные величины.

3. Программный комплекс

Для решения уравнений движения спускаемой капсулы и космического аппарата, а так же для проведения статистического моделирования, в среде программирования Delphi был разработан программный комплекс, включающий в себя следующие основные процедуры и их взаимодействия, которые представлены на рисунке 1:

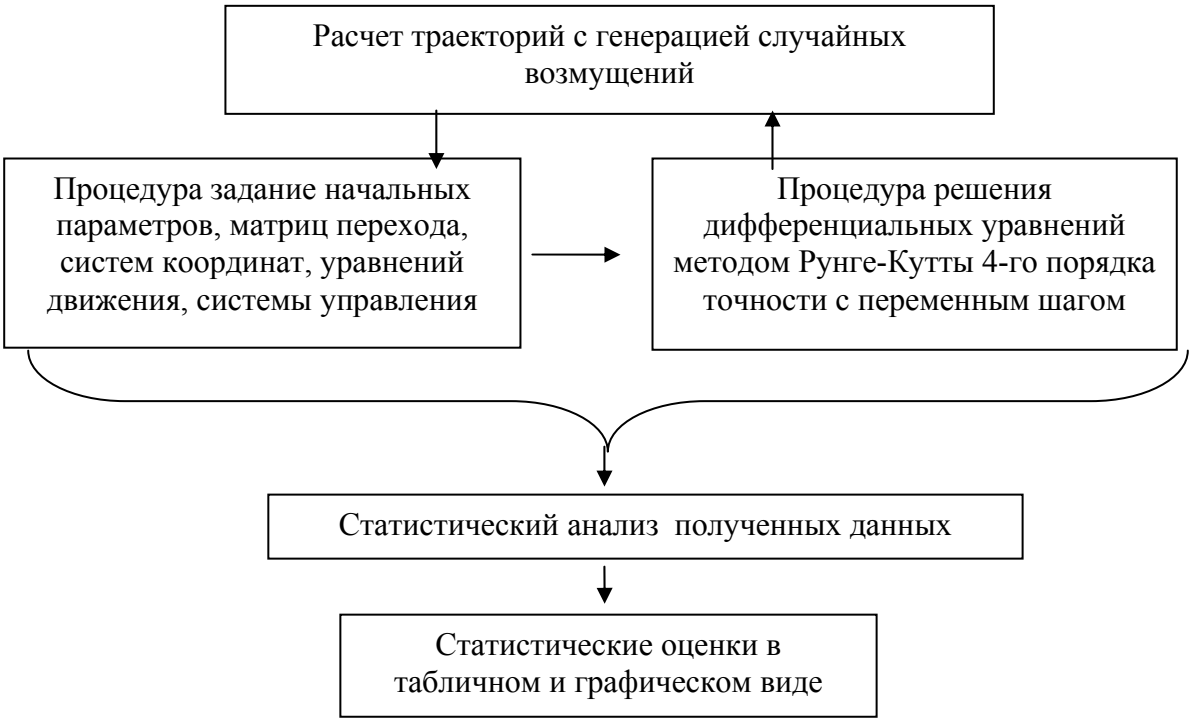


Рисунок 1. Схема основных процедур программы и их взаимодействия

4. Результаты моделирования

В результате моделирования были получены гистограммы распределения угла нутации (пространственный угол между продольной осью капсулы и направлением троса), угловых скоростей и угла атаки капсулы. На рисунке 2 построена гистограмма распределения угла нутации в конце развертывания тросовой системы и на нее наложена плотность распределения усеченного нормального закона.

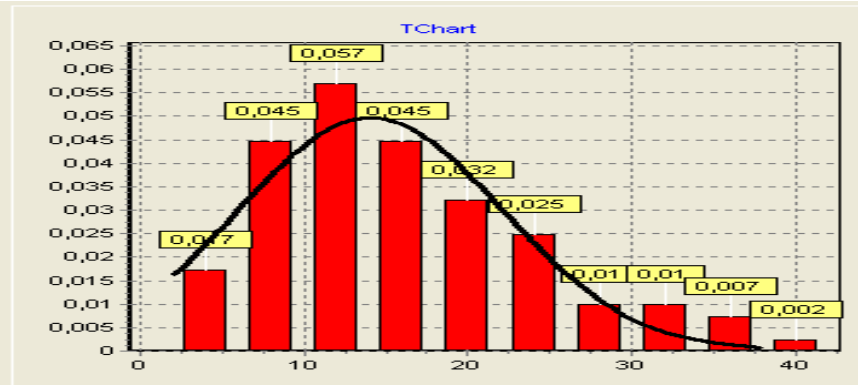


Рисунок 2. Гистограмма распределение угла нутации спускаемой капсулы.

В результате решения задачи о влияние возмущений на величину дисперсии угла нутации и модуля угловой скорости (использовались методы СЛ и НК), было определенно, что наибольший вклад в полную дисперсию дают погрешность работы системы управления и величина асимметрии точки крепления троса к капсуле (более 50%). Другие факторы также влияли на величину дисперсии, но существенно менее значительно. Поэтому для обеспечения безопасного спуска капсулы на тросе необходимо минимизировать параметры асимметрии точки крепления троса и ошибки системы управления. Асимметрия точки крепления это конструктивный параметр, его минимизация достижима за счет использования более точных методов проектирования капсулы и схем крепления троса. Качество системы управления сильно зависит от коэффициентов обратной связи, поэтому для минимизации её влияния на величину разброса основных выходных параметров нужно точнее подбирать коэффициенты обратной связи (с учетом как можно большего числа возмущающих факторов).

Литература

1. Лойцянский, Л.Г. Курс теоретической механики [Текст]. В 2т./ Л.Г. Лойцянский, А.И. Лурье// - М.: Дрофа, 2006.

2. Иващенко, Н.Н. Автоматическое регулирование. Теория и элементы систем [Текст]/ Н.Н. Иващенко // - М.: Машиностроение, 1978. – 736 с.

3. Потасов, К.В. Статистический анализ экспериментальных данных [Текст] / К.В. Потасов // - М.: Мир, 2005. – 142 с., ил.

Моделирование логопериодических колебаний с помощью иерархических структур

А.С. Пивоварова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

При изучении поведения различных систем перед катастрофами было выяснено, что катастрофическим событиям предшествуют не только возрастание соответствующих физических величин, но и ускорение колебаний, происходящих в системе. В частности колебания с увеличивающимся периодом и амплитудой наблюдаются на финансовых рынках вблизи точек краха, а также в химическом составе подземных вод перед землетрясением. Было выяснено, что такие колебания хорошо аппроксимируются гармониками логопериодических колебаний [1, 2, 3]. В данной работе подробно рассматривается моделирование предкрахового поведения финансового рынка.

В ряде работ [1, 4] был предложен подход к моделированию предкрахового поведения цен, основанный на предположении об иерархической организации агентов на рынке. Рассматриваются агенты, являющиеся одновременно фундаменталистами и подражателями. Каждый агент проводит свой собственный фундаментальный анализ, в результате которого выбирает свое время для покупки акций. Однако агент понимает, что его анализ может быть неполным или необъективным, и поэтому старается узнать больше о действиях других ближайших агентов. При этом агент не проводит никакого технического анализа, т.е. цена не несет для него никакой информации. Другое важное упрощение состоит в том, что на рынке существует некоторый постоянный уровень предложения и все рассматриваемые агенты заинтересованы только в покупке акций. Тогда задача сводится к определению количества активных покупателей среди всех агентов-покупателей.

Иерархическая модель финансовых крахов Джохансена-Сорнетта

Рассматривается иерархическая структура агентов, которая во многом отражает реальную иерархическую организацию фондового рынка. Такая организация может быть как постоянным свойством рынка, так и спонтанно возникшим [6, 7].

В данной модели индивидуальные трейдеры называются агентами нулевого уровня. Агенты 0 уровня объединяются в группы по m, и такие группы сами представляют собой агентов, но уже 1 уровня. Эти группы (агенты 1 уровня) также объединяются в группы по m агентов, формируя агентов 2 уровня и т.д. В таком случае получается иерархическая организация, в которой агенты порядка n состоят из m^n индивидуальных трейдеров. Для простоты можно положить $m = 2$.

В начальный момент времени, все индивидуальные трейдеры 0 уровня иерархии уже провели свой фундаментальный анализ, который определяет наиболее выгодное по их мнению время для покупки акций. Положим, что агент i обладает предпочтительным временем t_i для покупки акций, и что все такие времена t_i распределены по Пуассону с функцией плотности вероятности $p_0(t)$, которой соответствует следующая функция распределения

$$P_0(t) \equiv \int_0^t p_0(t')dt' = 1 - \exp\{-\lambda t\}. \qquad (1)$$

Здесь коэффициент λ характеризует ширину разброса значений времен t_i.

Один из способов описания взаимодействия агентов состоит в учете неуверенности агентов в своем собственном решении. При этом информация о том, что некоторый агент уже совершил покупку, является для еще не вступивших в торговлю агентов поводом пересмотреть принятое ранее решение о времени покупки. Запишем количественно, как влияет агент i в момент времени t_i на агента j, собственное время совершения покупки которого равно t_j

$$t_{ij} = t_i + 2^{-\beta}\left(t_j - t_i\right), \tag{2}$$

где $\beta \geq 0$ - коэффициент влияния. Т.к. при этом $2^{-\beta} \leq 1$, новое время действия удовлетворяет условию $t_i \leq t_{ij} \leq t_j$. Следовательно агент j сокращает свое собственное время покупки под влиянием информации о совершении покупки агентом i. При этом можно говорить о том, что функция распределения времени действия для агента j принимает вид

$$P(t) = 1 - \exp\left\{-\lambda\left[t_i + 2^{\beta}\left(t - t_i\right)\right]\right\}. \tag{3}$$

Для простоты вычислений положим данное влияние однородным в том смысле, что все агенты на всех уровнях имеют одинаковое β.

В виду ограниченного доступа к информации влиять друг на друга могут только агенты, находящиеся на одном уровне иерархии внутри одного кластера. При этом считается, что агент более высокого уровня иерархии совершил покупку, если совершили покупку все его дочерние элементы. Т.е. как только два агента уровня m совершат покупку, соответствующий им агент $m+1$ уровня, также совершит покупку и данная информация передастся его соседнему агенту. Т.о влияние каскадно передается по уровням дерева.

Данная модель поддается формализации и до некоторой степени точному аналитическому исследованию. Учитывая выражение (3), для случая двоичного дерева можно записать вероятность того, что агент уровня N совершит покупку в момент времени t, выраженную через соответствующие вероятности его дочерних агентов уровня $N-1$.

$$p_N(t) = \frac{2}{2^{-\beta}} \int_0^t p_{N-1}(t_1)\, p_{N-1}\left(\frac{t - \left(1 - 2^{-\beta}\right)t_1}{2^{-\beta}}\right) dt_1 \,. \tag{4}$$

В рамках предложенной модели можно дать четкое определение краха. В пределе бесконечного числа агентов (т.е. бесконечного числа уровней иерархии), наличие краха в некоторый момент времени t_c означает, что при $t < t_c$ число активных покупателей оставалось относительно небольшим. Со временем это число постепенно увеличивалось, и к моменту времени t_c произошло насыщение рынка, т.е. конечная доля агентов выставила заявки на продажу. Математически критическая точка, определенная как срабатывание всех агентов на рынке, означает, что при увеличении N, $p_N(t)$ стремиться к дельта-функции.

Таким образом, рассматривая данное выше определение краха, мы можем получить вероятность возникновения краха в данной иерархической структуре, применяя выражение (4) N раз, начиная с $p_0(t)$.

В случае $\beta = 1$ можно получить точное выражение для критической точки. Это самый простой случай, т.к. из уравнения (2) время t_{ij} является просто средним значений t_i и t_j. Тогда, итерируя начальное распределение $p_0(t) = \lambda \exp\{-\lambda t\}$ по формуле (4), получим следующее рекуррентное соотношение для вероятности совершения покупки агентом уровня N

$$p_N(t) = C_N t^{2^N-1} \lambda^{2^N} \exp\{-2^N \lambda t\}.$$

Данные распределения стремятся к δ - функции при $N \to \infty$. Найдем положение максимумов этих распределений

$$t_c{}^N = \frac{2^N - 1}{2^N \lambda} = [N \to \infty] = \frac{1}{\lambda}.$$

Полученное время краха в пределе бесконечного числа агентов является математическим ожиданием начального распределения времен действия, т.е. просто средним временем действия всех агентов.

В случае $\beta \neq 1$ пока не удалось аналитически получить значение критической точки. При численном исследовании данной модели строилась зависимость количества агентов выставивших заявки на покупку от времени (рис. 1), т.е. фактически спроса от времени При приближении к критической точке на разных временных масштабах наблюдаются логопериодические колебания, наложенные на степенную зависимость. Т.о можно сделать вывод о том, что логопериодические колебания и степенной рост цен перед обвалами является следствием дискретной иерархической структуры фондовых рынков и наличия на них положительной обратной связи.

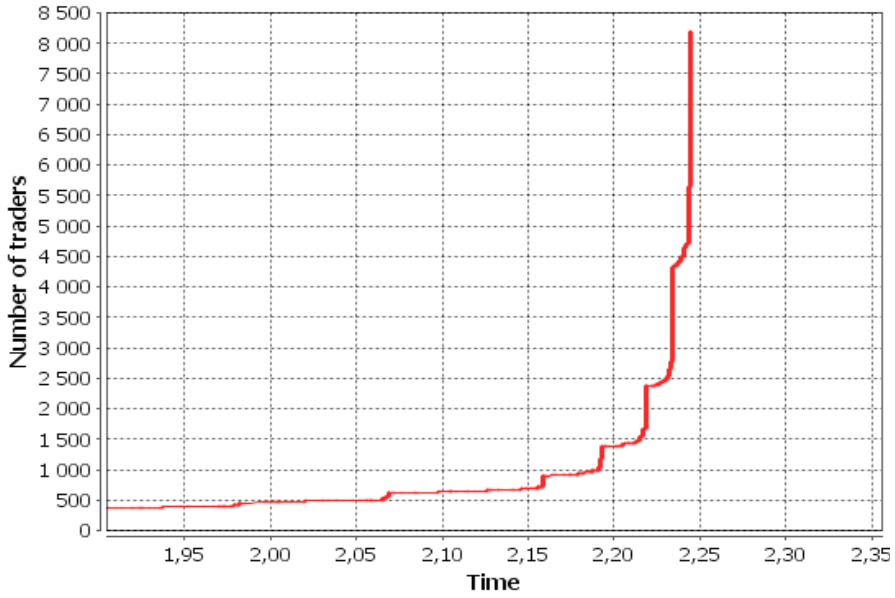


Рис. 1. Количество агентов, выставивших заявку на покупку акций к моменту времени t; параметры системы: $n = 13$ ($N = 8192$), $\lambda = 0.01$, $\beta = 5$.

В пределе при $N \to \infty$ в системе возникает крах, вызванный значительным превышением спроса над предложением, причем время краха t_c одно и тоже при любой начальной реализации распределения времен действий. При численном анализе системы количество агентов всегда ограничено, поэтому время краха t_c будет различным для различных начальных реализаций. Однако при увеличении числа агентов разброс таких значений должен уменьшаться, т.е. плотность распределения времен краха $t_c{}^j$ должна стремиться к дельта-функции. Для подтверждения этого факта были построены функции плотности распределения времен крахов для систем с различным числом агентов. Было показано, что распределение точек краха действительно сужается с увеличением числа агентов к некоторому критическому значению времени краха.

Ультраметрическое обобщение иерархической модели финансовых крахов

Одним из недостатков вышеприведенной модели является то, что агенты являются однородными по взаимодействию друг с другом. Такое упрощение не является вполне оправданным, т.к. введение иерархической структуры агентов требует явного учета их взаимного расположения при взаимодействии. Поэтому целесообразным выглядит введение зависимости степени влияния агентов друг на друга от некоторого расстояния между ними, введенного на данной иерархической структуре.

Пусть все агенты образуют некоторое множество. Т.к. все элементы этого множества образуют иерархическую структуру, можно ввести на этом множестве ультраметрику, т.е. метрику, удовлетворяющую усиленному неравенству треугольника $d_{ik} \le \max(d_{ij}, d_{jk})$. Тогда расстояние между двумя элементами множества вводятся в виде $d_{ij} = p^m$, где p - арность дерева, m - количество уровней дерева до первого общего предка от элементов i и j.

В данной модели предполагается, что агенты влияют друг на друга не одинаково, как в предыдущей модели, а их влияние уменьшается с расстоянием.

При численном исследовании были получены решения, представляющие собой самоподобные нелинейные колебания, наложенные на степенной закон. Построение плотностей распределения точек краха при разном числе агентов показало, что наблюдается поведение, аналогичное исходной иерархической модели. Т.е. плотность распределения также стремиться к δ - функции, при этом сужение происходит даже быстрее, чем в предыдущей модели.

Ультраметрическая модель финансовых крахов

Следует отметить, что случай влияния сработавшего кластера лишь на соседний кластер является довольно грубым приближением реальных процессов распространения информации на рынке. Ясно, что влияние агентов друг на друга сильно неоднородно и зависит от взаимного расположения агентов внутри иерархической структуры. Введем ультраметрическое расстояние между агентами аналогично тому, как это было сделано в предыдущем пункте, т.е. $d_{ij} = p^m$, где p - арность дерева, m - количество уровней дерева до первого общего предка от элементов i и j.

Пусть начальные времена действия агентов являются случайной величиной, распределенной по Пуассону как и в предыдущих моделях.

В момент, когда агент с наименьшим временем выставляет свою заявку, информация о его действиях начинает влиять на всех агентов на рынке, однако данное влияние тем меньше, чем дальше в ультраметрическом смысле находятся трейдеры от данного сработавшего агента.

Принципиальным отличием данной модели от иерархических моделей, описанных ранее, является то, что влияние агентов начинает учитываться не только при срабатывании некоторого кластера, а оно присутствует всегда. Каждый сработавший агент сокращает время срабатывания всех оставшихся агентов.

Численное исследование данной модели показало, что в малой окрестности критической точки, т.е. времени срабатывания всех агентов, наблюдается сильное пороговое изменение количества вступивших в торговлю агентов с увеличивающейся амплитудой и частотой. Стоит отметить, однако, что визуальное наблюдение такого поведения затруднено из-за быстрого убывания периода полученных колебаний.

Также были исследованы плотности распределения времен крахов в зависимости от числа агентов. Оказалось, что плотности распределения сужаются с увеличением числа участников рынка, а время, соответствующее их максимумам

уменьшается.

Литература

1. Sornette D., Johansen A., Bouchaud J.-Ph. Stock market crashes, Precursors and Replicas. Journal de Physique, France. V.6, 1996. P. 167-175.
2. Johansen A., Sornette D., Wakita H., Tsunogai U., Newman W.I., Saleur H. Discrete scaling in earthquake pre-cursory phenomena: Evidence in the Kobe earthquake, Japan// J. Phys. I (France). 1996. V.6, p.1391‑1402.
3. Подлазов А.В. Режимы с обострением с комплексными показателями. Лог-периодические колебания в модели разрыва пучка волокон // Препринты ИПМ им. М.В.Келдыша. № 35. 22 с. 2009.
4. Sornette D., Johansen A.. A hierarchical model of financial crashes. Physica A, V. 261, 1998, P. 581-598.
5. Huang Y., Ouillon G., Saleur H., Sornette D. Spontaneous generation of discrete scale invariance in growth models, Phys. Rev. E 55, 1997. P. 6433-6447.
6. Tostesen E. Dynamics of hierarchically clustured cooperative agents, Cand. Scient. Thesis, University of Copenhagen, 1995.

Использование HASH функций при моделировании стохастических процессов

Д.А. Рыбаков
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Параметры многих процессов имеют две составляющие: закономерность и случайность. Если закономерность часто можно выразить с помощью аналитических функций, то случайность можно выразить с помощью хэш функций (hash functions). Использование хэш функций открывает перспективы компактного представления данных для некоторых случаев. В качестве примера приведено моделирование трехмерной перколяционной системы на решетке.

Хэш функции в качестве входного параметра получают массив байтов. В результате работы алгоритма на выходе получается число, которое зависит только от входных параметров. Среди хэш функций существуют такие, свойства значений которых очень близки к свойствам равномерных распределений случайных чисел. Примером такой функции служит функция MD5, которая возвращает числа от 0 до $2^{128}-1$. Если в качестве параметра этой функции подавать случайный набор байтов, то возвращаемые значения будут иметь равномерное псевдослучайное распределение в указанном диапазоне. При этом возвращаемые значения будут некоррелированными, даже если входные массивы коррелированны. Например, если два входных массива отличаются только на один бит, то возвращаемые два числа будут независимыми.

Такая особенность делает возможным использовать хэш функций в качестве генератора псевдослучайных чисел. Что бы получать числа в диапазоне от 0 до 1 требуется нормализация. Нормализированная функция

$$f(s) = MD5(s)/2^{128}, \quad (1)$$

где s - массив байтов, возвращает псевдослучайное число с плавающей точкой с равномерным распределением в диапазоне от 0 до 1,.

Замечательным свойством этой функции является то, что если требуется получить и использовать большой массив псевдослучайных чисел $\{x_i\}$, то отпадает необходимость хранить этот массив где-либо. Достаточно каждый раз при необходимости производить вычисления $x_i = f(i)$.

Для того, что бы получать разные массивы псевдослучайных чисел следует сделать входной параметр зависимым от заранее заданной константы R,

$$x_i = f(s(R,i)), \quad (2)$$

где функция $s(R,i)$ производит символьное слияние текстового представления своих аргументов с разделителем. Например, $s(1,2) = "1,2"$, $s(1,2,3,4,5) = "1,2,3,4,5"$. В таком случае вся последовательность $\{x_i\}$ зависит только от величины R. Таким образом, достигается компактное представление последовательности псевдослучайных чисел любой длины с помощью только одного числа R.

Продемонстрируем эту особенность при моделировании перколяционных задач [1]. Теория перколяции изучает возникновение связанных структур в случайных средах. Примером такой среды может служить трехмерная кубическая решетка, в которой некоторые кубы являются проводящими с вероятностью p. Если две и более примыкающих ячеек являются проводящими, то они образуют проводящий кластер. Крайними случаями такой среды являются диэлектрик (при $p=0$) и сплошной проводник (при $p=1$). В промежуточных случаях возникает сложная стохастическая

среда с интересными характеристиками с точки зрения математики и физики. В более общих случаях математическая модель перколяции может описывать такие физические явления, как фазовые переходы второго рода, горение, протекания и многие другие явления в которых присутствует два и более компонента разной природы.

При моделировании такой системы в простом случае используется массив $P_{x,y,z}$, заполненный единицами с вероятность p и нулями с вероятностью $1-p$, где x,y,z - целые числа. Массив $P_{x,y,z}$ обычно располагается в ОЗУ вычислительной машины и занимаемый объем ОЗУ растет по кубическому закону от линейного размера массива. При вычислениях программа обращается к ячейкам памяти $P_{x,y,z}$, которое можно заменить вызовом функции $P(x,y,z)$, где

$$P(x,y,z)=\begin{cases}1, \text{если } MD5(s(x,y,z,R)) \le p \\ 0\end{cases}. \quad (3)$$

При фиксированных параметрах x,y,z функция $P(x,y,z)$ всегда будет возвращать фиксированное значение без обращения к ячейкам большого массива. Таким образом, формируется некоторая решетка, доступная только для чтения. При этом параметр R является идентификатором, который однозначно задает структуру этой решетки.

Есть случаи, когда этого бывает достаточно. В процессе работы можно выделять подмножества и записывать их в память компьютера. Эти подмножества будут требовать меньшее количество ОЗУ, чем вся решетка. В случае перколяционной задачи можно выделять кластеры, пути внутри кластеров и т.д.

Перколяционные кластеры являются фрактальными объектами с фрактальной размерностью $D \approx 2.5$ [2,3] (см. рис. 1). Из курса фрактальное геометрии известно, что объем фрактального объекта степенным образом зависит его линейного размера

$$V \sim L^D, \quad (4)$$

где V - объем, L - линейный размер объекта [4].

Таким образом, один кластер включает $N \sim L^{2.5}$ ячеек решетки. Например, весь массив $5\,000 \times 5\,000 \times 5\,000$ потребовал бы ~125 Гигабайт, в то время как перколяционный кластер – порядка 1.7 Гигабайт, а другие подмножества, например, пути внутри кластера – еще меньше. При этом сокращение использования ОЗУ компенсируется более длительным выполнением программы, так как обращение к функции $P(x,y,z)$ более длительное, чем обращение к ячейке памяти.

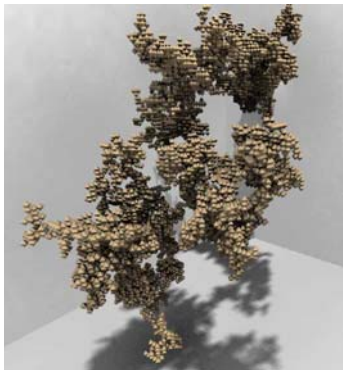


Рис 1. Перколяционный кластер, найденный алгоритмом с использованием хэш функций.

Следует отметить, что кроме *MD5* существует множество других хэш функций. При использовании других функций следует убедиться, что возвращаемые значения являются некоррелированными и имеют равномерное распределение.

Выводы

При определенных условиях хэш функция является аналогом массива псевдослучайных величин, доступного в программе только для чтения. При этом достигается компактное представление массива псевдослучайных величин любого размера. Различные приемы программирования могут значительно сократить использование ОЗУ и других хранилищ данных при моделировании стохастических процессов. При этом нехватка ОЗУ компенсируется бóльшим временем исполнения программы.

Литература

1. Тарасевич_Ю. Ю. Перколяция: теория, приложения, алгоритмы – М., 2002 – 122 с. –960 экз. – ISBN 5-354-00233-8.

2. Bagnich_S. A., Konash A. V., Computer investigation of the percolation processes in two- and three- dimensional systems with heterogeneous internal structure // Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment – April 2003 – Volume 502, Issues 2-3, 21 – P. 731-732.

3. Ламажапов Х.Д., Прохоров С.А, Рыбаков Д.А. Свойства трехмерных кластеров, составленных из параллелепипедов // Вестник Новосибирского государственного университета. Серия: Физика – 2009 – том 4 , номер 3. С. 67-73.

4. Мандельброт_Б. Фрактальная геометрия природы – М., 2002 – 656 с. – 2000 экз. – ISBN 5-93972-108-7.

Применение метода вычислительного эксперимента для исследования влияния параметров правой части нелинейного уравнения теплопроводности на его решение

А.А. Дегтярев, М.В. Силакова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В настоящей работе решается задача численного моделирования теплового процесса в тонкой плоскопараллельной оптической пластине, имеющей форму кругового диска, облучаемую лазерным пучком света с гауссовским законом распределения интенсивности. Лазерный пучок обладает круговой симметрией и падает нормально на поверхность пластины, причем оси пучка и пластины совпадают. Схема воздействия излучения на оптический элемент приведена не рисунке 1.

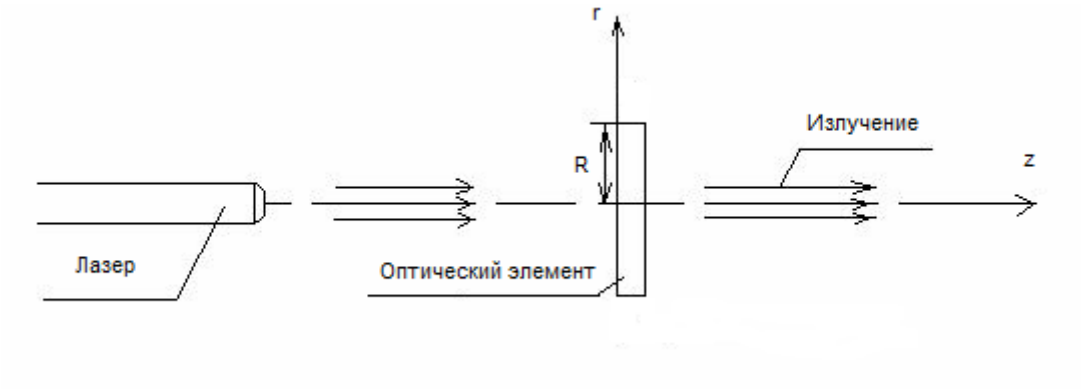


Рисунок 1. Схема прохождения лазерного излучения через оптический элемент

Термодинамический процесс в тонкой прозрачной плоскопараллельной оптической пластине (например, в лазерном окне) с учетом “интерференционных эффектов”, описывается следующей математической моделью [1,2,3]:

$$\begin{cases} c\dfrac{\partial u}{\partial t} = K\left[\dfrac{1}{r}\dfrac{\partial u}{\partial r} + \dfrac{\partial^2 u}{\partial r^2} - \dfrac{2\alpha}{Kl}(u-u_0)\right] + \dfrac{A(nl)}{l}I(r);\ 0 < r \le R,\ 0 \le t \le T; \\ c\dfrac{\partial u}{\partial t} = K\left[2\dfrac{\partial^2 u}{\partial r^2} - \dfrac{2\alpha}{Kl}(u-u_0)\right] + \dfrac{A(nl)}{l}I(r);\ r = 0,\ 0 \le t \le T; \\ u\big|_{t=0} = u_0; \\ \left.\dfrac{\partial u}{\partial r}\right|_{r=R} = 0. \end{cases} \quad (1)$$

где α - коэффициент теплообмена между гранями диска и окружающей средой; c - объемная теплоемкость материала диска; K - коэффициент температуропроводности; n-показатель преломления пластины; l - толщина диска; u_0-температура окружающей среды; $A(nl)$- поглощательная способность; $I(r) = \dfrac{P}{\pi a^2}\exp\left(-\left(\dfrac{r}{a}\right)^2\right)$-интенсивность излучения лазера; $n = n_0 + \alpha_n(u-u_0)$, α_n-коэффициент теплового изменения показателя преломления; $l = l_0(1+\alpha_l(u-u_0))$, α_l- коэффициент теплового расширения.

Функция поглощательной способности имеет вид [1,3]:

$$A(n,l) - A(u-u_0) - 4\varkappa kl\frac{1+n^2}{4n^2+(1-n^2)^2\sin^2\xi}\left[1+2\varkappa kl\frac{(1-n^2)^2\sin^2\xi-2(1+n^4)}{(1+n^2)(4n^2+(1-n^2)^2\sin^2\xi)}\right], \quad (2)$$

где $k=\frac{2\pi n}{\lambda}$ - волновое число; $\xi = kl$; $\varkappa$ - показатель поглощения материала пластины; λ – длина волны излучения.

Положим, что в начальный момент времени температура пластины совпадает с температурой окружающей среды. Особенностью рассматриваемой математической модели, описывающей тепловой процесс, является нелинейность уравнения теплопроводности, обусловленная учетом так называемых “интерференционных эффектов” [1], которые возникают в прозрачных пластинах с высокой степенью плоскопараллельности полированных граней. Поглощательная способность такой пластины становится весьма чувствительной к малым тепловым изменениям ее толщины и представляет собой периодическую. функцию от температуры. Так же существенное влияние на формирование теплового поля оптического элемента оказывает шероховатость граней. Для учета влияния шероховатости граней представим значение толщины элемента как сумму некоего среднего значения l_{cp} и отклонения δl от этого значения, которое характеризует случайность толщины оптического элемента. Считаем, что случайная величина δl имеет нулевое математическое ожидание и дисперсию $M(\delta l)^2=\sigma^2$. В данной работе рассмотрен нормальный закон распределения случайной величины. Получена формула, позволяющая перейти от случайного поля к детерминированному:

$$\overline{A}(n,l_{cp}) = \int A(n,l_{cp}+x)P_{\delta l}(x)dx, \quad (3)$$

где $P_{\delta l}(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-Mx)^2}{2\sigma^2}\right)$ - функция плотности распределения случайной величины δl. С помощью компьютерной программы проведена серия экспериментов, направленных на исследование влияния параметра σ, характеризующего шероховатость граней, на поглощательную способность оптического элемента. Графики полученные в ходе экспериментов приведены на рисунке 2.

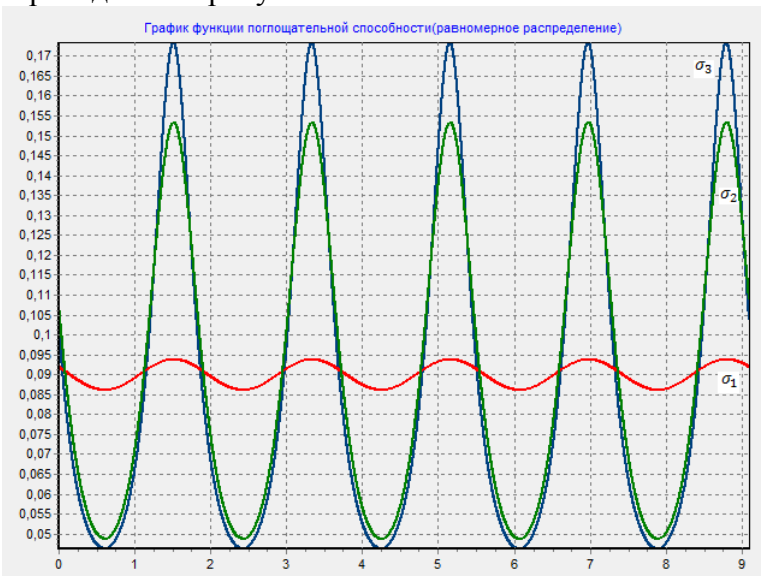


Рисунок 2. Сравнение графиков поглощательной способности для различных значений параметра σ: $\sigma_1 = 0{,}00004$; $\sigma_2 = 0{,}000008$; $\sigma_3 = 0{,}000006$

Для решения нелинейной краевой задачи теплопроводности, описывающей термодинамический процесс в оптическом элементе, была использована схема на равномерной сетке:

$r_i = ih_i,\ i = 0,..,I$, $\quad h_r = \dfrac{R}{I}$ - шаг разбиения по радиусу.

$t_k = kh_t,\ k = 0,..K$, $\quad h_t = \dfrac{T}{K}$ - шаг разбиения по времени.

$$
\begin{cases}
c\dfrac{u_i^{k+1}-u_i^k}{h_t} = k\left[\dfrac{u_{i+1}^k - 2u_i^k + u_{i-1}^k + u_{i+1}^{k+1} - 2u_i^{k+1} + u_{i-1}^{k+1}}{2h_r^2} + \dfrac{1}{ih_r}\dfrac{u_{i+1}^{k+1} - u_{i-1}^{k+1} + u_{i+1}^k - u_{i-1}^k}{4h_r} - \dfrac{2\alpha}{Kl}\left(u_i^k + u_i^{k+1}\right)\right] \\
+I_i\dfrac{A(nl)}{l}; \\
i = \overline{1, I-1};\ k = \overline{0, K-1}; \\
c\dfrac{u_0^{k+1}-u_0^k}{h_t} = k\left[\dfrac{u_1^k - 2u_0^k + u_{-1}^k + u_1^{k+1} - 2u_0^{k+1} + u_{-1}^{k+1}}{h_r^t} - \dfrac{\alpha}{kl}(u_0^k + u_0^{k+1})\right] + I_0\dfrac{A(nl)}{l}; \\
k = \overline{0, K-1}; \\
u_i^0 = 0;\ i = \overline{0, I}; \\
\dfrac{u_{I+1}^k - u_{I-1}^k}{2h_r} = 0;\ k = \overline{0, K};
\end{cases}
\tag{3}
$$

С помощью вычислительного эксперимента установлено, что термодинамический процесс существенно зависит от параметров правой части показателя поглощения и характеристики шероховатости σ. Для вычислительных экспериментов были использованы значения параметров, соответствующие лазерному окну, выполненному из германия.

Так, например, при учете шероховатостей граней, для различных значений СКО от среднего значения толщины, можно наблюдать следующую картину (рисунок 3):

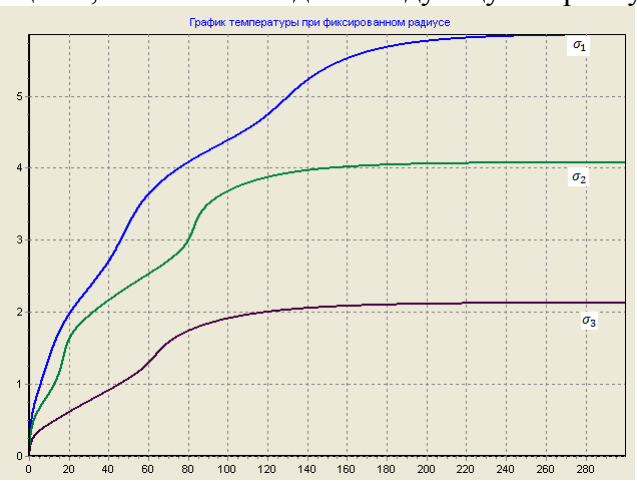


Рисунок 3. Графики зависимости температуры при фиксированном радиусе для различных значений σ : $\sigma_1 = 0{,}00002$; $\sigma_2 = 0{,}000002$; $\sigma_3 = 0{,}0000022$.

На рисунке 3 приведены полученные графики изменения температуры пластины при $r = 4\,мм$ для трех различных значений σ. Видно, что с увеличением σ интерференционная осцилляция ослабевает, вплоть до полного исчезновения.

Величина $\sigma_0 = \frac{\lambda}{4\pi}$ обеспечивает переход от максимума к минимуму поглощения. Установлено, что при значении σ меньше $0{,}1\sigma_0$ учет шероховатостей граней оказывается лишним и для моделирования теплового процесса можно использовать нелинейное уравнение теплопроводности, для которого функция поглощения рассчитывается через формулу (2). При значении $\sigma \in [0{,}1\sigma_0; 0{,}9\sigma_0]$, процесс нагревания оптического элемента описывается нелинейной моделью, учитывающую шероховатость граней через формулу (3). При параметре σ больше $0{,}9\sigma_0$ интерференционные эффекты практически не оказывают влияния на тепловой процесс в пластине, для описания которого вполне возможно использовать линейное уравнение теплопроводности.

Так же установлено, что показатель поглощения $\varkappa$ так же существенное влияние на динамику теплового поля в оптическом элементе. Графики зависимости температуры при фиксированном радиусе для различных значений показателя поглощения представлены на рисунке 4.

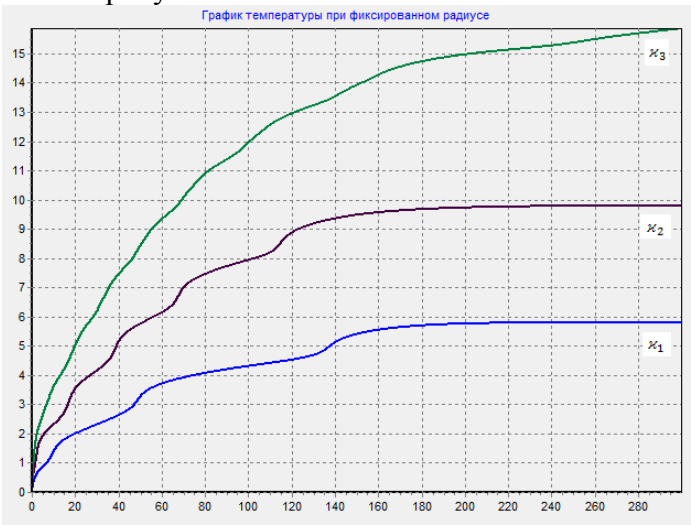


Рисунок 4. Графики зависимости температуры при фиксированном радиусе для различных значений показателя поглощения $\varkappa$: $\varkappa_1 = 0{,}0000085$; $\varkappa_2 = 0{,}000017$; $\varkappa_3 = 0{,}000034$.

Видим, что графики нагрева пластины при выбранных значениях показателя поглощения $\varkappa$ имеют принципиально качественные отличия.

Литература

1. Карлов, Н.В., Кузьмин, Г.П., Сисакян, Е.В. Интерференционные эффекты при измерении коэффициентов поглощения прозрачных материалов – Квантовая электроника, 1977,т.4,№ 8, 1816-1818 с.
2. Дегтярев, А.А., Сисакян, Е.В., Сойфер, В.А. Модифицированный лазерный калориметрический метод автоматизированного измерения малых оптических потерь – Препринт ИОФАН СССР, 1985,№51, 46с.
3. Дегтярев, А.А., Шамаров, П.А., Комбинированный алгоритм вычисления оптических оценок параметров оптических элементов. // Самара, СГАУ,1995.
4. Ландау, Л.Д., Лифшищ, Е.М. Электродинамика сплошных сред . // М.:Наука, 1982. – 620 с.
5. Самарский А.А., Тихонов А.Н. Уравнения математической физики // М: Наука, 1966.– 724 с.

Имитационное моделирование фондовых рынков

А.А. Стеряков
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Микроскопическая симуляция, или агентное моделирование, - относительно новое направление в имитационном моделировании, использующееся для исследования децентрализованных систем, динамика функционирования которых определяется не глобальными правилами и законами, а, наоборот, когда глобальные правила и законы являются результатом индивидуальной активности членов группы. Такой подход является альтернативой рассмотрению системы как множества одинаковых типичных объектов. В настоящее время микроскопическая симуляция применяется во многих отраслях науки (экология, социология и т.п.) с целью изучить макроскопическое поведение сложной системы, состоящей из множества относительно простых взаимодействующих элементов. Обычно при таком подходе микроскопические элементы и их свойства моделируются с помощью компьютера. Компьютерная симуляция позволяет изучать макроскопические системы, состоящие из большого числа неоднородных элементов, что невозможно сделать в рамках аналитических методов.

В связи с этим становится вполне обоснованным использовать именно микроскопическую симуляцию для изучения закономерностей функционирования финансовых рынков. В рамках этого подхода предлагается рассматривать финансовый рынок, как макроскопическую систему, состоящую из большого числа неоднородных взаимодействующих между собой агентов [1, 2].

В работе в рамках микроскопической симуляции разработана математическая модель биржи одного типа акций. В модели реализуется самоподдерживаемый процесс взаимной торговли между агентами с установлением цены, зависящей только от внутренних параметров биржи. Система реагирует на внешние воздействия и ее реакция в достаточной степени адекватна реальным аналогам. Модель дает адекватно отвечающие реальным временные ряды цен, характеристики которых хорошо согласуются с характеристиками реальных рынков. Математически строго формализованы модели поведения агентов и основные этапы их взаимодействия в системе.

На базе предложенной модели спроектирована и реализована программа «Симулятор биржи», представляющая собой компьютерную реализацию разработанной модели агентного типа. Разработанный программный пакет предоставляет набор удобных инструментов для детального изучения модели. Предусмотрена возможность введения пользовательского агента, позволяющая оказывать внешнее влияние на систему, тем самым исследуя ее реакцию, и использовать программу в качестве простейшего биржевого тренажера. Использованный при проектировании объектный подход допускает изменение отдельных этапов или объектов математического моделирования без существенной переделки всего программного пакета.

Модели агента и взаимодействия агентов в системе

Для выполнения микроскопической симуляции необходимо описать все элементы системы и задать способ их взаимодействия. Предлагаемая имитационная модель представляет собой совокупность биржевых агентов, действующих автономно, имеющих различные параметры и взаимодействующих только посредством биржи, а

именно с помощью одного из финансовых инструментов – биржевого стакана, или стека.

Предполагается, что каждый агент обладает портфелем, состоящим из безрискового актива (денежный банковский счет), характеризующийся постоянным банковским процентом и рискового актива (акции, приносящие, в общем случае, различные дивиденды). Также агент использует функцию полезности, зависящую от общей стоимости его портфеля. В таком случае поведение каждого агента во времени будет основано на периодическом изменении состава личного портфеля, состоящего из двух активов таким образом, чтобы ожидаемый в будущем прогноз стоимости этого портфеля (основанный на прогнозе цены и волатильности рискового актива) максимизировал функцию полезности агента [3, 4, 5]. Тогда решение агента на каждой итерации можно представить в виде желаемой доли стоимости акций в портфеле, которую в случае показательной функции полезности можно представить в виде

$$\pi_i(t) = \frac{\mu_i(t+1) + \delta / P(t) - r}{\beta W_i(t) \sigma_i^{\,2}(t+1)},$$

где $P(t)$ - цена одной акции в момент времени t, δ - дивиденды, выплачиваемые за одну акцию в единицу времени, r - безрисковая процентная ставка в единицу времени, $W_i(t)$ - стоимость портфеля агента в момент времени t, β - коэффициент, характеризующий функцию полезности. Чтобы принять решение, агенту также необходимо спрогнозировать математическое ожидание $\mu_i(t+1)$ и дисперсию $\sigma_i^{\,2}(t+1)$ относительного изменения цены.

Прогноз может выполняться различными способами. Например, для так называемого технического аналитика он определяется экспоненциально взвешенным средним предыдущих значений прогнозируемых величин. В случае если агент использует значения только на одном предыдущем шаге:

$$\mu_i(t+1) = \lambda_i \mu_i(t) + (1-\lambda_i)\rho(t),$$

$$\sigma_i(t+1)^2 = \lambda_i \sigma_i(t)^2 + (1-\lambda_i)\big(\rho(t) - \mu_i(t)\big)^2.$$

Всё взаимодействие агентов происходит посредством инструментов биржи, а именно, биржевого стакана и в результате удовлетворения их потребностей в продаже или покупке акций. Формально процесс взаимодействия можно разбить на четыре этапа.

На первом этапе происходит формирование заявок. По решению агента на каждом шаге строится выставляемая им заявка, состоящая из трёх элементов: количества акций, цены сделки и её направления (покупка или продажа). Следующий этап - заполнение биржевого стакана. В течение торгов, в каждый момент времени, в базе данных биржи находится очередь неисполненных заявок на продажу (offers) и на покупку (bids). Эта сторона фондового рынка хорошо отражена в предлагаемой модели. Подобно реальному, смоделированный биржевой стек собирает все заявки, сформированные на первом этапе, разделяет их по направлению на две группы и сортирует каждую в необходимом порядке. На третьем этапе выполняется переформирование стакана. В ходе торгов, продавец может немного снижать свою цену, а покупатель — повышать. Иными словами, разница между ценами торгующих сторон (т.е. спрэд в стакане) сужается. И когда продавец и покупатель сходятся в своих мнениях о цене, совершается сделка. Для отражения таких отношений в модели предлагается переформировывать полученный после второго этапа взаимодействия стакан. Такое действие будет отвечать процессу взаимных уступок между агентами. На последнем этапе происходит удовлетворение заявок. После переформирования в стакане наличествует некоторое количество перекрывающихся по цене заявок со

стороны покупателей и продавцов. Для того чтобы приблизить модель к реальности, условно торг начинается между самой лучшей заявкой на продажу (самой дешевой по цене для покупателей) и самой лучшей заявкой на покупку (самой дорогой по цене для продавцов). Сделка происходит по цене продавца, если цена лучшего покупателя выше и объем его заявки больше, он частично удовлетворяет свою заявку по цене лучшего продавца и переходит к следующему в стеке продавцов. При удовлетворении (полном или частичном) заявки происходит соответствующее изменение количества акций и денежного актива у «хозяина» заявки. Так происходит поочередное удовлетворение заявок из двух стеков продавцов и покупателей, пока лучшая цена покупателей (т.е. самая высокая) еще больше лучшей цены продавцов (т.е. самой низкой). Цена последней зафиксированной сделки заносится во временной ряд биржи. Оставшиеся неудовлетворенными заявки снимаются и на следующей итерации все этапы повторяются заново.

Компьютерное представление

В рамках работы спроектирована и реализована программа «Симулятор биржи», представляющая собой компьютерную реализацию разработанной модели агентного типа. Программа выполнена в концепции объектно-ориентированного программирования на языке Java SE v1.6 в свободной интегрированной среде разработки приложений NetBeans IDE. Примерный вид интерфейса приложения (все доступные окна) представлен на рисунке 1.

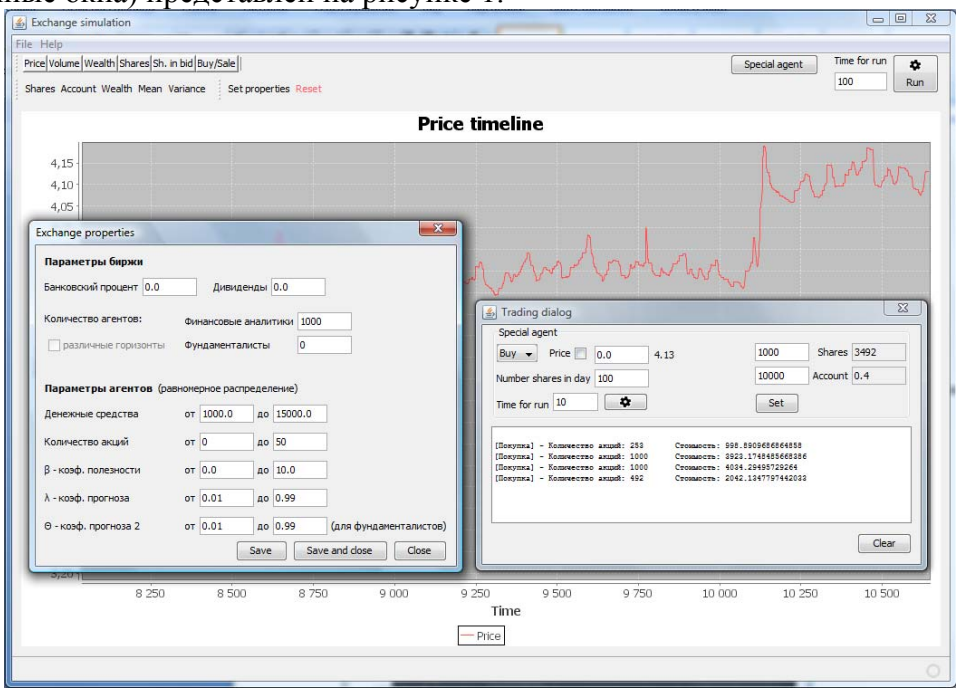


Рисунок 1. Общий вид приложения в ОС Vista

Таким образом, выполнена не только задача моделирования биржи одного типа акций с помощью микроскопической симуляции, но и реализован удобный и эффективный инструмент изучения данной модели, позволяющий кроме полного исследования системы при различной вариации параметров и внешних возмущений (имеется в виду реализованная возможность введения пользовательского агента в систему) легко вносить и значительные коррективы в модель на различных уровнях и этапах ее компьютерного представления, тем самым улучшая те или иные ее качества.

Численные исследования

Для проведения численных исследований была выбрана более конкретная модификация системы, а именно: наличие на рынке одного типа агентов (технических

аналитиков), одинаковый для всех агентов интервал начального распределения денежных средств и акций, нулевые дивиденды и процентная банковская ставка.

С помощью реализованной программы, визуализирующей процесс торговли в рассматриваемой модели, а также позволяющей проводить ее исследования, изменяя большинство параметров и отслеживая статистические данные, были получены различные режимы работы системы. Удалось установить некоторые закономерности в процессе ценообразования, а также выявить характер зависимости режимов торговли от различных параметров. В том числе получилось найти режимы, в которых модель генерирует временные ряды со статистическими характеристиками, близкими к характеристикам реальных временных рядов с фондовых рынков. Была проведена численная оптимизация стратегии торговли пользовательского агента на моделируемой бирже.

По результатам исследования работы системы ясно, что замкнутая система (т.е. без дивидендов, без притока извне денежных средств) имеет свою равновесную цену, вокруг которой совершаются хаотические колебания. Иллюстративные графики выхода системы на режим колебаний вокруг стационарного положения приведены на рис. 2.

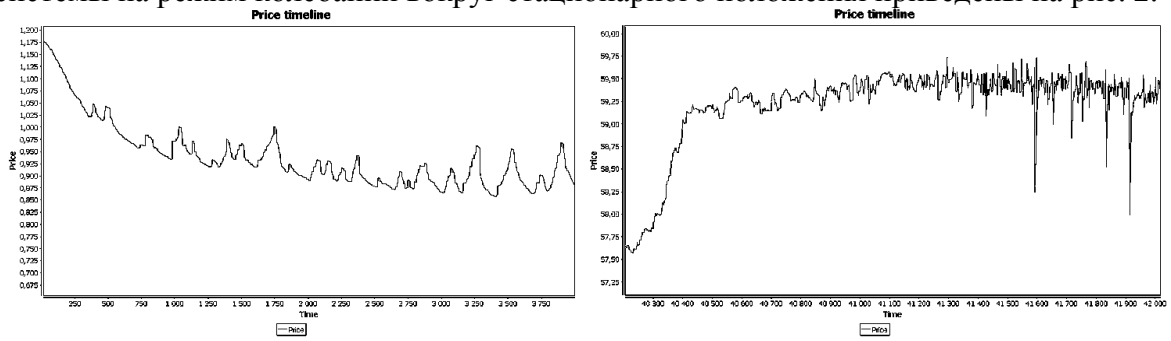


Рисунок 2. Выход системы на равновесный режим

Очень интересные результаты дало исследование модели на статистические характеристики. А именно при анализе автокорреляционной функции было выяснено, что она в значительной мере зависит от коэффициента отклонения цены в заявке, который характеризует разброс цен, по которым хотят торговать агенты в модели. При малых значениях параметра (около 0,01 и вплоть до 0,1-0,2), что соответствует малому шуму в ценообразовании, наблюдается довольно заметная автокорреляция вплоть до лага в 100 шагов (временных единиц системы). С увеличением параметра (около 0,3-0,4) автокорреляция уменьшается, а главное становится отрицательной при малых лагах. Дальнейшее увеличение (начиная с 0,7-0,8) приводит к исчезновению автокорреляции даже при малых лагах. Графические результаты автокорреляционной функции при типичных значениях параметра приведены на рисунке 3.

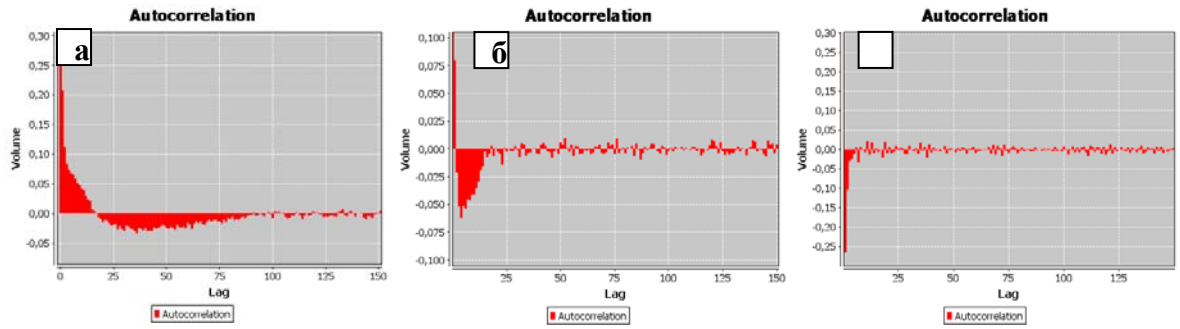


Рисунок 3. Зависимость автокорреляции от коэффициента отклонения цены в заявке (а. $\nu = 0.025$, б. $\nu = 0.35$, в. $\nu = 1.0$)

Таким образом, т.к. случай с отрицательной автокорреляцией на малых лагах и быстрым ее затуханием хорошо согласуется с характеристиками реальных рынков, модель при таком выборе коэффициента отклонения цены в заявке дает адекватно отвечающие реальным временные ряды цен.

Литература

1. M. Levy, H. Levy, and S. Solomon. A microscopic model of the stock market: Cycles, booms, and crashes. Economics Letters, 45:103–111, 1994.

2. M. Levy, H. Levy, and S. Solomon. Microscopic Simulation of Financial Markets. Academic Press, New York, 2000.

3. Brock, W. and Hommes, C. Heterogeneous beliefs and routes to chaos in a simple asset pricing model, Journal of Economic Dynamics and Control, 1998, v. 22, 1235–1274.

4. C. Chiarella, X. He, 2001. "A Non-Stationary Asset Pricing Model under Heterogeneous Expectations," Computing in Economics and Finance 2001 39, Society for Computational Economics.

5. M.Anufriev, G.Bottazzi, F.Pancotto. Journal of Economic Dynamics and Control, 2006, v.30, 1787-1835.

Исследование устойчивости закрученных течений со скачками скорости и плотности потока

С.С. Сугак, И.П. Завершинский
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Аннотация

Представлена модель возникновения прецессирующих вихревых ядер в закрученных струях, составлена и решена система уравнений для цилиндрического закрученного течения с перепадом скорости и плотности потока на границе разрыва.

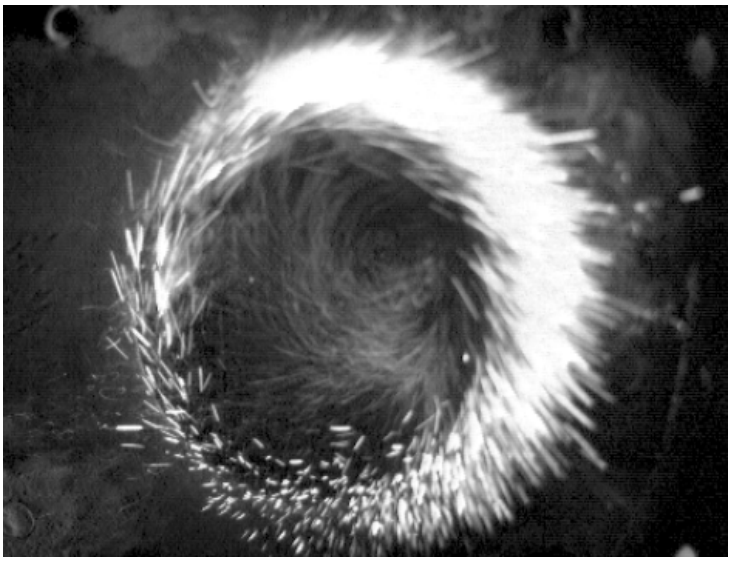
Содержание работы

Закрученные струи являются важным объектом изучения с точки зрения практических приложений в связи с широким распространением вихревых горелочных устройств. С другой стороны, в закрученных струях реализуется богатое разнообразие режимов течения и эффектов, представляющих большой научный интерес [1].

Основными гидродинамическими неустойчивостями являются неустойивостями Кельвина-Гельмгольца и Рэлея-Тэйлора. Неустойчивость Кельвина-Гельмгольца впервые была описана Кельвином для несжимаемой жидкости. Она проявлялась как цепочка вихрей расположенный на прямой, вдоль которой существует тангенциальный разрыв скорости. Впоследствии её стали называть “неустойчивостью тангенциального разрыва скорости” или просто “неустойчивостью тангенциального разрыва”.

В динамике различных сплошных сред неустойчивость Рэлея-Тейлора имеет собственные названия, связанные с историей её открытия в данной среде. В гидродинамике она известна как неустойчивость тяжёлой жидкости над лёгкой, в физике плазмы – желобковая неустойчивость и т.д. Пример развития этой неустойчивости можно наблюдать, если в стакан с жидким маслом налить немного воды. Можно заметить множество водяных желобков, опускающихся на дно стакана (отсюда и название – желобковая неустойчивость). [4]

В зависимости от соотношения перепада скорости и плотностипроявляются обе эти неустойчивости. Например желобковая неустойчивость может быть усилена или ослаблена центробежной силой, а неустойчивость Кельвина-Гельмгольца стабилизирована отрицательным градиентом плотности [4].



На рисунке представлена фотография прецессирующего ядра полученного в экспериментах описанных в работе [2].

В данной работе рассмотрена модель сжимаемого невязкого цилиндрического закрученного потока. Общие характеристики этой модели повторяют модель течений экспериментально изученных в работах [1] и [2]. В указанных работах были описаны структуры, возникающие в закрученных газовых потоках, в частности прецессирующие вихревые ядра (ПВЯ).

Система линеаризованных динамических уравнений для предложенной модели описывается уравнениями

$$\begin{cases} \dfrac{\partial \upsilon_r'}{\partial t} + \dfrac{\upsilon_{\varphi o}}{r}\dfrac{\partial \upsilon_r'}{\partial \varphi} + \upsilon_{zo}\dfrac{\partial \upsilon_r'}{\partial z} - \dfrac{2\upsilon_{\varphi o}\upsilon_\varphi'}{r} = -\dfrac{1}{\rho_o}\dfrac{\partial}{\partial r}(c_{so}^2\rho') \\ \dfrac{\partial \upsilon_\varphi'}{\partial t} + \upsilon_r'\dfrac{\partial \upsilon_{\varphi o}}{\partial r} + \dfrac{\upsilon_{\varphi o}}{r}\dfrac{\partial \upsilon_\varphi'}{\partial \varphi} + \upsilon_{zo}\dfrac{\partial \upsilon_\varphi'}{\partial z} + \dfrac{\upsilon_r'\upsilon_{\varphi o}}{r} = -\dfrac{1}{\rho_o}\dfrac{\partial}{r\partial \varphi}(c_{so}^2\rho') \\ \dfrac{\partial \upsilon_z'}{\partial t} + \upsilon_r'\dfrac{\partial \upsilon_{zo}}{\partial r} + \dfrac{\upsilon_{\varphi o}}{r}\dfrac{\partial \upsilon_z}{\partial \varphi} + \upsilon_{zo}\dfrac{\partial \upsilon_z'}{\partial z} = -\dfrac{1}{\rho_o}\dfrac{\partial}{\partial z}(c_{so}^2\rho') \\ \dfrac{\partial \rho'}{\partial r} + \dfrac{\rho_o}{r}\dfrac{\partial (r\upsilon_r')}{\partial r} + \upsilon_r'\dfrac{\partial \rho_o}{\partial r} + \dfrac{\partial}{r\partial \varphi}(\rho_o\upsilon_\varphi' + \rho'\upsilon_{\varphi o}) + \dfrac{\partial}{\partial z}(\rho_o\upsilon_z' + \rho'\upsilon_{zo}) = 0 \end{cases} \quad (1)$$

с условиями на сшивке

$$\left[\eta c_{so}^2 + R\Omega_o^2\xi\right]_{R-0}^{R+0} = 0,$$

$$\left[\xi\rho_o\right]_{R-0}^{R+0} = 0$$

и начальными условиями

$$\Omega_o(r) = \begin{cases} \Omega_1, r \le R \\ \Omega_2, r > R \end{cases} \qquad \upsilon_{zo}(r) = \begin{cases} \upsilon_1, r \le R \\ \upsilon_2, r > R \end{cases}$$

$$c_{so}(r) = \begin{cases} c_{s1}, r \le R \\ c_{s2}, r > R \end{cases} \qquad \rho_o(r) = \begin{cases} \rho_1, r \le R \\ \rho_2, r > R \end{cases}$$

где υ_r, υ_φ, υ_z, ρ - радиальная, тангенциальная и продольная компоненты скорости течения и плотность соотвественно, с индексом “0” – невозмущенная часть, с знаком “тильда” – возмущенная, c_{so} - скорость звука в жидкости, r - расстояние до оси вращения, $\eta = \dfrac{\rho'}{\rho_o}$ - относительный перепад плотности на границе разрыва, ξ - радиальное смещение, определяемое по возмущению радиальной скорости $-i\omega\xi = \upsilon_r'$, $\Omega_0(r) = \dfrac{\upsilon_{\varphi o}}{r}$.

В ходе предварительного исследования неустойчивостей была рассмотрена модель галактического газового диска со скачками скорости и плотности. Подробно она была рассмотрена в работах [3],[4] и [5]. В данных работах исследуется проблема спиральной структуры галактик. Основная трудность заключается в том, что галактические диски, в плоскости которых расположены спиральные рукава, вращаются дифференциально: с удалением от центра галактики угловая скорость вращения падает – в основной области диска обратно пропорционально расстоянию от центра. Это означает, что если спиральные рукава – уплотнения из газа и молодых звёзд – вращались бы так же дифференциально, как и газовый диск, в котором они наблюдаются, то со временем они растянулись бы настолько, что после 1-2 оборотов переферии диска спиральные рукава стали бы уже с трудом отличимы от фона. Между тем наша спиральная Галактика в окрестности Солнца совершила порядка 50 оборотов, сохранив четко выраженную спиральную структуру [3].

Решением проблемы оказалось рассмотрение модели вращающегося газового диска со скачком скорости и плотности. Полученное для такой модели дисперсионное уравнение и его решение является частным решением уравнений (1) если принять движение потока вдоль оси Oz стационарным.

Итогом проведённой работы является решение дисперсионного уравнения, которое имеет вид

$x = \frac{\omega}{\Omega_1} = \frac{b \pm i\sqrt{ac - b^2}}{a}$, где в общем случае

$$a = 1 + Q$$

$$b = m - 1 + qQ(m+1) + \frac{k}{S}(1 + aQ)$$

$$c = m(1 - q^2) + (-2m + m^2 + 2mq^2Q + 2\frac{k}{S}(m + mqQa - 1) + \frac{k^2}{S^2}(1 + a^2Q))$$

$$Q = \frac{\rho_1}{\rho_2}; \quad q = \frac{\Omega_2}{\Omega_1}; \quad S = \frac{R\Omega_1}{\upsilon_{z1}}; \quad a = \frac{\upsilon_{z2}}{\upsilon_{z1}}$$

где m - волновое число по оси φ в цилиндрических координатах, k - волновое число по оси z.

Из полученного решения можно извлечь критерий формирования неустойчивости имеющий вид прецессирующего вихревого ядра (ПВЯ).

Литература

1. Шторк С.И., Кала К.Э., Фернандес Э.К., Хейтор М.В.// Письма в ЖТФ. 2002, т.31. №15, с.62.

2. Anacleto P.M., Fernandes E.C., Heitor M.V., Stork S.I.// *CHARACTERIZATION OF A STRONG SWIRLING FLOW* WITH. PRECESSING VORTEX CORE BASED ON MEASUREMENTS OF VELOCITY AND LOCAL PRESSURE FLUCTUATIONS. International Symposia on Applications of Laser Techniques to Fluid Mechanics

3. Фридман А.М.// УФН. 2007. т.177. №2. с.121.

4. Фридман А.М.// УФН. 2008. т.178. №3. с.225.

5. Морозов А.Г., Хоперсков А.В.// Физика дисков. ВолГУ, 2005.

Термические напряжения в ледяном массиве при замерзании воды в цилиндрической трубе

А.Ф. Федечев[1], Н.Л. Федечева[2]
[1]*Самарский государственный университет, Самара*
[2]*Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*

Термические напряжения в слое льда, возникающее в процессе замерзания воды в следствие ее расширения, имеют большое значение в отраслях промышленности, связанных с замораживанием пищевых продуктов или производством льда. Кроме того, они могут привести к разрушению трубопроводов или резервуаров с водой. К сожалению, теоретический анализ возникновения термических напряжений затруднен наличием подвижной границы фазового перехода. Точные решения задач при наличии фазовых превращений из-за их нелинейности получить, как правило, не удается.

В настоящей работе методом последовательных приближений [1] получено приближенное аналитическое решение задачи о термических напряжениях в цилиндрическом слое льда при замерзании воды.

Пусть жидкость заполняет некоторую область $r \le R$. Температура жидкости поддерживается постоянной и равной $T_{ж}$, причем $T_{ж}$ выше температуры фазового перехода T_L. Будем считать, что теплофизические свойства воды и льда заданы и постоянны, а также известна теплота фазового перехода Q. В первом приближении лед можно рассматривать как поликристаллический материал с изотропными свойствами. В этом случае хрупкие свойства льда характеризуются упругой деформацией, сменяемой внезапным разрушением [2]. Поэтому до начала разрушения к поликристаллическому льду можно применять теорию упругости в квазистатическом приближении [3]. Будем считать, что касательными напряжениями на поверхности раздела лед- стенка трубы – можно пренебречь. При сделанных допущениях для определения термических напряжений приходим к осесимметричной задаче:

$$\rho c \frac{\partial T}{\partial \tau} = \lambda \frac{1}{r} \frac{\partial}{\partial r}\left(r \frac{\partial T}{\partial r} \right), \quad y(\tau) < r < R, \quad \tau > 0; \tag{1}$$

$$\frac{d^2 u}{dr^2} + \frac{1}{r}\frac{du}{dr} - \frac{1}{r^2} u = \frac{1+\upsilon}{1-\upsilon} \alpha_T \frac{d}{dr}\left(T + \Phi\right); \tag{2}$$

$$\lambda \frac{\partial T}{\partial r} = \alpha_1 \left(T_c - T\right), \quad u = 0, \quad \text{при } r = R; \tag{3}$$

$$\lambda \frac{\partial T}{\partial r} = \alpha_2 \left(T - T_{ж}\right) + \rho Q \frac{dy}{d\tau}, \quad T = T_L, \quad \sigma_{rr} = 0, \quad \text{при } r = y(\tau); \tag{4}$$

$\int_y^R r\sigma_{zz} dr = 0,$ - условие равновесия льда при отсутствии касательных напряжений. (5)

Здесь приняты следующие обозначения:

ρ - плотность льда, c - коэффициент теплоемкости, λ - коэффициент теплопроводности, $y(\tau)$ - координата поверхности раздела лед-вода, u - проекция вектора перемещения на радиус, υ - коэффициент Пуассона, α_T - коэффициент линейного теплового расширения, β_T - коэффициент увеличения объема при замерзании, α_1, α_2 - коэффициенты теплообмена,

$$\sigma_{rr} = \frac{E(1-\upsilon)}{(1+\upsilon)(1-2\upsilon)}\frac{du}{dr} + \frac{E\upsilon}{(1+\upsilon)(1-2\upsilon)}\left(\frac{u}{r}+k\right) + \frac{E\alpha_T}{(1-2\upsilon)}(T+\Phi),$$

$$\sigma_{zz} = \frac{E(1-\upsilon)}{(1+\upsilon)(1-2\upsilon)}k + \frac{E\upsilon}{(1+\upsilon)(1-2\upsilon)}\left(\frac{u}{r}+\frac{du}{dr}\right) + \frac{E\alpha_T}{(1-2\upsilon)}(T+\Phi),$$ - нормальные компоненты тензора напряжений, E - модуль упругости, $\Phi = \frac{\beta_T}{\alpha_T}$, $k = \frac{dw}{dz}$ - искомая const.

Переходя к безразмерным величинам $\xi = \frac{r}{R}, z = \frac{y(\tau)}{R}, Fo = \frac{\lambda}{\rho c R^2}\tau, \theta = \frac{(T-T_L)}{(T_c-T_L)}\frac{1}{b}$, $b = \frac{R\alpha_1}{\lambda}$, $\gamma = \frac{\alpha_2(T_L-T_{ж})}{\alpha_1(T_c-T_L)}$, $\beta = \frac{Q}{cb(T_c-T_L)}$, $\bar{u} = \frac{(1-\upsilon)}{(1+\upsilon)\alpha_T bR(T_c-T_L)}u$, $B = \frac{(\Phi+T_L)}{b(T_c-T_L)}$, $\bar{\sigma} = \frac{(1-2\upsilon)}{\alpha_T bE(T_c-T_L)}\sigma$, система (1-5) примет вид:

$$\frac{\partial\theta}{\partial Fo} = \frac{1}{\xi}\frac{\partial}{\partial\xi}\left(\xi\frac{\partial\theta}{\partial\xi}\right), \; z(Fo) < \xi < 1, \; Fo > 0; \tag{6}$$

$$\frac{d^2\bar{u}}{d\xi^2} + \frac{1}{\xi}\frac{d\bar{u}}{d\xi} - \frac{1}{\xi^2}\bar{u} = \frac{d}{d\xi}(\theta + B); \tag{7}$$

$$\frac{\partial\theta}{\partial\xi} = 1 - b\theta, \quad \bar{u} = 0, \quad \text{при } \xi = 1; \tag{8}$$

$$\frac{\partial\theta}{\partial\xi} = \gamma + \beta\frac{dz}{dFo}, \quad \theta = 0, \quad \bar{\sigma}_{rr} = 0, \text{ при } \xi = z(Fo); \tag{9}$$

$\int_{я}^{1} \xi\,\bar{\sigma}_{яя}\,d\xi = 0,$ - условие равновесия льда при отсутствии касательных напряжений.(10)

Проинтегрируем уравнение теплопроводности(6) в пределах от ξ до1.

$$\frac{\partial\theta}{\partial\xi}(Fo,\xi) = \frac{1}{\xi}(1 - b\theta|_{\xi=1}) - \frac{1}{\xi}\int_{\xi}^{1}\xi\frac{\partial\theta}{\partial Fo}d\xi. \tag{11}$$

Из последнего выражения с использованием граничного условия (9) будем иметь

$$\gamma + \beta\frac{dz}{dFo} = \frac{1}{z}(1 - b\theta|_{\xi=1}) - \frac{1}{z}\int_{z}^{1}\xi\frac{\partial\theta}{\partial Fo}d\xi. \tag{12}$$

Воспользуемся далее тем обстоятельством, что $\frac{\partial}{\partial Fo} = \frac{\partial}{\partial z}\frac{\partial z}{\partial Fo}$ и производная $\frac{\partial z}{\partial Fo}$ не зависит от координаты ξ. После несложных преобразований получаем выражение для скорости движения границы раздела фаз $\frac{\partial z}{\partial Fo}$ в виде

$$\frac{\partial z}{\partial Fo} = \frac{(1 - b\theta(Fo,1) - \gamma z)}{\left(\beta z + \int_{z}^{1}\xi\frac{\partial\theta}{\partial z}d\xi\right)} \tag{13}.$$

Откуда $$Fo = \int_z^1 \frac{\left(\beta z + \int_z^1 \xi \frac{\partial \theta}{\partial z} d\xi\right)}{\left(1 - b\theta(Fo,1) - \gamma z\right)} dz\,. \qquad (14)$$

Проинтегрируем уравнение (11) в пределах от z до ξ

$$\theta(z,\xi) = \ln\left(\frac{\xi}{z}\right)(1 - b\theta(z,1)) - \frac{dz}{dFo}\int_z^\xi \frac{1}{\xi}\left(\int_\xi^1 \xi \frac{\partial \theta}{\partial z} d\xi\right) d\xi. \qquad (15)$$

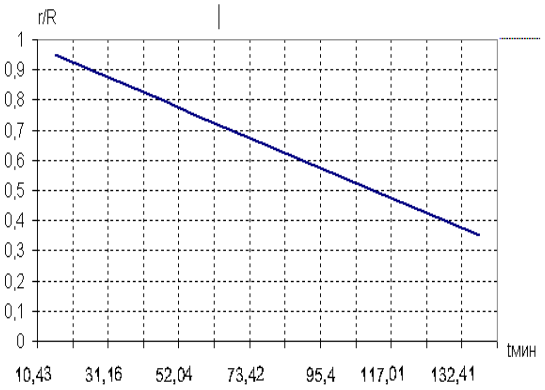
Уравнение (15) представляет собой нелинейное интегро-дифференциальное уравнение, которое решается методом итерации, вычислительная схема которого имеет следующий вид:

$$\theta_{j+1}(z,\xi) = \ln\left(\frac{\xi}{z}\right)(1 - b\theta_j(z,1)) - \frac{\left(1 - b\theta_j(z,1) - \gamma z\right)}{\left(\beta z + \int_z^1 \xi \frac{\partial \theta_j}{\partial z} d\xi\right)}\int_z^\xi \frac{1}{\xi}\left(\int_\xi^1 \xi \frac{\partial \theta_j}{\partial z} d\xi\right) d\xi. \qquad (16)$$

В качестве нулевого приближения берется выражение $\theta_0(z,\xi) = \ln\left(\frac{\xi}{z}\right)$.

Подставляя нулевое приближение в (16), находим первое приближение

$$\theta_1(z,\xi) = (1 + b\ln(z))\ln\left(\frac{\xi}{z}\right) - \frac{(1 - \gamma z + b\ln(z))}{2((1+2\beta)z^2 - 1)}\left(\xi^2 - z^2 + 2\ln\left(\frac{z}{\xi}\right)\right) \text{ и т д.}$$



фазового перехода в зависимости от времени.

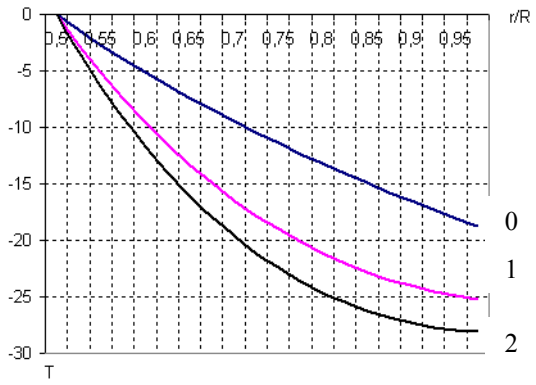


Рис.2. Изменение температуры по толщине слоя льда в зависимости от r/R(индексы соответствуют номерам приближений)

Итерационный процесс сходится достаточно быстро. Характер сходимости этого процесса представлен на рис.2. После определения θ интегрируем уравнение (7)

$$\bar{u} = \frac{1}{\xi}\int_\xi^1 (\theta_1)\xi d\xi + \frac{C_1\xi}{2} + \frac{C_2}{\xi} = \left[-(1 + b\ln z)(0{,}25 + 0{,}5\ln z) - a(z)(0{,}75 + \ln z - 0{,}5z^2)\right]\frac{1}{\xi} -$$

$$-\left[(1 + b\ln z)\left(\frac{\xi}{2}\ln\xi - \frac{\xi}{4} - \frac{\xi\ln z}{2}\right) - a(z)\left(\frac{\xi^3}{4} - \frac{\xi z^2}{2} + \frac{\xi}{2} + \xi\ln\frac{z}{\xi}\right)\right]. \qquad (17)$$

Из граничного условия (8) получаем равенство $C_2 = -0{,}5C_1$. (18)

Граничные условия $\bar{\sigma}_{rr} = \frac{d\bar{u}}{d\xi} + \frac{\upsilon}{1-\upsilon}\frac{\bar{u}}{\xi} + \bar{k} - B - \theta_1 = 0$ при $\xi = z$; и

$$\int_{я}^{1} \xi \bar{\sigma}_{яя} d\xi = \int_{я}^{1} \xi \left[\frac{1-\upsilon}{\upsilon} \bar{k} + \frac{\upsilon}{1-\upsilon} \left(\frac{\bar{u}}{\xi} + \frac{d\bar{u}}{d\xi} \right) - B - \theta \right] d\xi = 0$$ совместно с равенством (18)

позволяют определить три неизвестные постоянные C_1, C_2, k. После этого можно определить напряжения и деформации внутри льда.

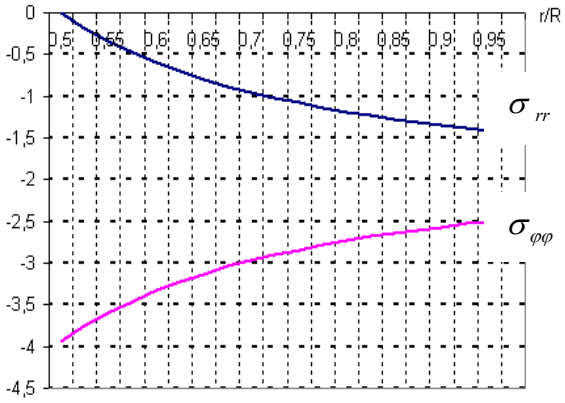


Рис.3. Изменение напряжений внутри льда при z=r/R=0,5.

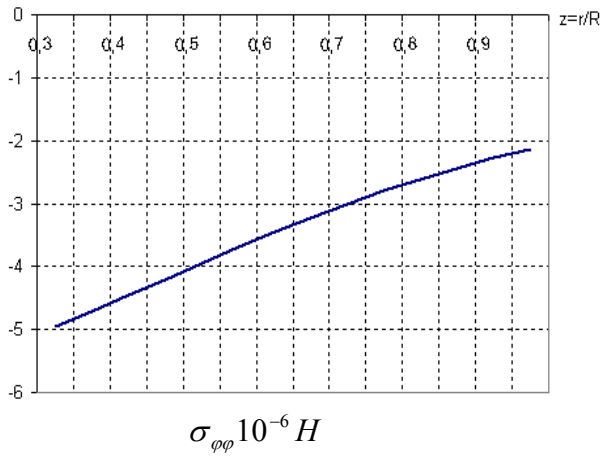


Рис.4. Изменение окружных напряжений на фронте фазовых превращений.

Из рис.3. видно, что окружные напряжения внутри льда намного превышают радиальные. Максимум окружного сжимающего напряжения находится на поверхности раздела лед-вода. Из рис.4. следует, что окружные напряжения на фронте фазового перехода возрастают с ростом толщины ледяного слоя.

В качестве примера приведен численный расчет при следующих данных для льда

$$\text{R}=0{,}15\,м;\ \ \lambda=2\ \frac{Вт}{м\cdot{}^0C};\ c_p=\frac{Дж}{кг\cdot{}^0C};\ \text{G}=333840\,\frac{Дж}{кг};\ \ \rho=910\,\frac{кг}{м^3};\ \ \text{E}=0{,}2\cdot10^9\,Па;$$

$$\upsilon=0{,}31;\ \ \beta=0{,}01;\ \alpha_1=15\,\frac{Вт}{м^2\cdot{}^0C};\ \alpha_2=1000\,\frac{Вт}{м^2\cdot{}^0C};\ \alpha_T=5{,}5\cdot10^{-7};\ T_c=\text{-}25\ {}^0C;$$

$$T_{ж}=4\ {}^0C;\ T_L=0\ {}^0C.$$

Литература

1. Беляев Н.М., Рядно А.А Методы теории теплопроводности. –М.: Высш.школа, Ч.2,1982.
2. Michel B. The Strength of Polycrystalline Ice. Can. J. of Civil Eng. Vol. 5, No. 3, pp.285-300.
3. Боли Б., Уэйнер Дж. Теория температурных напряжений. – М.: Мир, 1964.

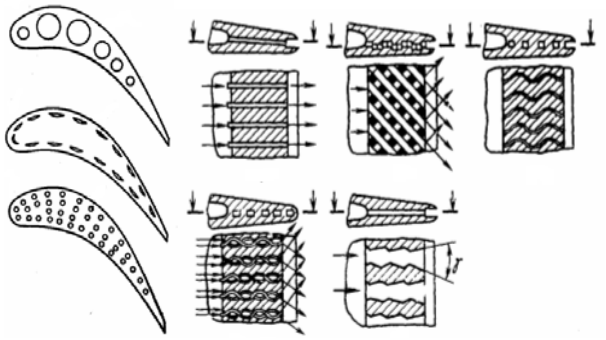
CFD-моделирование охлаждения лопатки с вихревой матрицей

Л.С. Шаблий
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Для повышения экономичности современных авиационных ГТД температуру газа на входе в турбину $T^*_г$ увеличивают до 1500...1800 К. Это существенно превосходит допустимую с точки зрения прочности температуру ($T_{доп}$ = 1150...1250 К) материала турбинных лопаток [1]. Чтобы предотвратить разрушение лопаток при высоких значениях $T^*_г$ применяют их охлаждение воздухом, продуваемым через внутренние каналы лопатки. Эти каналы должны обеспечивать эффективное равномерное охлаждение лопатки и не допускать местных перегревов материала, вызывающих повышенные напряжения и деформации.

Расход охлаждающего воздуха, отбираемого от компрессора ГТД, желательно иметь как можно меньше, так как его увеличение ухудшает экономичность двигателя в целом. Согласно [2] более высокой эффективностью охлаждения обладают каналы с меньшими значениями гидравлического диаметра. В связи с этим целесообразно уменьшать проходные сечения каналов (одновременно увеличивая их количество) (рисунок 1, а) и увеличивать смоченный периметр каналов за счет видоизменения формы каналов – делая их плоскими, сплющенными, гофрированными и т.д. (рисунок 1, б). Самой совершенной с этой точки зрения является так называемая схема «вихревой матрицы» – набор специально спроектированных, густо переплетенных каналов (рисунок 2). При этом задача оценки эффективности такой схемы охлаждения лопаток особенно на этапе проектирования становится весьма актуальной.



а б

Рисунок 1. Схемы различных каналов охлаждения лопаток: а - количество и размеры; б - форма;

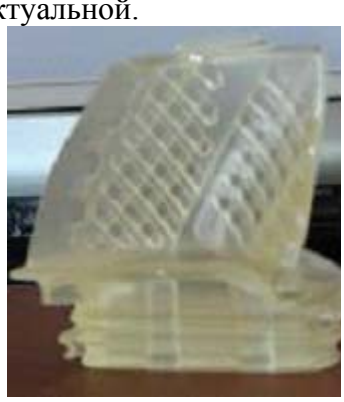


Рисунок 2. Вихревая матрица в стереолитографической модели лопатки

В связи с этим в данной работе была исследована лопатка перспективного ГТД с охлаждающей вихревой матрицей с целью определения минимального расхода охлаждающего воздуха, обеспечивающего допустимую температуру на поверхности лопатки.

Описание расчётной модели

Поставленная задача решалась с помощью программного комплекса Ansys CFX. Задача является сопряженной и включает в себя расчёт потоков, омывающих лопатку внутри и снаружи, а также расчёт тепловых потоков внутри металла лопатки.

Расчётная модель состоит из конечно-элементной зоны газовой фазы потока и соответствующей ей области твёрдого тела (рисунок 3). Для газовой среды рассчитываются уравнения неразрывности, количества движения и теплового баланса,

для твёрдого тела – только уравнение теплового баланса. Связь уравнения теплового баланса разнофазных расчётных зон выполнено с помощью интерфейса «Жидкость-Твёрдое тело». Для зоны газовой фазы на входе задан расход газа, омывающего перо лопатки, на выходе – статическое давление газа за венцом, а также задан расход охлаждающего воздуха, проходящего через вихревую матрицу лопатки. Он подаётся через отверстия в замке лопатки и выходит через четыре отверстия в выходной кромке (рисунок 4). Таким образом в рассмотренной расчётной зоне имеется два входа рабочего тела с разными параметрами и один выход их смеси.

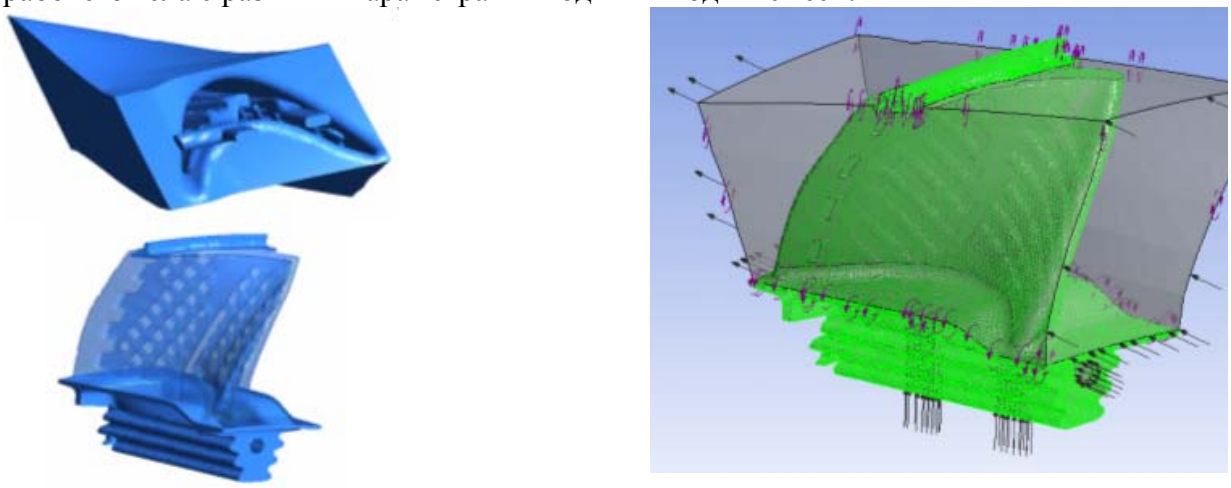


Рисунок 3. Расчётная модель (зелёным цветом выделена металлическая лопатка, охваченная полупрозрачной газовой зоной)

Для расчётной зоны тела лопатки задан тепловой поток через замок. Задача рассчитывается в стационарной постановке на модели одной лопатки, поэтому и для газовой фазы, и для металла на боковых поверхностях задаются условия периодичности, которые позволяют смоделировать расчёт полного лопаточного венца.

Любой канал, в котором происходит течение вязкого газа, для правильного расчёта должен быть разбит сеткой с таким количеством ячеек, которое требуется для разрешения пограничного слоя применяемой моделью турбулентности. Для различных моделей турбулентности требуется от 8-ми до 20 ячеек в пределах пограничного слоя. Применяемая модель турбулентности k-ω (SST) рассчитывает пограничный слой (в отличие, например, от k-ε, которая только моделирует его условно) и требует большое количество ячеек в области пограничного слоя – от 10 до 15. Поэтому при построении модели задавался размер ячеек каналов вихревой матрицы 1/50 эквивалентного диаметра канала. Поскольку вихревая матрица имеет сложную геометрическую форму, разбиение её сеткой конечных элементов с таким размером ячеек приводит к образованию большого количества элементов и «утяжелению» задачи до такого уровня, что решать её на персональном компьютере становится невозможно по причине нехватки памяти и неоправданно большого времени счёта. Кроме того, при объединении жидкой и твёрдой зон интерфейсом «Жидкость-Твёрдое тело», CFD-решатель при запуске упорядочивает области сетки на обеих зонах, ставя каждой ячейке в соответствие пару. На этот процесс, так называемую интерполяцию, также требуется дополнительное время и память, пропорциональные размерам и сложности интерфейса. Для обширного интерфейса разветвленной вихревой матрицы, имеющей большую площадь поверхности, это представляет существенные затраты по памяти и времени счёта. Поэтому решение задачи прямой оптимизации выполнено на суперкомпьютере «Сергей Королёв» в режиме параллельного счёта HP Parallel.

Результаты моделирования

Первоначально было получено распределение температур на поверхности лопатки при отсутствии охлаждающего газа (нулевой расход) и при вдуве охлаждающего газа (рисунок 4) с температурой 800 К и расходом 2 мг/с (6% от расхода рабочего газа). Максимальные значения температур в обоих случаях наблюдаются в области бандажной полки (рисунок 5) и составляют соответственно 1338 К и 1068 К. Температура лопатки без охлаждения выше допустимой температуры в 1150 К. При применении охлаждения температура снизилась ниже допустимого уровня.

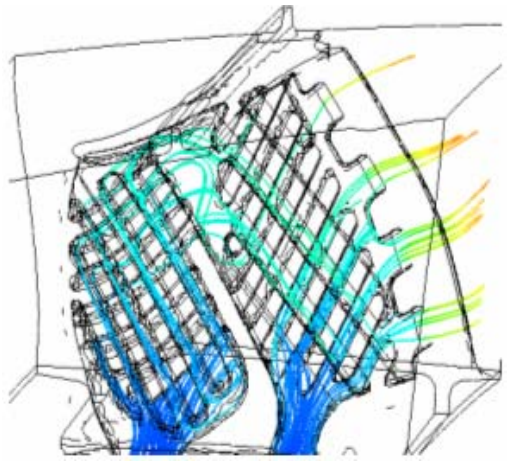


Рисунок 4. Вдув охлаждающего воздуха в вихревую матрицу лопатки

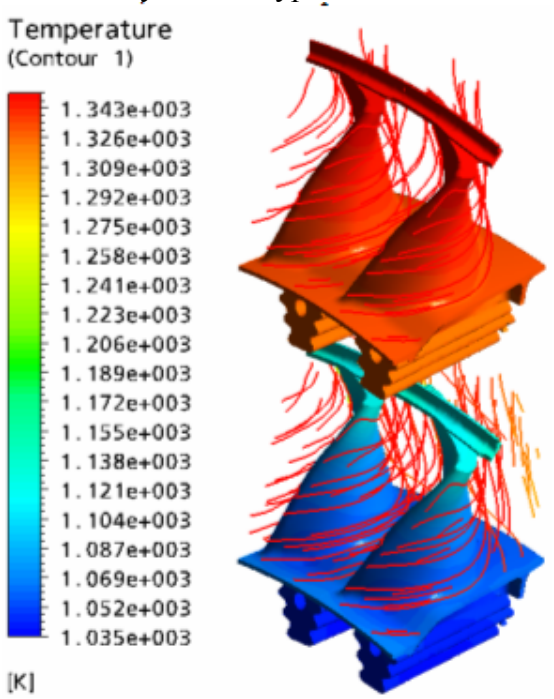


Рисунок 5. Распределение температур поверхности лопатки для случаев без охлаждения и с охлаждением

Таким образом, охлаждение дает возможность снизить температуру лопатки, позволяя ей оставаться работоспособной длительное время. В дальнейшем необходимо было выяснить минимальное значение расхода охлаждающего воздуха, достаточное для обеспечения температуры лопатки не превышающей допустимое значение. Для этого была проведена серия расчётов с изменением расхода охлаждающего воздуха от нуля до 2 мг/с через каждые 0,2 мг/с.

Полученная зависимость максимальной температуры поверхности лопатки от расхода охлаждающего воздуха представлена на рисунке 6. Из этого графика видно, что при расходе охлаждающего воздуха 1 мг/с температура лопатки не превышает 1150К. Интересно также отметить, что с увеличением расхода охлаждающего воздуха снижается эффективность охлаждения лопатки. Так подача первых 0,2 мг/с позволяет снизить температуру на 60К, а увеличение расхода с 1,8 мг/с до 2 мг/с снижает температуру только на 15К.

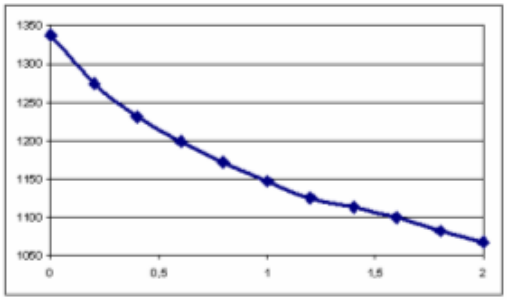


Рисунок 6. Зависимость максимальной температуры поверхности лопатки от расхода охлаждающего воздуха

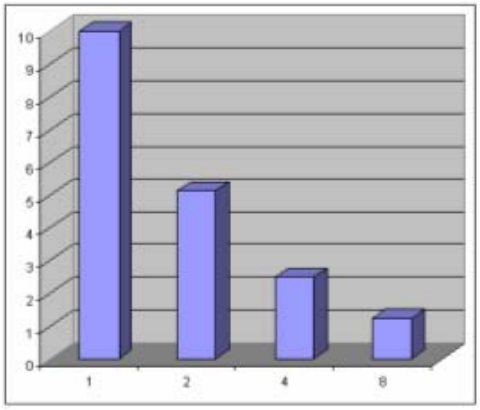


Рисунок 7. Зависимость времени счёта от количества расчётных узлов

Расчёты точек зависимости выполнялись преимущественно на восьми расчётных узлах (ядрах) суперкомпьютера. Для оценки эффективности распараллеливания счёта несколько режимов охлаждения лопатки были посчитаны на четырёх, двух и одном узле. Увеличение количества узлов привело (рисунок 7) практически к пропорциональному уменьшению времени счёта. Это естественно для задач, в которых основное время расчёта занимает решение уравнений, а не пересылка данных. При распараллеливании конечно-элементных задач, они декомпозируются на узлы по данным, то есть каждому узлу отводится своя расчётная подобласть, а соединяются они (интерполируются) по поверхностям этих подобластей. При этом расчёт уравнений во внутренних ячейках подобласти занимает гораздо больше процессорного времени узла, чем пересылка поверхностных данных между узлами. Поэтому увеличение количества узлов (увеличение времени пересылок) не оказывает заметного влияния на общее время процесса, в то время как дробление области на меньшие участки позволяет получить решение пропорционально быстрее.

Таким образом, в процессе расчётного исследования теплового состояния турбинной лопатки с охлаждающей вихревой матрицей на суперкомпьютере, был найден минимальный расход охлаждающего воздуха, обеспечивающий допустимый уровень температуры материала лопатки. Кроме того, было установлено, что при параллельных вычислениях в режиме HP Parallel увеличение числа ядер до восьми обеспечивает практически линейное сокращение времени счёта.

Литература

1. Холщевников К.В., Емин О.Н., Митрохин В.Т. Теория и расчёт авиационных лопаточных машин: Учебник. 2-е изд., перераб. и доп. - М.: Машиностроение, 1986. - 432 с.

2. Кочеров Е.П., Михеенков Е.Л. Охлаждение ступеней турбин авиационных ГТД // Учеб. пособие, Самара: РИО СГАУ, 2007, 168 стр.

Исследование динамики фондового рынка с применением радиально-базисных нейронных сетей

П.В. Левченко

Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Целью научной работы ставилась оценка применимости нейронных сетей радиально-базисного типа при решении задачи прогнозирования на фондовом рынке. Ставилась задача проследить теоретические закономерности между структурными параметрами сети (количество нейронов в скрытом слое, количество обучающих векторов) и точностью прогнозирования. Требовалось также определить влияние размерности векторов данных и длины временного отрезка обучающей выборки на точность получаемого прогноза.

Сети радиального типа имеют ряд существенных при решении задачи прогнозирования преимуществ. Во-первых, как уже сказано, они моделируют произвольную нелинейную функцию с помощью всего одного промежуточного слоя, и тем самым избавляют нас от необходимости решать вопрос о числе слоев. Во-вторых, параметры линейной комбинации в выходном слое можно полностью оптимизировать с помощью хорошо известных методов линейного моделирования, которые работают быстро и не испытывают трудностей с локальными минимумами. Обучаются радиально-базисные сети на порядок быстрее чем, например, многослойный персептрон.

С "групповым" подходом разделения данных связано неумение сетей радиально-базисного типа экстраполировать свои выводы за область известных данных. При удалении от обучающего множества значение функции отклика быстро спадает до нуля. Достоинство это или недостаток - зависит от конкретной задачи, однако в целом склонность к экстраполяции на данные, лежащие далеко от обучающего множества, - вещь, как правило, опасная и необоснованная. Наиболее существенным недостатком радиально-базисной архитектуры является повышенная чувствительность к "проклятию размерности", ситуации, в которой нейронная сеть испытывает значительные трудности, когда число входов велико.

Проведенный теоретический анализ показал возможность и обоснованность применения нейросетевых методов прогнозирования динамики финансового рынка. Основным преимуществом при таком подходе является свойство нейронной находить в поступающей совокупности данных скрытые закономерности, которые, возможно, не будут выявлены аналитическим методом. Недостатком же использования нейросетевой модели анализа является сильное влияние на финансовый рынок и на биржу внешних псевдослучайных факторов (принятие того или иного закона, изъятие или выпуск на рынок больших денежных масс, катастрофы и т.д.). Несмотря на возможность учета таких факторов, точность прогноза в период воздействия оных может существенно понизиться. Обоснованность применения радиально-базисной модели вытекает из соответствия ее качеств требуемым для построения различных оперативных прогнозов на финансовом рынке.

Процесс обучения радиально-базисной сети с учетом выбранного типа радиальной базисной функции сводится:

- к подбору центров c_i и параметров σ_i формы базисных функций;
- к подбору весов нейронов выходного слоя.

Процесс самоорганизации обучающих данных автоматически разделяет пространство на так называемые области Вороного, определяющие различающиеся

группы данных. Данные, сгруппированные внутри кластера, представляются центральной точкой, определяющей среднее значение всех его элементов. Центр кластера отождествляется с центром соответствующей радиальной функции. В нашей работе для корректировки центров радиальных функций мы использовали алгоритм *K-усреднений*. Согласно этому алгоритму центры радиальных базисных функций размещаются только в тех областях входного пространства, в которых имеются информативные данные. Если обучающие данные представляют непрерывную функцию, начальные значения центров в первую очередь размещают в точках, соответствующих всем максимальным и минимальным значениям функции.

Ширина радиальной функции σ_i определяет, насколько "острой" будет гауссова функция. Если эти функции выбраны слишком острыми, сеть не будет интерполировать данные между известными точками и потеряет способность к обобщению. Если же гауссовы функции взяты чересчур широкими, сеть не будет воспринимать мелкие детали (одна из форм проявления дилеммы пере- или недообучения). Как правило, отклонения выбираются таким образом, чтобы колпак каждой гауссовой функций захватывал "несколько" соседних центров. В настоящей работе мы использовали алгоритм *K- ближайших соседей*.

В качестве предметной области при решении задачи прогнозирования динамики фондового рынка нами были выбраны котировки акций компаний, представляющих различные сегменты фондового рынка на Московской Межбанковской Валютной Бирже. В тестировании использовались котировки акций следующих компаний:

- ОАО «Газпром» (ММВБ: GAZP);
- ОАО «Аэрофлот» (ММВБ: AFLT);
- ОАО «Татнефть» (ММВБ:TATN);
- ОАО «Мобильные ТелеСистемы» (МТС) (ММВБ: MTSI);
- ОАО «Газпромнефть» (ММВБ: SIBN).

Такой выбор заранее гарантирует определенную некоррелированность в поведении компаний при тех или иных воздействиях внешних и внутренних факторов даже с учетом того что 2 из 5 компаний относятся к нефтяному сегменту. Следует также отметить, что акции выбранных компаний относятся к так называемым «голубым фишкам» (акции или ценные бумаги наиболее крупных, ликвидных и надёжных компаний со стабильными показателями получаемых доходов и выплачиваемых дивидендов). Поведение акций таких предприятий на фондовой бирже в целом достаточно стабильно и хорошо подходит для проверки полученных с помощью разработанной нейронной сети гипотез.

На этапе подготовки данных задача кодирования входов и выходов нейронной сети при этом была решена уже исходя выбранной предметной области: котировки той или иной компании на фондовой бирже выражены в числах. В настоящей работе мы рассматривали пространство данных как метрическое пространство, которое является конечномерным векторным пространством $\mathbb{R}^n$ над полем вещественных чисел с евклидовой метрикой, введённой по формуле:

$$\rho(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \cdots + (x_n - y_n)^2}, \quad (1)$$

где $x = (x_1, x_2, x_3 \ldots x_n) \in \mathbb{R}^n$, $y = (y_1, y_2, y_3 \ldots y_n) \in \mathbb{R}^n$.

Значение компоненты вектора x_i в настоящей работе суть стоимость акций компании в определенный момент времени.

Основным же этапом в предварительной подготовке полученных выборок с торгов стало нормирование данных. Нормировка данных позволяет избавиться от

негативного влияния сильных шумов в выборке. Кроме того, решается проблема повышения информативности всех данных в целом, учет всей области пространства данных.

В нашей работе мы масштабировали данные на единичный отрезок нормированием исходных векторов на их евклидовы нормы:

$$\begin{cases} \hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3 \dots \hat{x}_n) \in \mathbb{R}^n \\ \hat{x}_i = \dfrac{x_i}{\sqrt{(x_1)^2+(x_2)^2+(x_3)^2+\dots+(x_n)^2}} \\ x = (x_1, x_2, x_3 \dots x_n) \in \mathbb{R}^n \end{cases}, \qquad (2)$$

где $\hat{x}$ - нормированный *вектор;* x - *исходный вектор.*

На этапе экспериментального подбора характеристик сети первым рассматривался вопрос подбора количества нейронов в скрытом слое. Следует отметить, что данная проблема тесно связана с задачей начальной инициализации центров нейронов c_i. В совокупности решение этих задач фактически представляют собой отображение реальной области данных нейронной сетью, что является одним из основополагающих факторов для получения адекватного прогноза. Наиболее эффективным с практической точки зрения оказалось случайное размещение радиальных нейронов по всей области данных, описывающиеся следующей формулой:

$$\begin{cases} c_j[i] = V[i] - (V[i] - v[i]) * rand, \ j = \overline{1 \dots m}, \ i = \overline{1 \dots n} \\ V[i] = \max_i V_i[i], \ i = \overline{1 \dots k} \\ v[i] = \min_i V_i[i], \ i = \overline{1 \dots k} \\ V_i \in \mathbb{R}^n \end{cases}. \qquad (3)$$

Проведенные нами тесты, представленные графиком зависимости точности прогноза от количества скрытых нейронов (рис.1), для постоянной обучающей выборки векторов полностью подтвердили теоретические зависимости между числом нейронов в скрытом слое и получаемой точностью прогноза.

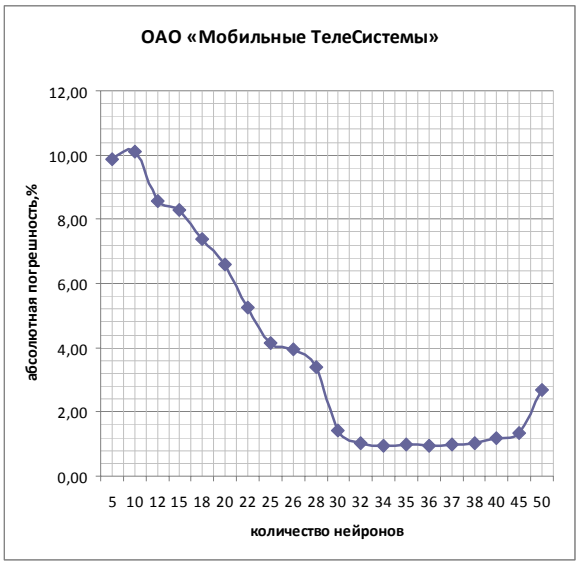


Рисунок 1. - График зависимости точностипрогноза количества нейронов в скрытом слое

Слишком малое число нейронов для конкретной области данных приводило к потере сетью свойств обобщения. Возникало сильное влияние шумов и нерегулярностей данных на результаты прогнозирования. Избыточное количество нейронов скрытого слоя приводило к случайному характеру установки весовых коэффициентов и центров нейронов и, следовательно, к уменьшению точности

прогноза. Практические исследования, представленные показали, что число радиальных нейронов должно быть в несколько раз меньше обучающих векторов.

В нашей работе мы также рассмотрели влияние размеров обучающей выборки на точность результата прогнозирования. Заметим, что, несмотря на существующие общие закономерности, для каждой конкретной задачи количество обучающих векторов должно выбираться на экспериментальной основе. При рассмотрении вопроса о размере обучающей выборки необходимо учитывать то, что меньший размер обучающей выборки означает в общем случае более слабое представление реальной области данных. Однако, как и в случае выбора количества скрытых нейронов, на практике допустимые для удовлетворительных прогнозов размеры обучающей выборки ограничены. При слишком большом количестве обучающих векторов возникает явление переобучения (оверфиттинга) нейронной сети. Причиной возникновения оверфиттинга является выявление нейронной сетью в обучающих данных случайных, локальных закономерностей, отсутствующих в генеральной совокупности. На графике (рисунок 2), отражающем зависимость между точностью прогноза и размером обучающей выборки для различных конфигураций сети, хорошо прослеживаются общие закономерности при переходе от недостаточного количества обучающих данных к избыточному.



Рисунок 1. - График зависимости точности прогноза от размера обучающей выборки

В настоящей работе нами также был проведен анализ применимости радиально-базисной сети для решения задачи прогнозирования в рамках диверсификационного портфеля. В качестве компаний были выбраны ОАО «Газпром», ОАО «Аэрофлот», ОАО «Татнефть» (ММВБ:TATN), ОАО «Мобильные ТелеСистемы»,ОАО «Газпромнефть». Основной задачей здесь являлось получение прогноза стоимости акции компании, входящий в портфель, на основании динамики изменения цен акций остальных компаний, образующих портфель. Иными словами, входной вектор составляли стоимости акций четырех компаний, ожидаемое значение представлялось котировкой пятой компании. Подобный подход позволил проверить гипотезу о существовании определенной коррелированности в поведении различных компаний, характерной для российского фондового рынка. Тестирование проводилось в различных масштабах времени(минута, час, день) для равных отрезков временного диапазона как для периода относительно устойчивого состояния рынка (2007-2009 г.), так и для периода динамики рынка в условиях кризиса (конец 2008-2009 г.).

Исследование проводилось на примере решения задачи подбора оптимального размера обучающей выборки при оптимальном количестве радиальных нейронов. На рисунке 3 представлен график зависимости усредненной по выбранным компаниям погрешности прогнозирования с шагом по времени в один час.

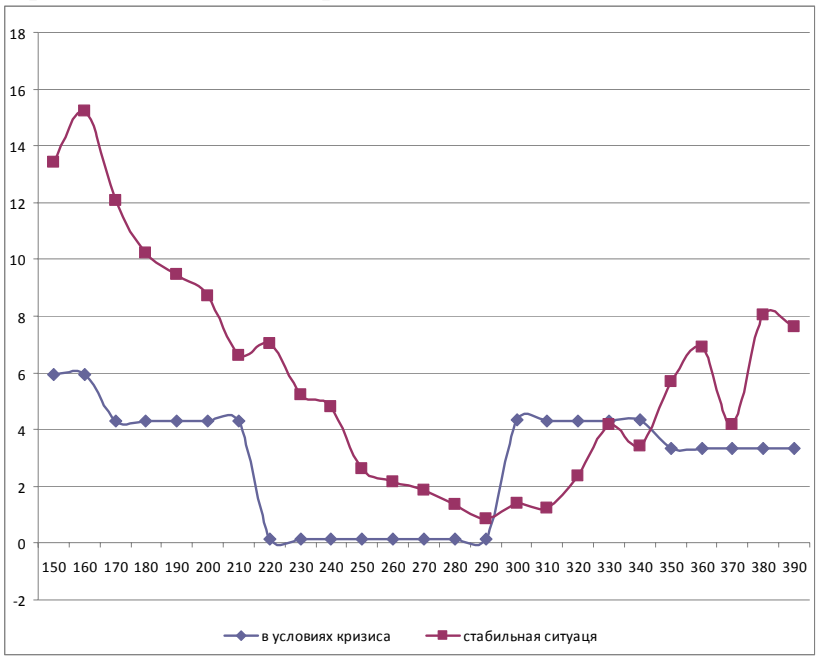


Рисунок 3. Прогнозирование котировки акции по 4 иным компаниям

Возможность подбора оптимальных параметров обучения сети и минимизация погрешности прогноза показали справедливость гипотезы о существовании коррелированности в поведении компаний, представленных на фондом рынке. На основе результатов тестирования также с определенной долей уверенности можно судить и о степени коррелированности компаний в различных экономических условиях. В условиях относительно стабильного рынка акции компаний, принадлежащих различным сегментам рынка, ведут себя схоже, согласно некоторым скрытым закономерностям. В условиях же кризисной ситуации поведение того же набора компаний становится более предсказуемым на основании динамики диверсификационного портфеля, что отображает существенно меньшая погрешность прогнозирования.

Модель движения материальной точки в поле ядерных сил для построения алгоритма глобальной многоэкстремальной оптимизации

А.Н. Коварцев, Д.С. Оплачко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Решение задач глобальной оптимизации является, несомненно, актуальной проблемой. Подобные задачи возникают в самых различных областях человеческой деятельности (экономике, геологии, химии, физике и т.д.). Это могут быть задачи проектирования, задачи распределения ограниченных ресурсов, задачи расчета траектории полета ракеты и т.п.

В работе рассматривается модель движения материальной точки в поле ядерных сил для построения алгоритма глобальной многоэкстремальной оптимизации .

Глобальная оптимизация

Пусть в двумерном пространстве в единичном квадранте $\Omega_x = [0;1]$x$[0;1]$ задана непрерывная функция $f(x_1, x_2)$ двух переменных.

Положим $X = (x_1, x_2)^T$.

В области Ω_x построим регулярную сетку точек $F = \{X_1,X_2,...,X_m\}$.

Точкам X_i присвоим массы $m_1,m_2,...,m_m$ равные значению функции в соответствующей точке $m_i = f(X_i)$.

Будем использовать евклидову метрику для измерения расстояний между точками: $||X_i - X_j|| = \sqrt{(X_i - X_j)^2}$.

Введем потенциал взаимодействия между точками X_i и X_j, при этом положим, что потенциал не зависит от номеров точек i и j: $\varphi(X_i, X_j) = \varphi(||X_i - X_j||)$.

В этом случае $\varphi(r)$ является потенциальной функцией, зависящей только от расстояния между точками.

Энергию взаимодействия между точками определим следующим образом:

$U_{ij} = m_i m_j \varphi(||X_i - X_j||)$.

Тогда суммарная энергия взаимодействия равна: $U = \frac{1}{2}\sum_{i^1 j} m_i m_j f(\left\| X_i - X_j \right\|)$.

Будем считать, что частицы движутся в вязкой среде, так что инерция движения гасится средой, тогда уравнение движения каждой из частиц можно описать уравнением движения первого порядка: $\frac{\partial X_i}{\partial t} = -\frac{\partial U}{\partial X_i}, \quad i = 1,\ldots,m.$ (1)

Предположим, что частицы, перемещаясь в потенциальном поле, изменяют свою массу в соответствии с оптимизируемой функцией $m_i = f(X_i)$, причем положим, что частица, имеющая наибольшую массу является неподвижной, т.е. не изменяет своих координат под воздействием других частиц. Очевидно, что наиболее «подвижными» будут частицы, имеющие большую массу, т.е. находящиеся либо в локальных окрестностях максимума функции, либо рядом с глобальным экстремумом. Более того, многие из них будут двигаться в направление к наиболее весомой, неподвижной точке.

В процессе движений частицы исследуют оптимизируемую функцию, набирая или уменьшая свою массу, передавая «эстафету самой упитанной, неподвижной точке» и собираются в окрестностях глобального экстремума. При этом «худосочные» точки остаются в областях с наименьшими значениями оптимизируемой функции. Все это

вместе и определяет **основную идею предлагаемого алгоритма** глобальной оптимизации.

Выбор модели потенциального поля

Если за основу модели потенциального поля взять ньютоновский потенциал, то возникает опасность сгущения всех частиц вокруг неподвижной точки, при этом могут остаться неисследованными многие участки области оптимизируемых параметров.

Для данной задачи более естественной выглядит потенциал ядерных сил: $\varphi(r) = \frac{-\alpha \exp(-r / r_0)}{r + \varepsilon^2}$, где r_0 – эффективный радиус взаимодействия (на расстояниях больших чем r_0 частицы практически не замечают друг друга), α – определяет мощность поля (при $\alpha < 0$ частицы притягиваются, при $\alpha > 0$ – отталкиваются), ε^2 – регуляризирующая постоянная, необходимая для того, чтобы точки не испытывали слишком сильные перемещения вблизи $r = 0$.

Дискретизация уравнения движения

Уравнение (1) проще всего решать по схеме Эйлера, что дает следующую итерационную формулу: $X_i^{(t+1)} = X_i^{(t)} - \frac{\partial U}{\partial X_i}\Delta t$. (2)

$$\frac{\partial U}{\partial X_k} = \frac{1}{2}\sum_{i \neq j} m_i m_j \frac{\varphi'(\|X_i - X_j\|)}{\|X_i - X_j\|}(\delta_{ik}(X_i - X_k) - \delta_{jk}(X_k - X_j)) = m_k \sum_{i \neq k} m_i \frac{\varphi'(\|X_i - X_k\|)}{\|X_i - X_k\|}(X_i - X_k),$$

причем

$$\varphi'(r) = \left(\frac{-\alpha \exp(-r / r_0)}{r + \varepsilon^2}\right)' = \frac{\frac{\alpha}{r_0}\exp(-r / r_0)\cdot(r + \varepsilon^2) + \alpha e^{-(r/r_0)}}{(r + \varepsilon^2)^2} = \alpha \exp(-r / r_0)\frac{r + \varepsilon^2 + r_0}{r_0 (r + \varepsilon^2)^2}.$$

Таким образом, уравнение движения приобретает вид:

$$X_k^{(t+1)} = X_k^{(t)} - \{m_k \sum_{k \neq i}\left[m_i \alpha e^{-\frac{\|X_i - X_k\|}{r_0}} \frac{\|X_i - X_k\| + \varepsilon^2 + r_0}{r_0 \|X_i - X_k\|(\|X_i - X_k\| + \varepsilon^2)^2}(X_i - X_k)\right]\}\Delta t \qquad (3)$$

Критерий остановки алгоритма глобальной оптимизации

Сходимость процесса уточнения приближенного значения $X_{\max}^{(t)}$ носит во многом случайный характер, так что в процессе работы алгоритма мы наблюдаем последовательное уточнение координат глобального максимума:

$X_{\max}^{(t_1)}, X_{\max}^{(t_2)}, \ldots, X_{\max}^{(t_n)}$. Причем справедливо $\lim_{n \to \infty} X_{\max}^{(t_n)} = X^*$.

Сформируем «окно» размера L и вычислим величину $e = (\frac{1}{L-1}\sum_{k=2}^{L} \|X_{\max}^{k} - X_{\max}^{k-1}\|) / abs(f(X_{\max}^{L})$, которая пропорциональна величине относительной ошибке вычисления глобального экстремума функции.

Тогда в качестве критерия остановки алгоритма можно рассмотреть условие $e \le e_{зад}$, где $e_{зад}$ – заданная относительная точность поиска глобального экстремума оптимизируемой функции.

Заключение

Задача разработки новых методов глобальной оптимизации остается открытой. Для решения задачи глобальной оптимизации не существует универсального по эффективности алгоритма. В настоящей статье рассмотрена модель движения материальной точки в поле ядерных сил для построения алгоритма глобальной многоэкстремальной оптимизации. Метод является точным и гарантирует глобальное приближение к решению задачи.

Метод повышения устойчивости среднеквадратичного оценивания коэффициентов разностного уравнения колебаний систем с турбулентным трением

А.А. Егорова, В.Е. Зотеев
Самарский государственный технический университет, Самара

Рассматривается метод повышения устойчивости среднеквадратичного оценивания коэффициентов разностного уравнения колебаний систем с турбулентным трением.

Установлено, что уменьшение периода дискретизации τ при обработке виброграмм колебаний приводит к ухудшению устойчивости вычисления оценок коэффициентов стохастического разностного уравнения [1]. Для того чтобы выяснить причины подобного явления, были проведены численно-аналитические исследования зависимости коэффициентов корреляций R_{ij} между столбцами матрицы обобщенной регрессионной модели [1] при среднеквадратичном оценивании от периода дискретизации τ. На рис.1 представлен график зависимости парного коэффициента корреляции R_{34} от $\frac{\tau}{T}$ для систем с турбулентным трением при применении итерационного метода.

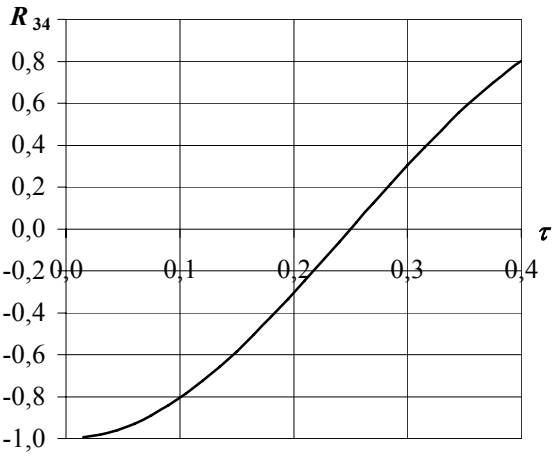


Рис. 1. Зависимость коэффициента корреляции R_{34} от периода дискретизации τ

Видно, что имеет место существенная корреляция между третьим и четвертым столбцами матрицы при отклонении значения τ от оптимальной величины $0{,}25T$. Эта корреляция ухудшает обусловленность матрицы системы нормальных уравнений. «Наилучшая» устойчивость вычислений обеспечивается при «расстоянии» между столбцами матрицы обобщенной регрессионной модели равном $0{,}25T$. Однако для описанной линейно-параметрической дискретной модели обеспечение оптимального «расстояния» между столбцами матрицы означает фиксацию величины периода дискретизации τ, что в свою очередь приводит к ограничению объема выборки N при заданном времени наблюдения динамического процесса. Поэтому повышение устойчивости вычислений при уменьшении периода дискретизации можно обеспечить

только за счет построения модели, в которой «расстояние» между регрессорами изменяется независимо от τ.

Рассмотрим временную последовательность отсчетов y_k в виде:

$\widetilde{y}_0, \widetilde{y}_1, \widetilde{y}_2, ..., \widetilde{y}_{l-1}, \widetilde{y}_l, \widetilde{y}_{l+1}, ..., \widetilde{y}_{2l-1}, \widetilde{y}_{2l}, \widetilde{y}_{2l+1}, ..., \widetilde{y}_k, \ldots$.

В соответствии с построенным в [1] разностным уравнением для любой тройки отсчетов $\widetilde{y}_0, \widetilde{y}_l, \widetilde{y}_{2l}$, отстоящих друг от друга на «расстоянии» $\tau_1 = l\tau$, $l \in N$, можно получить:

$$\lambda_0 \widetilde{y}_l - 2\lambda_1 \widetilde{y}_{2l} + \lambda_0 \lambda_1 \widetilde{y}_l = \widetilde{y}_0 + \widetilde{y}_{2l}, \tag{1}$$

где $\lambda_0 = 2\cos\omega\tau l$, $\lambda_1 = \dfrac{\delta_0 \tau l}{T}$.

При сдвиге уравнения (1) на k отсчетов вправо имеем:

$$\lambda_0 \widetilde{y}_{k+l} - \lambda_1 \left[\frac{k}{l} \widetilde{y}_k + \left(\frac{k}{l} + 2 \right) \widetilde{y}_{k+2l} \right] + \lambda_0 \lambda_1 \left(\frac{k}{l} + 1 \right) \widetilde{y}_{k+l} = \widetilde{y}_k + \widetilde{y}_{k+2l}. \tag{2}$$

Сдвигая последнее соотношение на $2l$ отсчетов влево, получаем модифицированную линейно-параметрическую дискретную модель колебаний системы с турбулентным трением:

$$\lambda_0 \widetilde{y}_{k-l} - \lambda_1 \left[\frac{k}{l} \widetilde{y}_k + \left(\frac{k}{l} - 2 \right) \widetilde{y}_{k-2l} \right] + \lambda_0 \lambda_1 \left(\frac{k}{l} - 1 \right) \widetilde{y}_{k-l} = \widetilde{y}_k + \widetilde{y}_{k-2l}, \tag{3}$$

где $k = 2l,\ 2l+1, 2l+2, ..., N-1$.

С учетом того, что в практике эксперимента результаты наблюдений y_k содержат аддитивную помеху ε_k, модифицированная линейно-параметрическая дискретная модель принимает вид стохастических разностных уравнений:

$$\begin{cases} y_0 = \lambda_2 + \varepsilon_0, \\ y_1 = -\dfrac{y_1}{l}\lambda_1 + \lambda_3 + \left(1 + \dfrac{\lambda_1}{l}\right)\varepsilon_1, \\ y_2 = -\dfrac{2y_2}{l}\lambda_1 + \lambda_4 + \left(1 + \dfrac{2\lambda_1}{l}\right)\varepsilon_2, \\ \ldots \\ y_{2l-1} = -\dfrac{(2l-1)y_{2l-1}}{l}\lambda_1 + \lambda_{2l+1} + \left(1 + \dfrac{(2l-1)\lambda_1}{l}\right)\varepsilon_{2l-1}, \\ y_k + y_{k-2l} = \lambda_0 \left[1 + \lambda_1 \left(\dfrac{k}{l} - 1\right)\right] y_{k-l} - \lambda_1 \left[\dfrac{k}{l} y_k + \left(\dfrac{k}{l} - 2\right) y_{k-2l}\right] + \eta_k, \\ \eta_k = \left[1 + \left(\dfrac{k}{l} - 2\right)\lambda_1\right]\varepsilon_{k-2l} - \lambda_0 \left[1 + \left(\dfrac{k}{l} - 1\right)\lambda_1\right]\varepsilon_{k-l} + \left[1 + \dfrac{k}{l}\lambda_1\right]\varepsilon_k, \end{cases} \tag{4}$$

$k = 2l, 2l+1, ..., N-1;\ N \geq 2l+2,$

где $\lambda_0 = 2\cos\omega\tau l$, $\lambda_1 = \dfrac{\delta_0 \tau l}{T}$, $\lambda_2 = a_0 \cos\psi_0$, $\lambda_3 = a_0 \cos(\omega\tau + \psi_0)$, $\lambda_4 = a_0 \cos(2\omega\tau + \psi_0), \ldots, \lambda_{2l+1} = a_0 \cos((2l-1)\omega\tau + \psi_0)$.

В матричной форме модифицированная линейно-параметрическая дискретная модель, описывающая свободные колебания диссипативной системы с турбулентным трением, имеет вид обобщенной регрессионной модели:

$$\begin{cases} b = F_\lambda \lambda + \eta, \\ \eta = P_\lambda \varepsilon. \end{cases} \tag{5}$$

Здесь $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_{2l+1})^{\mathrm{T}}$ – вектор неизвестных коэффициентов; $\varepsilon = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{N-1})^{\mathrm{T}}$ – N –мерный вектор случайной помехи в результатах наблюдений; $\eta = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2l-1}, \eta_{2l}, \eta_{2l+1}, \ldots, \eta_{N-1})^{\mathrm{T}}$ – N–мерный вектор случайного эквивалентного возмущения; $b = (y_0, y_1, ,\ldots, y_{2l-1}, y_0 + y_{2l}, \ldots, y_{k-1} + y_{k-2l-1}, \ldots, y_{N-2l-1} + y_{N-1})^{\mathrm{T}}$ – N –мерный вектор, где N – объем выборки результатов измерений.

Матрица регрессоров F_λ размера $N \times (2l+2)$ имеет столбцы:

$$f_{k1} = \left(0, 0, \ldots, 0, (1+\lambda_1) y_l, \left[1+\lambda_1\left(1+\frac{1}{l}\right)\right] y_{l+1}, \ldots, \left[1+\lambda_1\left(\frac{k-1}{l}-1\right)\right] y_{k-l-1}, \ldots, \left[1+\lambda_1\left(\frac{N-1}{l}-1\right)\right] y_{N-l-1}\right)^{\mathrm{T}},$$

$$f_{k2} = \left(0, -\frac{y_1}{l}, -\frac{2y_2}{l}, \ldots, -\frac{(2l-1)y_{2l-1}}{l}, -2y_{2l}, -\frac{1}{l} y_1 - \left(2+\frac{1}{l}\right) y_{2l+1}, \ldots, -\left(\frac{N-1}{l}-2\right) y_{N-2l-1} - \frac{N-1}{l} y_{N-1}\right)^{\mathrm{T}},$$

$$f_{k3} = (1, 0, \ldots, 0)^{\mathrm{T}}, \quad f_{k4} = (0, 1, 0, \ldots, 0)^{\mathrm{T}}, \ \ldots, \ f_{k\times(2l+2)} = (0, 0, \ldots, 1, \ldots, 0)^{\mathrm{T}}.$$

Элементы матрицы P_λ размера $N \times N$ в стохастическом разностном уравнении случайного эквивалентного возмущения находятся по следующим формулам: $p_{k+1,k+1} = 1 + \dfrac{k\lambda_1}{l}$, $k = 0, 1, \ldots, 2l-1$; $p_{k+1,k-2l+1} = 1 + \left(\dfrac{k}{l} - 2\right)\lambda_1$, $p_{k+1,k-l+1} = -\lambda_0\left[1 + \left(\dfrac{k}{l} - 1\right)\lambda_1\right]$, $p_{k+1,k+1} = 1 + \dfrac{k}{l}\lambda_1$, $k = 2l, 2l+1, \ldots, N-1$. Остальные элементы $p_{kj} = 0$.

По найденным оценкам $\hat{\lambda}_j$ вычисляются оценки частоты $\hat{\omega}$, начальных декремента $\hat{\delta}_0$, амплитуды $\hat{a}_0$ и фазы колебаний $\hat{\psi}_0$ соответственно по формулам:

$$\hat{\omega} = \frac{1}{\tau l}\arccos\frac{\hat{\lambda}_0}{2}, \quad \hat{\delta}_0 = \frac{2\pi\hat{\lambda}_1}{\arccos\dfrac{\hat{\lambda}_0}{2}}, \quad \hat{a}_0 = \frac{\sqrt{\hat{\lambda}_2^2 - 2\hat{\lambda}_2\hat{\lambda}_3\cos\left(\dfrac{1}{l}\arccos\dfrac{\hat{\lambda}_0}{2}\right) + \hat{\lambda}_3^2}}{\sin\left(\dfrac{1}{l}\arccos\dfrac{\hat{\lambda}_0}{2}\right)},$$

$$\hat{\psi}_0 = \operatorname{arctg}\frac{\cos\left(\dfrac{1}{l}\arccos\dfrac{\hat{\lambda}_0}{2}\right) - \dfrac{\hat{\lambda}_3}{\hat{\lambda}_2}}{\sin\left(\dfrac{1}{l}\arccos\dfrac{\hat{\lambda}_0}{2}\right)}.$$

Таким образом, построена модель, в которой «расстояние» между регрессорами можно варьировать с помощью параметра l. При любом малом значении периода дискретизации τ оптимальная величина $\tau_1 = 0{,}25T$ обеспечивается значением параметра

$$l_{\mathrm{opt}} = \left[\frac{T}{4\tau}\right]. \tag{6}$$

Итерационная процедура среднеквадратичного оценивания коэффициентов обобщенной регрессионной модели (10) описывается формулой

$\hat{\lambda}^{(i+1)} = \left(F^{\mathrm{T}}\Omega^{-1}_{\hat{\lambda}^{(i)}}F_{\hat{\lambda}^{(i)}}\right)^{-1}F^{\mathrm{T}}_{\hat{\lambda}^{(i)}}\Omega^{-1}_{\hat{\lambda}^{(i)}}b$, $i = 0,1,2,\ldots$, где $\Omega_{\hat{\lambda}} = P_{\hat{\lambda}}P^{-1}_{\hat{\lambda}}$. За начальное приближение вектора $\hat{\lambda}^{(0)}$ можно принять оценку $\hat{\lambda}^{(0)} = \left(F^{\mathrm{T}}\Omega^{-1}_{\hat{\lambda}^{(n)}}F\right)^{-1}F^{\mathrm{T}}\Omega^{-1}_{\hat{\lambda}^{(n)}}b$, где матрицы F, $\Omega_{\hat{\lambda}}$ и вектор b соответствуют следующей модели:

$$\begin{cases} y_0 = \lambda_3 + \varepsilon_0, \\ y_1 = \lambda_4 + \varepsilon_1, \\ \ldots \\ y_{2l-1} = \lambda_{2l+2} + \varepsilon_{2l-1}, \\ y_k + y_{k-2l} = \lambda_0 y_{k-l} - \lambda_1\left[\dfrac{k}{l}y_k + \left(\dfrac{k}{l}-2\right)y_{k-2l}\right] + \lambda_2\left(\dfrac{k}{l}-1\right)y_{k-l} + \eta_k, \\ \eta_k = \left[1+\left(\dfrac{k}{l}-2\right)\lambda_1\right]\varepsilon_{k-2l} - \lambda_0\left[1+\left(\dfrac{k}{l}-1\right)\lambda_1\right]\varepsilon_{k-l} + \left[1+\dfrac{k}{l}\lambda_1\right]\varepsilon_k, \end{cases}$$

$k = 2l,\ 2l+1,\ ...,\ N-1;\ N \geq 2l+3,$

где $\lambda_0 = 2\cos\omega\tau l$, $\lambda_1 = \dfrac{\delta_0\tau l}{T}$, $\lambda_2 = \lambda_0\lambda_1$,

$\lambda_3 = a_0\cos\psi_0$,

$$\lambda_4 = \frac{a_0 l}{l+\lambda_1}\cos(\omega\tau+\psi_0),\ldots,\lambda_{2l+2} = \frac{a_0 l}{l+(2l-1)\lambda_1}\cos((2l-1)\omega\tau+\psi_0). \quad (7)$$

В матричной форме данная модель имеет вид (5), где $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_{2l+2})^{\mathrm{T}}$ – вектор неизвестных коэффициентов; $\varepsilon = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{N-1})^{\mathrm{T}}$ – N–мерный вектор случайной помехи в результатах наблюдений; $\eta = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2l-1}, \eta_{2l}, \eta_{2l+1}, \ldots, \eta_{N-1})^{\mathrm{T}}$ – N–мерный вектор случайного эквивалентного возмущения; $b = (y_0, y_1, ,\ldots, y_{2l-1}, y_0 + y_{2l}, \ldots, y_{k-1} + y_{k-2l-1}, \ldots, y_{N-2l-1} + y_{N-1})^{\mathrm{T}}$ – N–мерный вектор, где N – объем выборки результатов измерений.

Матрица регрессоров F размера $N \times (2l+3)$ имеет столбцы:

$$f_{k1} = (0,0,\ldots,0,y_l,y_{l+1},\ldots,y_{k-l-1},\ldots,y_{N-l-1})^{\mathrm{T}},$$

$$f_{k2} = \left(0,0,\ldots,0,-2y_{2l},-\frac{1}{l}y_1-\left(2+\frac{1}{l}\right)y_{2l+1},\ldots,-\left(\frac{N-1}{l}-2\right)y_{N-2l-1}-\frac{N-1}{l}y_{N-1}\right)^{\mathrm{T}},$$

$$f_{k3} = \left(0,0,\ldots,0,y_l,\left(1+\frac{1}{l}\right)y_{l+1},\ldots,\left(\frac{k-1}{l}-1\right)y_{k-l-1},\ldots,\left(\frac{N-1}{l}-1\right)y_{N-l-1}\right)^{\mathrm{T}},$$

$$f_{k4} = (1,0,\ldots,0)^{\mathrm{T}},\quad f_{k5} = (0,1,0,\ldots,0)^{\mathrm{T}},\ \ldots,\ f_{k\times(2l+3)} = (0,\ 0,\ldots,1,\ldots,0)^{\mathrm{T}}.$$

Элементы матрицы P_λ размера $N \times N$ в стохастическом разностном уравнении находятся по следующим формулам: $p_{k+1,k+1} = 1$, $k = 0,1,...,2l-1$;

$p_{k+1,k-2l+1} = 1+\left(\dfrac{k}{l}-2\right)\lambda_1$, $p_{k+1,k-l+1} = -\lambda_0\left[1+\left(\dfrac{k}{l}-1\right)\lambda_1\right]$, $p_{k+1,k+1} = 1+\dfrac{k}{l}\lambda_1$,

$k = 2l, 2l+1, ..., N-1$. Остальные элементы $p_{kj} = 0$.

На рис. 2, 3 представлены зависимости погрешности вычисления частоты и декремента колебаний от случайной аддитивной помехи ($\dfrac{\tau}{T} = 0{,}01$, число усреднений равно 100) при использовании при использовании базового метода [1], а также

модифицированного метода. Видно, что применение модифицированной модели позволяет в несколько раз уменьшить погрешность вычисления декремента колебаний за счет повышения устойчивости среднеквадратичного оценивания. На рис. 3, 4 точки 1 соответствуют результатам вычисления оценок динамических характеристик системы с турбулентным трением на основе базового метода, точки 2 – на основе модифицированного метода.

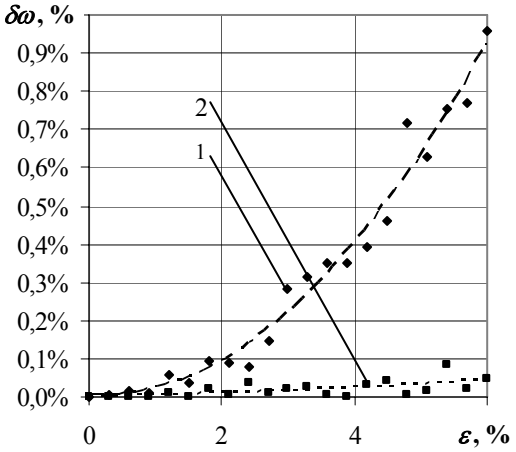


Рис. 2. Зависимость погрешности вычисления частоты колебаний

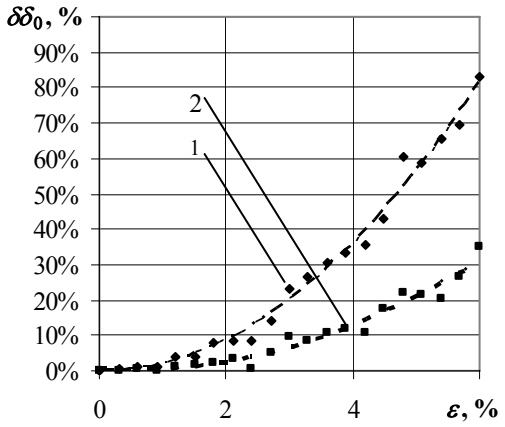


Рис. 3. Зависимость погрешности вычисления декремента колебаний

Таким образом, предложен эффективный способ повышения вычислительной устойчивости среднеквадратичного оценивания коэффициентов линейно-параметрической дискретной модели, который заключается в преобразовании стохастического разностного уравнения, описанного в [1]. Применение модифицированной модели при оптимальном значении параметра l позволяет в несколько раз повысить точность вычисления динамических характеристик систем с турбулентным трением по результатам наблюдений мгновенных значений ее свободных колебаний.

Литература

1. *Зотеев В. Е.* Параметрическая идентификация диссипативных механических систем на основе разностных уравнений. – М.: Машиностроение, 2009. – 344 с.

Принцип сведения в задаче об устойчивости при постоянно действующих возмущениях

Д.М. Щепакин
Самарский государственный университет, Самара

В работе рассматриваются сингулярно возмущенные системы с постоянно действующими возмущениями. Для их исследования применяется метод геометрической декомпозиции, в основе которого лежит теория интегральных многообразий быстрых и медленных движений. Показано, что для таких систем справедлив принцип сведения, суть которого состоит в том, что рассматриваемая система устойчива по отношению к постоянно действующим возмущениям тогда и только тогда, когда таким свойством обладает подсистема, описывающая движение по медленному интегральному многообразию.

Сингулярно возмущенные системы

Рассмотрим систему

$$\begin{cases} \dfrac{dx_1}{dt} = f_1(t, x_1, x_2, \varepsilon), \\ \varepsilon \dfrac{dx_2}{dt} = f_2(t, x_1, x_2, \varepsilon), \end{cases} \tag{1}$$

$f_i(t,0,0,\varepsilon) = 0$, $(i = 1,2)$,

ε - малый параметр. Предполагается, что для системы (1) выполняются следующие условия.

(i) Уравнение $f_2(t, x_1, x_2, 0) = 0$ имеет изолированное решение $x_2 = h_0(t, x_1)$ при $t \in R,\ x \in R$.

(ii) В области $\Omega_0 = \{(t, x_1, x_2, \varepsilon) \mid x \in R^m, \| x_2 - h_0(t, x_1) \| < \rho,\ t \in R, \varepsilon \in [0; \varepsilon_0]\}$ векторные функции f, g, h_0 равномерно непрерывны и ограничены вместе с частными производными по всем переменным до $(k+2)$-го порядка включительно $(k \geq 0)$.

(iii) Собственные значения $\lambda_i(t,x)(i = 1,...,n)$ матрицы

$B(t, x_1) = \frac{\partial f_2}{\partial x_2}(t, x_1, h_0(t, x_1), 0)$

подчиняются неравенству

$\operatorname{Re} \lambda_i(t,x) \leq -2\gamma < 0.$

Из условия (ii) следует, что в области Ω функции f_1, f_2 удовлетворяют условию Липшица по переменной x_2 с некоторой константой $N > 0$.

При выполнении условий (i)–(iii) возможно провести следующую замену переменных [1-3]:

$$\begin{cases} x_1 = u + \varepsilon H(t, u, v, \varepsilon), \\ x_2 = v + h(t, x_1, \varepsilon) = v + h(t, u + \varepsilon H(t, u, v, \varepsilon), \varepsilon), \end{cases} \tag{2}$$

где $x_2 = h(t, x_1, \varepsilon)$ — уравнение медленного многообразия, $y = \varepsilon H(t, u, z)$ — уравнение быстрого многообразия, переменная $u(t)$ описывает движение по медленному многообразию, $y = x_1 - u$, а $z = x_2 - h(t, x_1)$.

При этом функция $H(t, u, v, \varepsilon)$ удовлетворяет неравенствам:

$| H(t, u, v, \varepsilon) | \leq a | v |,$

$| H(t, u, v, \varepsilon) - H(t, u, \tilde{v}, \varepsilon) | \leq b | v - \tilde{v} |,$

$$|H(t,u,v,\varepsilon)-H(t,\tilde{u},v,\varepsilon)|\le b|v||u-\tilde{u}|.$$

Заметим также, что $h(t,0,\varepsilon)=0$. В результате получим:

$$\begin{cases} \dfrac{du}{dt}=F(t,u,\varepsilon), \\ \varepsilon\dfrac{dv}{dt}=B(t,u)v+G(t,u,v,\varepsilon), \end{cases} \tag{3}$$

где

$$F(t,u,\varepsilon)=f_1(t,u,h(t,u,\varepsilon),\varepsilon)G=Z(t,u,H(t,u,v,\varepsilon),v,\varepsilon),$$

$$Z(t,u,y,v,\varepsilon)=f_2(t,u+\varepsilon y,v+h(t,u+\varepsilon y,\varepsilon),\varepsilon)-\frac{\partial f_2}{\partial x_2}(t,u,h_0,0)-$$

$$-\varepsilon\frac{\partial h}{\partial t}-\varepsilon\frac{\partial h}{\partial x_1}f_1(t,u+\varepsilon y,v+h(t,u+\varepsilon y,\varepsilon),\varepsilon),$$

$$B(t,u)=\frac{\partial f_2}{\partial x_2}(t,u,h_0,0).$$

Для функции G справедливо неравенство:

$$|G(t,u,v,\varepsilon)|=|Z(t,u,H(t,u,v,\varepsilon),v,\varepsilon)|\le A|v|^2.$$

Наряду с системами (1) и (3) рассмотрим следующие системы:

$$\begin{cases} \dfrac{dx_1}{dt}=f_1(t,x_1,x_2,\varepsilon)+g_1(t,x_1,x_2,\varepsilon), \\ \varepsilon\dfrac{dx_2}{dt}=f_2(t,x_1,x_2,\varepsilon)+g_2(t,x_1,x_2,\varepsilon), \end{cases}$$

и

$$\begin{cases} \dfrac{d\tilde{u}}{dt}=F(t,\tilde{u},\varepsilon)+g_3(t,\tilde{u},v,\varepsilon), \\ \varepsilon\dfrac{dv}{dt}=B(t,\tilde{u})v+G(t,\tilde{u},v,\varepsilon)+g_4(t,\tilde{u},v,\varepsilon), \end{cases}$$

где функции g_3, g_4 имеют вид

$$g_3=g_1(t,u+\varepsilon H(t,u,v,\varepsilon),v+h(t,u+\varepsilon H(t,u,v,\varepsilon),\varepsilon),\varepsilon)-$$

$$-\frac{\partial H}{\partial v}[g_2(t,u+\varepsilon H(t,u,v,\varepsilon),v+h(t,u+\varepsilon H(t,u,v,\varepsilon),\varepsilon),\varepsilon)-$$

$$-\varepsilon\frac{\partial h}{\partial x_1}g_1(t,u+\varepsilon H(t,u,v,\varepsilon),v+h(t,u+\varepsilon H(t,u,v,\varepsilon),\varepsilon),\varepsilon)],$$

$$g_4=g_2(t,u+\varepsilon H(t,u,v,\varepsilon),v+h(t,u+\varepsilon H(t,u,v,\varepsilon),\varepsilon),\varepsilon)-$$

$$-\varepsilon\frac{\partial h}{\partial x_1}g_1(t,u+\varepsilon H(t,u,v,\varepsilon),v+h(t,u+\varepsilon H(t,u,v,\varepsilon),\varepsilon),\varepsilon).$$

Вспомогательные предложения

Рассмотрим систему

$$\frac{du}{dt}=F(t,u,\varepsilon). \tag{4}$$

и систему

$$\frac{du}{dt}=F(t,u,\varepsilon)+g_5(t,u), \tag{5}$$

где $g_5(t,u)$ — достаточно малые возмущения. Допустим, что решение $u=0$ системы (4) устойчиво при постоянно действующих возмущениях.

Для второй подсистемы системы (3) построим функцию Ляпунова. Так как мы можем найти функцию $u(t)$ из её первой подсистемы, т.е. из уравнения (4), то можно считать, что матрица $B(t,u)=B(t)$. Для уравнения $\frac{dv}{dt}=\frac{1}{\varepsilon}B(t)v$ существует функция Ляпунова в виде положительно определенной квадратичной формы с коэффициентами, зависящими от времени [4, 5]: $V(t,v)=(V(t)v,v)$, где V(t) — матрица этой квадратичной формы. При вычислении полной производной функции $V(t,v)$ по времени в силу этой системы мы опять получаем квадратичную форму с матрицей $-K(t)$, где $K(t)$ — положительно определенная матрица. Покажем, что эта функция Ляпунова подходит для доказательства устойчивости нулевого решения при постоянно действующих возмущениях для второй подсистемы системы (3).

Частные производные квадратичной формы ограничены в ограниченной области. При вычислении полной производной этой функции по времени в силу второй подсистемы системы (3) мы получаем

$$\frac{dV}{dt}(t,v)=v^T[\frac{\partial V}{\partial t}(t)+(V(t)B(t)+B^T(t)V(t))\frac{1}{\varepsilon}]v+$$
$$+G^T(t,u(t),v,\varepsilon)V(t)v+v^TV(t)G(t,u(t),v,\varepsilon)=$$
$$=v^T\frac{\partial V}{\partial t}(t)v+\frac{1}{\varepsilon}[v^T(V(t)B(t)+B^T(t)V(t))v+G^T(t,u(t),v,\varepsilon)V(t)v+v^TV(t)G(t,u(t),v,\varepsilon)].$$

Так как $|G(t,u,v,\varepsilon)|\le A|v|^2$, то

$$|G^T(t,u(t),v,\varepsilon)V(t)v+v^TV(t)G(t,u(t),v,\varepsilon)|\le 2A|v|^3\|V\|.$$

Если мы теперь в качестве $K(t)$ возьмем единичную матрицу, умноженную на $2A\|V(t)\|$, то $|v^T(V(t)B(t)|=2\|V\|A|v|^2$, и тогда $\frac{dV}{dt}$ — отрицательно определенная функция.

Таким образом, по достаточному признаку [5] условие устойчивости при постоянно действующих возмущениях выполняется для решения $v=0$ второй подсистемы системы (3). В силу произвольности $g_5(t,u)$ из системы (5) можно утверждать следующее.

Предложение. В предположении устойчивости при постоянно действующих возмущениях нулевого решения системы (4) нулевое решение системы (3) устойчиво при постоянно действующих возмущениях.

Принцип сведения при постоянно действующих возмущениях

Если ограничить $|x_1(t_0)|$ и $|x_2(t_0)|$ так, что $|x_1(t_0)|$ и $|x_2(t_0)|$ будут меньше δ_1, а функции g_1, g_2 ограничить в малой окрестности нуля так, что из этого будет следовать ограниченность функций g_3, g_4 в малой окрестности нуля числом δ_2, то это будет означать устойчивость при постоянно действующих возмущениях нулевого решения системы (1).

Введем следующее обозначение:

$$\xi=\min\left\{\frac{\delta_1}{\varepsilon a},\delta_1\right\}.$$

Найдем такое $\sigma>0$, что из условия $|\theta|\le\sigma$ следовало бы, что $|h(t,\theta,\varepsilon)|\le\frac{\xi}{3}$ для любого $\varepsilon\in(0;\varepsilon_0]$.

Если $|x_1(t_0)| \le \min\left\{\sigma, \frac{\delta_1}{3}\right\}$, $|x_2(t_0)| \le \frac{\xi}{3}$, то $|u(t_0)| \le \delta_1$, $|v(t_0)| \le \delta_1$.

Пусть $\bar{\xi} = \min\left\{\frac{\varepsilon}{a\varepsilon}, \varepsilon\right\}$.

Найдем такое $\sigma > 0$, что из условия $|\theta| \le \sigma$ следовало бы, что $|h(\theta, t, \varepsilon)| \le \frac{\bar{\xi}}{3}$ для любого $\varepsilon \in (0; \varepsilon_0]$. Пусть $\bar{\varepsilon} = \min\left\{\bar{\sigma}, \frac{\varepsilon}{3}, \frac{\bar{\xi}}{3}\right\}$.

Если мы теперь по числу $\bar{\varepsilon}$ выберем число δ_2, то аналогичными рассуждениями получим, что из условий

$|u + \varepsilon H(t, u, v, \varepsilon)| \le \bar{\varepsilon}$,

$|v + h(t, u + \varepsilon H(t, u, v, \varepsilon), \varepsilon)| \le \bar{\varepsilon}$,

следует, что

$|g_i(t, x_1, x_2, \varepsilon)| \le \delta_2$, $(i = 1, 2)$,

если

$t > t_0$, $|x_1| \le \varepsilon$, $|x_2| \le \varepsilon$, $\varepsilon \in (0; \varepsilon_0]$.

Для функций $g_3(t, u, v, \varepsilon)$ и $g_4(t, u, v, \varepsilon)$ такую же оценку можно получить, выбрав δ_2 не по $\bar{\varepsilon}$, а по $\min\left\{\frac{\bar{\varepsilon}}{3}, \frac{\bar{\varepsilon}}{3c^2}\right\}$, где

$$c = \max\left\{\varepsilon\left|\frac{\partial h}{\partial x_1}(t, u, v, \varepsilon)\right|, \left|\frac{\partial H}{\partial v}(t, u, v, \varepsilon)\right|\right\}, \quad (|u| \le \varepsilon, |v| \le \varepsilon).$$

Из неустойчивости нулевого решения при постоянно действующих возмущениях первой подсистемы системы (3) следует аналогичная неустойчивость нулевого решения системы (1) в силу того, что первая подсистема системы (3) описывает поведение решений полной системы (1) по интегральному многообразию $x_2 = h(t, x_1, \varepsilon)$.

Таким образом, доказана теорема.

Теорема (Принцип сведения). Нулевое решение системы (1) устойчиво при постоянно действующих возмущениях тогда и только тогда, когда устойчиво при постоянно действующих возмущениях нулевое решение системы (4).

Результаты работы могут быть применены для решения задач устойчивости по отношению к постоянно действующим возмущениям для моделей летательных аппаратов и навигационных приборов.

Автор выражает глубокую благодарность своему научному руководителю профессору В.А. Соболеву за постановку задачи и полезные обсуждения.

Литература

1. Sobolev V.A. Integral manifolds and decomposition of singularly perturbed system /V.A. Sobolev //System and Control Lett.-1984.-No. 5.-P. 169-179.

2. Sobolev V.A. Decomposition of control systems with singular perturbations system /V.A. Sobolev // Proc. 10th Congr. IFAC. Munich-1987.-Vol. 8.-P. 172-176.

3. Воропаева Н.В. Геометрическая декомпозиция сингулярно возмущенных систем / Н.В. Воропаева, В.А. Соболев – М,: ФИЗМАТЛИТ, 2009. - 256 с.

4. Боровских А.В. Лекции по обыкновенным дифференциальным уравнениям / А.В. Боровских, А.И. Перов – Москва-Ижевск: НИЦ «Регулярная и хаотическая динамика», Институт компьютерных исследований, 2004. – 504 с.

5. Малкин И.Г. Теория устойчивости движения / И.Г. Малкин - М: Наука, 1966. - 530 с.

Исследование устойчивости динамической системы с задержкой в модели рынка одного товара с учетом спекулятивной активности

И.П. Завершинский, М.Р. Майлян
Самарский государственный аэрокосмический университет имени академика С.П.Королева (национальный исследовательский университет), Самара

Введение

В настоящей работе излагаются некоторые результаты изучения устойчивости положения равновесия в модели рынка одного товара и множества участников. В отличие от классической модели в работе рассмотрена модель рынка одного товара, обладающего не только потребительской, но и спекулятивной ценностью. Возможность присутствия у некоторых из участников спекулятивных мотивов при решении вопроса о степени индивидуальной заинтересованности в сбыте или приобретении данного товара может привести к потере устойчивости рыночного равновесия.

В такой ситуации представляет интерес исследование параметров, при которых на рынке устанавливается устойчивое положение равновесия. Для решения этой задачи использовалась классическая модель установления рыночного равновесия, гипотеза о спекулятивном поведении участников рынка, описанная Разжевайкиным В.Н. [1], а также модель формирования предложения и спроса созданная А. А. Шананиным [2]. Полученная модель, предполагающая ограниченную устойчивость рынка, была модифицирована в непрерывную по времени.

Классическая модель рыночного равновесия

Классическая модель рыночного равновесия (1) описывает рынок однородного товара. В соответствии с традиционными представлениями теории равновесия [3,4] считаются выполненными следующие предположения: в каждый момент времени товар продается по единой цене p; поведение потребителей и производителей описывается функциями спроса $C(p)$ и предложения $Y(p)$; характерное время изменения функции спроса и предложения много больше характерного времени изменения цены.

$$\frac{dp}{dt} = g\frac{C(p)-Y(p)}{Y(p)}p, \qquad (1)$$

где $g > 0$ – параметр, имеющий размерность 1/*время* и характеризующий скорость реакции цены на различие спроса и предложения.

Модель рыночного равновесия с учетом спекулятивной активности

Рассмотрим рынок одного товара, способного обладать как потребительской, так и спекулятивной ценностью. Согласно [1] возможность использования товара в спекулятивных целях отражается в паутинообразной модели ценообразования добавлением в правую часть уравнения слагаемого вида (здесь и далее $\partial_u f(u) = \frac{df(u)}{du}$) :

$$\alpha\beta D(P)\int_0^{\infty}\frac{\partial_t P(t-\tau)}{P(t-\tau)}e^{-\tau\beta}d\tau, \qquad (2)$$

где $\alpha \geq 0$ - показатель спекулятивной активности, $\beta > 0$ - коэффициент "дисконтирования" информации, $D(P)$ - функция избыточного спроса.

Схожие постановки можно найти, например, в [5].

Содержательный смысл подынтегрального выражения, отнесенного к параметру $\tau > 0$, заключается в относительной скорости изменения цены товара в τ-давний момент

с учетом индивидуального коэффициента "дисконтирования" $\beta > 0$ информации, являющейся достоянием отдаленного прошлого (перед интегралом этот коэффициент стоит из соображений нормировки).

В случае когда относительная скорость изменения цены в условиях отсутствия спекулятивных мотивов описывается моделью ценообразования Вальраса, добавление в модель членов, учитывающих спекулятивные мотивы, приводит к модели динамики цены, описывающейся одним интегродифференциальным уравнением вида

$$\partial_t P(t) = \alpha\beta P(t) \int_0^\infty \frac{\partial_t P(t-\tau)}{P(t-\tau)} e^{-\tau\beta} d\tau + g \frac{C(P(t)) - Y(P(t))}{Y(P(t))} P(t)^2, \quad (3)$$

где $C(P(t))$ - функция зависимости спроса от цены, $Y(P(t))$ - функция зависимости предложения от цены.

Для конкретизации модели определим вид функций спроса и предложения. В рамках непрерывной модели функция предложения будет выглядеть так

$$Y(P(t)) = M(t_0)\left[1 - \left(\frac{SV}{P(t-\tau_0)}\right)^{\frac{\gamma+\mu}{\mu}}\right], \quad (4)$$

где $\tau_0 > 0$ – время задержки предложения.

Запаздывание в реакции производителя на изменение цены отражает реально происходящие на рынке процессы и допускает экономическую интерпретацию – например, как характерная продолжительность производственного цикла [6].

Потребительский спрос описываем модельной функцией

$$C(P(t)) = C_0 \left(\frac{P(t)}{SV}\right)^{\rho-1}, \quad (5)$$

где C_0 - характеристика спроса, ρ - степень необходимости товара [2].

В переменных

$$u(t) = \ln\left(\frac{P}{P_s}\right), \quad v(t) = \beta \int_0^\infty u(t-\tau)\, e^{-\tau\beta} d\tau$$

уравнение (3) эквивалентно следующей системе обыкновенных дифференциальных уравнений:

$$\begin{cases} \partial_t u = \alpha\beta(u - v) + g \dfrac{C_0 - M(\tau_0)\left[1 - \left(\frac{SV}{P_s e^u}\right)^{\frac{\gamma+\mu}{\mu}}\right]}{M(t_0)\left[1 - \left(\frac{SV}{P_s e^u}\right)^{\frac{\gamma+\mu}{\mu}}\right]} P_s e^u, \\ \partial_t v = \beta(u - v), \end{cases} \quad (6)$$

с тривиальным положением равновесия, соответствующим рыночному равновесию P_s уравнения (3).

С системой (6) был проведен *аналитический анализ* и получены следующие результаты. При $\alpha = 0$ положение равновесия P_s уравнения (3) является глобально

асимптотически устойчивым. При $\rho < 1$ равновесная цена P_s существует всегда, а при $\rho=1$ необходимым и достаточным условием существования рыночной цены является выполнение неравенства $M(t_0) > C_0$. При $\alpha \neq 0$ на динамику цены оказывает влияние первый член правой части уравнения (3). При увеличении α с прохождением через пороговое значение $\alpha_0 = 1 + \delta P_s/\beta$ корни пересекают мнимую ось, переходя в правую полуплоскость. Для определения характера возникающей бифуркации была вычислена первая ляпуновская величина,которрая, как показали расчеты, в окрестности критической точки, то есть при $|\alpha - \alpha_0| \ll \alpha_0$, принимает отрицательные значения, а следовательно будет наблюдаться суперкритическая бифуркация Андронова-Хопфа. Численное решение системы уравнений (6) подтверждает полученные результаты.

Модель рыночного равновесия с учетом спекулятивной активности нескольких участников рынка

Переход к модели рынка одного товара и нескольких участников был осуществлен следующим образом: динамика каждого участника рынка описывается полученной выше моделью, а взаимодействие осуществляется через слагаемые, учитывающие адаптацию производителей к рынку. Таким образом, задаваясь индивидуальными коэффициентами адаптации $\gamma_{ij} > 0$, $j = i \pm 1$, можно записать динамическую модель рынка в виде

$$\partial_t P_i(t) = \alpha_i \beta P_i(t) \int_0^\infty \frac{\partial_t P_i(t-\tau)}{P_i(t-\tau)} e^{-\tau\beta} d\tau - \gamma_{ij} P_i(t) P_j(t) + g \frac{C(P_i(t)) - Y(P_i(t))}{Y(P_i(t))} P_i(t)^2 . \quad (7)$$

Вычислительные эксперименты

С моделью ценообразования (7) производились численные эксперименты с целью выявления зависимости устойчивости рыночных механизмов от степени корреляции между производителями. При проведении вычислительных экспериментов варьировались параметр γ_{ij} и число производителей $N \in [5,50]$, при этом критическое значение параметра r = 50% (r – процент участников рынка, чьи показатели спекулятивной активности первышают критическое значение α_0) не менялось, что говорит о его универсальности.

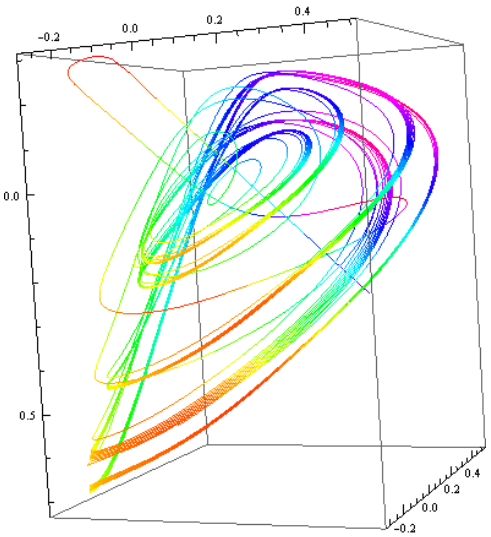
При r < 50% пока значения коэффициентов адаптации не слишком велики, рыночные механизмы регулируют цены товара, так что они сходятся к равновесным, балансирующим спрос и предложение. Если γ_{ij} превысит критическое значение, цены на рынке начинают колебаться, а затем в системе возбуждаются автоколебания.

Если r ~ 60%, то при малых значениях коэффициентов адаптации автоколебания цен сохраняются, но с ростом их величины происходит хаотизация колебаний.

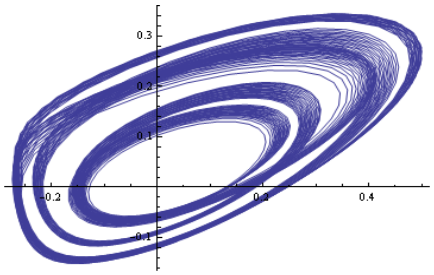
Если r ~ 85%, то в этом случае все участники, у которых показатели спекулятивной активности не превышают критического значения теряют устойчивость.

Начиная с γ_{ij} = 0.7232 в системе в результате бифуркации удвоения периода пятиоборотного цикла возникает хаотический аттрактор, имеющий слоистую структуру (рисунок 1). Слоистая структура пропадает при γ_{ij} = 0.733.

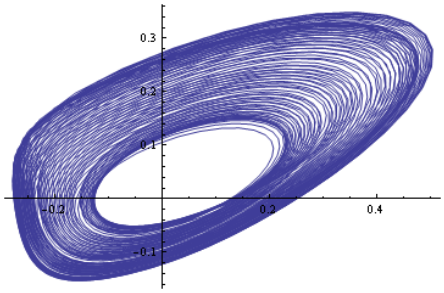
На интервале (4.45,4.63) также реализуется бесконечный каскад бифуркаций удвоения периода. На рисунке 2 приведен такой предельный цикл и бифуркации удвоения его периода. Этот бесконечный каскад бифуркаций приводит при γ_{ij} = 4.65 к рождению хаотического аттрактора.



γ_{ij} = 0.7232

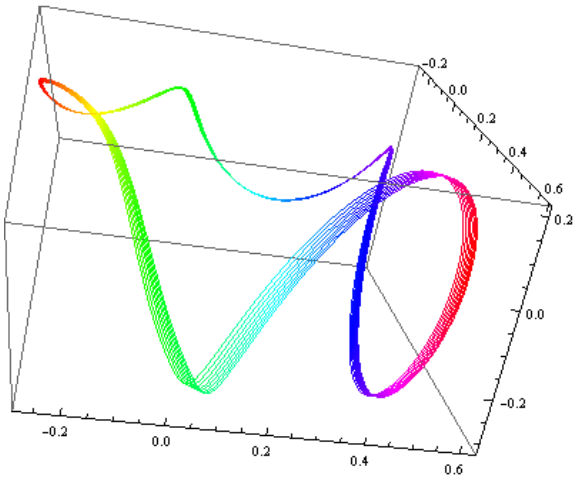


γ_{ij} = 0.7245

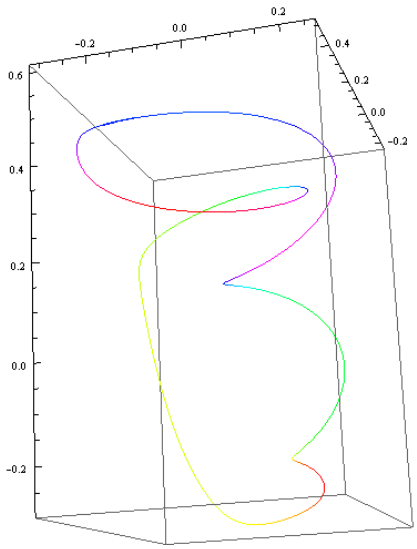


γ_{ij} = 0.733

Рисунок 1.



γ_{ij} = 4.6



γ_{ij} = 4.452

Рисунок 2.

Вблизи значения параметра γ_{ij} = 2 наблюдается следующая картина: с течением времени чередуются стадии регулярного и хаотического поведения.

Заключение

В работе разработана модель ценообразования для исследования устойчивости рыночных механизмов на рынке одного товара для одного и нескольких участников. Полученная модель учитывает наличие у участников рынка спекулятивных мотивов.

Для упрощенной модели динамики цены проведено исследование устойчивости рыночного равновесия аналитическим способом, и полученные результаты подтверждены численным решением динамической системы. Наличие спекулятивных мотивов является дестабилизирующим фактором для рыночного равновесия, причем потеря устойчивости происходит в колебательном режиме.

С моделью рыночного равновесия с учетом спекулятивной активности нескольких участников рынка проведены вычислительные эксперименты и получены следующие результаты:

1. Если величина спекулятивной активности более 50% участников рынка превышает критическое, то при малых значениях коэффициента адаптации «соседние» участники рынка колебательно теряют устойчивость, а при дальнейшем увеличении бифуркационного параметра происходит хаотизация колебаний цен.
2. Система взаимодействующих производителей, с математической точки зрения, описывается уравнениями связанных осцилляторов, причем за связь отвечают слагаемые учитывающие адаптацию производителей к ценам на рынке.
3. В полученной системе реализуется сценарий перехода к хаосу через каскад бифуркаций удвоения периода.

Литература

1. Разжевайкин В.Н. Методы устойчивости в модели рынка при наличии у его участников спекулятивных мотивов// Иссл.операций. М.: ВЦ РАН, 1999. С.46-60.

2. Шананин А.А. Об устойчивости рыночных механизмов //Математическое моделирование. 1991. Т. 3, № 2. С. 42 - 62.

3. Мэнкью Н.Г. Принципы экономикс. М.: Питер,2009. 762 с.

4. Вальрас Л. Элементы чистой политической экономии. М.: Изограф, 2000. 448 с.

5. Прасолов А.В. Математические методы экономической динамики: Учебное пособие. СПб.: Издательство «Лань», 2008. 352 с.

6. Обросова Н.К. Потеря устойчивости равновесной цены в модели ценообразования вальрасовского типа // Математическое моделирование. 1998. Т.10, №5. С.47 – 57.

Распространение звука в неоднородной тепловыделяющей среде

Н.Е. Молевич[1,2], Н.В. Трошкин[1]
[1]*Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*
[2]*Самарский филиал Учреждения Российской академии наук Физического института им. П.Н. Лебедева РАН, Самара*

Стационарная неравновесная среда является неоднородной, причем ее неоднородность в основном определяется способом организации теплоотвода. Влияние поперечной неоднородности колебательно-возбужденного газа, формируемой теплопроводностью, на условия акустической устойчивости исследовалось в [1,2]. Показано, что при малой степени неравновесности область акустической неустойчивости в колебательно возбужденном газе (на плоскости мощности накачки – обратное время релаксации) расширяется, а при большей степени неравновесности – сужается по сравнению с однородным случаем.

Влияние продольной неоднородности колебательно – возбужденного газа, вызванной продольным конвективным теплоотводом, на инкременты нарастания газодинамических возмущений ранее не рассматривалось. Некоторые аспекты этой проблемы обсуждались в [3], где экспериментально была обнаружена конвективная неустойчивость быстропроточных лазеров с конвективным теплоотводом по увеличению структурной характеристики пульсаций показателя преломления активной среды, и, к примеру, в [4,5], где показано, что продольная неоднородность магнито – газодинамического течения плазмы приводит к его конвективной неустойчивости.

Исходная система одномерных уравнений газодинамики, описывающая распространение возмущений в потоке тепловыделяющего газа, имеет вид

$$\begin{aligned} \frac{d\rho}{dt} + \rho\frac{\partial \mathrm{v}}{\partial x} &= 0, \\ \rho\frac{d\mathrm{v}}{dt} &= -\frac{\partial P}{\partial x}, \\ P &= \rho T / m, \\ C_{V\infty}\frac{dT}{dt} - \frac{T}{\rho}\frac{d\rho}{dt} &= Q(T,\rho), \end{aligned} \qquad (1)$$

где $d/dt = \partial/\partial t + \mathrm{v}\partial/\partial x$; ось x выбрана в направлении скорости U потока, сдвиговой вязкостью и теплопроводностью пренебрежено.

В исходную систему подставим разложения функций на стационарную часть (профили характеристических величин потока) и нестационарную часть (акустические возмущения): $\mathrm{v} = U(x) + \tilde{\mathrm{v}}(x,t), T = T_0(x) + \tilde{T}(x,t)$ и так далее.

Градиенты стационарных величин сильно связаны друг с другом:

$$\begin{aligned} \frac{1}{\rho}\frac{d\rho}{dx} &= -\frac{1}{(1-\gamma_\infty M_\infty^2)}\frac{1}{T}\frac{dT}{dx}, \\ \frac{1}{T}\frac{dT}{dx} &= \frac{S(1-\gamma_\infty M_\infty^2)}{M_\infty c_\infty C_{P\infty}(1-M_\infty^2)\tau}, \\ \frac{1}{U}\frac{dU}{dx} &= -\frac{1}{\rho}\frac{d\rho}{dx}, \end{aligned} \qquad (2)$$

где $M_\infty = \dfrac{U}{c_\infty}$ - число Маха в высокочастотном пределе.

Эволюцию акустического возмущения частоты ω исследуем в приближении слабой неоднородности среды (геометрической акустики), для применимости которой должно выполняться условие на градиенты: $\frac{1}{T_0}\frac{dT_0}{dx} << \frac{\omega}{u_\infty(1+M_\infty)}$.

Это условие выполнимо только в высокочастотном спектральном диапазоне

$$\omega\tau_Q >> \left|\frac{1-\gamma_\infty M_\infty^2}{(1-M_\infty)M_\infty C_{P\infty}}\right|. \qquad (3)$$

Система после линеаризации и пренебрежении малыми слагаемыми сводится к замкнутым уравнениям второго порядка для каждого из параметров, характеризующих поток. Для скорости уравнение записывается:

$$\frac{\partial^2\widetilde{\mathrm{v}}}{\partial t^2}-(c_\infty^2-U^2)\frac{\partial^2\widetilde{\mathrm{v}}}{\partial x^2}+2U\frac{\partial^2\widetilde{\mathrm{v}}}{\partial x\partial t}=-2\alpha_\infty c_\infty\left(\frac{\partial\widetilde{\mathrm{v}}}{\partial t}+U\frac{\partial\widetilde{\mathrm{v}}}{\partial x}\right)-$$
$$-\frac{M_\infty c_\infty}{(1-\gamma_\infty M_\infty^2)}\left[(1+2\gamma_\infty+M_\infty^2)M_\infty c_\infty\frac{\partial\widetilde{\mathrm{v}}}{\partial x}+(\gamma_\infty+M_\infty^2)\frac{\partial\widetilde{\mathrm{v}}}{\partial t}\right]\frac{1}{T_0}\frac{dT_0}{dx}. \qquad (4)$$

Из дисперсионного соотношения для (4) получается мнимая часть волнового вектора равной:

$$G_{\mathrm{v}\infty}=\frac{\alpha_{\infty Q}^{Q\infty}}{1+M_\infty}-\frac{M_\infty(M_\infty-\gamma_\infty)(1-M_\infty)}{2(1-\gamma_\infty M_\infty^2)(1+M_\infty)}\cdot\frac{1}{T_0}\frac{dT_0}{dx}+\frac{1}{2k}\frac{dk}{dx}(1-M_\infty). \qquad (5)$$

Это и есть высокочастотный инкремент нарастания возмущений скорости в слабонеоднородной среде.

Третье слагаемое в (5) выражается через градиент температуры:

$$\frac{1}{k}\frac{dk}{dx}=-\frac{1}{(1+M_\infty)}\left(\frac{M_\infty}{U}\frac{dU}{dx}+\frac{1}{2T_0}\frac{dT_0}{dx}\right)=-\frac{(1+2M_\infty-\gamma_\infty M_\infty^2)}{2T_0(1+M_\infty)(1-\gamma_\infty M_\infty^2)}\frac{dT_0}{dx}, \qquad (6)$$

И окончательно, декремент возмущений скорости (7):

$$G_{\mathrm{v}\infty}=\frac{\alpha_{\infty Q}^{Q\infty}}{1+M_\infty}+G_{\mathrm{v}\infty}^{\mathrm{i}}. \qquad (7)$$

Второе слагаемое в (7) - поправка к инкременту, обусловленная продольной неоднородностью стационарной тепловыделяющей среды:

$$G_{\mathrm{v}\infty}^{\mathrm{i}}=\frac{-[1-2M_\infty(\gamma_\infty-1)+M_\infty^2(2-\gamma_\infty)]}{4\tau_Q M_\infty u_\infty C_{P\infty}(1+M_\infty)^2}. \qquad (8)$$

Подобной процедурой определяются инкременты нарастания возмущений давления, плотности и температуры (9).

$$G_{P\infty}=G_{\mathrm{v}\infty}+\frac{(1+\gamma_\infty M_\infty^2)}{2T_0(1-\gamma_\infty M_\infty^2)}\frac{dT_0}{dx},$$
$$G_{\rho\infty}=G_{\mathrm{v}\infty}+\frac{(3-\gamma_\infty M_\infty^2)}{2T_0(1-\gamma_\infty M_\infty^2)}\frac{dT_0}{dx}, \qquad (9)$$
$$G_{T\infty}=G_{\mathrm{v}\infty}-\frac{1}{2T_0}\frac{dT_0}{dx}.$$

Как видно из (9), в неоднородном потоке, в отличие от однородного, возмущения газодинамических величин имеют разные инкременты. Однако

относительные возмущения ($\tilde{\mathrm{v}}/u_\infty, \tilde{P}/P_0, \tilde{\rho}/\rho_0$ и так далее) нарастают с одним инкрементом:

$$G_\infty = G_{\mathrm{v}\infty} + \frac{1}{2T_0}\frac{dT_0}{dx}. \qquad (10)$$

В предельных случаях больших и малых чисел Маха инкремент G_∞ совпадает с инкрементами $G_{\rho\infty}$ и $G_{P\infty}$, соответственно.

Следует подчеркнуть, что вклады дополнительных слагаемых G^i в инкременты наиболее существенны в диапазоне малых чисел Маха:

$$G_{\mathrm{v}\infty} = \alpha_{\infty Q}^{Q\infty} - \frac{1}{4\tau_Q M_\infty u_\infty C_{P\infty}},\ G_{\mathrm{P}\infty} = \alpha_{\infty Q}^{Q\infty} + \frac{1}{4\tau_Q M_\infty u_\infty C_{P\infty}},$$

$$G_{\rho\infty} = \alpha_{\infty Q}^{Q\infty} + \frac{5}{4\tau_Q M_\infty u_\infty C_{P\infty}},\ G_{\mathrm{T}\infty} = \alpha_{\infty Q}^{Q\infty} - \frac{3}{4\tau_Q M_\infty u_\infty C_{P\infty}}. \qquad (11)$$

В сверхзвуковых потоках все волны малой амплитуды распространяются в одну сторону, но с разной фазовой скоростью. При $M_\infty >> 1$ для «быстрой» волны неоднородные поправки увеличивает инкремент нарастания возмущений скорости, давления и температуры, но, по-прежнему, уменьшают инкремент нарастания возмущений плотности:

$$G_{\mathrm{v}\infty} = \frac{\alpha_{\infty Q}^{Q\infty}}{M_\infty} - \frac{(2-\gamma_\infty)}{4\tau_Q M_\infty u_\infty C_{P\infty}},\ G_{\mathrm{P}\infty} = \frac{\alpha_{\infty Q}^{Q\infty}}{M_\infty} - \frac{(2+\gamma_\infty)}{4\tau_Q M_\infty u_\infty C_{P\infty}} = G_{\mathrm{T}\infty},$$

$$G_{\rho\infty} = \frac{\alpha_{\infty Q}^{Q\infty}}{M_\infty} + \frac{(3\gamma_\infty - 2)}{4\tau_Q M_\infty u_\infty C_{P\infty}}. \qquad (12)$$

Для волны, бегущей против потока, в (5) скорость потока надо брать с обратным знаком. Из (2) следует, что все градиенты тоже сменят знак. Таким образом для этой волны в (11) и (12) ситуация обратная.

Полученные результаты естественным образом обобщаются на потоки газа с иным типом неравновесности. Рассмотрим колебательно-возбужденный газа, поддерживаемый в неравновесном состоянии теплоисточником, зависящим от параметров среды (температуры и плотности). Система (1) видоизменяется:

$$C_{V\infty}\frac{dT}{dt} + \frac{dE_K}{dt} - \frac{T}{\rho}\frac{d\rho}{dt} = Q(T,\rho),$$

$$\frac{dE_K}{dt} = -\frac{E_K - E_{KE}(T)}{\tau_K(\rho,T)} + \alpha\cdot Q, \qquad (13)$$

где $\tau_K(\rho,T)$ - время релаксации, E_{KE} - равновесная колебательная энергия, Q - источник энергии в расчёте на одну молекулу,а α - показатель эффективности его работы. При $\alpha = 0$ среда перестает быть колебательно-неравновесной.

Формулы (5)-(12), можно обобщить на этот случай: $\alpha_{\infty Q}^{Q\infty}$ заменить суммой $\alpha_K + (1-\alpha)\alpha_{\infty Q}^{Q\infty}$, а $\tau_Q^{-1} \to \left(\frac{1-\alpha}{\tau_Q} + \frac{S}{\tau_K}\right)$, где $\alpha_K = \frac{C_{V0}(c_\infty^2 - c_0^2)}{2c_\infty^3 C_{V\infty}\tau_K}$.

Низкочастотный предел или поток с сильной неоднородностью рассмотрим на примере среды, описанной в [3]: (*CO2:N2:He = 1:4:22; T(0) = 300 K, U(0) = 270 м/с, ρ(0) = 0.042 кг/м3, Q= 30 Вт/см3*). Процесс колебательно релаксации молекул активной среды газа опишем блочной схемой. Необходимые функции равновесной колебательной энергии и времени колебательной релаксации зададим согласно [6,7]:

$$\tau^{-1} = 10^{-6} N_A \frac{\rho}{M}[\delta_{CO_2} 1{,}31\cdot 10^{-10} \exp(-62{,}75/T^{1/3}) + \delta_{N_2} 2{,}97\cdot 10^{-10} \exp(-75{,}46/T^{1/3}) +$$

$$+\delta_{He} 1{,}1\cdot 10^{-12}\sqrt{T}\exp(-58{,}82/T^{1/3})],$$

где N_A – число Авогадро; $\delta_{CO_2}, \delta_{N_2}, \delta_{He}$ - относительные концентрации углекислого газа, азота и гелия в смеси соответственно;

$$\frac{E_{KE}}{k} = \delta_{CO_2} \frac{\theta_3}{\exp\left(\frac{\theta_3}{T}\right) - 1} + \delta_{N_2} \frac{\theta}{\exp\left(\frac{\theta}{T}\right) - 1},$$

где k - постоянная Больцмана, $\theta_3 = 3383, \theta = 3357\ K$ - характеристические колебательные температуры колебательных мод молекул.

Счет производится разностной схемой второго порядка точности как по координате, так и по времени. Из-за того, что вязкость в (1) не учитывается, а численная вязкость не вводится, необходимо гладкое включение питающего источника. Оно хорошо видно на участках подъема [-0.5, 0], и спада [1.2, 2.5] на графике декремента абсолютных и относительных возмущений скорости (рис. 1).

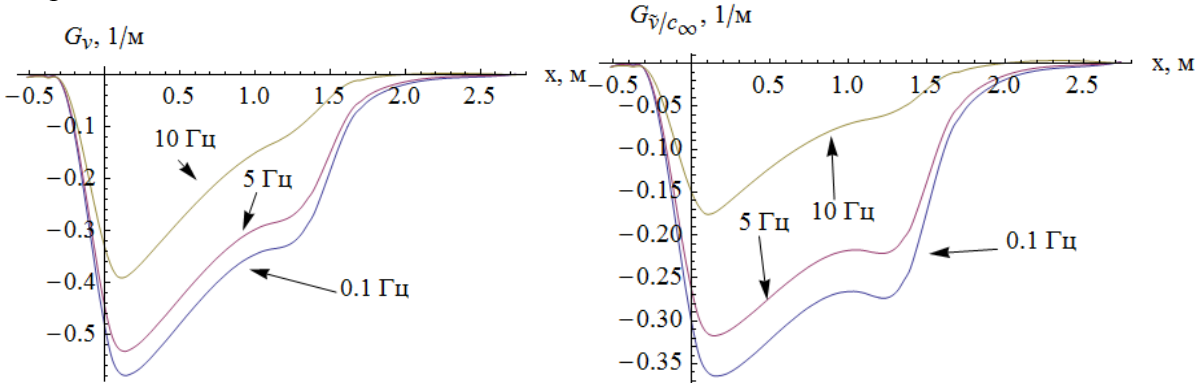


Рисунок 1. Декремент скорости

Так же как и в случае слабой неоднородности, декременты абсолютных возмущений разные, а относительных совпадают. Подобные декременты могут привести к двухкратному росту амплитуды пульсаций.

Работа частично поддержана аналитической целевой программой «Развитие научного потенциала высшей школы» (2009 – 2010 гг.), проект 2.1.1/309, Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009 – 2013 годы, НК-410 (10) и НК -422 (22).

Литература

1. Кольцова Е.В., Осипов А.И., Уваров А.В. Акустические возмущения в неравновесном неоднородном газе // Акустический журнал. 1994. Т. 40. № 6. С. 969-973.

2. Осипов А.И., Уваров А.В. Неравновесный газ: проблемы устойчивости // УФН. 1996. Т. 166. № 6. С. 639-650.

3. Панченко В.Я., Буяров С.А., Галушкин М.Г., Голубев В.С., ГришаевР.В., Дубров В.Д.,Завалов Ю.Н. Исследование мелкомасштабных неоднородностей активной среды прокачных CO2-лазеров // Сборник трудов / ИПЛИТ РАН, М., 2005. С. 101-109.

4. Колесников В.К., Хаит В.Д. Об акустической устойчивости дозвукового потока в МГД канале // ТВТ. 1976. Т. 14. N6. С.20-31.

5. Руткевич И.М. Акустическая неустойчивость МГД течения плотной плазмы // МЖГ. 1977. N.4. С. 3-11.

6. Гридин А.Ю., Климов А.И., Ходатаев К.В. Исследование распространения ударной волны в неоднородном поперечном импульсном разряд // ТВТ. 1994. Т. 32. № 4. С. 486-490.

7. Басаргин И.В., Мишин Г.И. Эволюция аномальных динамических свойств распадающейся плазмы тлеющего разряда // ЖТФ. 1996. т.66. № 7. С. 198 - 203.

Критические явления в модели кубического автокаталатора

Н.М. Шемонаева, Е.А. Щепакина
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Работа посвящена исследованию критических явлений в математической модели кубического автокаталатора. Интерес к критическим явлениям особенно возрос в настоящее время, связано это не только с вопросами безопасности, но и с тем, что часто именно критические режимы являются целью технологического процесса. Поэтому определение условий протекания критических режимов становится особенно актуально.

Автокаталатор – реакционный сосуд, в котором протекает автокаталитическое взаимодействие нескольких реагентов. Особенностью автокаталитической реакции является то, что один из реагентов выступает в роли катализатора, т.е. одновременно является и реагентом, и катализатором. В безразмерном виде математическая модель процесса описывается системой обыкновенных дифференциальных уравнений [1]:

$$\begin{cases} \varepsilon \dfrac{dp}{dt} = q - p - (p+q)(q-p)^2, \\ \dfrac{dq}{dt} = \dfrac{\mu - q + p}{2}, \end{cases} \qquad (1)$$

где p, q – безразмерные переменные, отражающие концентрации двух реагентов; ε, μ – параметры, характеризующие соотношения скоростей химических реакций. Отметим, что параметр ε мал, следовательно, данная система является сингулярно возмущенной, и для ее исследования можно применить аппарат теории интегральных многообразий сингулярно возмущенных систем [2].

Медленным интегральным многообразием сингулярно возмущенной системы дифференциальных уравнений

$$\begin{cases} \dot{x} = f(x, y), \\ \varepsilon \dot{y} = g(x, y), \end{cases} \qquad (2)$$

называется инвариантная поверхность, движение по которой осуществляется со скоростью медленной переменной x, т.е. со скоростью порядка $O(1)$. Отметим, что в системе есть также скорости порядка $O(\varepsilon^{-1}), \varepsilon \to 0$. Уравнение $g(x, y) = 0$ описывает медленную кривую системы (2). Участок медленной кривой, для всех точек которого выполняется неравенство

$g'_y(x, y) < 0 \quad (g'_y(x, y) > 0)$,

называется устойчивым или притягивающим (неустойчивым или отталкивающим). Устойчивые и неустойчивые участки медленной кривой являются нулевыми приближениями соответствующих устойчивых (притягивающих) и неустойчивых (отталкивающих) интегральных многообразий. Данные интегральные многообразия расположены в ε-окрестности медленной кривой, за исключением окрестностей точек срыва, в которых $g'_y(x, y) = 0$.

В случае системы (1) медленная кривая описывается уравнением

$f = q - p - (p+q)(q-p)^2 = 0$

и состоит из двух ветвей, см. рис. 1. Участки F_1, F_3 медленной кривой являются устойчивыми, а участок F_2 - неустойчивым. Отметим, что система (1) имеет особую точку с координатами

$$p = -\frac{\mu^2 - 1}{2\mu}, \quad q = \frac{1 + \mu^2}{2\mu}.$$

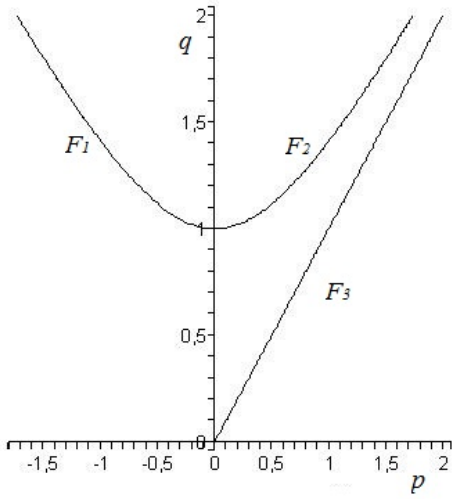


Рис.1. Медленная кривая системы (1)

При $\mu > 1$ особая точка устойчива и лежит на устойчивом участке F_1 медленной кривой, при $\mu < 1$ особая точка неустойчива и расположена на неустойчивом участке. Таким образом, в зависимости от значения параметра μ можно выделить три основных режима автокаталитической реакции: медленный, быстрый и разделяющий – критический.

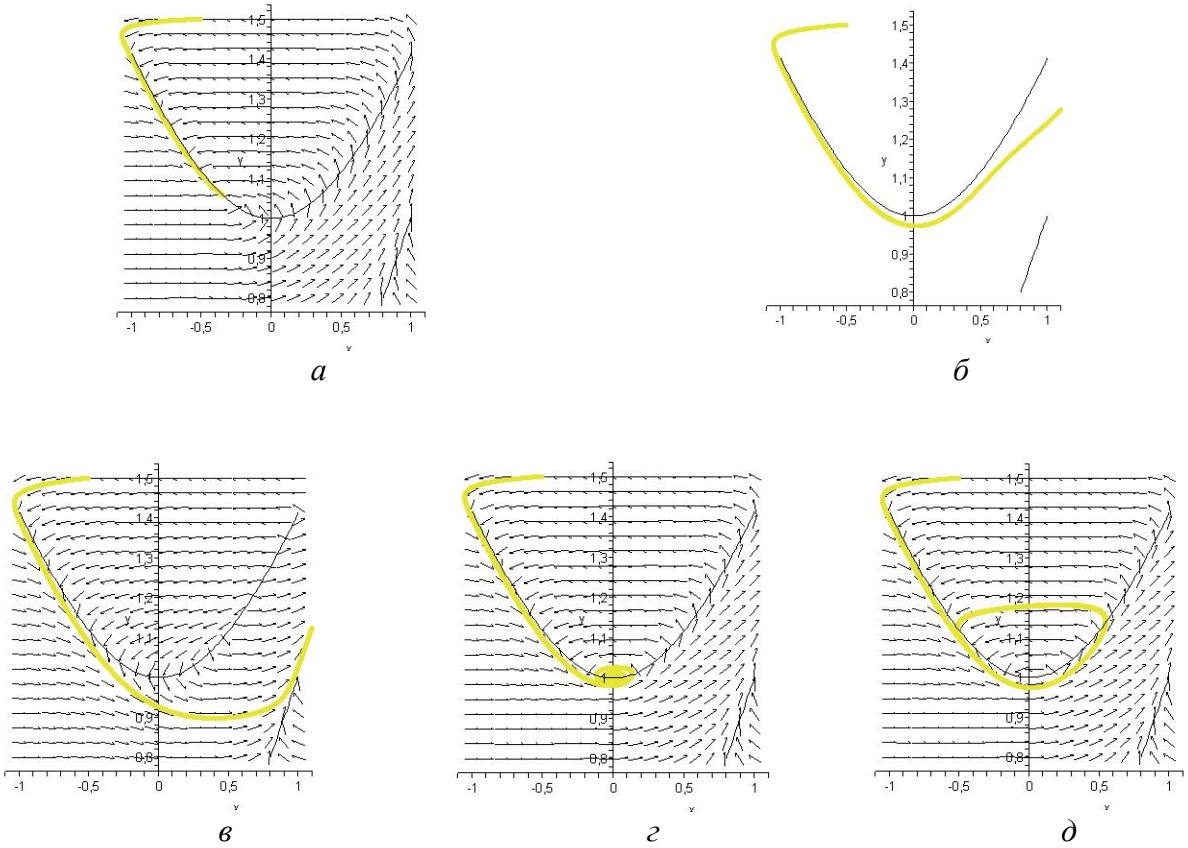
При $\mu > 1$ наблюдается медленный режим, см. рис. 2 (а). В этом случае траектория, выходящая из начальной точки, лежащей в области притяжения F_1, за короткий промежуток времени притягивается к F_1 и далее следует вдоль нее со скоростью медленной переменной, стремясь к устойчивой особой точке при $t \to \infty$.

При $\mu < 1$ наблюдается быстрый режим, когда особая точка лежит на неустойчивом участке F_2 или на участке F_3. В первом случае траектория после участка движения вдоль F_1 достигает точку срыва и срывается с медленного интегрального многообразия, далее следует участок движения со скоростью быстрой переменной, см. рис. 2 (б). Во втором случае после участка быстрого движения траектория, достигнув F_3, следует вдоль нее со скоростью медленной переменной, стремясь к устойчивой особой точке при $t \to \infty$, см. рис. 2 (в).

В окрестности значения $\mu = 1$ можно наблюдать критический режим, который является некоторым водоразделом между описанными выше режимами. При $\mu = 1$ особая точка сливается с точкой срыва, теряя при этом устойчивость, одновременно с этим от нее отделяется устойчивое периодическое решение, т.е. в системе наблюдается бифуркация Андронова-Хопфа [2, 4], см. рис. 2 (г). При уменьшении параметра μ особая точка отодвигается от точки срыва на неустойчивую ветвь, но находится в достаточно малой окрестности. При этом наблюдается эволюция предельного цикла: размер цикла постепенно увеличивается, и можно говорить о существовании

траектории-утки [1, 4, 5], см. рис. 2 (д). Под траекторией-уткой сингулярно возмущенной системы понимается траектория, содержащая непрерывный участок, который решение системы проходит сначала вдоль устойчивой части медленной кривой, а затем вдоль неустойчивой, причем оба раза расстояния порядка $O(1)$.

Следуя работе [4], найдены значения параметра μ, при которых предельный цикл рождается (бифуркационное значение) и становится траекторией уткой («уточное» значение) в виде асимптотического разложения по целым степеням малого параметра.



а *б* *в* *г* *д*

Рис. 2. Медленная кривая (изображена тонкой линией) и траектория системы (1) (жирная линия) в случае: (а) медленного режима, $\mu = 1.4$, $\varepsilon = 0.1$; (б) быстрого режима, $\mu = 0.95789999$, $\varepsilon = 0.1$; (в) $\mu = 0.5$, $\varepsilon = 0.1$; (г) бифуркации Андронова-Хопфа, $\mu = 1{,}0125$, $\varepsilon = 0.1$; (д) траектории-утки, $\mu = 0.975$, $\varepsilon = 0.1$

Так, значение параметра μ, при котором наблюдается бифуркация Андронова-Хопфа имеет представление

$$\mu = 1 + \frac{1}{8}\varepsilon + \varepsilon^2 \ldots,$$

значение параметра μ, при котором система (1) имеет траекторию-утку

$$\mu = 1 - \varepsilon\frac{1}{4} + \varepsilon^2 \ldots .$$

Найдена траектория-утка в виде системы (1) в виде асимптотического разложения

$$q(p,\varepsilon) = \sqrt{p^2+1} - \varepsilon\frac{\sqrt{p^2+1}}{4} + \varepsilon^2 \ldots,$$

отвечающая «уточному» значению параметра μ. При других значениях μ в системе нет траектории-утки.

В рассматриваемой системе наблюдается бифуркация Андронова-Хопфа с мягкой потерей устойчивости, когда система под действием постоянно присутствующих малых возмущений переходит сначала из неустойчивого состояния равновесия на «малую» устойчивую периодическую траекторию. В экспериментах при значениях параметра, близких к бифуркационному, возникающее периодическое решение мало отличается от стационарного решения, поскольку его амплитуда очень мала и может теряться в экспериментальном шуме. Однако при достижении параметром «уточного» значения ситуация резко меняется: незначительное изменение значений параметра приводит к так называемому уточному взрыву, когда амплитуда концентрационных колебаний практически мгновенно принимает достаточно большие значения. Это означает, что именно «уточное» значение параметра может рассматриваться как граница безопасного протекания процесса.

Работа выполнена при поддержке гранта РФФИ 10-08-00154а.

Литература

1. Brons M. Explosion of limit cycles and chaotic waves in a simple nonlinear chemical system /M. Brons, J. Sturis //Physical Review. E.–2001.–Vol. 64.–026209.

2. Стрыгин В.В. Разделение движений методом интегральных многообразий / В.В. Стрыгин, В.А. Соболев. – М.: Наука, 1988. – 256 с.

3. Марсден Дж. Бифуркация рождения цикла и ее приложения / Дж. Марсден, М. Мак-Кракен – М.: МИР, 1980. – 368 с.

4. Щепакина Е.А. Периодические колебания в модели каталитического реактора /Е.А. Щепакина //Известия РАЕМ. Серия ММИУ.–1998.–Т. 2.–№4.–С. 108-116.

5. Теория бифуркаций/ В.И. Арнольд, В.С. Афраймович, Ю.С. Ильяшенко, Л.П. Шильников //Современные проблемы математики: Фундаментальные направления. – М.: ВИНИТИ, 1986.–Т. 5.–С. 5-218.

Об одном методе выделения информативных фрагментов сигнала

П.С. Сидоров
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Алгоритмы решения многих задач обработки сигналов включают в себя этап выделения информативных фрагментов. Как правило, за счет отбора фрагментов удается повысить вычислительную устойчивость алгоритмов и сократить общее время работы последующих этапов алгоритма. Понятие информативности плохо формализуемо и его значение индивидуально для каждой конкретной задачи обработки. В работе под информативностью фрагмента понимается спектральное число обусловленности задачи обращения дискретного оператора свертки, построенного по отсчетам этого фрагмента. Чем меньше соответствующее число обусловленности, тем более информативен фрагмент и наоборот. Фрагменты, информативные в указанном смысле, могут использоваться в алгоритмах решения задачи идентификации систем.

Информативность фрагмента сигнала и число обусловленности

Информативность, в определенном выше смысле, фрагмента сигнала (части изображения)

$$X_{i,j} = \begin{pmatrix} a_{i-1,j-1} & a_{i-1,j} & a_{i-1,j+1} \\ a_{i,j-1} & a_{i,j} & a_{i,j+1} \\ a_{i+1,j-1} & a_{i+1,j} & a_{i+1,j+1} \end{pmatrix}, \tag{17}$$

где $a_{i,j}$ - компоненты двумерного сигнала, может определять число обусловленности матрицы для (17) представимой в виде

$$A = \begin{pmatrix} A_1 & A_0 & C \\ A_2 & A_1 & A_0 \\ C & A_2 & A_1 \end{pmatrix},\ A_i = \begin{pmatrix} x_{i,1} & x_{i,0} & c \\ x_{i,2} & x_{i,1} & x_{i,0} \\ c & x_{i,2} & x_{i,1} \end{pmatrix}, \tag{18}$$

где блок C состоит из отсчетов c, доопределяющих фрагмент (17) за границей, а $x_{i,j}$ - элементы фрагмента (17). Далее используется обозначение $A\left(X^{l\times l}\right)$ - матрица размера $l^2 \times l^2$, составленная из отсчетов фрагмента X размера $l \times l$.

Матрица (18) имеет блочно-теплицеву структуру с теплицевыми блоками и изоморфна дискретному оператору двумерной линейной свертки. В работе рассматриваются фрагменты малых размеров $X_{i,j} \in R^{3\times 3}$ и $A \in R^{9\times 9}$, так как задача вычисления спектрального числа обусловленности для матриц больших размеров представляет серьезную вычислительную сложность. В качестве доопределяющего коэффициента c предлагается выбирать среднее значение по фрагменту (17). В докладе приводятся различные важные свойства матриц вида (18).

Число обусловленности матрицы A суть нормированная производная Фреше [1] отображения $A \to A^{-1}$ и зависит от выбора матричной нормы. Таким образом, обусловленность матрицы в 2-норме

$$k(A) = \|A\|_2 \cdot \left\|A^{-1}\right\|_2 = \sigma_{\max} \cdot \sigma_{\min}^{-1}, \tag{19}$$

где $\sigma_{\max}$ и $\sigma_{\min}$ - максимальное и минимальное сингулярные числа матрицы A.

Пусть отсчеты сигнала $\left\{a_{i,j}\right\}_{i,j=0}^{N-1}$ принимают значения из множества $B=\{0,1\}$, тогда фрагмент (17) может определять один из $2^9=512$ шаблонов, некоторые из которых приведены на рис.1 с указанием соответствующего числа обусловленности $k(A)$.

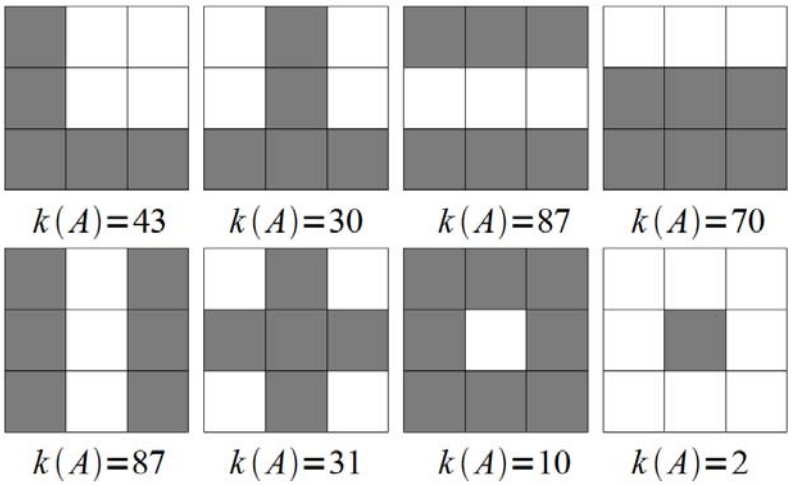


Рис.1. Информативность некоторых фрагментов бинарных изображений.

Фрагмент, изображенный на рис.1. последним во втором ряду (правый нижний), оказывается самым информативным (с числом обусловленности равным 2) при решении задачи идентификации параметров искажающей системы типа свертки, так как решение данной задачи – импульсная характеристика суть реакция системы на δ-импульс.

Каждому фрагменту размера 3 на 3 отсчета можно поставить в соответствие число от 0 до 511, рассматривая соединение строк фрагмента в вектор как двоичное представление десятичного числа. Например, фрагменту с наименьшим числом обусловленности на рис.1 соответствует число $(000010000)_2=2^4=16$ и теплицева матрица A^{16}. Обозначим множество матриц (18) составленных по фрагментам вида (17) для бинарных изображений как $\Omega=\left\{A: A\in B^{9\times 9}\ \ u\ \ c=0\right\}$, тогда имеет место факт 1.

Факт 1. Минимум на множестве чисел обусловленности для матриц из Ω достигается на матрице A^{16}.

Алгоритм выделения информативных фрагментов

Естественно, что фрагменты малых размеров невозможно использовать при решении практических задач идентификации, поэтому важно учитывать при исследовании информативности также и размер фрагмента. Авторы предлагают следующий алгоритм:

1. Изображение обрабатывается скользящим окном малого размера $l\times l$ и для каждого фрагмента-окна рассчитывается информативность. В результате первого этапа известно поле H информативности точек сигнала.

2. В наиболее информативных по некоторому порогу точках запускается процесс увеличения размера фрагмента до тех пор, пока не выполнится условие

$$k\left(A\left(X^{l+1\times l+1}\right)\right) > \gamma \cdot k\left(A\left(X^{l\times l}\right)\right), \gamma > 0, \quad (20)$$

где γ - некоторое число определяемое экспериментальным путем.

3. Найденные фрагменты проверяются на отсутствие включения и пересечения между собой. Малые фрагменты, входящие во фрагменты большего размера, исключаются из дальнейшего анализа.

На рис.2 представлены результаты выделения информативных фрагментов на модельном и реальном изображениях.

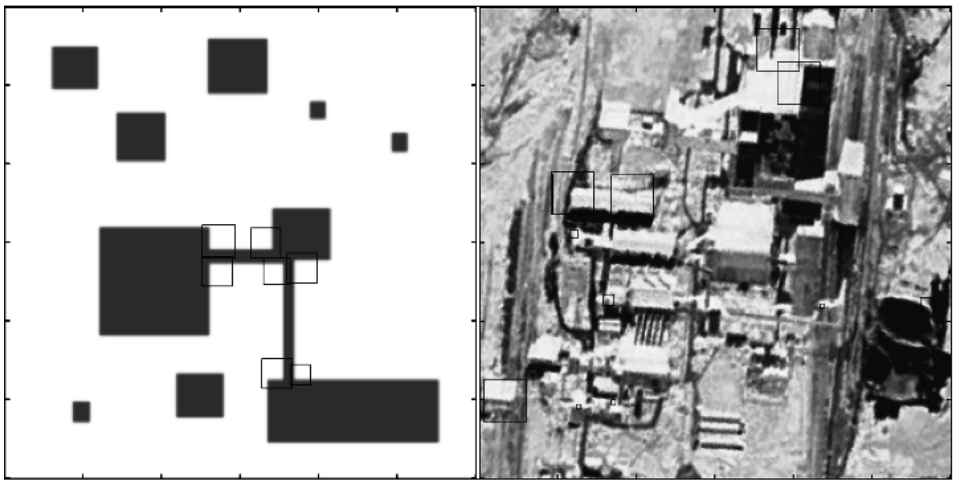


Рис.2. Результат работы предлагаемого алгоритма выделения информативных областей.

Заключение

Как показали вычислительные эксперименты, предлагаемый алгоритм обладает следующими недостатками:

1. Задача вычисления точного значения спектрального числа обусловленности матрицы размерности больше $10^2 \times 10^2$, представляет серьезную вычислительную сложность и для ее решения необходимо применять итерационные или вероятностные методы оценки.

2. Как правило, исследуемые сигналы оказываются зашумленными, таким образом, и матрица A является возмущенной. Высокая чувствительность спектра [1] матрицы A к возмущениям не позволяет получить устойчивую оценку числа обусловленности. В этом случае необходима фильтрация изображения, в работе авторы используют предварительное размытие изображения окном размера 3×3.

В докладе приводятся результаты исследования алгоритма выделения информативных областей использующего иные определения нормы матрицы – отличные от спектральной, при этом качество найденных фрагментов ухудшается, однако вычислительная сложность уменьшается на порядки.

Литература

1. Голуб Дж., Ван Лоун Ч. Матричные вычисления: Пер. с англ. – М.: Мир, 1999. – 548 с.

Бесконтактное измерение концентрации газа по анализу спектров излучения синглетного кислорода

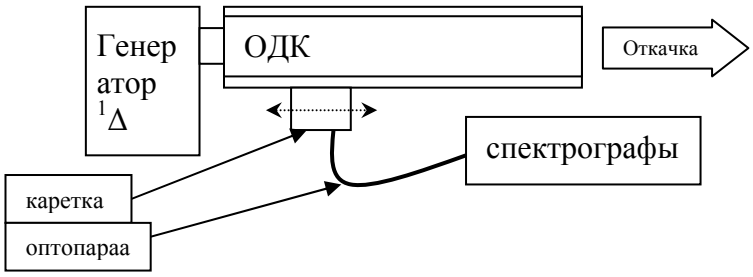
А.С. Инсапов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В некоторых областях существует необходимость бесконтактного измерения параметров неравновесного газового потока, так как контактные методы работают не всегда достаточно корректно и могут вносить сильные изменения в поток. Одна из таких областей это химические кислородно-йодные лазеры. Контактные методы измерения в процессе транспортировки газа от генератора допускают сильные ошибки по причине отдачи возбужденным кислородом энергии датчикам температуры. Данная работа посвящена измерению концентрации синглетного кислорода по потоку и нахождении константы скорости реакции самотушения.

Экспериментальная установка

Установка состоит из генератора синглетного кислорода, Оптической Диагностической Камеры (ОДК), 2х спектрографов и каретки с оптопарой, закрепленной на реостате.



Определение концентрации синглетного кислорода

Поток синглетного кислорода излучает в 4х основных диапазонах:

$^1\Delta + {}^1\Delta \rightarrow {}^3\Sigma + {}^3\Sigma + h\nu$ (λ= 634 нм) (1)

$^1\Delta + {}^1\Delta \rightarrow {}^3\Sigma + {}^3\Sigma_{V=1} + h\nu$ (λ= 703 нм) (2)

$^1\Sigma \rightarrow {}^3\Sigma$ (λ= 760 нм) (3)

$^1\Delta \rightarrow {}^3\Sigma$ (λ=1270 нм) (4)

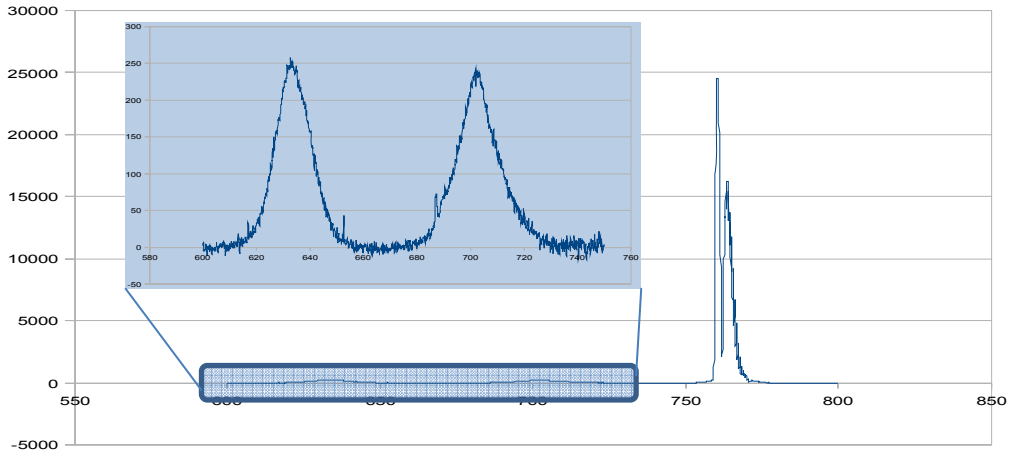


Рис1. Излучение синглетного кислорода в диапазоне 600-800нм

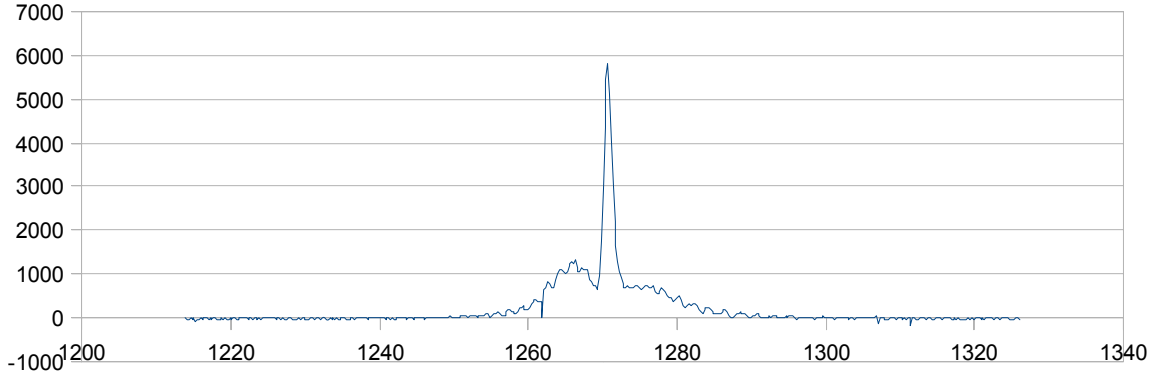


Рис2. Излучение синглетного кислорода в диапазоне 1220-1320нм

В данной работе для измерения концентрации были использованы пики на 634нм и 1270нм. Первый является димольным излучением (р-я (1)), а второй спонтанным излучением (р-я (4)). Димольное излучение возникает за счет столкновения двух молекул кислорода в первом электронно-возбужденном состоянии, после которых обе молекулы переходят в основное состояние при этом обе могут перейти в основное колебательное состояние, либо одна из молекул в состоянии V=1. Две этих реакции обуславливают 2 пика (634нм и 703нм). Форма и ширина пиков является функцией температуры и используется для ее определения. Второй пик, используемый для нахождения концентрации, является следствием спонтанного излучения синглетного кислорода из первого электронно-возбужденного состояния и тоже может служить для определения температуры методом сравнения синтезированного спектра и спектра, снятого монохроматором, однако из-за сильной зашумленности, как правило, лучше использовать димольное излучение.

Для вычислений концентраций использовалась формулы:

$$n_a = \frac{4}{A_a T L t_e} \int_{\text{a-X}} \mathrm{C}(\lambda)\mathrm{R}(\lambda)d\lambda$$ для спонтанного излучения (4) и

$$n_a = \left(\frac{4}{k_D T L t_e} \int_{(1)} \mathrm{C}(\lambda)\mathrm{R}(\lambda)d\lambda \right)^{0.5}$$ для димольного излучения (1),

где A_a – коэффициент Эйнштейна для спонтанного перехода (4), k_D=(6.06±0.19)×10-23 см3/с - константа скорости димольного излучения, L - ширина ОДС, T - пропускание оптического окна, t_e -время экспозиции спектров, C(λ)-число отсчетов пиксела на длине волны λ, R(λ)-(фотон×(отсчет)-1см-2нм-1)-абсолютная спектральная чувствительность системы оптоволокно-спектрометр. Для обработки данных была написана программа, которая обрабатывала пакеты данных и выдавала информацию о концентрации вдоль потока. Полученные сведения позволяют определить константу скорости реакции.

Определение скорости реакции самотушения синглетного кислорода

Реакция самотушения имеет вид : $^1\Delta + {}^1\Delta \rightarrow$ продукты реакции.

Зная давление, температуру, расход O_2, сечение канала ОДК, определяем скорость потока. Зная координаты каретки и скорость находим время $t_п$ перемещения газа от начала ОДК до каретки. Согласно стехиометрии реакции самотушения концентрация $^1\Delta$ изменяется по формуле : $\frac{1}{n_0} - \frac{1}{n} = k_s t_n$. По обработке массивов спектра

было получено : $k_s = (5.07 \pm 1{,}24) * 10^{-17} cm^3 / c; \varepsilon = 0.25$, по инфракрасной части спектра- $k_s = (7.19 \pm 0.94) * 10^{-17} cm^3 / c; \varepsilon = 0.13$. Эти значения являются усреднением по диапазону температур 320-380К, и корректность усреднения еще предстоит проверить на более широких диапазонах для выявления зависимости от температуры или опровержения таковой. Для оценки температуры использовалась зависимость ширины пиков реакции (1) от температуры. Ширина пиков получалась путем аппроксимации их функцией псевдо-Фойхта по алгоритму Левенберга-Марквардта, представленному в свободной библиотеке ALGLIB.

Исходные данные и инструменты

Используемые в работе спектры были получены на генераторе синглетного кислорода в лаборатории химических и газоразрядных лазеров СФ ФИАН им. П.Н. Для написания программ использовалась среда разработки Microsoft Visual Studio 2008 (ЯП C#), а также использовался OpenOffice для обработки полученных значений концентраций. Также использовалась свободная библиотека математических алгоритмов ALGLIB (http://alglib.sources.ru/)

Заключение

В ходе данной работы был рассмотрен метод измерения концентрации синглетного кислорода в первом возбужденном состоянии по потоку и, как следствие, нахождение константы скорости реакции самотушения. В дальнейшем необходимо повысить точность определения концентрации и проверить зависимость константы скорости реакции самотушения на зависимость от температуры.

Благодарности

Особую благодарность автор статьи приносит д. ф.-м. н. М.В. Загидуллину за постановку задачи и внимание к работе.

Бесконтактное измерение температуры газа по анализу спектров излучения кислорода в диапазоне 600-800 нм

М.С. Малышев

Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Измерение температуры газов, содержащих достаточное количество молекул, находящихся в возбуждённом состоянии, обычными средствами практически невозможно. Это является следствием того, что молекулы при столкновении с поверхностью термометра передают ему не только тепловую энергию, но также и энергию возбуждения, вследствие чего показания термометра оказываются завышенными.

Однако, необходимость измерения температуры подобных газов возникает во многих научных и инженерных задачах, одной из которых является измерение параметров выходного потока генератора синглетного кислорода. Вышеупомянутый генератор позволяет получить на выходе кислород с высокой долей молекул в возбуждённых состояниях и является одним из компонентов химического кислородно-йодного лазера. В этой работе освещается измерение температуры данного выходного потока, хотя нижеследующие методики могут быть применены к любым газам, содержащим большое количество молекул кислорода в возбуждённом состоянии.

Аппроксимация спектра излучения одиночных молекул

У молекулы кислорода имеется три электронных энергетических уровня: основной уровень $^3\Sigma$, и возбуждённые уровни $^1\Sigma$ и $^1\Delta$. Переходу $^1\Sigma \rightarrow {}^3\Sigma(0\text{-}0)$ соответствует длина волны $\lambda_1 \approx 760$ нм; переходу $^1\Delta \rightarrow {}^3\Sigma(0\text{-}0)$ – длина волны $\lambda_2 \approx 1270$ нм. Вследствие расщепления данных уровней на вращательные подуровни, линии имеют сложную структуру и занимают довольно широкую область спектра. Линия λ_2 по сравнению с λ_1 имеет более сложную структуру, а также слабее меняет свой вид с изменением температуры, поэтому в данной работе не используется.

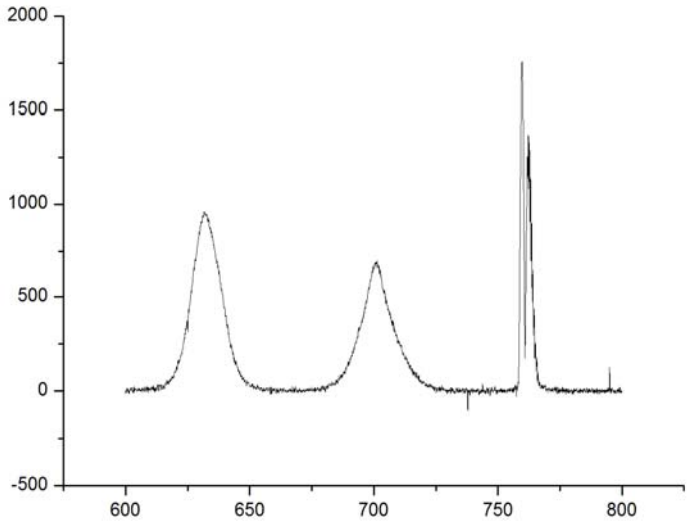


Рис.1. Структура спектра кислорода в диапазоне 600-800 нм.

Методика измерения температуры, предложенная в этой работе заключается в сравнении экспериментально полученного спектра со спектром, рассчитанным на основании заранее измеренных параметров спектральных линий, взятых из таблиц. Зависимость интенсивности отдельной спектральной линии от температуры имеет вид:

$$I_{JJ'}(T) = A_{JJ'}C(2J+1)e^{-\frac{E_J}{kT}}$$

Здесь J – вращательное число верхнего уровня, E_J – энергия верхнего уровня, J' – вращательное число нижнего уровня, C – некая константа пропорциональности.

Конечный расчётный спектр получается путём сложения умноженных на интенсивность и смещённых на соответствующую величину аппаратных функций прибора, использованного для снятия исследуемого спектра. Для аппроксимации аппаратной функции был выбран псевдо профиль Фойхта, как дающий наименьшие расхождения с реальной функцией.

В ходе работы была написана программа, осуществляющая вышеупомянутую методику. В качестве входных данных программа требует экспериментальный спектр, параметры аппаратной функции, а также таблицы, по которым будет осуществлено построение аппроксимирующего спектра. Программа осуществляет несколько итераций, в каждой из которых вычисляется среднеквадратичное отклонение между экспериментальным и расчётным спектрами. Путём изменения температуры, для которой строится расчётный спектр, производится поиск минимума СКО. Температура, при которой был осуществлён этот минимум и является искомой температурой газа.

Данная методика не лишена недостатков. Одним из них является то, что получение достаточно чистого спектра излучения на длине $\lambda_1 \approx 760$ нм является довольно сложной задачей по причине быстрого гашения $^1\Sigma$ кислорода на парах воды. Также в процессе работы с программой было выяснено, что данная методика даёт плохую сходимость экспериментального и аппроксимирующего спектров в случае комнатных и превышающих их температур. Эти недостатки частично исправлены в методике, описанной ниже.

Аппроксимация спектра димольного излучения

Молекулы синглетного кислорода также могут излучать кванты и при столкновениях друг с другом, это так называемое димольное излучение. При этом обе молекулы либо могут перейти в основное состояние, либо одна из молекул будет колебательно возбуждённой. В результате этих процессов на спектре появляются два довольно широких пика с вершинами $\lambda_1 \approx 633$ нм и $\lambda_2 \approx 702$ нм. Большая ширина этих пиков обуславливается различием в скоростях сталкивающихся молекул, подчиняющихся распределению Максвелла. Как видно, ширина пиков должна зависеть от температуры.

Была написана программа, которая осуществляет аппроксимацию вышеупомянутых пиков. В качестве основной части аппроксимирующей функции был выбран тот же псевдо профиль Фойхта, что использовался для аппроксимации аппаратной функции. Также в аппроксимирующую функцию была добавлена базовая линия, соответствующая уровню шумов. Аппроксимация осуществлялась при помощи алгоритма Левенберга-Марквардта, реализованного в свободной библиотеке алгоритмов ALGLIB. После аппроксимации, программа вычисляет ширину пиков и прочие их параметры, такие как координаты вершин и параметры аппроксимирующей функции.

Были обработаны спектры, полученные в ходе нескольких запусков в диапазоне температур 170-380K. Для каждого спектра из этого набора, при помощи программы, описанной в предыдущем пункте, была вычислена температура. Затем, с помощью

программы, описанной в текущем пункте, была получена ширина пиков димольного излучения. Совмещением полученных данных, была получена зависимость ширины пиков от температуры газа, которая, как оказалось, довольно хорошо ложится на прямую. Аппроксимирующие зависимости имеют следующий вид:

$\lambda_1 \approx 633$ нм: $w_1(T) = (199 \pm 14) + (0{,}53 \pm 0{,}05)T$ см$^{-1}$

$\lambda_2 \approx 702$ нм: $w_2(T) = (152 \pm 17) + (0{,}49 \pm 0{,}06)T$ см$^{-1}$

Таким образом, имея зависимость ширины пиков от температуры достаточно измерить ширину пиков димольного излучения для определения температуры газа.

Вследствие того, что погрешность измерения температуры при получении данной зависимости довольно велика и составляет 5-10K, погрешность аппроксимации тоже довольно велика. Как следствие, большой оказывается и погрешность определения температуры при помощи этой зависимости. Также, очевидно, что данные зависимости верны только в пределах температурного диапазона, в котором проводилось снятие исследуемых спектров. Однако, такой метод определения температуры достаточно прост и удобен и при соответствующем усовершенствовании программного комплекса может быть осуществлён непосредственно при получении спектра с монохроматора.

В процессе вычисления ширины пиков также была получена зависимость координат вершин пиков от температуры. Она тоже достаточно хорошо ложится на линию, причём для двух пиков коэффициенты наклона аппроксимирующих прямых примерно совпадают. Данные зависимости также могут использоваться для измерения температуры, но в данной работе такая возможность подробно не рассматривается.

Исходные данные и инструменты

Упомянутые в данной работе спектры были получены на генераторе синглетного кислорода в лаборатории химических и газоразрядных лазеров СФ ФИАН им. П.Н. Лебедева и предоставлены автору д. ф.-м. н. М.В. Загидуллиным. Для построения расчётного спектра молекул кислорода использовалась база данных HITRAN (http://www.cfa.harvard.edu/HITRAN/HITRAN2008). Для написания программ использовалась среда разработки Microsoft Visual Studio 2005, язык разработки C#. Также использовалась свободная библиотека математических алгоритмов ALGLIB.

Заключение

В данной работе было предложено два метода бесконтактного измерения температуры газовой смеси, содержащей кислород в возбуждённом состоянии. К сожалению, эти методы пока обладают недостаточно высокой точностью. В дальнейшем, необходимо повышать количество обрабатываемых данных для уменьшения погрешности аппроксимации пиков димольного излучения, а также совершенствовать сами методы аппроксимации.

Фильтрация шумов, вызванных нелинейными колебаниями модели в аэродинамическом эксперименте при помощи искусственных нейронных сетей

Д.В. Ливерко
Московский физико-технический институт (национальный исследовательский университет), г. Жуковский

Введение

Важной задачей при проведении экспериментальных исследований летательных аппаратов является наиболее полное извлечение полученной информации о поведении аэродинамических коэффициентов исследуемого объекта. Используемые в настоящее время измерительные системы позволяют записывать в режиме реального времени большие массивы экспериментальных параметров. Однако, зачастую при проведение экспериментальных исследований невозможно обеспечить постоянство параметров, которые существенно влияют на аэродинамические характеристики исследуемого ЛА, например число М, скорость полета или режим работы двигательной установки. В этом случае при проведении обработки экспериментальных данных для получения математической модели аэродинамических характеристик ЛА необходимо решать задачу сведения, т.е. корректировки измеренных аэродинамических сил и моментов с целью приведения их к заданным постоянным параметрам. При этом необходимо решать задачу сведения в многомерном пространстве параметров, что существенно усложняет решение.

Как правило, для решения таких задач необходимо иметь какие-либо математические модели, описывающие поведение аэродинамических характеристик от исследуемых параметров. Таким образом, получается некий заколдованный круг, для построения математической модели аэродинамики ЛА по данным экспериментальных исследований необходимо иметь эту модель.

Удобным математическим аппаратом для решения подобных задач аппроксимации данных в многомерных пространствах являются искусственные нейронные сети.

Целью работы является исследование эффективности использования искусственных нейронных сетей для решения задач такого рода.

Описание математической модели аэродинамического эксперимента

В качестве примера использования данного подхода рассмотрена задача построения математической модели стационарных аэродинамических характеристик самолета по результатам эксперимента в аэродинамической трубе периодического действия Т109 ЦАГИ с закрытой рабочей частью. В этой трубе рабочий поток создается с помощью сжатого воздуха из системы газгольдеров. Аэродинамические силы и моменты определяются по показаниям тензовесов расположенных в хвостовой державке модели, рис. 1.

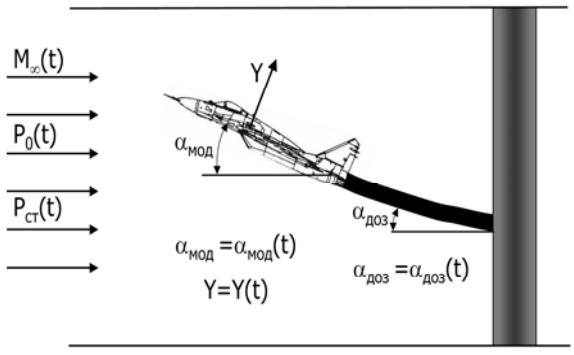


Рис. 1. Схема модели в АДТ Т-109 ЦАГИ

Наличие колебаний, обусловленных упругостью системы модель – весы, и нестационарность потока в аэродинамической трубе приводят к зашумлению экспериментальных результатов. На рис. 2 показано изменение угла атаки модели по времени в течение одного пуска аэродинамической трубы

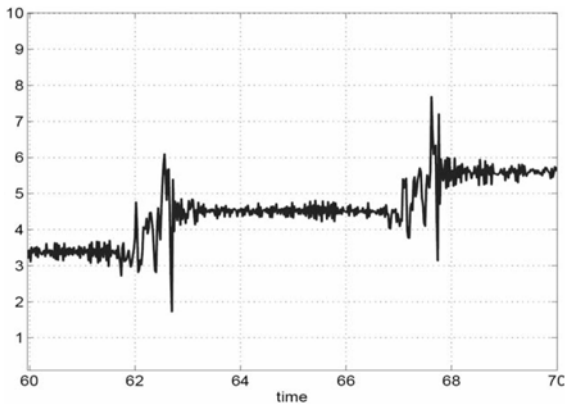


Рис. 2. Изменение угла атаки в течение пуска

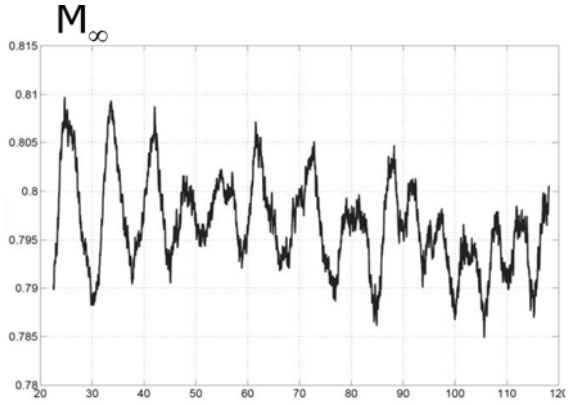


Рис. 3. Изменение числа Маха в течение пуска

Из графика видно, упругие колебания системы приводят к высокочастотному изменению угла атаки модели во время проведения эксперимента.

Кроме высокочастотных колебаний вызванных упругими колебаниями модели в рабочей части наблюдаются низкочастотные колебания параметров набегающего потока связанные с газодинамическим трактом АДТ. Для примера на рис. 3 зависимости изменения числа М.

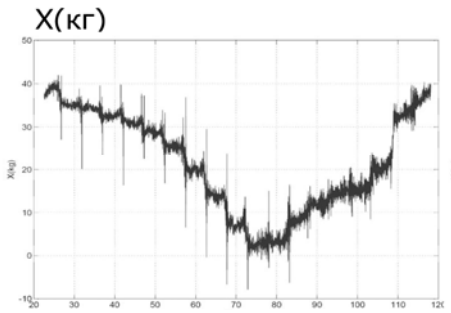


Рис. 4. Изменение продольной компоненты аэродинамической силы по времени пуска

При этом отклонения сил и моментов, определяемых по весам, от их стационарных значений являются сложными функциями многих переменных. Это параметры потока в рабочей части АДТ и их производные, положение и конфигурация модели, производные сил и моментов, измеряемых весами АДТ. В целом число значимых переменных достигает несколько десятков.

Зависимость измеренной тензовесами продольной силы от времени приведена на рис. 4.

2. Упругая модель крыла

Для оценки работоспособности и точности предложенного метода была решена модельная задача по определению стационарных аэродинамических характеристик профиля колеблющегося в нестационарном потоке.

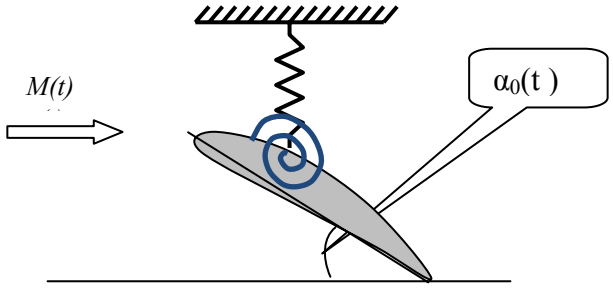


Рис. 5. Схема модели

3. Описание метода фильтрации

В данной работе были построены и обучены нейронные сети для аппроксимации аэродинамических коэффициентов сил и моментов модели с учетом нестационарных эффектов. Применение нейронной сети в данной задаче схематично представлено на рис. 5.

$$\vec{F} = \vec{\Phi}(\{\alpha, \beta, \delta_{мех}, M_\infty, Re\} = const, ..., 0, 0, ...);$$

$$\vec{F} = (X, Y, Z, Mx, My, Mz)^T;$$

$$\vec{P} = (\alpha, \beta, \delta_{мех}, M_\infty, Re, ..., 0, 0, ...)^T;$$



Рис. 5.

4. Результаты

Был поставлен численный эксперимент, и обработаны его данные при помощи Искусственных нейронных сетей и метода наименьших квадратов. В результате мы убедились, что нейронные сети являются удобным математическим аппаратом для обработки и аппроксимации экспериментальных данных, позволяя наиболее полно использовать полученную информацию; так как с помощью нейронных сетей из эксперимента при среднем Махе 0.8 можно получить данные для Маха 0.805 и 0.795.

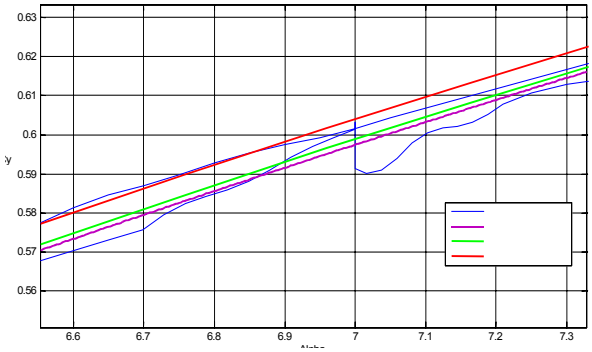


Рис. 6.

Среднеквадратичное отклонение для метода наименьших квадратов: 9.8299e-004
Среднеквадратичное отклонение для искусственной нейронной сети: 5.2388e-004

Заключение

В целом по полученным результатам можно сформулировать следующие выводы: нейронные сети являются удобным математическим аппаратом для обработки и аппроксимации экспериментальных данных, позволяя наиболее полно использовать полученную информацию.

Литература

1. Hertz J., Krogh A., Palmer R.G. Introduction to the theory of neural computation. Lecture Notes Volume I the Santa FE Institute studies in the sciences of complexity, 1991.

2. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. -Новосибирск, Наука, 1996.
3. Ереза А.Г., Павленко А.А., Свириденко Ю.Н. Учет аэроупругих колебаний при испытании модели в аэродинамической трубе, материалы XVII школы-семинара «Аэродинамика летательных аппаратов», ЦАГИ, 2006
4. Greenman R.M. Two-Dimensional High-Lift Aerodynamic Optimization Using Neural Networks NASA/TM-1998-112233
5. Аникин В.А., Свириденко Ю.Н. Применение искусственных нейронных сетей в задачах проектирования и определения аэродинамических характеристик профилей вертолетного винта. 6 форум Российского вертолетного общества, Москва, 2004
6. Свириденко Ю.Н. Применение искусственных нейронных сетей для создания модели аэродинамики ЛА по данным аэродинамического эксперимента, сб. докладов 5-научной конференции по гидроавиации «Гидроавиасалон-2004» , Москва, 2004
7. Ереза А.Г., Павленко А.А., Свириденко Ю.Н. Применение искусственных нейронных сетей для обработки данных аэродинамического эксперимента

Секция 4

ГЕОИНФОРМАТИКА И ОБРАБОТКА ИЗОБРАЖЕНИЙ

Спектрально-энергетический метод идентификации линейной модели наблюдения при дистанционном зондировании Земли

В.В. Сергеев, А.Ю. Денисова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Линейная модель наблюдений является распространенным способом описания систем формирования изображений. Так, например, она хорошо описывает действие видеоинформационного тракта систем дистанционного зондирования Земли. Идентификация, т.е. определение параметров модели наблюдения, является предварительным этапом решения многих важных задач, в том числе, восстановления сигналов, оценки качества систем формирования изображений и т.п. В статье исследуется один из возможных методов решения задачи идентификации, а именно, спектрально-энергетический метод.

Постановка задачи

Линейная модель наблюдения [1] представляет собой свертку двумерной последовательности входного сигнала $x(n_1,n_2)$ с импульсной характеристикой искажающей системы $h(k_1,k_2)$ с добавлением аддитивного статистически независимого от сигнала «белого» шума $v(n_1,n_2)$.

$$y(n_1,n_2)=\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}h(k_1,k_2)x(n_1-k_1,n_2-k_2)+v(n_1,n_2),\quad 0\le n_1,n_2\le N-1 \qquad (1)$$

где n_1,n_2 - целочисленные аргументы. Пусть входной сигнал в модели (1) является стационарным случайным процессом с нулевым средним. Автоковариационная функция (АКФ) «белого» шума имеет вид:

$$B_V(m_1,m_2)=\begin{cases}D_V, & m_1=m_2=0;\\ 0, & \textit{иначе.}\end{cases} \qquad (2)$$

где D_V – дисперсия шума. Задача идентификации состоит в том, чтобы найти неизвестную импульсную характеристику $h(k_1,k_2)$ искажающей системы по выходному сигналу $y(n_1,n_2)$.

Будем решать задачу при следующих ограничениях на модель входного сигнала и искажающие воздействия:

1. Импульсная характеристика искажающей системы является "узкой", ее можно считать отличной от нуля в некотором окне:

$$\begin{cases} h(k_1,k_2) \neq 0, & |k_1| \le K,\ |k_2| \le K \\ h(k_1,k_2) = 0, & \textit{иначе} \end{cases}, \tag{3}$$

выполняется условие нормировки:

$$\sum_{k_1=-K}^{K} \sum_{k_2=-K}^{K} h(k_1,k_2) = 1, \tag{4}$$

и частотная характеристика искажающей системы (преобразование Фурье импульсной характеристики):

$$H\left(e^{i\omega_1}, e^{i\omega_2}\right) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} h(k_1,k_2) e^{-i(\omega_1 k_1 + \omega_2 k_2)} \tag{5}$$

является вещественной и неотрицательной функцией.

2. Входной сигнал имеет экспоненциальную неразделимую АКФ, типичную для космических изображений [2]. Ограничимся случаем, когда АКФ обладает радиальной симметрией, т е. одинаковым образом экспоненциально спадает по всем направлениям:

$$B_X(m_1,m_2) = D_X \rho^{\sqrt{m_1^2+m_2^2}}. \tag{6}$$

где D_X - дисперсия сигнала, ρ - коэффициент корреляции между соседними пикселами. Можно получить, что энергетический спектр входного сигнала (преобразование Фурье от АКФ) в этом случае имеет вид:

$$\Phi_X\left(e^{i\omega_1}, e^{i\omega_2}\right) = \sum_{l_1=-\infty}^{\infty} \sum_{l_2=-\infty}^{\infty} \frac{-2\pi \ln \rho D_X}{\left((\ln\rho)^2 + (\omega_1 + 2\pi l_1)^2 + (\omega_2 + 2\pi l_2)^2\right)^{3/2}}. \tag{7}$$

Спектрально-энергетический метод идентификации

Известно, что энергетические спектры входного и выходного сигналов в модели наблюдения (1) связаны соотношением [1]:

$$\Phi_Y\left(e^{i\omega_1}, e^{i\omega_2}\right) = \left|H\left(e^{i\omega_1}, e^{i\omega_2}\right)\right|^2 \Phi_X\left(e^{i\omega_1}, e^{i\omega_2}\right) + D_V, \tag{8}$$

где $\Phi_X\left(e^{i\omega_1}, e^{i\omega_2}\right)$, $\Phi_Y\left(e^{i\omega_1}, e^{i\omega_2}\right)$ – энергетические спектры входного и выходного сигнала соответственно. Из формулы (8) можно выразить модуль частотной характеристики системы, но при соблюдении ограничения 1 знак модуля снимается, т.е. получаем однозначный результат:

$$H\left(e^{i\omega_1}, e^{i\omega_2}\right) = \sqrt{\frac{\Phi_Y\left(e^{i\omega_1}, e^{i\omega_2}\right) - D_V}{\Phi_X\left(e^{i\omega_1}, e^{i\omega_2}\right)}}. \tag{9}$$

Искомая импульсная характеристика вычисляется через обратное преобразование Фурье от частотной характеристики [1]:

$$h(k_1,k_2) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H\left(e^{i\omega_1}, e^{i\omega_2}\right) e^{i(\omega_1 k_1 + \omega_2 k_2)} d\omega_1 d\omega_2. \tag{10}$$

Из (9) видно, что для получения оценки импульсной характеристики искажающей системы при применении описанного спектрально-энергетического метода необходимо иметь оценки энергетических спектров входного и выходного сигналов, а также дисперсии шума.

Оценка энергетического спектра выходного сигнала

В качестве оценки энергетического спектра выходного сигнала используется периодограмма [3], непосредственно получаемая из наблюдаемого выходного сигнала:

$$\Phi_Y\left(e^{i\omega_1},e^{i\omega_2}\right)=\frac{1}{N^2}\left|Y\left(e^{i\omega_1},e^{i\omega_2}\right)\right|^2,\quad Y\left(e^{i\omega_1},e^{i\omega_2}\right)=\sum_{n_1=0}^{N}\sum_{n_2=0}^{N}y\left(n_1,n_2\right)e^{-i\left(\omega_1 n_1+\omega_2 n_2\right)}.$$

(11)

Полученная оценка $\Phi_Y\left(e^{i\omega_1},e^{i\omega_2}\right)$ применяется для расчета АКФ выходного сигнала (12):

$$B_Y\left(m_1,m_2\right)=\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\Phi_Y\left(e^{i\omega_1},e^{i\omega_2}\right)e^{i\left(\omega_1 m_1+\omega_2 m_2\right)}d\omega_1 d\omega_2 \tag{12}$$

Энергетический спектр реального изображения имеет сильно изрезанный вид, а выражение (7) для энергетического спектра входного сигнала дает гладкую функцию, поэтому для повышения точности расчетов на основе выражения (9), требуется сглаживание периодограммы. В рамках данной работы это сглаживание реализуется умножением АКФ (12) на гауссовскую маску:

$$\tilde{B}_Y\left(m_1,m_2\right)=B_Y\left(m_1,m_2\right)\cdot W\left(m_1,m_2\right),\qquad W\left(m_1,m_2\right)=\begin{cases}\exp\left(-\dfrac{\left(m_1^2+m_2^2\right)}{2\left(Q/6\right)^2}\right), & |m_1|\le Q,|m_2|\le Q\\ 0, & \textit{иначе}\end{cases}, \tag{13}$$

где Q - размер окна, выбираемый так, чтобы охватить все значимые отсчеты АКФ. Далее через обратное преобразование Фурье вычисляется сглаженная оценка энергетического спектра $\tilde{\Phi}_Y\left(e^{i\omega_1},e^{i\omega_2}\right)$.

Оценка дисперсии шума

Энергетический спектр является убывающей функцией и убывание происходит тем быстрее, чем сильнее коррелированность входного сигнала и сглаживающие свойства искажающей системы. В реальных изображениях, получаемых в результате дистанционного зондирования Земли, коэффициент корреляции соседних отсчетов составляет не менее 0,9, и энергетический спектр успевает убыть в высокочастотной области до нуля в условиях отсутствия шума на изображении. Если шум присутствует, к энергетическому спектру добавляется константа, равная дисперсии шума. Поэтому вычисление дисперсии шума проводится путем усреднения оценки энергетического спектра зашумленного сигнала по области высоких частот.

Оценка параметров энергетического спектра входного сигнала

Энергетический спектр входного сигнала (7) имеет всего два параметра: дисперсия D_X и коэффициент корреляции ρ. Их можно оценить непосредственно по отсчетам АКФ выходного сигнала. Поскольку импульсная характеристика искажающей системы предполагается быстро убывающей, она искажает только небольшую область вокруг центра АКФ. Вне этой области влиянием импульсной характеристики на АКФ можно пренебречь, то есть принять, что

$$B_Y\left(m_1,m_2\right)\approx B_X\left(m_1,m_2\right)=D_X\rho^{\sqrt{m_1^2+m_2^2}}. \tag{14}$$

Учитывая экспоненциальный характер убывания АКФ, можно вычислить ряд значений ρ_i точечных оценок коэффициента корреляции, например, по горизонтальному сечению АКФ.

$$\rho_i=\frac{B_Y\left(0,i+1\right)}{B_Y\left(0,i\right)},\qquad 1\le i\le I. \tag{15}$$

Можно построить гистограмму значений ρ_i, и за оценку величины ρ взять наиболее вероятное значение. Заметим, что при больших значениях индекса i величины ρ_i начинают хаотически меняться, что вызвано статистической

неустойчивостью оценки малых значений АКФ выходного сигнала по периодограмме. Поэтому вычисления по формуле (15) следует проводить только в области больших значений АКФ, например, приняв предельное значение индекса I в формуле (15) как номер последнего отсчета АКФ для которого выполняется соотношение:

$$\frac{B_Y(0,0)}{B_Y(0,i)} > 0.5. \tag{16}$$

Дисперсия входного сигнала оценивается с использованием рассчитанного значения ρ:

$$D_X = \frac{B_Y(0,i')}{(\rho)^{i'}} \tag{17}$$

где i' – номер наиближайшего по абсолютному отклонению от значения ρ отсчета последовательности ρ_i.

Для повышения точности указанную процедуру оценки параметров можно проводить по нескольким сечениям АКФ, взятым в разных направлениях с последующим усреднением результатов.

Исследование точности оценки параметров модели АКФ и восстановления импульсной характеристики

Для исследования предложенного метода был проведен ряд вычислительных экспериментов. Было принято, что искажающая система имеет импульсную характеристику в виде двумерной гауссоиды:

$$h(k_1,k_2) = Ae^{-\frac{1}{2R^2}\left(k_1^2+k_2^2\right)}, \ |k_1| \le K, \ |k_2| \le K. \tag{18}$$

где R – радиус пятна размытия (использовались значения $R \in \{0{,}5;\ 1;\ 2\}$), A – коэффициент, обеспечивающий выполнение условия нормировки (4). Сигнал и белый шум имели нормальное распределение, отношение шум/сигнал (D_V/D_X) принималось равным нулю и 0,01. Входной сигнал моделировался в виде стационарного случайного поля с изотропной неразделимой АКФ и мозаичного поля. Эти поля имеют одинаковую АКФ вида (6). Размер использованных в эксперименте изображений составляет 4096×4096, коэффициент корреляции был равен 0,9.

В качестве показателя точности восстановления импульсной характеристики использовалось нормированная среднеквадратичная ошибка восстановления:

$$\xi = \frac{1}{h(0,0)}\sqrt{\frac{\sum_{k_1,k_2=-K}^{K}\left(\tilde{h}(k_1,k_2)-h(k_1,k_2)\right)^2}{(2K+1)^2}}. \tag{19}$$

В результате экспериментов установлено, что:

1) если искажающая система имеет узкую импульсную характеристику, то дисперсия шума оценивается с большой погрешностью, так как в спектре сигнала остаются существенные высокочастотные компоненты;

2) для искажающих систем с радиусом пятна размытия от 0,5 до 2 погрешность восстановления коэффициента корреляции и дисперсии не превосходит 10%;

3) для радиусов пятна размытия от 0,5 до 2 и отношении $D_V/D_X = 0.01$ при заданной корреляции и размерах изображения погрешность (19) восстановления импульсной характеристики не превышает 10%.

Заключение

Предлагаемый метод идентификации модели наблюдения продемонстрировал свою работоспособность. Достоинством метода является возможность сравнительно

быстрой обработки больших объемов данных и высокая точность восстановления в рассмотренных пределах искажений. Дальнейшее развитие метода предполагает обобщение для неизотропной модели АКФ и применение других способов оценки дисперсии и коэффициента корреляции входного сигнала.

Литература

1. Методы компьютерной обработки изображений. Под редакцией В.А. Сойфера / М.В. Гашников, Н.И. Глумов, Н.Ю.Ильясова, В.В. Мясников [и др.]; Под ред. В.А. Сойфера. – 2-е изд., испр. – М.: Физматлит, 2003. –784 с.

2. Сергеев Г.А., Янутш Д.А. Статистические методы исследования природных объектов. – Л.: Гидрометеоиздат, 1973. – 300 с.

3. Марпл-мл. С. Л. Цифровой спектральный анализ и его приложения.– М: МИР, 1990. – 584 с.

Разработка геоинформационной системы агропромышленного комплекса Самарской области для учета и контроля земель сельхозназначения

Н.С. Воробьева, Е.И. Тимбай
Институт систем обработки изображений РАН, Самара

Введение

В докладе рассматриваются задачи и принципы разработки и внедрения геоинформационной системы агропромышленного комплекса (далее - ГИС АПК), предназначенной для нужд Министерства сельского хозяйства и продовольствия Самарской области. Цель рассматриваемой работы – создание системы, предоставляющей средства ведения реестра и контроля сельскохозяйственных земель, находящихся в ведении Министерства и обеспечивающей мониторинг сельскохозяйственных полей на основе данных дистанционного зондирования Земли.

Процесс построения ГИС АПК имеет следующие особенности:

1) Система является распределенной – информация, вводимая в систему в территориальных управлениях, передается в центральный узел Министерства.

2) Небольшое количество персонала, обслуживающего систему, обуславливает разработку автоматизированных методов для ускорения анализа и обработки данных.

3) В Самаре находится собственная станции приема космической информации «УниСкан», обеспечивающая своевременность и оперативность получения ДДЗ, а также доступность достаточного количества снимков для проведения анализа.

Основные области применения рассматриваемой системы ГИС АПК: контроль реальных площадей посевов и выданных субсидий; оценка состояния культур; планирование и контроль севооборота; оценка земель сельхозназначения; выявление земель, неиспользуемых в течение длительного периода; снижение рисков и информационная поддержка при страховании урожая и оценке ущерба; контроль объектов сельскохозяйственного назначения.

Структура ГИС АПК

ГИС АПК состоит из двух подсистем: «Учет и контроль земель сельхозназначения» и «Мониторинг состояния сельхозугодий на основе ДДЗ».

Подсистема учета и контроля земель сельхозназначения предоставляет средства для ведения реестра сельскохозяйственных земель Самарской области. Реестр содержит графическую и семантическую информацию об объектах. Структура рассматриваемой подсистемы была разработана с использованием инструментальной геоинформационной системы «ИнГЕО».

Подсистема мониторинга состояния сельхозугодий на основе космических снимков предоставляет пользователю возможность создания карт реальных границ полей, оценки их характеристик по ДДЗ, проверки наличия сельскохозяйственных культур на заданном наборе полей, выявления расхождения между декларированными сведениями и реальными данными. Для осуществления мониторинга сельхозугодий используются снимки среднего разрешения (SPOT, Landsat) и низкого разрешения (Terra, Aqua). Данные среднего разрешения являются наиболее подходящими для решения задачи мониторинга, так как удовлетворяют и требованиям периодичности, и достаточного разрешения. Данные с космических аппаратов Terra и Aqua используются для ежедневных наблюдений, их недостаточное разрешение компенсируется высокой периодичностью. По результатам обнаружения посевов составляется сводная ведомость расхождения декларированных и реальных данных, а также список полей, подлежащих наземной проверке. Для всех полей, внесенных в реестр сельхозугодий

Самарской области, с использованием метода обнаружения посевов была проведена проверка наличия озимых культур на поле, а также с использованием ДДЗ были определены истинные границы полей. В результате были установлены расхождения между декларированными сведениями и реальными данными.

В связи с большим объемом информации, сжатыми сроками для ее интерпретации (2-4 недели на всю область), малой возможной численностью персонала, только визуальное обнаружение посевов оказывается неэффективным. Поэтому встает задача разработки автоматизированных методов обнаружения посевов. Для решения поставленной задачи были разработаны методы обнаружения и распознавания культур по космическим снимкам.

Методы обнаружения и распознавания культур по космическим снимкам

В системе ГИС АПК используются следующие методы для осуществления мониторинга состояния сельхозугодий на основе космических снимков: обнаружение озимых культур, распознавание типов культур на основе временных рядов вегетационных индексов.

Задачи классификации, решаемые в рамках ГИС АПК, обладают следующими особенностями: принятие решения на основании временной последовательности снимков; наличие исходных данных в виде векторных границ полей в ГИС; неоднородность посевов, а вследствие, изображения, внутри поля.

Метод обнаружения озимых культур основан на использовании снимков среднего разрешения и разработан для контроля полей, засеянных озимыми культурами. В качестве исходных данных используются снимки за осенний и весенний временные периоды. Результатом работы метода должно быть отображение *поле – класс* (есть озимые, неопределенно, нет озимых). Реализация метода делится на два этапа: построение классификатора полей и применение классификатора. Вектор признаков строится с использованием средних значений вегетационного индекса NDVI на поле для последовательности весенних и осенних снимков.

Метод распознавания типов культур на основе временных рядов вегетационных индексов основан на использовании последовательности космических снимков низкого разрешения и позволяет принять решение о принадлежности рассматриваемого объекта-поля к одному из следующих классов: озимые культуры, яровые культуры, технические культуры, однолетние/многолетние травы, неиспользуемые земли, пары, не определен. Классификация осуществляется на основе анализа признаков временного ряда вегетационного индекса. Выделяют следующие признаки: начало цикла, длительность, максимум временного ряда NDVI, площадь под графиком.

Основные результаты проведенной работы следующие:

1) Разработана начальная структура региональной ГИС АПК.

2) Сформирована и привязана к картоснове пространственная база данных границ посевных площадей.

3) Разработан метод обнаружения культур по космическим снимкам, позволяющий в автоматическом режиме достаточно эффективно определять наличие озимых культур на поле. Работа метода подтверждена результатами наземных проверок.

4) Проводятся исследовательские работы по улучшению метода распознавания типов культур на основе временных рядов вегетационных индексов.

5) Проводятся научно-исследовательские работы по оценке состояния сельхозкультур и прогнозирования урожайности на основе совместного использования снимков низкого, среднего разрешения и наземных наблюдений.

6) Министерством сельского хозяйства Самарской области принято решение в 2010 году внедрить систему для всех 27 районов Самарской области.

Детектирование бликов в системах видеонаблюдения с выделением последовательности опорных кадров

Е.Ю. Минаев, А.В. Никоноров, В.А. Фурсов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

1. Введение и постановка задачи

В связи с угрозой террористических атак стала актуальной проблема обнаружения оставленных подозрительных предметов. В существующих системах видеонаблюдения для решения задачи обнаружения небольших предметов, оставленных в зоне наблюдения, присутствует детектор объектов. Зона наблюдения, как правило, подразделяется на зону стоянки и проезжую часть, и детектор объектов настроен на обнаружение предметов в зоне стоянки. Стекла машин, отражающее покрытие номерных знаков в солнечный день вызывают появление нестационарных бликов различной яркости и меняющейся формы, что увеличивает частоту ложных срабатываний детектора.

Для устранения ложных срабатываний необходимо распознать в появляющихся объектах блики, определить их контур и исключить из зоны наблюдения на время их существования. Будем в дальнейшем называть алгоритм устранения бликов *детектором бликов*, действующим параллельно с основным детектором объектов. На рисунке 1 приведен кадр с камеры видеонаблюдения с участком улицы состоящей из стояночной области и проезжей части, задача детектирования объектов и, соответственно, бликов актуальна для стояночного участка.

Блики должны быть локализованы и отделены от объектов не являющихся бликами. Так, например, на рисунке 1 яркий блик должен детектироваться, а яркая крыша автомобиля – нет. Таким образом, можно сформулировать следующие требования к бликам.

1. Детектирование бликов имеет смысл на неподвижных объектах, так, например, блик на лобовом стекле движущегося автомобиля не полежит детектированию.
2. Высокая абсолютная яркость блика.
3. Высокая яркость относительно окружения.
4. Площадь блика находится внутри некоторых пределов.
5. «Рваная» или «звездообразная» форма блика.

Аналогичная задача идентификации множественных бликов на изображении решалась в работе [7], в качестве идентифицирующего алгоритма применяется адаптивный оконный алгоритм с невысокой вычислительной сложностью. Однако требование 5 требует дополнительного анализа формы объекта, для того, чтобы провести разделение объектов и бликов. Алгоритм из [7] не позволяет провести такой анализ.

2. Метод активных контуров с анализом формы контура

Для обнаружения бликов используется метод активных контуров[1], как один из наиболее точных методов выделения замкнутых контуров на растровом изображении. Модели активного контура имеют ряд преимуществ по сравнению с другими методами сегментации [1], [3].



Рис. 1. Исходное изображение



Рис. 2. Изображение после применения активного контура

Впервые метод активного контура (далее - АК) был предложен в [5]. Согласно [5] активный контур определяется как параметризованная кривая на плоскости, которая движется к границе объекта и при этом минимизирует функционал энергии внутренних и внешних сил, действующих на кривую. Внутренние силы зависят от первой и второй производных радиуса-вектора точки на кривой r[x(s),y(s)] по параметру s и определяют форму кривой, а внешние силы характеризуются полутоновым рельефом изображения и заставляют кривую смещаться в направлении границ объекта. Минимум функционала достигается на границе объекта. Недостатками метода являются необходимость подбора весовых коэффициентов функционала и задания начального приближения контура в достаточно малой окрестности границы выделяемого объекта [3].

Для обнаружения бликов на первом этапе обработки из исходного изображения методом АК выделяется набор контуров, из которых будут выбираться блики. Для получения затравочного контура [1] используется пороговая обработка, в результате которой составляется функция $\phi_0(x,y)$, такая, что $\phi_0 > 0$ внутри контура, $\phi_0 < 0$ снаружи контура. Эволюцию затравочного контура определяет следующее уравнение:

$$\frac{\partial \phi}{\partial t} = \delta_\varepsilon(\phi)\left[\mu \cdot div\left(\frac{\nabla \phi}{|\nabla \phi|}\right) - \nu - \lambda_1(u_0 - c_1)^2 + \lambda_2(u_0 - c_2)^2\right] = 0, \quad (1)$$

$$\phi(0,x,y) = \phi_0(x,y),$$

где $u_0(x,y)$ - функция яркости исходного изображения , $\mu, \nu, \lambda_1, \lambda_2$ - параметры АК,

$$\begin{cases} c_1(\phi(x,y)) = average(u_0(x,y)), & \phi(x,y)0 \\ c_2(\phi(x,y)) = average(u_0(x,y)), & \phi(x,y) < 0 \end{cases}.$$

Для исходной задачи параметры λ_1, λ_2, характеризующие действие внутренних и внешних сил на контур, определены как $\lambda_1 = 1$, $\lambda_2 = 35$, чтобы внешними силами отделить контуры связанные 1-2 пикселями. Уравнение эволюции контура решатся численно по следующей схеме:

$$\frac{\phi_{i,j}^{n+1} - \phi_{i,j}^{n}}{\Delta t} = \delta_h(\phi_{i,j}^{n})\left[\frac{\mu}{h^2}\Delta_-^x\left(\frac{\Delta_+^x \phi_{i,j}^{n+1}}{\sqrt{(\Delta_+^x \phi_{i,j}^{n})^2/(h^2) + (\phi_{i,j+1}^{n} - \phi_{i,j-1}^{n})^2/(2h^2)}}\right) + \right.$$

$$+\frac{\mu}{h^2}\Delta_-^y\left(\frac{\Delta_+^x \phi_{i,j}^{n+1}}{\sqrt{(\Delta_+^y \phi_{i,j}^{n})^2/(h^2) + (\phi_{i+1,j}^{n} - \phi_{i-1,j}^{n})^2/(2h^2)}}\right) \quad . \quad (2)$$

$$\left. -\nu - \lambda_1(u_{0,i,j} - c_1(\phi^n))^2 + \lambda_2(u_{0,i,j} - c_2(\phi^n))^2\right]$$

Решение уравнения (2) итерационным методом [2] позволяет точно обозначить контура соседних бликов в случаях касания, что не всегда позволяют сделать классические градиентные методы активного контура [3].

На рисунке 2 приведен набор контуров полученных в результате применения метода АК для изображения с рисунка 1. Предварительно, для повышения производительности, из рассмотрения исключена часть исходного кадра, не относящаяся к стояночной зоне, согласно требованию 1.

Из набора полученных в результате работы АК областей исключаются области с недостаточной относительной яркостью и области неудовлетворяющие характерному диапазону площадей бликов согласно требованиям 2, 3 и 4. Результат применения указанных требований к контурам с рисунка 2 показан на рисунке 3.

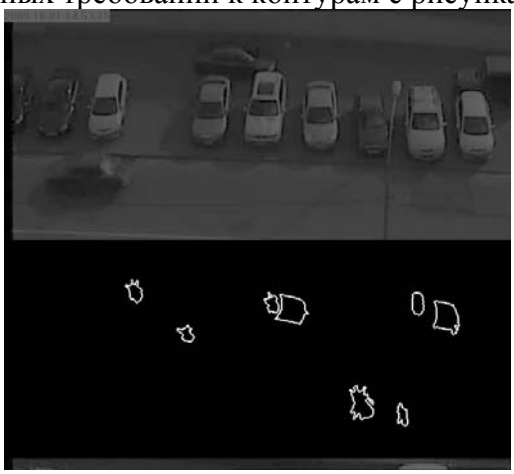


Рис. 3. Отсеивание контуров по критерию площади и яркости



Рис. 4. Изображение с выделенными бликами

Для того, чтобы отсеять яркие элементы изображения, не являющиеся бликами применяется дополнительный анализ формы контура, согласно требованию 5. Производится отсев оставшихся контуров по значению критерия, описывающего форму контура:

$$\frac{S}{l^2} > a, \tag{3}$$

где S – площадь фигуры внутри контура, l – длина контура, a – некоторое пороговое значение.

По критерию (3), с учетом того, что маленькие по площади блики имеют круглую форму, а большие звездчатую, производится отсев контуров, не являющихся бликами. В результате получается окончательный набор контуров бликов. Пример окончательного выделения бликов приведен на рисунке 4.

3. Выделение последовательности опорных кадров с использованием функции контраста

В системе видеонаблюдения работающей в реальном времени требуется скорость обработки порядка 400 fps (кадров в секунду). Такие требования обусловлены тем, что, как правило, один сервер видеонаблюдения обрабатывает поток с 16 камер в обычном разрешении с частотой 25 fps. При таких к требованиях к производительности, обработка каждого кадра методом АК не представляется возможной. Для получения требуемых результатов за приемлемое время предлагается использовать тот факт, что контурная информация меняется не на каждом кадре, следовательно, обработка каждого кадра методом АК не требуется. Следовательно, производительность системы может быть повышена за счет прореживания последовательности кадров, обрабатываемых детектором бликов. В общем потоке кадров с одной камеры необходимо выделить последовательность кадров, на которых

происходит появление новых контурных объектов. Такие кадры далее будем называть *опорными кадрами*. Опорные кадры обрабатываются методом АК для максимально точного обнаружения контурных объектов.

Для выделения последовательности опорных кадров целесообразно использовать менее точные, чем МАК, но более производительные алгоритмы. В настоящей работе предлагается использовать алгоритм выделения контуров, основанный на так называемой функции контраста [4]. Также проводится сравнение эффективности предлагаемого метода с известным локальным методом выделения контуров – методом Превитт [6].

Для локализации перепада яркости фрагмент реального изображения можно аппроксимировать идеальным перепадом, меняя его параметры. Если такую аппроксимацию удаётся сделать достаточно точно в данном месте изображения, то считается, что в данном месте имеется перепад с найденными параметрами[6]. Введем следующую аппроксимацию перепада яркости, рисунок 5:

$$L(x)=\begin{cases} b, x\in[x_0-s, x_0], \\ b+h, x\in[x_0+a, x_0+s+a] \end{cases}, \tag{4}$$

где s – количество учитываемых отсчетов, слева и справа от перепада яркости,a – ширина перепада.

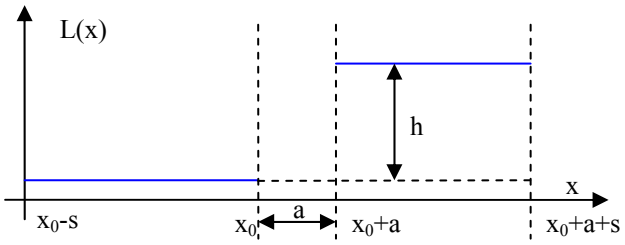


Рис. 5. Модель идеального перепада яркости

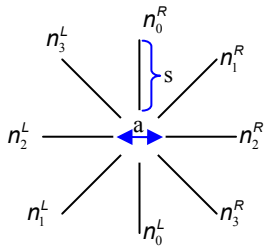


Рис. 6. Направления расчета двумерной функции контраста

В рамках данной модели, для изображения вводится в рассмотрение так называемая *функция контраста* [4]:

$$f(s,a)=\frac{1}{s}\sum_{i=s+a}^{2s+a}\breve{L}_k(x_i),-\frac{1}{s}\sum_{i=1}^{s}\hat{L}_k(x_i). \tag{5}$$

Двумерная модификация функции контраста [4]:

$$F(s=s_0,a=a_0,x,y)=\frac{1}{s}\sum_{n_j}\left(\sum_{x,y\in n_j^R}L(x,y)-\sum_{x,y\in n_j^L}L(x,y)\right). \tag{6}$$

Вычисление проводится по четырем направлениям, как показано на рисунке 6.

Полученная функция контраста $F(x,y)$ отображает структуру контуров изображений, значение функции указывает величину перепада яркости в данной точке. Данный метод контрастирования обеспечивает сглаживание шумов, параметр s определяет степень сглаживания. Отметим, что использование функции (6) для обнаружения контуров в двумерном случае является развитием идеи впервые предложенной А. Розенфельдом в работе [8].

Тестовую последовательность из 1000 кадров удалось сократить до 233 кадров с использованием функции контраста и до 586 кадров с использованием окна Превитт, что подтверждает эффективность использование функции контраста для прореживания видеоданных. При использовании функции контраста количество запусков алгоритма

АК сокращается на 77%, что позволяет сократить количество опорных кадров в 2,4 раза, соответственно скорость обработки видеопотока возрастает более чем вдвое.

4. Заключение

В работе предложен алгоритм детектирования бликов в видеопотоке на основе модифицированного метода активных контуров и критерия анализа формы контура. Данный алгоритм позволяет избавить от ложных срабатываний основной детектор объектов. Для повышения производительности алгоритма используется прореживание последовательности кадров с использованием двумерной функции контраста.

Работа выполнена при поддержке российско-американской программы «Фундаментальные исследования и высшее образование» и РФФИ (проект № 09-07-00269-а).

Литература

1. Chan, T.F.; Vese, L.A., Active contours without edges, Image Processing, IEEE Transactions on Volume 10, Issue 2, Date: Feb 2001, pp: 266-277

2. Aubert G., Vese L., A variational method in image recovery, *SIAM J. Numer. Anal.*, vol. 34, no. 5, pp. 1948–1979, 1997.

3. Мурашов Д.М., Метод автоматизированной сегментации изображений цитологических препаратов на основе модели активного контура, Труды МФТИ, 2009, Том 1, № 1, стр. 80-89.

4. Bibikov S.A., Fursov V.A., Nikonorov A.V., Shadow Artifacts Correction on Fine Art Reproductions, 3rd International Workshop on Image Mining Theory and Applications in conjunction with VISIGRAPP 2010, Angers, France, May 2010. – pp. 3-12.

5. Kass M., Witkin A., Terzopoulos D., Snakes: Active contour models // Int. J. Computer Vision. — 1987. — N. 1. — P. 321–331.

6. Прэтт У., Цифровая обработка изображений – М.: МИР, 1982 – 791 с.

7. Фурсов В.А., Никоноров А.В., Распределенный алгоритм коррекции точечных артефактов на цветных изображениях, Труды четвертой международной конференции «Параллельные вычисления и задачи управления» (PACO’2008), 1087.

8. Rosenfeld A., Thurston M., Edge and Curve Detection for Visual Scene Analysis, IEEE, Trans. Computers, pp-20, 5, 562-569 (May 1971).

Алгоритм определения периодов повышенной сейсмической опасности по «шаблонам малой динамики»

А.Н. Коварцев [1], О.А. Никонорова [2]
[1] *Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*
[2] *Оренбургский государственный университет, Оренбург*

Введение

Для повышения вероятности более точного определения времени землетрясения в комплексе с долгосрочными, среднесрочными и краткосрочными предвестниками необходимо использовать и те природные процессы, которые могут опосредованно инициировать землетрясение, если тектонические напряжения достигли критического уровня. В качестве таковых могут быть рассмотрены процессы изменения приливных сил за определенный период времени и, согласно динамической модели определения периодов повышенной сейсмической опасности можно составить базу данных «шаблонов землетрясений». Каждый из шаблонов – это фрагменты функций, описывающие в окрестности времени сильного землетрясения процесс изменения первых двух главных компонент приливных факторов, являющихся линейной комбинацией исходных признаков (вертикальная компонента приливной силы тяжести и ее производные до четвертого порядка). Днем повышенной сейсмической опасности будет тот день от начала года, в который развитие приливного процесса происходит по сценарию, схожему с шаблоном из базы данных. При реализации этого подхода возникают трудности формализации подбора «шаблонов землетрясений» для обучающей выборки в рассматриваемой географической области. В связи с этим интересным является составление рейтинга «шаблонов землетрясений», участвующих в определении периодов повышенной сейсмической опасности по динамической модели.

Основная часть

Рассмотрим сейсмический регион, территория которого разбивается на области исследования [1]. Например, для южной части Курильских островов рассмотрим сектора размером $1^0 \times 1^0$. Составим базу данных «шаблонов землетрясений» для одного из выделенных секторов на базе основного каталога землетрясений за период с 1983 по 2003 годы по методике, описанной в работе [2].

Для полученной базы данных можно составить рейтинг «шаблонов землетрясений» по некоторому критерию $K_{эфф}$, который рассчитывался из следующих предположений. Первоначально, учитывая количество использований образцов, были подсчитаны коэффициенты использования

$$k_i = \frac{n_i}{\sum n_i},$$

где n_i - количество использований шаблонов с номером *i*.

Далее, было определена доля каждой из групп шаблонов, участвующих в 1 кратном, 2 кратном, …, 6 кратных прогнозах

$$s_j = \sum_{\left(\substack{i=1 \\ n_i = j}\right)}^{N} k_i .$$

Критерий $K_{эфф}$ подсчитывался по формуле $\forall i\,(n_i = j) \quad K_{эфф}(i) = s_j k_i$.

На рисунке 1 представлено распределение долей каждой из групп шаблонов, участвующих в 1 кратном, 2 кратном, …, 6 кратных прогнозах.

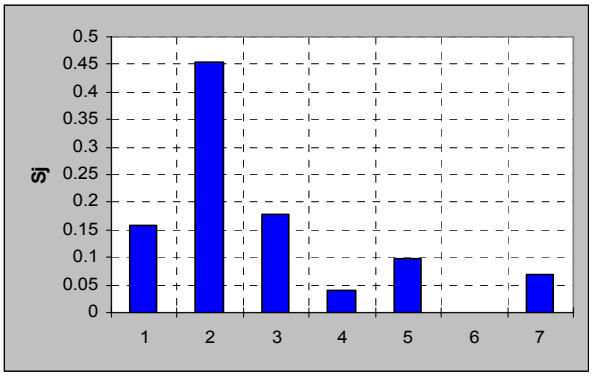


Рис.1. - Распределение долей участия «шаблонов землетрясений»

Наиболее существенный вклад в процесс определения периодов повышенной сейсмической опасности вносят шаблоны, используемые 2 и 3 раза. Рассмотрим фазовые портреты «шаблонов землетрясений», имеющих наибольший рейтинг, представленные на рисунке 2 в первой строке (использовались два раза) и во второй строке шаблоны, которые участвовали в прогнозе три раза.

Несложно заметить, что высокий рейтинг имеют шаблоны с небольшой амплитудой размаха фазового портрета в координатах главных компонент и наоборот образцы, сработавшие один раз, имеют большой размах фазовой кривой. То есть это позволяет предположить, что при минимальных суточных вариациях приливных сил, развитие геодинамического процесса усиливается, и, наоборот, при больших амплитудах ослабляется.

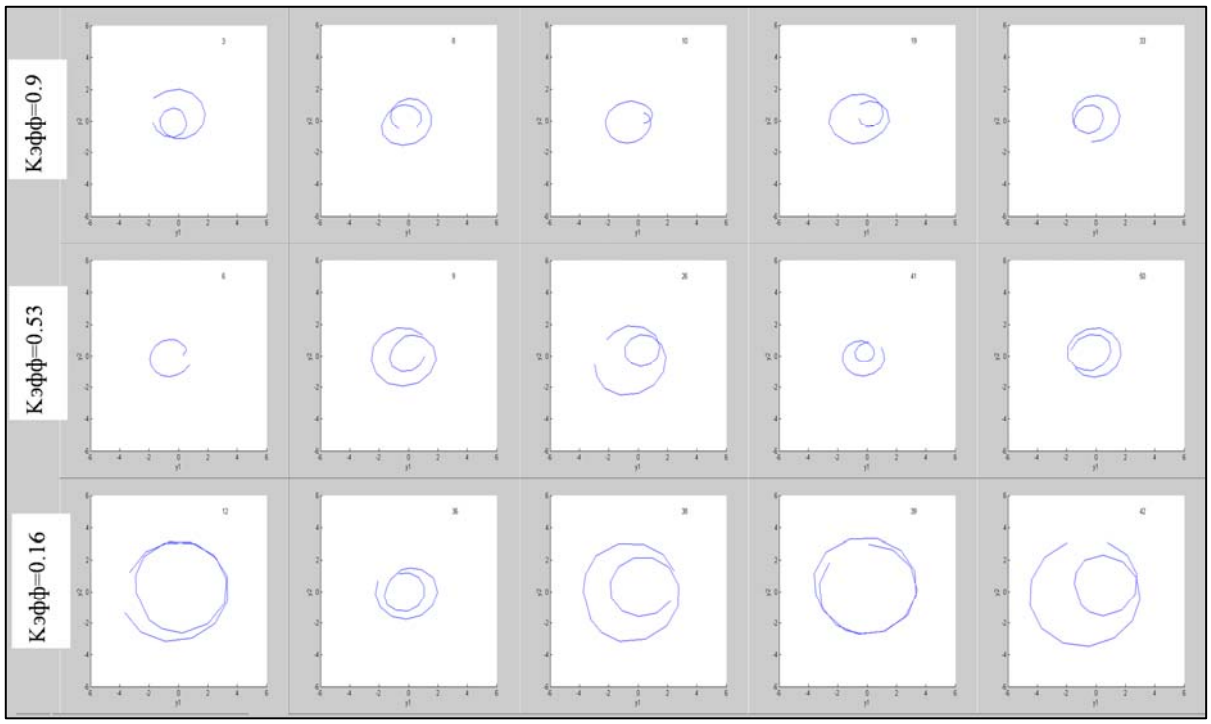


Рис.2. – Визуализация образцов по рейтингу

Тогда для исследуемого года можно рассчитать значения вектора $X = \left(x, \dot{x}, \ddot{x}, \dddot{x}, \ddddot{x}\right)'$ на каждый час и составить матрицу признаков $F_{new} = \left(x_i, \dot{x}_i, \ddot{x}_i, \dddot{x}_i, \ddddot{x}_i\right)$, где $i = \overline{1, K}$, K – количество часов года. После центрирования

и нормирования можно провести факторный анализ, и среди набора главных факторов, рассчитанных на каждый час исследуемого года выделить так называемые «шаблоны малой динамики» приливных сил, т.е. такие участки, которые принадлежат «минимальному окну» их вариации $\begin{cases} y_{min1} \le y_1^{(i)} \le y_{max1}, \\ y_{min2} \le y_2^{(i)} \le y_{max2}, \end{cases} \quad i = \overline{1,24}$.

Методами интервального корреляционного анализа исследуем близость статистической связи между наблюдаемами землетрясениями и рассчитанными «шаблонами малой динамики».

Пусть $\breve{x}(t_s)$ - бинарный процесс прогнозируемых моментов землетрясений, в котором нулями обозначается отсутствие землетрясения, 1 – прогноз землетрясения на момент времени t_s. $\breve{y}(t_i)$ - аналогичный процесс для реальных землетрясений ($\breve{x}(t_s), \breve{y}(t_i) \in \{0,1\}$). Тогда оценкой интервальной корреляционной функции, характеризующей распределение отсчетов в потоке, находящихся на временном интервале $J \cdot \Delta$ можно оценить по формуле

$$C_{\breve{x}\breve{y}}(J) = \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{L} \delta_{i,s},$$

где Δ - ширина временного интервала ($\Delta \approx 7$ дней), $\delta_{i,s}$ индикатор состояния, который однозначно определяет отсчет $\breve{x}(t_s)$, отстоящий от отсчета $\breve{y}(t_i)$ на временном интервале $J \cdot \Delta$ [3]

$$\delta_{i,s} = \begin{cases} 1, & \textit{если} \quad ceil[\dfrac{t_s^{(\breve{y})} - t_i^{(\breve{x})}}{\Delta} + 0.5] = J. \\ 0, & \textit{иначе}. \end{cases}$$

Усреднение на серию экспериментов, используя статистику разных лет, можно реализовать по формуле

$$C_{\breve{x}\breve{y}}(J) = \frac{1}{\sum_{k=1}^{n} N_k} \sum_{k=1}^{n} \left(\sum_{i_k=1}^{N_k} \sum_{s_k=1}^{L_k} \delta_{i_k,s_k}^{(k)} \right).$$

Усредненная оценка интервальной корреляционной функции за период с 1983 по 2003 годы для нескольких секторов южной части Курильской гряды показана на рисунке 3.

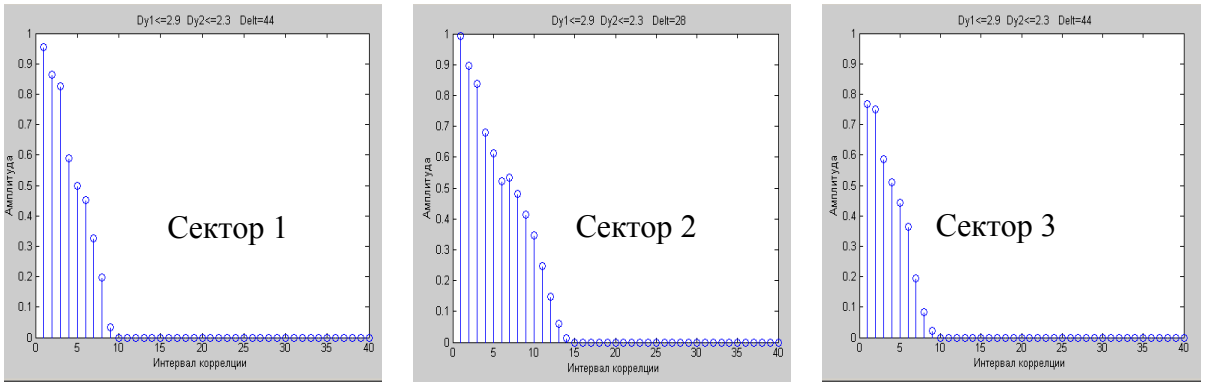


Рис.3. – Оценка корреляционной функции для тестовой выборки

Анализ статистической обработки временных рядов на тестовой выборке для

1983-2003 годов указывает на существенную временную корреляционную связь между землетрясениями и днями повышенной сейсмической опасности, полученными по методу выявления шаблонов малой динамики. На основе которой несложно построить алгоритм определения зон повышенной сейсмической опасности, без использования базы данных «образцов землетрясений», который схематически изображен на рис. 4.

Основными шагами данного алгоритма являются:

1) На каждый час прогнозируемого года рассчитать вектор приливных сил и составить матрицу признаков F.

2) Вычислить матрицу $\widetilde{F}$ путем центрирования и нормирования матрицы F.

3) Рассчитать для матрицы $\widetilde{F}$ ковариационную матрицу S.

4) Для матрицы S вычислить матрицу факторных нагрузок и выделить первые две главные компоненты.

5) Получить новую матрицу признаков Y, пересчитав матрицу $\widetilde{F}$ в главных компонентах.

6) На каждые 24 часа рассматриваемого года определить вариации главных компонент приливных сил, геометрически представляющие собой диаметры шаблонов.

7) Проверить условие принадлежности паттерна окну с задаваемыми параметрами $\varepsilon_1, \varepsilon_2$. Если оно истинно, то рассматриваемый участок является «шаблоном малой динамики» и соответственно день от начала рассматриваемого года, для которого был определен этот «шаблон» является днем повышенной сейсмической опасности.

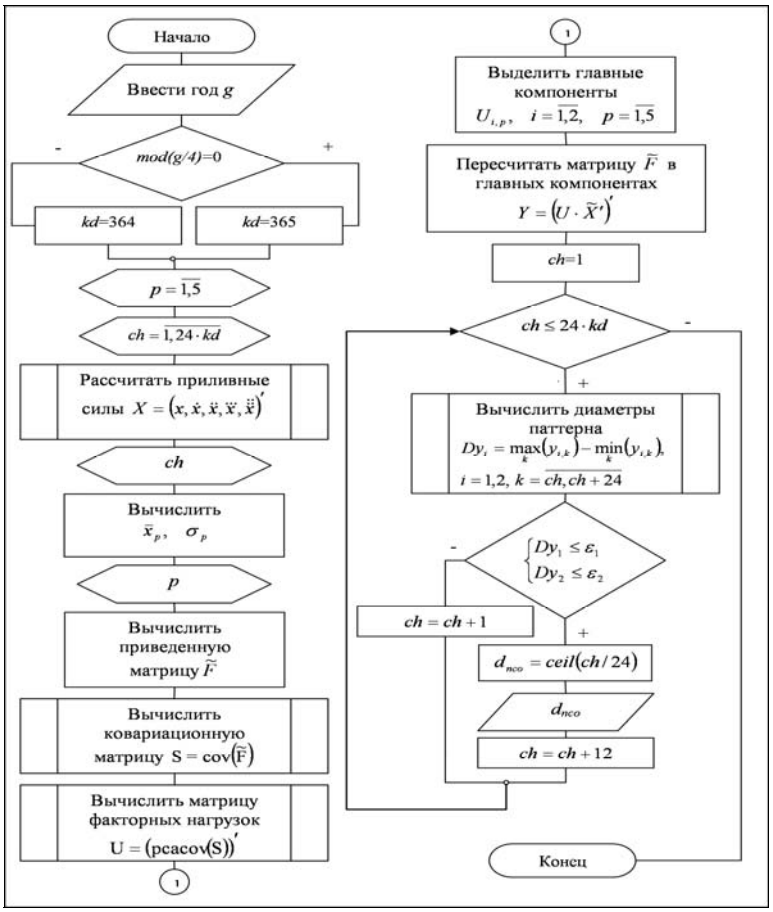


Рис.4. - Схема алгоритма определения ППСО по шаблонам малой динамики

Кроме того, по данным о сильных землетрясениях мира, произошедших в первом квартале 2010 года, для секторов с центром координат в эпицентрах этих

событий построена визуализация размещения «шаблонов малой динамики» главных компонент приливных сил в зависимости от Dy_1 (Dy_2). Получены обнадеживающие результаты, а именно времена землетрясений на Гаити, в Чили, Тайване и Турции принадлежат интервалам времени тревог, рассчитанным по методу «шаблонов малой динамики» приливных сил.

Заключение

1) Проведена оценка рейтинга «образцов землетрясений» по критерию количества срабатывания их при определении периодов повышенной сейсмической опасности с 1983 по 2003 годы, которая показала, что фазовые портреты «образцов землетрясений» с наибольшим рейтингом имеют небольшую амплитуду размаха в координатах главных компонент.

2) Предложен метод определения «шаблонов малой динамики», который позволяет избавиться от необходимости ведения базы данных «образцов землетрясений» выбранных для многих лет в исследуемом секторе, что снимает проблему отбора образцов в динамической модели определения периодов повышенной сейсмической опасности.

3) Оценка результатов по определению периодов повышенной сейсмической опасности на основе метода выделения «шаблонов малой динамики» приливных сил для вариаций размеров окна с позиций корреляционного анализа показала, что между реальными землетрясениями и «шаблонами малой динамики» приливных сил существует заметная статистическая связь, на основе которой построен алгоритм определения зон повышенной сейсмической опасности.

Литература

1. Алгоритмы прогноза землетрясений. – М.:ГЕОС, 2006. – 292 с. – (Вычислительная сейсмология; Вып.37)

2. Коварцев А.Н., Никонорова О.А. Динамическая модель определения периодов повышенной сейсмической опасности // Научное творчество XXI века - 2010: Материалы II Всероссийской научной конференции. – Красноярск: ПГТА, 2010. - С. 79 – 83.

3. Прохоров С.А. Прикладной анализ случайных процессов. / под ред. С.А.Прохоров, Самара: СНЦ РАН. – 2007. – 582 с.

Метод идентификации минных полей по аэрокосмическим снимкам

Н.С. Булдаков
Саратовский государственный технический университет

Введение

Стремительное развитие космических и информационных технологий открывает новые сферы использования космического зондирования в мирных целях. Одной из жизненно важных задач, где дистанционное зондирование может дать значительный эффект, является идентификация минных полей [1]. Современные информационные технологии позволяют автоматизировать процесс обработки снимков и существенно повысить вероятность обнаружения.

Предлагаемый в настоящей работе метод реализует комбинированный алгоритм обнаружения мин на снимках. На первом этапе обработки используется технология клеточных автоматов [2]. Она позволяет выполнить контрастирование, сегментацию и формирование бинарного изображения в виде системы точечных элементов. На втором этапе осуществляется фактическое распознавание минного поля.

Особенности задачи распознавания образов минных полей

Изображения минных полей могут быть получены методом зондирования поверхности земли в различных диапазонах длин волн. При этом фотосъемка может вестись с различных носителей: космический аппарат, самолет, дистанционно пилотируемый летательный аппарат. Однако, несмотря на различные методы получения изображений минных полей, основные этапы обработки снимка остаются традиционными и включают оцифровку изображения исследуемого участка местности и ввод его в ЭВМ; предварительную обработку с целью выделения малоразмерных объектов одинаковой природы (предположительно мин), формирование бинарного изображения; распознавание мин и минных полей на исследуемом участке местности.

В данной работе предполагается, что изображение местности с установленными минами уже имеется, поэтому основное внимание уделено разработке алгоритмов предварительной обработке снимков и распознаванию образов минных полей. На рис 1 приведен пример изображения минного поля.

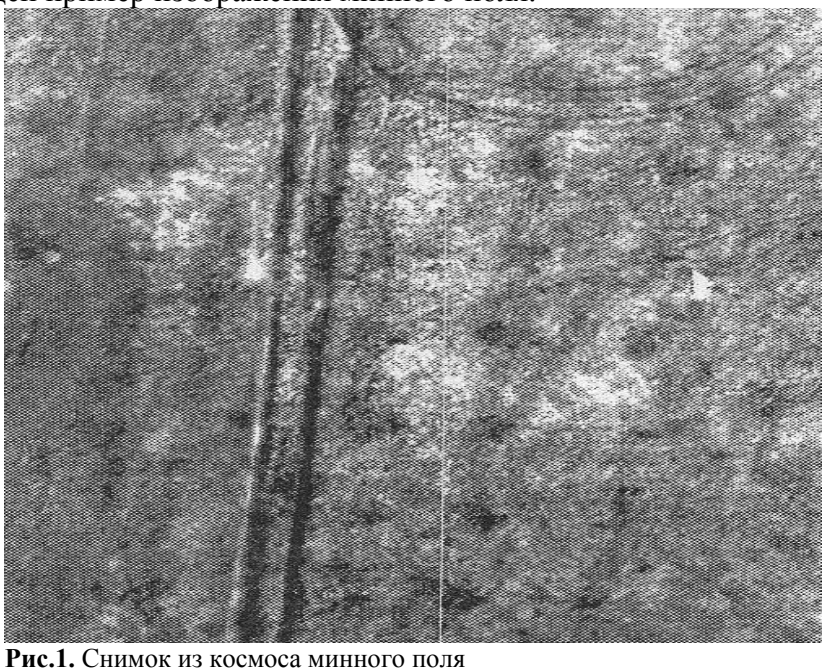


Рис.1. Снимок из космоса минного поля

Заглубленные в грунт мины проявляются в виде темных пятен неправильной формы, которые представляют собой вариации отражающей способности подстилающей поверхности. С помощью сканера данное изображение было преобразовано в цифровую форму и введено в память ЭВМ.

Математически цифровая форма изображения имеет вид матрицы $N \times M$, каждый элемент которой характеризуется величиной яркости $S_{x_i y_j}$ и координатами x_i и y_j ($i = 1, N$; $j = 1, M$). Предполагается, что с учетом масштаба изображения, средний диаметр изображения мины занимает 12 пикселей. Следует отметить, что этот диаметр включает размер самой мины и измененный вокруг нее подстилающий слой.

Особенностью данной задачи является то, что необходимо выделить малоразмерные объекты, изображения которых зависят от типа мин, способов их установки, условий наблюдения и типа регистрирующих датчиков. Другая особенность заключается в том, что даже в пределах одного снимка объекты могут находиться на различных подстилающих поверхностях. Это хорошо видно на рис.1, где в пределах одного минного поля подстилающая поверхность имеет существенно разную яркость. Кроме того, решение задачи распознавания затрудняется из-за относительно большой переходной области, окружающей объекты. Постепенно переходя в фон, эти области «смазывают» интересующие объекты.

Предварительная обработка снимков

Предварительный анализ аэрокосмического снимка минного поля показывает, что имеется много мелких разбросанных случайным образом элементов изображения, которые создают помехи для идентификации мин. Дополнительные сложности возникают из-за того, что яркости этих элементов существенно отличаются друг от друга. Поэтому, прежде чем сегментировать изображение, основываясь на яркости точек, необходимо получить сглаженную версию с более однородной внутренней структурой. Причем желательно, чтобы размеры большинства элементов изображения были не меньше фактических размеров мин. Поэтому необходимо провести операцию «размывания яркости» или низкочастотную фильтрацию изображения (рис. 2).

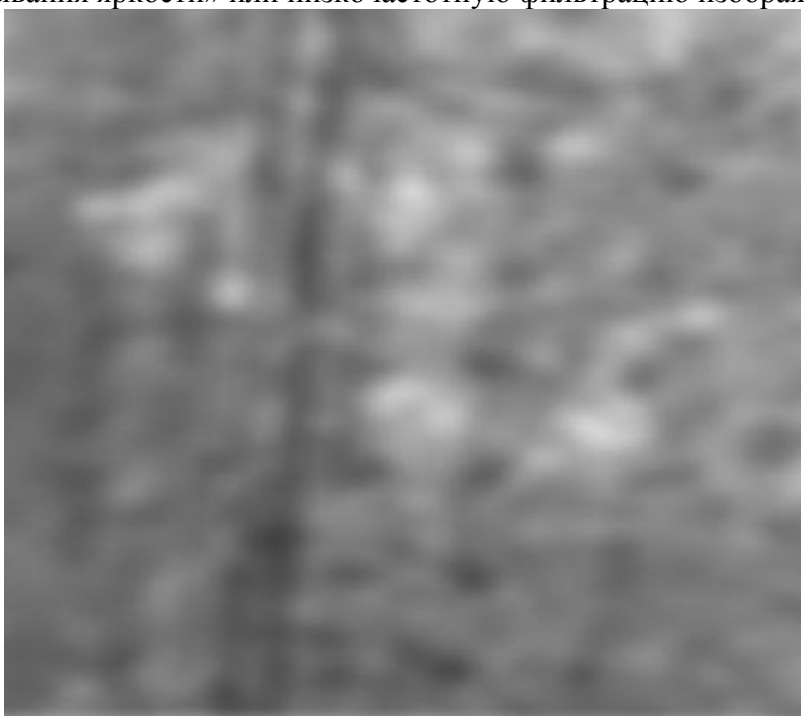


Рис.2. Результат работы фильтра Гаусса

Низкочастотная фильтрация позволила устранить мелкие помехи. Однако и контуры искомых объектов стали менее заметны. Поэтому следующая операция заключается в контрастировании изображения.

Традиционно контрастирование выполняется с помощью дифференциальных фильтров, работа которых основана на локальной разности яркостей. Но в данном случае они малоэффективны из-за отмеченных выше особенностей. Их применение приводит к чрезмерной детализации изображения, которая разрушает искомые объекты. Поэтому был применен алгоритм интегрального контрастирования.

Получение бинарной формы изображения является весьма ответственной задачей, поскольку позволяет выделить нужную информацию, устранив излишнюю. Ошибки в этом алгоритме могут привести к совершенно противоположному результату. В работе был разработан и исследован алгоритм стягивания с последующим компарированием. Этот алгоритм реализуется на трех последовательных шагах. На первом шаге осуществляется операция стягивания. Обычно эта стандартная операция выполняется над бинарными образами. В нашем случае эта операция используется особым образом. Вначале вычисляется коэффициент охвата R_{xy} точки с координатами (x, y). Здесь $R_{xy} = Int\left(\frac{1}{S^t_{xy}} \sum_{p=x-1}^{x+1} \sum_{q=y-1}^{y+1} S^t_{pq}\right)$ является целым числом и отношением суммарной яркости пикселей, окружающих данную точку, к яркости самой точки. В алгоритме было установлено правило: если $R_{xy} \geq 6$, то S^t_{xy} принимает значение яркости равное 255. Эта процедура позволила стянуть объекты, уменьшив разброс в яркости (Рис. 3).

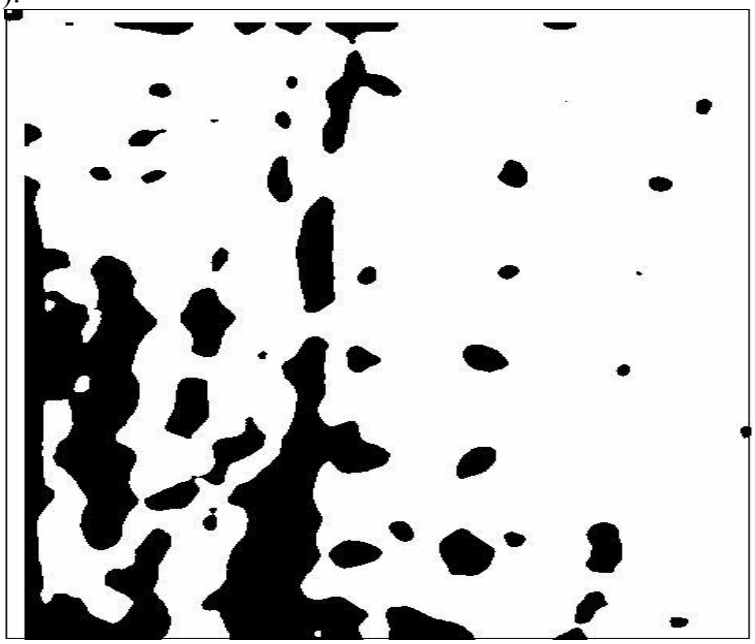


Рис. 3. Операция «Стягивание», после устранения помех

Однако при этом возникли высокочастотные помехи. Поэтому на следующем шаге было использовано равномерное сглаживание со следующими параметрами L1 = L2 = K1 = K2 =2. Эта операция позволяет устранить мелкие детали (точки и штрихи) изображения, которые возникли на предыдущем шаге. На третьем шаге производится приведение изображения к бинарной форме. Яркость текущего пикселя сравнивается с пороговым значением. Величина порога в данном случае была выбрана экспериментально и равна 180. Яркости со значением, меньшим 180, присваивалось нулевое значение, а яркости, большей или равной 180, - значение 255 (логическая единица).

Изображения мин имеют замкнутую форму, сравнительно небольшие размеры. Эти изображения не соприкасаются друг с другом. Поэтому логично на следующем этапе обработки выделить замкнутые области (кластеры), которые не соприкасаются между собой.

Предложенный алгоритм сначала вычисляет число соприкосновений по формуле $CNUM_{xy} = \sum_{p=x-1}^{x+1} \sum_{q=y-1}^{y+1} S_{pq}^{t}$. Если точка с координатами (x, y) имеет нулевое значение (данный пиксель имеет нулевую яркость) и имеется два соседних элемента также с нулевой яркостью, то CNUM будет равно 3. В этом случае две соседние точки включаются в кластер исходной точки. Здесь реализуется метод «заражение клетки», когда точка (клетка) заражает (присваивает номер своего кластера) всем соседям, с которыми она контактирует. Если точка принадлежит сразу нескольким кластерам, то последние объединяются. После того, как будут определены все кластеры, находятся максимальные и минимальные значения координат для каждого кластера. Эти значения позволяют судить о размере и форме кластера. Кластеры, имеющие форму и размеры, существенно отличающиеся от аналогичных характеристик искомых объектов, удаляются.

Поскольку форма и размер оставшихся элементов изображения не несут полезной информации, целью данной операции является получение одинаковых размеров кластеров. В результате выполнения нормализации размеров маленькие кластеры подросли, а большие ужались. Результат представлен на рис. 4.

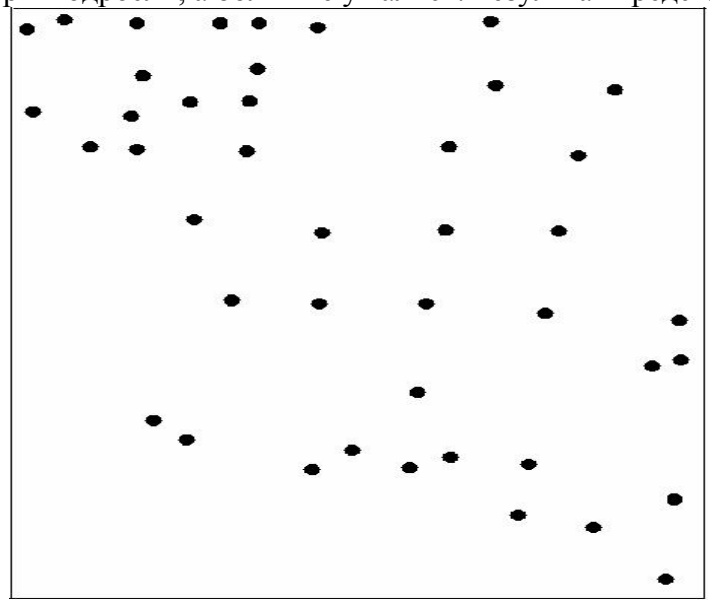


Рис. 4. Нормализованные по размеру элементы изображения

На этом заканчивается этап предварительной обработки изображения. В результате него изображение было структурировано и удалены объекты, не представляющие интерес. Следует отметить, что на этом этапе возможна потеря и искомых объектов.

Алгоритм распознавания мин

В зависимости от способа постановки минных полей, они имеют разную конфигурацию. Например, минные поля, устанавливаемые с наземных минных заградителей, имеют правильную форму в виде клеточного поля, а минное поле, установленное системой залпового огня представляет совокупность эллипсов рассеяния. Следует отметить, что из тактических соображений минное поле, независимо от способа его постановки, имеет определенную плотность. Это связано с

тем, что изреженное поле оказывается малоэффективным, а слишком плотное - ведет к большому перерасходу боеприпасов. Это обстоятельство позволяет задать предположительные размеры фигур минных полей.

В данной работе предлагается алгоритм для распознавания клеточного поля, установленного минным заградителем. Алгоритм основан на том, что известны примерные размеры клеток, а угол между сторонами клеток равен примерно 90^o. Алгоритм реализует формулу, хорошо известную из школьного курса математики: $\cos\alpha = \frac{x_1 \cdot x_2 + y_1 \cdot y_2}{\sqrt{x_1^2 + y_1^2} \cdot \sqrt{x_2^2 + y_2^2}}$, где x_1, x_2, y_1,y_2 – координаты векторов, между которыми вычисляется угол. Векторы проводятся через любые 3 точки (центры мин), одна из которых является их общим началом. Диапазон для косинуса: от -0.15 до 0.15. Расстояние между центрами мин вычисляется по формуле $\rho = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$, где x_1, x_2, y_1,y_2 – координаты центров мин. Диапазон для расстояния: от 80 до 96 пикселей включительно. Для минных полей с клеточной структурой в качестве признаков идентификации используются числовые значения, характеризующие взаимное расположение мин (рис. 5).

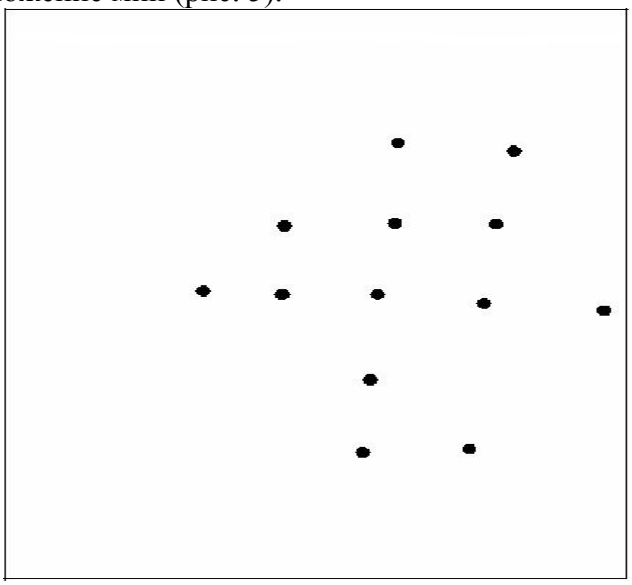


Рис. 5. Идентифицированное минное поле

Дальнейшие исследования предполагается проводить в направлении повышения «интеллектуальности» алгоритмов с целью распознавания различных форм минных полей вне зависимости от их ориентации.

Литература

1. Ивашов С.И., Разевиг В.В., Илющенко Р.Р. и др. Использование методов дистанционного зондирования в операциях по гуманитарному разминированию //Наукоёмкие технологии. – 2002. - №6. - С. 27-42.

2. Тоффоли Т., Марголус Н. Машины клеточных автоматов. - М.: Мир, 1991. - 280 с.

3. Fukushima K., Miyake S., Takayuki I. Neocognitron: A neural network model for a mechanism of visual pattern recognition // IEEE Transaction on Systems, Man and Cybernetics. - 1983. - Vol. SMC-13(5). - Pp.826–834.

Автоматическая географическая привязка спутниковых изображений с пиксельной точностью в задаче минимизации объема хранимой информации

Р.К. Газизов, И.Р. Фатхулисламов
Уфимский Государственный Авиационный Технический Университет, г. Уфа

Введение

Для повышения эффективности использования систем хранения картографических данных телекоммуникационных систем (ТС) необходимо совершенствовать методы и алгоритмы обработки изображений за счет уменьшения избыточности исходных данных и хранения динамической разницы изображений, что является перспективным направлением развития ТС. Одним из таких направлений является сжатие данных дистанционного зондирования Земли (ДЗЗ), которые хранятся на серверах баз данных. Объемы требуемого пространства измеряются здесь уже в петабайтах. Большой интерес к созданию оптимальных алгоритмов сжатия аэрокосмических данных проявляют большое количество стран, университетов, образовательных и коммерческих структур.

На первой стадии процесса получения продуктов обработки спутниковых данных выполняется первичная обработка цифровых изображений. Обязательным этапом первичной обработки является географическая привязка спутниковых изображений, который предваряет другие этапы: коррекция геометрических искажений и синтез разнородных спутниковых изображений. В зависимости от точности решения задачи привязки изображений зависит в дальнейшем эффективность решения задачи минимизации объема хранимых данных.

Существующие методы «точной» географической привязки спутниковых изображений прошлых лет показали, что не позволяют получить пиксельной точности привязки изображений. Поэтому при привязке изображений необходимо использовать информацию, которую несет в себе и само изображение.

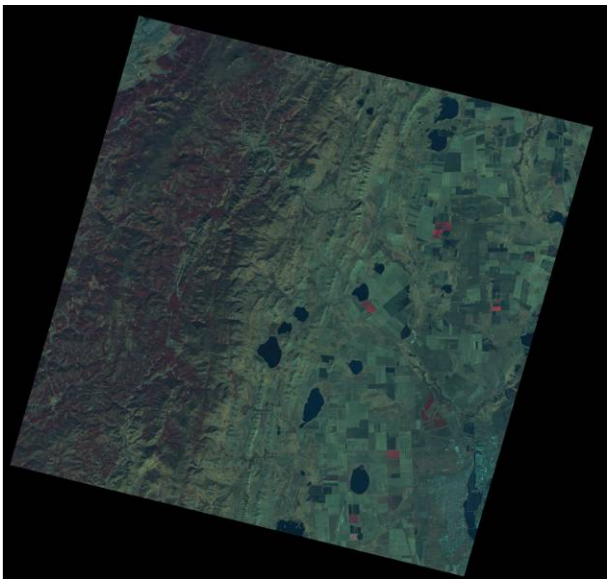
В тоже время задача точной привязки невозможна без вычисления контрольных точек на изображениях. Таким образом, одной из составляющей процедуры точной географической привязки является метод автоматического расчета контрольных точек.

В данном докладе рассмотрены существующие методы автоматического поиска контрольных точек на изображениях и сделаны выводы по применимости каждого из методов. Также, в докладе представлен разработанный метод автоматической привязки изображений с пиксельной точностью, получаемых с метеорологических ИСЗ, который в дальнейшем используется для решения задачи минимизации объема хранимой информации.

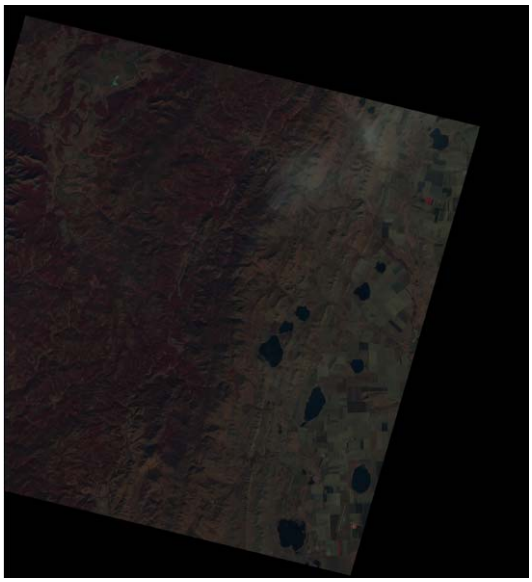
Методы автоматического поиска контрольных точек

На сегодняшний день существует множество алгоритмов поиска контрольных точек. В данном докладе рассмотрены следующие: Canny detector [1], Sobel detector [2], Laplacian of Gaussian (LoG) [3], Shi/Tomasi detector [4], Harris detector [5], Difference of Gaussians (DoG) [6], FAST [7], Affine shape adaptation [8].

Для иллюстрации работы алгоритмов рассмотрены следующие изображения, полученных со спутников NOAA/AVHRR.



(а)



(б)

Рис. 1. Исходные изображения (а) и (б). Разрешение 3885×3737.



(в)



(г)

Рис. 2. Исходные изображения (в) и (г). Разрешение 639×435.

Таблица 1. Результаты эксперимента

	Изобра-жение	**Canny**	**Sobel**	**Harris**	**Shi / Tomasi**	**FAST**	**DoG**	**LoG**	**Affine shape adaptation**
T, сек	а	3.49	1.82	4.00	3.51	13.22	2.30	2.81	-
	б	3.58	1.84	4.29	3.65	17.48	2.34	2.84	-
	в	0.07	0.09	0.64	0.65	5.40	0.08	0.04	54.09
	г	0.07	0.07	0.67	0.63	5.27	0.08	0.05	56.13
N	а	182	55079	102826	244591	2107	397398	97235	1000*
	б	153	43707	72365	93718	1316	404120	34204	1000*
	в	92	6380	5891	7846	711	6398	1622	186
	г	172	5351	5795	7663	711	6323	1452	189

В таблице 1 показаны результаты работы каждого из алгоритмов на рассматриваемых изображениях, где t-время работы алгоритма, N – количество найденных особых точек, * - указано максимальное число контрольных точек, которые могут быть найдены на изображении.

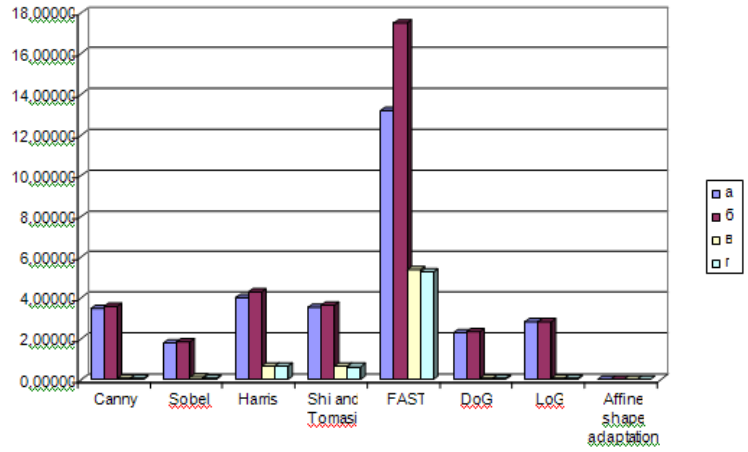


Рис. 3. Распределение времени выполнения каждого из алгоритмов.

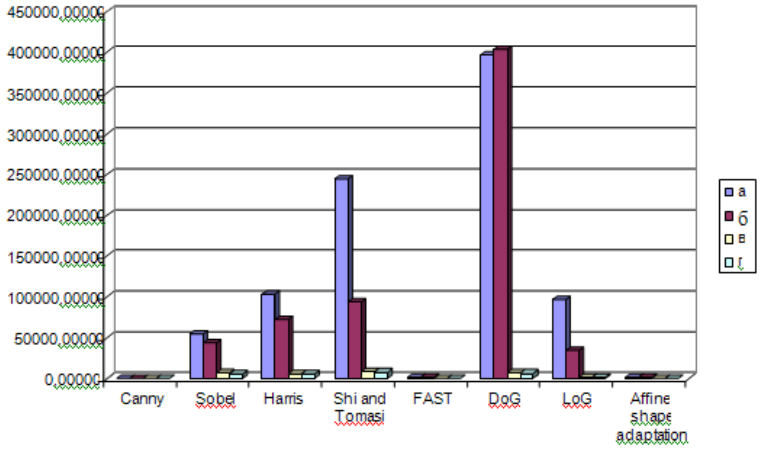


Рис. 4. Количество найденных контрольных точек.

Итак, по полученным данным видно, что детектор Canny дает невысокое количество найденных особенностей, при этом время счета достаточно велико. Детектор Sobel является стандартным детектором для поиска границ. Время счета алгоритма практически в 2 раза меньше на изображениях высокого разрешения по сравнению с детектором Canny. Количество найденных контрольных точек при этом достаточно велико. Детектор Харриса является одним из самых простых в плане реализации, он так же позволяет регулировать количество искомых особых точек путем изменения порога чувствительности (по умолчанию 0.04). Время его работы при этом практически совпадает со временем счета по алгоритму Canny, количество найденных контрольных точек на порядок превышает алгоритм Canny. Время счета детектора Shi/Tomasi практически совпадает с временем счета по алгоритму Харриса. При этом количество найденных контрольных точек больше. Детектор FAST, вопреки своему названию, работает дольше всех, но при этом демонстрирует невысокую точность обнаружения контрольных точек. Его результаты можно регулировать, изменяя некоторые параметры, к примеру параметры построения окружности, описывающей «подозрительную» точку. Детектор DoG показывает хорошие результаты при правильном наборе параметров, т.е. требует их корректный выбор для каждого конкретного изображения. К примеру, при изменении освещенности, для получения аналогичных результатов следует изменить окрестность поиска локального максимума. Таким образом, для сравнения двух изображений особые точки на них придется искать по-разному. Детектор LoG показывает различные результаты с изменением порога чувствительности, который отбрасывает «несущественные» детали. Даёт высокие показатели точности и быстродействия. Affine shape adaptation детектор удобен в

использовании, имеет множество настраиваемых параметров. При правильной настройке дает результаты высокой точности и надежности в случае использования их для сравнения изображений, т.к. метод устойчив к трансформации. Но такая точность сильно влияет на быстродействие – из всех рассмотренных детекторов этот работает дольше всех.

Минимизация объема хранимой информации

Для решения задачи минимизации объема хранимой информации предлагается следующий алгоритм:

1. Автоматический поиск точечных особенностей на аэрокосмических снимках.
2. Поиск возможных взаимных соответствий точек одного снимка с другим.
3. Приведение снимков к единой системе координат.
4. Исключение избыточной информации.

Предложенный алгоритм позволит хранить не весь снимок, а лишь различия снимков. Алгоритм реализован на языке программирования Matlab с использованием библиотеки vlfeat 0.9.7 [9].

Наиболее известным и простым в реализации алгоритмом поиска контрольных точек является детектор Харриса, который позволяет находить особенности с высокой скоростью. В данной работе был использован данный алгоритм поиска контрольных точек.

Сопоставление найденных точечных особенностей может быть вычислено, например, путем сравнения их окрестностей кросскорреляционным [10] методом. Однако на аэрокосмических снимках высокого разрешения сразу несколько особенностей на первом снимке могут оказаться близкими к одной и той же особенности на втором. Следовательно, отличить верные соответствия от ложных не всегда представляется возможным. Поэтому более эффективным по сравнению с кросскорреляционным методом является использование алгоритма Scale Invariant Feature Transform (SIFT) [11], который, однако, обладает более высокой сложностью обработки. Поиск возможных соответствий производится функцией VL_UBCMATCH в составе библиотеки vlfeat 0.9.7, которая вычисляет евклидовое расстояние между дескрипторами изображений.

После этого производится фильтрация ложных соответствий между снимками с использованием алгоритма RANSAC, который подробно описан в [12].

После нахождения взаимных соответствий контрольных точек на изображениях, вычисляется матрица преобразования функцией cp2tform в составе пакета Image Processing Toolbox среды программирования Matlab. Изображений приводится к единой системе координат функцией imtransform.

После приведения изображений к единой системе координат, производится вычисление разницы рассматриваемых снимков. Вместо исходного изображения сохраняются лишь различия снимков.

В вычислительных экспериментах использовались серии снимков со спутника, предоставленные кафедрой телекоммуникационных систем УГАТУ.

Таблица 2. Результаты работы алгоритма

№	Разреше-ние снимка	Количество найденных особенных точек на каждом снимке	Количество найденных пар соответствий	Ложных	Степень сжатия	Время (сек)
1	5396x5132	58501/44489	10296	113	41%	1279
2	2698x2566	9132/6336	2183	52	35%	73
3	1349x1283	1752/1339	433	39	29%	14

Как видно из полученных результатов (таблица 2), при уменьшении разрешения снимков в 2 раза количество найденных точечных особенностей уменьшается в 5-6 раз. Как следствие, пропорционально уменьшается и количество найденных пар соответствий. Степень сжатия изображений при этом ухудшается.

Заключение

В статье выполнена реализация существующих методов автоматического поиска контрольных точек. Алгоритмы реализованы с использованием среды программирования Matlab. Произведен сравнительный анализ и сделаны выводы по работе каждого их них.

Также представлена возможность реализации алгоритма, исключающего избыточность информации. Алгоритм был реализован с использованием среды программирования Matlab и библиотеки vlfeat 0.9.7. Проведенные вычислительные эксперименты позволяют говорить об эффективности предлагаемого подхода.

Учитывая большое время счета алгоритма в дальнейшем предполагается его реализация с использованием графических процессоров на языке программирования Matlab и пакета Jacket [13]. Т.к. современные графические процессоры очень эффективно обрабатывают и отображают компьютерную графику, благодаря специализированной конвейерной архитектуре они намного эффективнее в обработке графической информации, чем типичный центральный процессор.

Литература

1. Canny, J.F. A computational approach to edge detection. IEEE Trans Pattern Analysis and Machine Intelligence, 8(6): 679-698, 1986.
2. Sobel, I. An Isotropic 3×3 Gradient Operator, Machine Vision for Three – Dimensional Scenes, Freeman, H., Academic Pres, NY: 376-379, 1990.
3. Sotak, G. E., Boyer, K. L. "The Laplacian-of-Gaussian kernel: A formal analysis and design procedure for fast, accurate convolution and full-frame output," Ohio State Univ. Signal Analysis and Machine Perception Lab., Tech. Rep. SAMPL-88-01, 1988.
4. J. Shi and C. Tomasi. "Good Features to Track". 9th IEEE Conference on Computer Vision and Pattern Recognition, 1994.
5. C. Harris and M. Stephens. A combined corner and edge detector. In Proceedings, 4th Alvey Vision Conference: 147–151, 1988.
6. Marr, D., Hildreth, E. "Theory of Edge Detection". Proceedings of the Royal Society of London. Series B, Biological Sciences: 215–217, 1980.
7. M. Trajkovic and M. Hedley. "Fast corner detection". Image and Vision Computing 16 (2): 75–87, 1998.
8. Baumberg, A. "Reliable feature matching across widely separated views". Proceedings of IEEE Conference on Computer Vision and Pattern Recognition: 1774-1781, 2000
9. http://www.vlfeat.org/
10. Basri, R., and Jacobs, D.W. Recognition using region correspondences. International Journal of Computer Vision, 25(2):145-166, 1997.
11. David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints International Journal of Computer Vision: pp. 91-110, 2004.
12. Fischler, Bolles. "RANdom SAmpling Consensus: a paradigm fro model fitting with application to image analysis and automated cartography" Commun.Assoc.Comp. Mach.: 381-95, 1981.
13. http://www.accelereyes.com/products/jacket

Анализ пространственного взаимодействия оптических шкал с использованием преобразования Радона

В.М. Гречишников, А.А. Юдин, О.Ю. Борисов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Взаимодействие оптических шкал при их взаимном смещении относительно друг друга широко используется при создании оптоэлектронных преобразователей информации для авиационно-космической техники. При этом характер этого взаимодействия непосредственно определяет их основную информативную характеристику-функцию модуляции излучения под действием контролируемой величины. Примером таких устройств являются, например, оптоэлектронные цифровые преобразователи угловых перемещений (ОЦПУ). Независимо от способа формирования выходного сигнала в основе их функционирования лежит модуляция оптического излучения одной (инкрементальные ОЦПУ) или несколькими (абсолютные ОЦПУ) кодовыми дорожками (КД) в пределах некоторой области, определяемой конфигурацией считывающего элемента (СЭ).Для получения и исследования функции модуляции излучения используются различные методы, основанные на применении двумерного пространственного преобразования Фурье, теории R- функций, а также классического аппарата аналитической геометрии на плоскости /2,3/. Однако существующие методы не обеспечивают необходимой адекватности модели и объекта, отличаются громоздкостью и большими затратами машинного времени даже при использовании современных ПК. В связи с этим представляется актуальным разработка метода получения функции модуляции излучения в ОЦПУ с использованием аппарата преобразования Радона, широко применяемого в задачах обработки изображений /1/.

Описание объекта исследований

Как правило, при обработке изображений, получаемых с помощью ПЗС- матриц, оперируют с координатно – ориентированным числовым массивом, каждый элемент которого (пиксель) характеризуется координатами x_i, y_i, а также значением яркости объекта в данной точке B_i. В задаче построения математической модели ОЦПУ каждую кодовую дорожку можно представить в виде оптического изображения, состоящего из чередующихся участков с резко отличающимися яркостными характеристиками (рис.1). Например, прозрачные участки КД можно интерпретировать как элементы изображения с относительной яркостью, равной 1, а непрозрачные – с яркостью, равной 0. При этом форма прозрачных участков кодовой шкалы представляет собой, как правило, сектор кольца, образованного окружностями с радиусами r_1, r_2. Форма отверстий СЭ может либо повторять форму отверстий кодовой шкалы, либо иметь форму окружности при считывании информации с помощью волоконных световодов круглого сечения (рис.1а, б).

С формальной точки зрения задача получения математической модели взаимодействия кодовой шкалы с одним или несколькими считывающими элементами сводится к отысканию значения оптического пропускания $F(\varphi)$ в каждой точке диапазона преобразования. Величина пропускания определяется отношением площади взаимного перекрытия $S(\varphi) = \Omega_k \cap \Omega_{сэ}$ прозрачных участков КД и СЭ к полной площади СЭ $S_{сэ}$.

$$F(\varphi) = S(\varphi) / S_{сэ} \qquad (1)$$

В реальных преобразователях КД, расположенные на измерительном валу, вращаются относительно СЭ. Для построения математической модели удобнее неподвижно зафиксировать изображение КД в некоторой плоскости XOY и перемещать относительно нее считывающий элемент. При этом начало координат совместим с центром вращения СЭ. Изображение к-той прозрачной области $\Omega_к$ может быть задано координатами точек

$$M_k(x_M, y_M), N_k(x_N, y_N), P_k(x_P, y_P), Q_k(x_Q, y_Q),$$

соединенными между собой отрезками прямых или окружностей

$$\Omega_к : \left[y_{кл}, y_{кв}, y_{кП}, y_{кН} \right],$$

где величины, указанные в скобках представляют собой уравнения левой, верхней, правой и нижней границ к-той области соответственно.

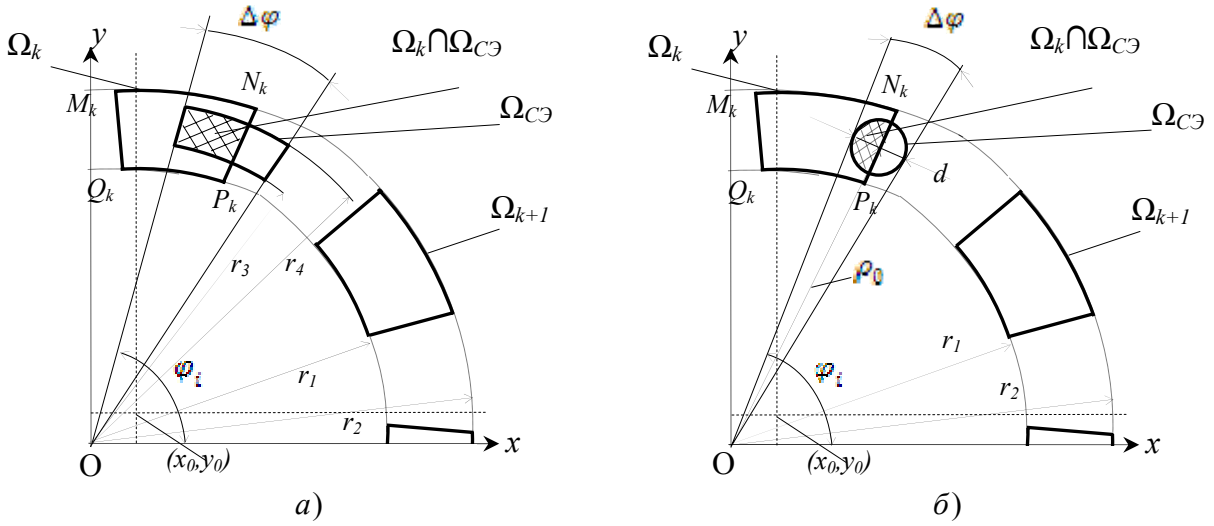


Рис. 1. Расчетны имодействия КД с секториальным (а) и круг

В соответствии с методикой /3/ на этапе математического описания областей $\Omega_к$ можно учесть комплекс инструментальных погрешностей изготовления КД: неточность угловых границ областей, эллиптичность и эксцентриситеты образующих окружностей, неплоскостность КД и СЭ, влияние дифракции и другие факторы. Для учета влияния эксцентриситета кодового диска относительно оси вращения в координаты точек изображения КД в плоскости XOY вводятся соответствующие приращения (x_0, y_0), обеспечивающие эксцентричное вращение области считывающего элемента $\Omega_{сэ}$ относительно кодовой дорожки

$$M_k(x_M + x_0, y_M + y_0), N_k(x_N + x_0, y_N + y_0), P_k(x_P + x_0, y_P + y_0), Q_k(x_Q + x_0, y_Q + y_0).$$

Секториальная форма СЭ

Пусть СЭ имеет форму сектора кольца, ограниченного окружностями с радиусами r_4, r_3 и двумя лучами, выходящими из начала координат, угол между которыми равен $\Delta\varphi$. При отсутствии инструментальных погрешностей развертку изображения КД по отрезку луча, выходящего из начала координат системы XOY ($x_0 = y_0 = 0$) можно представить в виде преобразования Радона

$$\omega(\varphi) = \int_{r_3}^{r_4} e(t\cos\varphi, t\sin\varphi)dt \qquad (2)$$

в котором подынтегральное выражение представляет собой единичную функцию вида :

$$e = \begin{cases} 1, при\ t\cos\varphi, t\sin\varphi \in \Omega_k \\ 0, при\ t\cos\varphi, t\sin\varphi \notin \Omega_k \end{cases}, \qquad (3)$$

заданную в диапазоне изменения параметра $t \in [r_4, r_3]$

Поскольку яркость прозрачных участков равномерно распределена по площади КД, то усредняя (4) по диапазону $r_4 - r_3$, получим также единичную функцию

$$e^*(\varphi) = \omega(\varphi)/(r_4 - r_3), \qquad (4)$$

описывающую развертку идеального изображения КД по углу φ и представляющую собой сигнал типа «меандр» с периодом $360/m$, m – целое число, кратное степени числа 2. Поскольку область перекрытия имеет конечный угловой размер $\Delta\varphi_П \in [0, \Delta\varphi]$, то для вычисления области взаимного перекрытия прозрачных областей КД и СЭ, площадь которой определяет значение пропускания в данной точке, необходимо проинтегрировать (4) по углу:

$$\int_{\varphi}^{\varphi+\Delta\varphi_П} e^*(\varphi)d\varphi = \Delta\varphi_П(\varphi) \qquad (5)$$

График этой функции в общем случае представляет собой последовательность трапецеидальных импульсов, определенных на интервале 0-360^0, длительность фронтов которых определяется величиной $\Delta\varphi$. Интегрируя (4)в пределах $0 - \Delta\varphi$, получим

$$\int_{0}^{\Delta\varphi} e^*(\varphi)d\varphi = \Delta\varphi \qquad (6)$$

Отношение углов

$$\Delta\varphi_П(\varphi)/\Delta\varphi = F(\varphi) \qquad (7)$$

равно отношению площади взаимного перекрытия КД и СЭ к полной площади СЭ и представляет собой искомую функцию модуляции (пропускания). Однако в результате влияния инструментальных погрешностей изготовления и установки КД и СЭ функция модуляции (6) будет изменяться от одного периода КД к другому. Поэтому поучить единое аналитическое выражение для функции модуляции во всем диапазоне изменения угла не представляется возможным, что приводит к необходимости перехода к численным методам анализа. В связи с этим рассмотрим методику получения функции модуляции, основанную на применении дискретного веерного преобразования Радона по лучу, выходящему из начала координат под углом φ_i:

$$\omega(\varphi_i) = \sum_{l=1}^{m} e(t_l cos\varphi_i, t_l \sin\varphi_i); \qquad (8)$$

где выражение под знаком суммы представляет собой единичную функцию вида

$$e = \begin{cases} 1, при\ t_l\cos\varphi_i, t_l\sin\varphi_i \in \Omega_k \\ 0, при\ t_l\cos\varphi_i, t_l\sin\varphi_i \notin \Omega_k \end{cases} \qquad (9)$$

где $t_l = r_3 + \frac{l}{m}(r_4 - r_3)$ - шаг дискретизации изображения $\Omega_к$ по лучу. Значение параметра *m* определяет шаг дискретизации преобразования по лучу и выбирается исходя из требуемой точности вычислений.

Условие принадлежности произвольной точки на луче к области $\Omega_к$ можно задать в виде базовой матрицы знаковых функций для точки, заведомо принадлежащей этой области. В качестве такой точки можно выбрать, например точку (x_{0k}, y_{0k}) пересечения диагоналей области $\Omega_к$. Базовая матрица в этом случае имеет вид:

$$\begin{vmatrix} sign[y_{0k} - y_л(x_{0k})] \\ sign[y_{0k} - y_П(x_{0k})] \\ sign[y_{0k} - y_в(x_{0k})] \\ sign[y_{0k} - y_Н(x_{0k})] \end{vmatrix} = \begin{vmatrix} sign\Delta y_{лk} \\ sign\Delta y_{Пk} \\ sign\Delta y_{вk} \\ sign\Delta y_{Нk} \end{vmatrix} \qquad (10)$$

Если при подстановке в левую часть (10) координат анализируемой точки правая часть будет отличаться от базовой матрицы хотя бы в одной позиции то, данная точка не принадлежит области $\Omega_к$ и ее следует отбросить.

Суммирование количества точек (пикселей) изображения, одновременно принадлежащих отрезку луча $\Delta r = r_4 - r_3$ и прозрачной области $\Omega_к$ производится по формуле

$$N(\varphi_i) = \sum_{j=1}^{p}\sum_{l=1}^{m} e[t_l \cos(\varphi_i + j\Delta\varphi_0), t_l \sin(\varphi_i + j\Delta\varphi_0), \qquad (11)$$

где $\Delta\varphi_0 = \Delta\varphi / p$ - шаг дискретизации по углу в пределах области СЭ; $\varphi_i = i\Delta\varphi_0$ - дискретные значения угла поворота. Выражение (11) представляет собой по существу двумерное преобразование Радона по области СЭ. Значение параметра p также определяется требуемой точностью формирования суммы (11).Очевидно, что найденная сумма пикселей $N(\varphi_i)$ пропорциональна площади перекрытия областей $\Omega_к$ и $\Omega_{сэ} : S(\varphi_i) = S_0 N(\varphi_i)$, где S_0 - площадь пикселя. Полная площадь СЭ $S_{сэ}$ легко вычисляется по известным координатам вершин четырехугольника с учетом всех инструментальных погрешностей разметки КД.

Круглая форма СЭ

При считывании информации с КД с помощью круглого торца волоконного световода (рис.1,б) определение значения функции модуляции в $i-$той точке диапазона преобразования сводится отысканию суммы пикселей, которые одновременно принадлежат лучу, области $\Omega_к$ и области $\Omega_{сэ}$, ограниченной окружностью радиусом $d = 2r_c$ (r_c -радиус поперечного сечения световода). При этом диапазон изменения параметра t лежит в пределах от r_3 до $r_4 = r_3 + 2r_c$, а диапазон сканирования изображения по углу равен :

$$\Delta\varphi = 2arctg(r_c / \rho_0) \qquad (12)$$

где $\rho_0 -$ расстояние от начала координат до центра окружности.

Для этого необходимо задать уравнение окружности, эксцентрично вращающейся относительно КД одновременно с лучом, выходящим из начала координат XOY.

Уравнение такой окружности имеет вид:

$$\left[(t-\rho_0)\cos\varphi\right]^2+\left[(t-\rho_0)\sin\varphi\right]^2\le r_c^2 \qquad (13)$$

Таким образом, условие одновременной принадлежности токи на луче с координатой t_l областям $\Omega_к$ и $\Omega_{сэ}$ запишется в виде:

$$\begin{cases}\left[(t_l-\rho_0)\cos\varphi_i\right]^2+\left[(t_l-\rho_0)\sin\varphi_i\right]^2\le r_c^2\\ (t_l\cos\varphi_i, t_l\sin\varphi_i)\in\Omega_k:\left[y_л, y_в, y_п, y_н\right]\end{cases} \qquad (14)$$

Выполнение второго условия системы (14) проверяется по (10). В этом случае функция модуляции в i – той точке диапазона преобразования равна

$$F(\varphi_i)=S_0N(\varphi_i)/\pi r_c^2 \qquad (15)$$

В этом выражении $N(\varphi_i)$ также представляет собой двумерное преобразование Радона по сечению круглого световода, которое определяется выражением (11) с учетом (12). В многоразрядных ОЦПУ в соответствии с описанной методикой формируется параллельный формат сигналов, отображающей функций модуляции по каждой кодовой дорожке.

Заключение

В результате проведенных исследований показана принципиальная возможность построения двумерного дискретного локального веерного преобразования Радона по сектору кольца и окружности с использованием преобразования на отрезке луча для получения функции модуляции излучения при взаимном перемещении оптических шкал. Практическое применение результаты работы могут найти при математическом моделировании и оценке технических возможностей вновь создаваемых конструкций оптоэлектронных преобразователей информации различного назначения. Дальнейшее развитие работы связано с учетом в разработанных моделях влияния неоднородности и нестационарности излучения, вызванными влиянием дифракционных эффектов и внешних эксплуатационных факторов.

Литература

1.Баранов В.Г., Храмов А.Г. Дискретное веерное преобразование Радона в задаче выделения центров ветвей сетчатых структур //Компьютерная оптика,№23,2002, с.44-47.

2.Гречишников В.М., Конюхов Н.Е. Оптоэлектронные цифровые датчики перемещений со встроенными волоконно-оптическими линиями связи.-М: Энергоатомиздат,1992.-160с.:ил.-(Б-ка по автоматике, Вып.677)

3. Гречишников В. М., Лофицкий И. В., Дорошева С. И. Получение функции модуляции излучения в оптоэлектронных цифровых преобразователях перемещения с учетом производственных погрешностей их изготовления //Сб. трудов студентов и аспирантов РТФ СГАУ, Вып. 10, 2006, с..114-120.

Модифицированный усредняющий фильтр. Определение мощности шума на видеоизображении

Р.Р. Диязитдинов, В.А. Буцких
Поволжский государственный университет телекоммуникаций и информатики, Самара

Введение

При обработке видеоизображений одной из задач является фильтрация для уменьшения уровня шума. При этом немаловажным фактором является не только полученный результат, но и скорость работы фильтра. Для работы рассматриваемого фильтра нам необходимо оценить мощность шума на видеоизображении.

Оценка мощности шума

Исходя из предположения, что природа шума, накладываемая на каждый пиксель изображения одна и та же, оценку мощности шума можно проводить как по одному пикселю, так и по всем пикселям изображения.

Сделаем также два предположения. Во-первых, шум имеет нормальное распределение. Во-вторых, при переходе от кадра к кадру условия освещенности практически не меняются. Если найти разность яркостей пикселей соседних кадров, то полученная случайная величина будет иметь математическое ожидание близкое к нулю $M[X] \approx 0$ и дисперсию близкую к удвоенной дисперсии шума $D[X] \approx 2\sigma_n^2$ [1].

Для вычисления дисперсии воспользуемся формулой:

$$D[X] = M[X^2] - M[X]^2,$$

где $D[X]$ - это дисперсия, $M[X^2]$ - математическое ожидание квадратов случайной величины, $M[X]$ - математическое ожидание случайной величины.

Для иллюстрации алгоритма вычисления дисперсии приведем пример. Пусть мы оцениваем дисперсию по N отсчетам, где значения случайной величины представлены отсчетами $x_1, x_2, ..x_{N-1}, x_N, x_{N+1}...$ Будем использовать буфер памяти $buffer$ для хранения N значений отсчетов и три переменные для хранения текущего значения суммы sum и суммы квадратов sum_sqr и вычисленной дисперсии D.

Первый шаг оценивания дисперсии:

$$buffer \leftarrow [x_1, x_2, ..x_{N-1}, x_N],\ sum = \sum_{i=i}^{N} x_i\ ,\ sum\_sqr = \sum_{i=i}^{N} x_i^2\ ,\ D = \frac{sum\_sqr}{N} - \left(\frac{sum}{N}\right)^2$$

Второй шаг оценивания дисперсии:

$$buffer \leftarrow [x_{N+1}, x_2, ..x_{N-1}, x_N],\ sum = sum - x_1 + x_{N+1},\ sum\_sqr = sum\_sqr - x_1^2 + x_{N+1}^2,$$

$$D = \frac{sum\_sqr}{N} - \left(\frac{sum}{N}\right)^2$$

Третий шаг оценивания дисперсии:

$$buffer \leftarrow [x_{N+1}, x_{N+2}, ..x_{N-1}, x_N],\ sum = sum - x_2 + x_{N+2}\ ,\ sum\_sqr = sum\_sqr - x_2^2 + x_{N+2}^2,$$

$$D = \frac{sum\_sqr}{N} - \left(\frac{sum}{N}\right)^2$$

И так далее.

Исходя из предположений, что $M[X] \approx 0$ и $D[X] \approx 2\sigma_n^2$, можем записать:

$$\sigma_n^2 \approx \frac{D[X]}{2} = \frac{sum\_sqr}{2N}.$$

Реализация фильтра

Рассмотрим модифицированный усредняющий фильтр [2]. Эта реализация фильтра отвечает двум нашим потребностям – высокое быстродействие и приемлемый результат фильтрации.

Классический усредняющий фильтр находит среднее значение яркости внутри блока (с нечетным числом столбцов и строк) и полученное значение вставляет в центр блока. Модификация состоит в том, что для всех пикселей, попавших внутрь блока, находится среднее значение M и дисперсия D. Если значение дисперсии меньше, чем дисперсия шума σ_n^2, то значение пикселя заменяется значением вычисленного среднего:

$$y = M .$$

Если же оно больше, чем дисперсия шума, то используется формула:

$$y = M + \frac{D - \sigma_n^2}{D} \cdot (x - M),$$

где x – значение яркости центрального пикселя, y – значение яркости центрального пикселя после обработки фильтром.

Для оценки мощности шума до и после фильтрации, будем использовать значение $h^2 = \frac{E}{N}$ – отношение энергии сигнала к мощности флуктуационной помехи [3]. На практике же значение энергии сигнала вычисляется, как среднее значение энергии (то есть мощность сигнала), а мощность помехи для аддитивного гауссова шума равна его дисперсии:

$$E = \frac{\sum_i^M x_i}{M};\ N = \sigma_n^2,$$

где x_i – значение яркости *i*-того пикселя, M – количество пикселей изображения.

После применения фильтра значение мощности помехи оценивается как:

$$N = \frac{\sum_i^M (y_i - x_i)^2}{M},$$

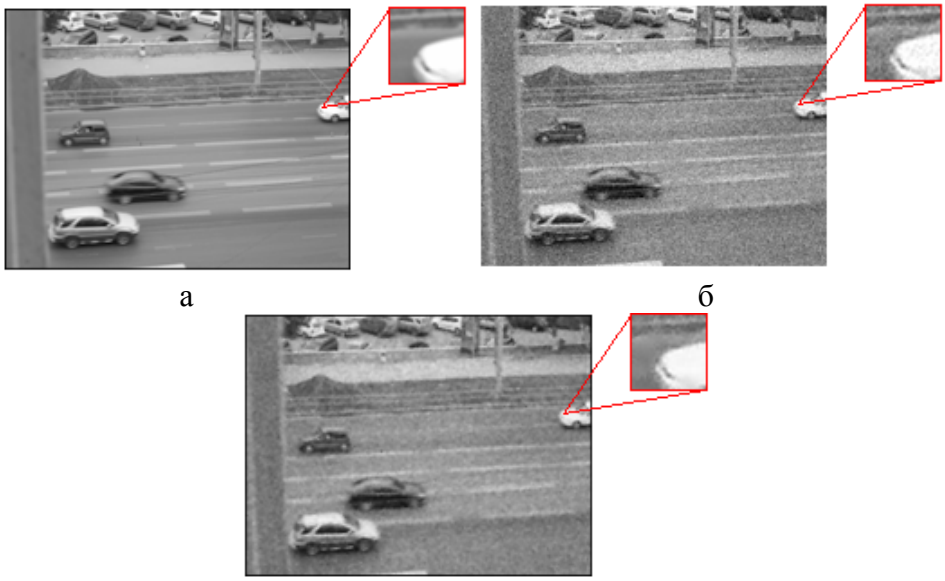
где x_i, y_i – значение яркости *i*-того пикселя на входе и выходе фильтра соответственно.

Результаты

Примеры изображений представлены на рисунке 1, результаты измерений h^2 сведены в таблицу 1 для девяти различных изображений. Для удобства представления использовалось не абсолютное h^2, а его логарифмическое значение $10 \lg h^2$.

На практике существует проблема влияния движущихся объектов. При переходе от кадра к кадру значение разности яркости пикселя, занимаемого подвижным объектом, может сильно отличаться от типичного значения. Чтобы избежать этого, можно выбрать порог. Если абсолютная разность яркости превышает этот порог, то эта разность не будет учитываться. Тем самым, мы отбрасываем часть значений, которые вносят погрешности в оценку дисперсии (мощности) шума. Для оценки этого порога, воспользуемся методикой: выберем значение p от 0 до 1, которое отражает наше предположение относительно того какую долю пикселей, участвующих в эксперименте мы будем использовать для оценки дисперсии шума. Затем строим гистограмму значений разности. И относительно центрального значения (то есть нуля) выберем

такой интервал $[-B;B]$, для которого отношения количества пикселей, попавших в интервал, к общему отношению пикселей не превышает выбранного значения p. В экспериментах полагалось $p = 0{,}9$.



а б в

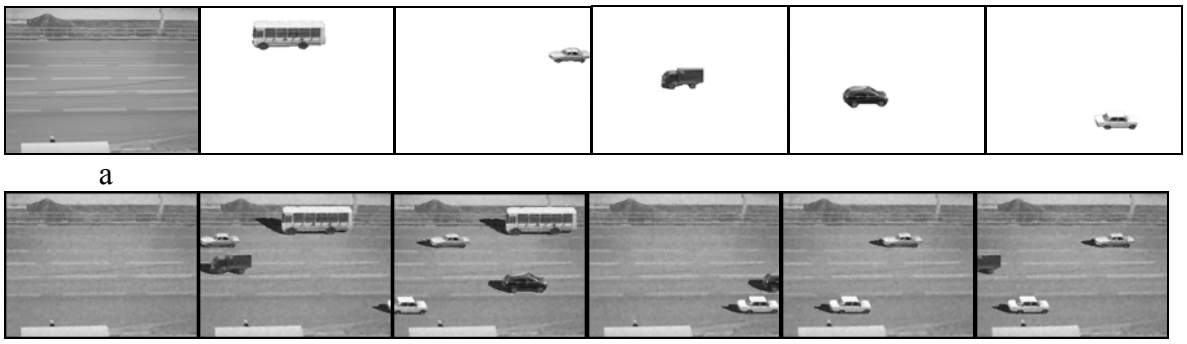
Рис. 1. Примеры изображений: (а) исходное, (б) с шумом, (в) после фильтрации

Таблица 1. Значения h^2 после применения модифицированного усредняющего фильтра для девяти различных изображений

Номер изображения	1	2	3	4	5	6	7	8	
h^2 до фильтрации, *дБ*	h_ϕ^2 после фильтрации, *дБ*								
10	20,03	20,95	20,51	20,56	20,21	18,61	19,96	19,80	0,23
13	21,70	22,71	22,23	22,26	21,85	20,16	21,61	21,42	1,87
17	24,05	25,10	24,59	24,61	24,17	22,42	23,91	23,75	4,19
20	25,91	26,93	26,44	26,45	26,01	24,32	25,76	25,63	6,04
23	27,90	28,84	28,40	28,41	27,98	26,42	27,76	27,63	8,01
27	30,70	31,53	31,16	31,20	30,81	29,44	30,63	30,49	0,82
30	32,94	33,66	33,35	33,40	33,07	31,86	32,92	32,77	3,05
33	35,23	35,82	35,56	35,63	35,38	34,34	35,25	35,10	5,34
37	38,27	38,67	38,47	38,53	38,41	37,61	38,30	38,18	8,35
40	40,49	40,75	40,59	40,64	40,61	39,99	40,53	40,43	0,55

Для проверки эффективности работы фильтра при определении мощности шума по видеоизображению, использовалась моделирование. На фоновое изображение, накладывались изображения посторонних объектов, затем полученное изображение

умножалось на значение контрастности, которое имеет вид $\alpha = \alpha_н + \alpha_c$, где $\alpha_н$ – номинальное значение контрастности, α_c – отклонение контрастности от номинального значения. В наших экспериментах $\alpha_н = 1$, α_c – случайная величина с равномерным распределением в интервале $[-\alpha_{c\max}; \alpha_{c\max}]$. После этого полученное изображение складывалось с гауссовым шумом. Примеры изображений представлены на рисунке 2. Результаты среднего значения оценки $\overline{h^2}$ и среднего значения $\overline{h^2}_ф$ после применения фильтра сведены в таблицах 2, 3 при отсутствии и наличии движущихся объектов.



а

б

Рис. 2. Изображения, использованные при моделировании (а) и полученные в ходе эксперимента (б)

Таблица 2. Значения $\overline{h^2}$ и $\overline{h^2}_ф$ при различных значениях контрастности

h^2 до фильтрации, *дБ*	Отклонение контрастности от номинального значения, $\alpha_{c\max}$, 10^{-3}	0	5	10	15	20	25	30	35	40	45	50
10	$\overline{h^2}$, *дБ*	10,33	10,33	10,32	10,32	10,32	10,32	10,32	10,31	10,31	10,30	10,30
	$\overline{h^2}_ф$, *дБ*	20,60	20,60	20,61	20,61	20,61	20,61	20,62	20,63	20,63	20,64	20,65
13	$\overline{h^2}$, *дБ*	13,23	13,23	13,23	13,22	13,22	13,21	13,21	13,20	13,19	13,18	13,17
	$\overline{h^2}_ф$, *дБ*	22,48	22,49	22,49	22,49	22,50	22,50	22,51	22,52	22,53	22,55	22,56
17	$\overline{h^2}$, *дБ*	17,17	17,17	17,17	17,16	17,15	17,14	17,12	17,10	17,08	17,05	17,02
	$\overline{h^2}_ф$, *дБ*	24,93	24,93	24,94	24,94	24,95	24,96	24,98	24,99	25,01	25,03	25,06
20	$\overline{h^2}$, *дБ*	20,17	20,17	20,16	20,14	20,12	20,09	20,06	20,02	19,98	19,93	19,88
	$\overline{h^2}_ф$, *дБ*	26,75	26,76	26,76	26,77	26,78	26,80	26,82	26,84	26,86	26,89	26,91
23	$\overline{h^2}$, *дБ*	23,18	23,17	23,15	23,12	23,08	23,03	22,96	22,89	22,80	22,72	22,62
	$\overline{h^2}_ф$, *дБ*	28,65	28,65	28,66	28,67	28,69	28,71	28,73	28,76	28,78	28,80	28,82
27	$\overline{h^2}$, *дБ*	27,16	27,14	27,08	27,01	26,91	26,79	26,65	26,49	26,32	26,15	25,96
	$\overline{h^2}_ф$, *дБ*	31,33	31,34	31,35	31,37	31,40	31,42	31,44	31,46	31,47	31,47	31,46
30	$\overline{h^2}$, *дБ*	30,16	30,12	30,01	29,88	29,69	29,47	29,23	28,98	28,73	28,48	28,23
	$\overline{h^2}_ф$, *дБ*	33,50	33,51	33,53	33,55	33,58	33,60	33,60	33,60	33,58	33,55	33,51
33	$\overline{h^2}$, *дБ*	33,18	33,10	32,91	32,67	32,36	32,01	31,66	31,34	31,03	30,76	30,52

	$\overline{h^2_{\text{ф}}}$, дБ	35,72	35,73	35,76	35,79	35,81	35,81	35,79	35,75	35,71	35,66	35,61
37	$\overline{h^2}$, дБ	37,13	36,94	36,52	36,05	35,53	35,03	34,62	34,30	34,05	33,87	33,72
	$\overline{h^2_{\text{ф}}}$, дБ	38,66	38,68	38,71	38,72	38,70	38,65	38,59	38,54	38,49	38,45	38,41
40	$\overline{h^2}$, дБ	39,97	39,60	38,83	38,09	37,39	36,84	36,46	36,20	36,02	35,89	35,79
	$\overline{h^2_{\text{ф}}}$, дБ	40,83	40,85	40,85	40,79	40,69	40,59	40,51	40,45	40,40	40,37	40,34

Таблица 3. Значения $\overline{h^2}$ и $\overline{h^2_{\text{ф}}}$ при различных значениях контрастности и наличии движущихся объектов

h^2 до фильтрации, дБ	Отклонение контрастности от номинального значения, $\alpha_{c\max}$, 10^{-3}	0	5	10	15	20	25	30	35	40	45	50
10	$\overline{h^2}$, дБ	10,12	10,11	10,11	10,11	10,11	10,11	10,11	10,10	10,10	10,09	10,09
	$\overline{h^2_{\text{ф}}}$, дБ	20,15	20,16	20,16	20,16	20,16	20,16	20,17	20,17	20,17	20,18	20,18
13	$\overline{h^2}$, дБ	12,91	12,91	12,91	12,90	12,90	12,89	12,89	12,88	12,87	12,87	12,86
	$\overline{h^2_{\text{ф}}}$, дБ	22,01	22,01	22,02	22,02	22,02	22,03	22,03	22,04	22,04	22,05	22,06
17	$\overline{h^2}$, дБ	16,85	16,85	16,84	16,84	16,83	16,82	16,80	16,78	16,76	16,74	16,72
	$\overline{h^2_{\text{ф}}}$, дБ	24,36	24,36	24,37	24,37	24,38	24,38	24,39	24,40	24,41	24,42	24,43
20	$\overline{h^2}$, дБ	19,92	19,92	19,91	19,89	19,87	19,85	19,82	19,78	19,74	19,70	19,65
	$\overline{h^2_{\text{ф}}}$, дБ	26,19	26,19	26,19	26,20	26,21	26,22	26,23	26,24	26,26	26,27	26,28
23	$\overline{h^2}$, дБ	22,96	22,95	22,93	22,91	22,87	22,82	22,76	22,69	22,62	22,53	22,44
	$\overline{h^2_{\text{ф}}}$, дБ	28,15	28,16	28,16	28,17	28,18	28,20	28,21	28,22	28,24	28,25	28,26
27	$\overline{h^2}$, дБ	26,97	26,95	26,90	26,83	26,74	26,62	26,49	26,34	26,18	26,01	25,84
	$\overline{h^2_{\text{ф}}}$, дБ	30,96	30,96	30,97	30,99	31,00	31,02	31,03	31,04	31,05	31,04	31,03
30	$\overline{h^2}$, дБ	29,98	29,94	29,85	29,72	29,55	29,33	29,10	28,86	28,62	28,38	28,15
	$\overline{h^2_{\text{ф}}}$, дБ	33,20	33,21	33,23	33,24	33,26	33,27	33,27	33,27	33,25	33,22	33,18
33	$\overline{h^2}$, дБ	33,03	32,96	32,78	32,55	32,25	29,33	31,58	31,26	30,97	30,71	30,48
	$\overline{h^2_{\text{ф}}}$, дБ	35,49	35,50	35,53	35,55	35,56	35,56	35,54	35,51	35,47	35,42	35,38
37	$\overline{h^2}$, дБ	37,00	36,82	36,41	35,96	35,45	34,97	34,58	34,28	34,04	33,86	33,72
	$\overline{h^2_{\text{ф}}}$, дБ	38,52	38,53	38,56	38,56	38,54	38,50	38,45	38,40	38,36	38,32	38,29
40	$\overline{h^2}$, дБ	39,83	39,47	38,73	38,02	37,34	36,81	36,45	36,20	36,02	35,90	35,81
	$\overline{h^2_{\text{ф}}}$, дБ	40,73	40,75	40,74	40,69	40,60	40,52	40,44	40,39	40,35	40,32	40,30

Заключение

Оценка мощности шума тем достовернее, чем меньше диапазон возможных значений случайной составляющей контрастности. Для работы фильтра она подходит, но для статистических алгоритмов такая оценка будет неудовлетворительной.

Литература

1. Elgammal A, Harwood D, Davis LS. Non-parametric model for background subtraction. European Conference on Computer Vision 2000;2:751–67.

2. Håkan Norell, Bengt Oelmann and Youshi Xu "Spatio-Temporal Noise Reduction ASIC for Real-Time Video Processing". In Proc. IEEE Nordic Signal Processing Symposium, Kolmården, Sweden, 13-15 June, 2000.

3. Теория электрической связи. Под редакцией Кловского Д.Д., М. «Радио и связь», 1998, с. 188.

Байесовский подход для оценивания заднего плана и обнаружения посторонних объектов на видеоизображении

Р.Р. Диязитдинов, В.А. Буцких
Поволжский государственный университет телекоммуникаций и информатики, Самара

Введение

Задача обнаружения посторонних объектов является хорошо изученной и практически решенной в настоящее время. Однако, для достижения приемлемых для практики результатов, очень часто используется цветное видеоизображение и дополнительные параметры, которые настраиваются исследователем вручную [1] [2] [3] [4] [6] (обычно это порог, при достижении которого регистрируется посторонний объект). Исключением из этого является метод, описанный в работе [5], в которой используется модифицированный фильтр Калмана.

Основываясь на критерии байесовского риска, мы предлагаем методы оценивания заднего плана видеоизображения и обнаружения посторонних объектов, которые в некоторой степени снимают проблему выбора параметров и могут работать с серыми изображениями.

Прежде чем описывать методы, введем понятие заднего плана (фоновое изображение, background) и переднего плана (посторонний объект, foreground) и опишем математическую модель исследуемого процесса. Сигнал заднего плана представляет собой неменяющийся во времени уровень яркости, если он является частью статичного объекта (например, асфальта, столба, ствола дерева и т.д.). Либо сигнал заднего плана, который представляет собой совокупность нескольких устойчивых состояний, характеризуемых своими уровнями яркости и вероятностями появлений, если он является частью динамического объекта (например, плывущие облака, колышущиеся ветки деревьев и т.д.). О сигнале переднего плана известно достаточно немного информации. Он, как правило, слабо коррелирован с сигналом переднего плана и сравним с ним по мощности.

Разделять сигналы переднего и заднего плана по энергии или их взаимной корреляции затруднительно или невозможно. Положим, что сигнал переднего плана имеет равномерное распределение, где границы распределения определяются максимальным и минимальным значением яркости, которые способна фиксировать видеокамера. Подобный подход используется в работе [6].

Обнаружение посторонних объектов

Критерий минимума среднего (байесовского) риска имеет вид [7]:

$$\frac{p(x \mid H_1)}{p(x \mid H_0)} \underset{<}{\geq} \frac{(R_{01} - R_{00}) \cdot \mathbf{P}\{H_0\}}{(R_{10} - R_{11}) \cdot \mathbf{P}\{H_1\}}, \qquad (1)$$

где H_1, H_0 – гипотеза о наличии и отсутствии сигнала в принятом сигнале x соответственно; $p(x \mid H_1)$, $p(x \mid H_0)$ – плотность вероятности наблюдаемого сигнала x при условии, что справедлива гипотеза H_1 и H_0 соответственно; $\mathbf{P}\{H_1\}$, $\mathbf{P}\{H_0\}$ – априорные вероятности наличия и отсутствия сигнала на выходе источника; R_{ij} – риск справедливости гипотезы H_i при принятии обнаружителем решения H_i, $i, j = \overline{0,1}$.

Обозначим:

$$\Lambda(x) = \frac{p(x \mid H_1)}{p(x \mid H_0)}, \ \eta = \frac{(R_{01} - R_{00}) \cdot \mathbf{P}\{H_0\}}{(R_{10} - R_{11}) \cdot \mathbf{P}\{H_1\}}.$$

Формула (1) перепишется в виде:

$$\Lambda(x) \gtrless \eta, \tag{2}$$

где $\Lambda(x)$ – отношение правдоподобия; η – порог обнаружителя.

Чтобы найти отношение правдоподобия, необходимо описать математическую модель наблюдаемых процессов, которую можно представить в виде:

$$x_I(t) = \theta\alpha_I s_I(t) + (1-\theta)\alpha_I r_I(t) + b_I + n_I(t), \tag{3}$$

где $x_I(t)$ – яркость I-того пикселя в момент времени t; $\theta = 1$, если верна гипотеза H_1, $\theta = 0$, если верна гипотеза H_0; $s_I(t)$ – сигнал заднего плана I-го пикселя; $r_I(t)$ – сигнал переднего плана (постороннего объекта) I-го пикселя; α_I – контрастность I-го пикселя; b_I - яркость I-го пикселя; $n_I(t)$ – шум в I-том пикселе.

Модель можно упростить, положив $b = 0$ (такое предположение оправдано, если солнце не светит в объектив камеры). Таким образом, можно упростить выражение (3):

$$x_I(t) = \theta\alpha_I s_I(t) + (1-\theta)\alpha_I r_I(t) + n_I(t). \tag{4}$$

В дальнейшем в обозначениях мы будем опускать символ I, подразумевая, что все вычисления применимы для каждого пикселя изображения.

Запишем плотности вероятности всех составляющих, записанных в формуле (4):

$$s(t):\quad p(s) = \sum_{i=1}^{K} p_i \cdot \delta(s - m_i);\ r(t):\quad p(r) = \begin{cases} \dfrac{1}{r_{\max} - r_{\min}}, r \in [r_{\min}, r_{\max}] \\ 0, r \notin [r_{\min}, r_{\max}] \end{cases};$$

$$n(t):\quad p(n) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{n^2}{2\sigma_n^2}\right);$$

$$\alpha:\quad p(\alpha) = \begin{cases} \dfrac{1}{\alpha_{\max} - \alpha_{\min}}, \alpha \in [\alpha_{\min}, \alpha_{\max}] \\ 0, \alpha \notin [\alpha_{\min}, \alpha_{\max}] \end{cases},$$

где p_i и m_i – вероятность и уровень яркости i-го состояния; K – количество состояний; $r_{\min}, r_{\max}$ – минимальный и максимальный уровень яркости, фиксируемых камерой, соответственно; $\alpha_{\min}, \alpha_{\max}$ – минимальный и максимальный уровень контрастности соответственно. $\alpha_{\min}$ – контрастность, соответствующая теням, $\alpha_{\max}$ – контрастность, учитывающая случайное изменение освещенности сцены.

Рассмотрим важный для практики случай, когда сигнал заднего плана характеризуется одним состоянием: $s(t):\quad p(s) = \delta(s - m_s)$. Это имеет место при съемке городских панорам, когда в поле зрения камеры не попадают деревья и облака. Найдем плотности вероятностей $p(x \mid H_1)$, $p(x \mid H_0)$ [8].

Вывод плотности вероятности для гипотезы H_1: $x(t) = \alpha \cdot s(t) + n(t)$.

$$p(x \mid H_1, \alpha) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(x - \alpha m_s)^2}{2\sigma_n^2}\right),\ p(x \mid H_1) = \int_{\alpha} p(x \mid H_0, \alpha) p(\alpha) d\alpha,$$

$$p(x \mid H_1) = \frac{1}{\alpha_{\max} - \alpha_{\min}} \int_{\alpha_{\min}}^{\alpha_{\max}} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(x - \alpha m_s)^2}{2\sigma_n^2}\right) d\alpha =$$

$$= \left| \begin{array}{l} t = \dfrac{x - \alpha m_s}{\sigma_n}; \\ \alpha = \dfrac{x - t \cdot \sigma_n}{m_s}; \\ d\alpha = -dt\dfrac{\sigma_n}{m_s}; \end{array} \begin{array}{l} \\ \alpha \to \alpha_{\min}; t \to \dfrac{x - \alpha_{\min} m_s}{\sigma_n}; \\ \alpha \to \alpha_{\max}; t \to \dfrac{x - \alpha_{\max} m_s}{\sigma_n}. \end{array} \right| = \frac{1}{\alpha_{\max} - \alpha_{\min}} \cdot \frac{1}{\sqrt{2\pi}\sigma_n} \cdot \left(-\frac{\sigma_n}{m_s} \right) \int\limits_{\frac{x - \alpha_{\min} m_s}{\sigma_n}}^{\frac{x - \alpha_{\max} m_s}{\sigma_n}} \exp\left(-\frac{t^2}{2} \right) dt =$$

$$= \frac{1}{2(\alpha_{\max} - \alpha_{\min}) m_s} \cdot \frac{2}{\sqrt{2\pi}} \int\limits_{\frac{x - \alpha_{\max} m_s}{\sigma_n}}^{\frac{x - \alpha_{\min} m_s}{\sigma_n}} \exp\left(-\frac{t^2}{2} \right) dt = \frac{1}{2(\alpha_{\max} - \alpha_{\min}) m_s} \left\{ \Phi\left(\frac{x - \alpha_{\min} m_s}{\sigma_n} \right) - \Phi\left(\frac{x - \alpha_{\max} m_s}{\sigma_n} \right) \right\},$$

где $\Phi(x) = \dfrac{2}{\sqrt{2\pi}} \cdot \int\limits_0^x \exp\left(-\dfrac{t^2}{2} \right) dt$ - функция Крампа.

Вывод плотности вероятности для гипотезы H_0 : $x(t) = \alpha \cdot r(t) + n(t)$.

Сначала найдем выражение плотности вероятности для величины $\alpha \cdot r(t)$:

$$X: \quad f(x) = \begin{cases} \dfrac{1}{r_{\max} - r_{\min}}, x \in [r_{\min}, r_{\max}] \\ 0, x \notin [r_{\min}, r_{\max}] \end{cases}, \; Y = \alpha X,$$

$$g(y) = f(\psi(y)) \cdot |\psi'(y)|, \; y = \alpha \cdot x, \; \psi(y) = \frac{y}{\alpha}, \; |\psi'(y)| = \frac{1}{\alpha}.$$

$$g(y) = f(\psi(y)) \cdot |\psi'(y)| = \begin{cases} \dfrac{1}{\alpha(r_{\max} - r_{\min})}, y \in [\alpha \cdot r_{\min}, \alpha \cdot r_{\max}], \\ 0, y \notin [\alpha \cdot r_{\min}, \alpha \cdot r_{\max}]; \end{cases}$$

Теперь найдем выражение плотности вероятности $p(x \mid H_0, \alpha)$:

$$X: f_1(x) = \begin{cases} \dfrac{1}{\alpha(r_{\max} - r_{\min})}, x \in [\alpha \cdot r_{\min}, \alpha \cdot r_{\max}], \\ 0, x \notin [\alpha \cdot r_{\min}, \alpha \cdot r_{\max}]; \end{cases}, \qquad Y: \quad f_2(x) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{y^2}{2\sigma_n^2} \right),$$

$Z = X + Y$.

$$g(z) = \int\limits_{-\infty}^{+\infty} f_1(x) f_2(z - x) dx = \int\limits_{\alpha \cdot r_{\min}}^{\alpha \cdot r_{\max}} \frac{1}{\alpha(r_{\max} - r_{\min})} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(z - x)^2}{2\sigma_n^2} \right) dx =$$

$$= \frac{1}{2\alpha(r_{\max} - r_{\min})} \cdot \left\{ \Phi\left(\frac{x - \alpha \cdot r_{\max}}{\sigma_n} \right) - \Phi\left(\frac{x - \alpha \cdot r_{\min}}{\sigma_n} \right) \right\},$$

$$p(x \mid H_0, \alpha) = \frac{1}{2\alpha(r_{\max} - r_{\min})} \cdot \left\{ \Phi\left(\frac{x - \alpha \cdot r_{\max}}{\sigma_n} \right) - \Phi\left(\frac{x - \alpha \cdot r_{\min}}{\sigma_n} \right) \right\}.$$

Выражение плотности вероятности $p(x \mid H_0)$:

$$p(x \mid H_0) = \int\limits_{\alpha} p(x \mid H_0, \alpha) p(\alpha) d\alpha,$$

$$p(x \mid H_0) = \frac{1}{2(\alpha_{\max} - \alpha_{\min}) \cdot (r_{\max} - r_{\min})} \cdot \int\limits_{\alpha_{\min}}^{\alpha_{\max}} \frac{\Phi\left(\dfrac{x - \alpha \cdot r_{\max}}{\sigma_n} \right) - \Phi\left(\dfrac{x - \alpha \cdot r_{\min}}{\sigma_n} \right)}{\alpha} d\alpha.$$

Таким образом,

$$p(x \mid H_1) = \frac{1}{2(\alpha_{\max} - \alpha_{\min})m_s} \left\{ \Phi\left(\frac{x - \alpha_{\min} m_s}{\sigma_n} \right) - \Phi\left(\frac{x - \alpha_{\max} m_s}{\sigma_n} \right) \right\}. \quad (5)$$

$$p(x \mid H_0) = \frac{1}{2(\alpha_{\max} - \alpha_{\min}) \cdot (r_{\max} - r_{\min})} \cdot \int_{\alpha_{\min}}^{\alpha_{\max}} \frac{\Phi\left(\frac{x - \alpha \cdot r_{\max}}{\sigma_n} \right) - \Phi\left(\frac{x - \alpha \cdot r_{\min}}{\sigma_n} \right)}{\alpha} d\alpha. \quad (6)$$

Значения функции Крампа описываются в виде таблицы. Интеграл в формуле (6) оформляется также в виде таблицы, так как все параметры нам известны: $[r_{\min}, r_{\max}]$ – при стандартной съемке по числу градаций серого оттенка – это [0; 255], $[\alpha_{\min}, \alpha_{\max}]$ – определяется из соображений здравого смысла ($0 \le \alpha_{\min} \le 1$, $\alpha_{\max} > 1$, для большинства случаев приемлем диапазон [0,2; 1,3]).

Задавая порог обнаружителя η и подставляя выражения (5) и (6) в формулу (2) мы получаем алгоритм обнаружения посторонних объектов.

Оценивание заднего плана

Как можно видеть в выражении (5) единственный параметр, который мы не определили – это m_s. Он определяет значение яркости заднего плана. Для его оценки обратимся опять к выражению:

$$x(t) = \theta \alpha s(t) + (1 - \theta) \alpha r(t) + n(t). \quad (7)$$

Положим, что вероятность появления сигнала постороннего объекта много меньше, чем вероятность появления заднего плана. Такое предположение является оправданным, например,9 для мест, где наблюдается непрерывное движение автотранспорта. Тогда можем записать, что $x(t) = \alpha s(t) + n(t)$.

Здесь необходимо сделать два замечания. Во-первых, мы оцениваем параметры на упрощенной модели, так как анализ выражение (7) приводит нас к так называемой задаче обучения без учителя [9], которая потребует слишком много машинного времени для ее решения. Во-вторых, задачу оценки параметра m_s мы сводим к задаче различения возможных значений [7], которые мы заранее определяем, так как задача оценки также является достаточно ресурсоемкой процедурой.

Пусть m_s может принимать одно из дискретных значений: m_1, m_2, $.. m_i$, $..$ m_M. Каждому значению m_i ставиться в соответствие гипотеза H_i. Плотность вероятности по результатам L измерений запишется как:

$$p(x \mid H_i, \alpha) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(- \frac{\sum_{j=1}^{L} (x_j - \alpha m_i)^2}{2\sigma_n^2} \right) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(- \frac{\left(\sum_{j=1}^{L} x_j^2 - 2\alpha m_i \cdot \sum_{j=1}^{L} x_j + L\alpha^2 m_i^2 \right)}{2\sigma_n^2} \right)$$

Обозначим $E_x = \sum_{j=1}^{L} x_j^2$, $S_x = \sum_{j=1}^{L} x_j$. В результате:

$$p(x \mid H_i, \alpha) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(- \frac{\left(E_x - 2\alpha m_i \cdot S_x + L\alpha^2 m_i^2 \right)}{2\sigma_n^2} \right), \; p(x \mid H_i) = \int_{\alpha} p(x \mid H_i, \alpha) p(\alpha) d\alpha.$$

$$p(x \mid H_i) = \frac{1}{\alpha_{\max} - \alpha_{\min}} \int_{\alpha_{\min}}^{\alpha_{\max}} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(- \frac{\left(E_x - 2\alpha m_i \cdot S_x + L\alpha^2 m_i^2 \right)}{2\sigma_n^2} \right) d\alpha.$$

$$p(x \mid H_i) = \frac{1}{2(\alpha_{\max} - \alpha_{\min})} \cdot \frac{1}{m_i \sqrt{L}} \cdot \exp\left(- \frac{E_x \cdot L - S_x^2}{2\sigma_n L} \right) \cdot \left\{ \Phi\left(\frac{S_x - \alpha_{\min} m_i L}{\sigma_n \sqrt{L}} \right) - \Phi\left(\frac{S_x - \alpha_{\max} m_i L}{\sigma_n \sqrt{L}} \right) \right\} \quad (8)$$

Алгоритм различения сигналов, оптимальный по критерию минимума среднего байесовского риска, можно записать в виде [11]:

$$\hat{k} = \arg\min_l \sum_{i=1}^{M} R_{il} \mathbf{P}\{H_i\} p(x \mid H_i), \quad (9)$$

где $\mathbf{P}\{H_i\}$ – априорная вероятность наличия сигнала m_i; R_{il} – риск справедливости гипотезы H_i при принятии решения H_l; $\hat{k}$ – характеризует принятое решение $m_{\hat{k}}$.

Для нашего случая: $\mathbf{P}\{H_i\} = \frac{1}{M}$, $R_{il} = (m_i - m_l)^2$. Также заметим, что в выражении (8) множитель $\frac{1}{2(\alpha_{\max} - \alpha_{\min})} \cdot \frac{1}{\sqrt{L}} \cdot \exp\left(- \frac{E_x \cdot L - S_x^2}{2\sigma_n L} \right)$ не влияет на результат вычисления операции аргумента минимума функции, поэтому выражение (9) можно записать:

$$\hat{k} = \arg\min_l \sum_{i=1}^{M} \frac{(m_i - m_l)^2}{m_i} \left\{ \Phi\left(\frac{S_x - \alpha_{\min} m_i L}{\sigma_n \sqrt{L}} \right) - \Phi\left(\frac{S_x - \alpha_{\max} m_i L}{\sigma_n \sqrt{L}} \right) \right\}. \quad (10)$$

Заключение

Хочется отметить, что в описанной методике не оговаривается каким образом определять дисперсию шума σ_n^2, которая входит во все формулы. Для определения этого параметра можно использовать либо паспортные данные на используемую видеокамеру, в которой прописана мощность шумов при съемке, либо в лабораторных условиях определить этот параметр.

Литература

1. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 780–785 (1997)
2. R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts and shadows in video streams”, IEEE Trans. on Patt. Anal. and Machine Intell., vol. 25, no. 10, Oct. 2003, pp. 1337-1342.
3. Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking. IEEE International Conference on Computer Vision and Pattern Recognition 1999;2:246–52.
4. Elgammal A, Harwood D, Davis LS. Non-parametric model for background subtraction. European Conference on Computer Vision 2000;2:751–67.
5. C. Ridder, O. Munkelt, H. Kirchner, Adaptive background estimation and foreground detection using kalman-filtering, in: Proceedings of International Conference on Recent Advances in Mechatronics, 1995, pp. 193–199.
6. K. Kim, T.H. Chalidabhongse, D. Harwood and L. Davis. Real-time Foreground-Background Segmentation using Codebook Model. Real-Time Imaging, vol.11, issue 3, pp.167-256, June 2005.
7. Горячкин О.В. – Лекции по статистической теории систем радиотехники и связи // 2008. - с.37-43, с.46-50.
8. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М. Высшая школа // 2004 – с. 121-122., с. 132.
9. Дуда Р., Харт П. Распознавание образов и анализ сцен. – М. Издательство «МИР» // 1976 – с. 224-231.

Интеграция модулей системы зашиты распределённых сетей

А.А. Иванов, М.В. Кузнецов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Общие положения

В рамках большой сети становиться проблематичным отслеживание состояния зашиты каждого узла, отсюда вытекает необходимость создания центра мониторинга и управления защитой. Для обеспечения комплексного подхода к организации информационной безопасности сети, необходима интеграция компонентов системы зашиты в рамках единого защитного контура.

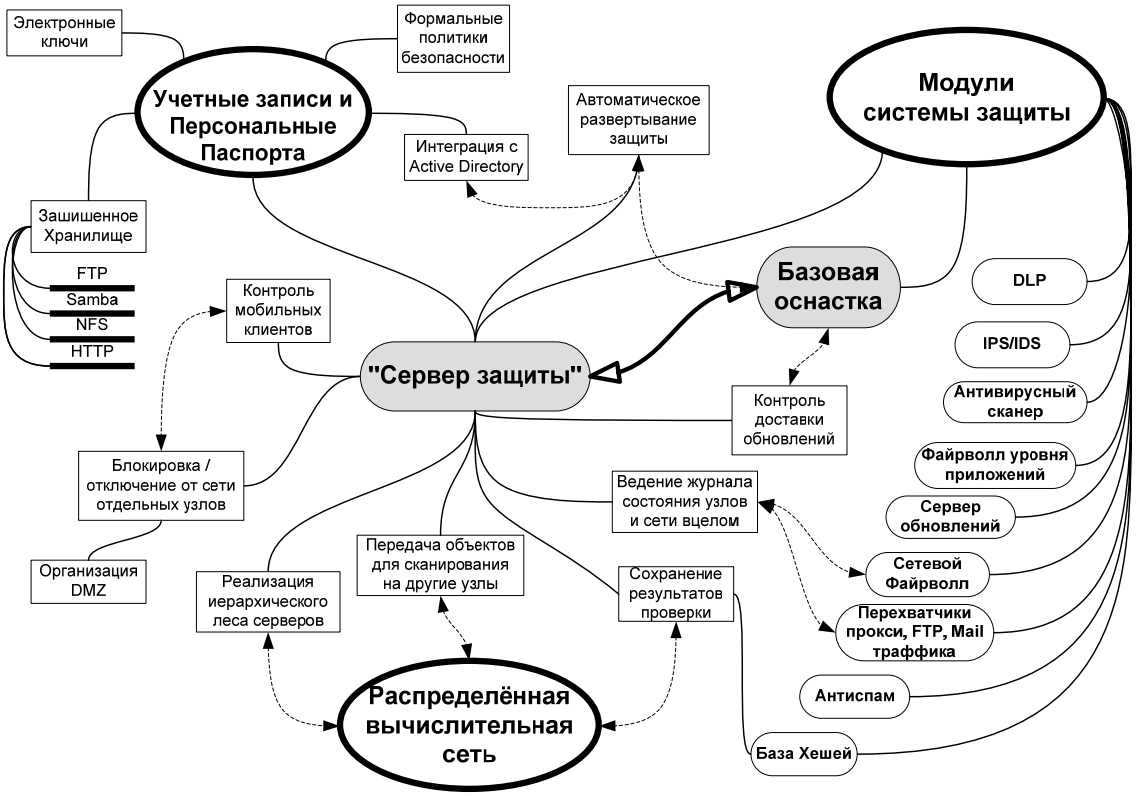


Схема 1. Структура функционирования системы защиты

Централизация управления

С целью облегчения контроля сложной структуры сети и повышения эффективности работы системы защиты предлагается ввести иерархию управляющих серверов, что позволяет реализовать:

- Автоматическое развертывание защиты в новых узлах сети

Набор правил, установленных на сервере защиты, может облегчить развертывание системы на новых узлах сети. Как показано на схеме 1, после того как оснастка установит связь с центром управления защитой, произойдет автоматическое импортирование наборов правил, в соответствии с ролью узла в сети;

- Если в сети присутствует контроллер домена, то возможно дополнение службы Active Directory функционалом по развертке системы защиты;
- Распределение нагрузок сегментов сети;

- Возможность автономной работы удаленных участков сети, сохраняя возможность централизованного контроля узлов сети.

Сервер защиты, помимо стационарных рабочих станций, может контролировать работу мобильных клиентов, согласно принятым политкикам безопасности, регулируя уровень доступа в сеть и к конфиденциальной информации.

Кроме того, система может реагировать на доступность управляющего сервера и, например, в случае его недоступности ограничить доступ к защищенному хранилищу на узле, или даже полностью блокировать его работу.

Для обеспечения скорости и надежности работы комплексной защиты необходимо реализовать:

- Систему распределения нагрузок между серверами, обеспечивающими функционирование защиты;
- Распределение системы аудита безопасности;
- Распределение объектов для проверки на сторонние доверенные клиенты;
- Систему автоматического резервирования состояния серверов.

Для повышения надёжности сети должна быть предусмотрена возможность подключения резервного управляющего сервера на случай неполадок с основным.

Также следует реализовать возможность использования относительно свободных клиентских машин для параллельного решения ресурсоёмких задач.

Возможности использования централизованного управления

В условиях корпоративной сети особо остро встает вопрос разграничения прав доступа, как к отдельным машинам, так и к информационным ресурсам.

В качестве решения этой задачи предлагается ввести систему контроля по учетным записям системы и по Персональным Цифровым Паспортам.

Как показано на схеме 1, использование системы учетных записей и Персональных Паспортов дает возможность, в том числе использовать формальные политики безопасности, как навесную защиту информации.

Также возможно создание интерфейса для использования электронных ключей для авторизации пользователей в рамках системы Персональных Паспортов.

В то время как учетная запись имеет привязку к учетной записи системы, персональный паспорт привязывается к конкретному человеку, с учетом его прав доступа и выполняемых обязанностей.

Режим работы системы защиты определяется в соответствии с той учётной записью, с которой произведен вход в систему.

На основе этого имеем возможность, без дополнительных манипуляций, на лету изменять режимы работы системы защиты, а именно:

- Корректировать список разрешенных/запрещенных программ для запуска;
- Контролировать уровень персонального доступа к сети;
- Производить фильтрацию пакетов на транспортном уровне OSI;
- Определять узлы, на которые данный пользователь имеет право доступа;
- Осуществлять контентную фильтрацию (доступ к определённым сайтам, получение файлов определенного типа);
- Ограничивать пропускную способность сети;
- Учитывать потребленный объем трафика.

Аутентификация и идентификация

Для более чёткого определения роли узла в сети и прав доступа конкретного пользователя наряду с учётной записью вводится Персональный Паспорт. Как и

«учетка» так и паспорт определяют поведение системы защиты, но в виду персональной аутентификации. Персональный Паспорт реализует следующие права:

- Имеет приоритет над настройками учетной записи;
- Выдается персонально, что дает возможность, работая в одном и том же программном окружении, иметь различный уровень доступа к объектам сети;
- В общем случае, можно предоставлять большие привилегии пользователю, авторизованному с помощью Персонального Паспорта;
- Возможность использования настроек «учетки» и ПП;
- Может являться одним из важнейших рубежей обороны против инсайдеров, введя разграничение доступа по персональным паспортам, мы можем контролировать и предотвращать попытки несанкционированного доступа к данным;
- Позволяет работать в одном программном окружении с теми же файлами, и в тоже время, иметь разные настройки зашиты, без привлечения администратора к реконфигурированию системы;
- При использовании на сервере RDP позволяет более гибко настраивать права доступа различных пользователей;
- Контроль по Персональному Паспорту позволяет уйти от использования ограничений операционной системы и использовать разграничение доступа методами системы защиты, независимо от используемой платформы.

Защищенное хранилище

Используя технологию авторизации по учётным записям и по Персональным Паспортам, организуется защищенное хранилище файлов.

Таким образом, организуется очень гибкая система контроля доступа:

- Уровень доступа определяется политикой безопасности установленной СЗ
  - можно использовать как канонические модели, так и «скриптовый» язык для описания собственной политики доступа к файлам;
- Можно следить за движением документов по сети предприятия
  - например, ввести разрешение на доступ к определенным документам только в соответствующем кабинете;
  - запретить доступ к сменным носителям и принтерам при работе с определёнными документами;
  - запретить доступ к конфиденциальной информации мобильным клиентам вне зоны предприятия (без связи с центральным сервером);
- К важным файлам вводится контроль на доступ только по ПП;
- Контроль запуска защищаемой информации на сменные носители;
- Регулирование прохождения файлов через сеть.

Обновления

Одним из важнейших условий функционирования системы защиты является своевременная загрузка обновлений необходимых программных модулей системы информационной безопасности. В крупной сети возникает проблема многократного выкачивания обновлений каждым отдельным компьютером в сети, это подводит к необходимости создания локального сервера обновлений.

Кроме того необходимо обеспечить удобный интерфейс опроса узлов сети на счет актуальности обновлений. А также добавить механизм принудительной инициации процедуры обновления при включении нового или временно отсутствовавшего абонента, в случае несоответствия установленной у пользователя версии системы защиты.

Таблица Хешей

Имея в виду однородность используемого программного обеспечения в сети можно ускорить работу антивирусного сканера на рабочих станциях с помощью введения в его состав сводной таблицы, в которой сопоставляются хеши файлов с результатами проверки.

Таким образом, если файл будет проверен на одном из компьютеров в сети, то, для приема решения о вредоносности или безопасности файла достаточно будет прохешировать файл и по таблице хешей получить готовый результат проверки. Если файл был проверен на шлюзе, или в файловом хранилище, он не будет повторно проверяться на рабочей станции.

Для работы такой системы необходимо предусмотреть надежный алгоритм зашиты самой системы информационной безопасности от атак, нацеленных на подделку результатов проверки. В случае, если вредоносный программный код будет занесен в таблицу как надежный, может произойти моментальное заражение всей сети.

Таблица, помимо классификации объекта согласно проверке, должна содержать информацию о методе проверки. Может быть полезным проводить поверхностное сканирование на шлюзе, а эвристический анализ уже на рабочих станциях.

Кроме того необходимо контролировать версию сигнатур, на модуле, средствами которого производится проверка поэтому, как видно на схеме 1, эта система должна координироваться с центральным сервером и системой обновлений.

Немаловажным является механизм, посредством которого будет происходить добавление результатов проверки в сводную таблицу хешей, и доставка самой таблицы на рабочие станции.

Удаленное сканирование

Работу антивирусов, занимающихся проверкой больших массивов данных, можно ускорить распределяя объекты для проверки на свободные в данный момент узлы в сети. Это позволит распределить нагрузку на системы занимающиеся проверкой файловых хранилищ, или трафика на, прокси-сервере или почтовом шлюзе.

Для этого надо реализовать:

- Механизм для передачи объектов на сканирование удаленный компьютер;
- Систему определения вычислительных ресурсов узла;
- Систему оценки целесообразности передачи файла на сторонние узлы. Может оказаться, что проверить файл локально окажется быстрее, чем передать его по сети и дождаться результатов проверки.

Таким образом, чётко координируя работу модулей системы зашиты можно добиться обеспечения информационной безопасности автоматизированных систем с учетом специфики функционирования конкретной сети. Это позволит увеличить как надежность, так и производительность системы. С введением защищенных хранилищ и контроля по персональным паспортам можно решить проблему инсайдеров и уменьшить влияние человеческого фактора на уровень защиты системы.

Предлагаемая система информационной безопасности позволяет контролировать как работу программ, так и работу персонала с сетью, оборудованием, и данными. Такая концепция построения системы безопасности поможет организовать защиту сетевых ресурсов как от внешних угроз и атак, основанных на уязвимостях прикладных программ и операционных систем, так и предотвратить угрозы, связанные с организационными вопросами политики безопасности предприятия.

Технология автоматического наполнения банка данных космических снимков в режиме постоянного приема

В.Н. Копенков [1], А.В. Чернов [2], Н.В. Чупшев [2]
[1] *Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*
[2] *Институт систем обработки изображений РАН, Самара*

Разработана технология, позволяющая в автоматическом режиме осуществлять радиометрическую коррекцию и межканальное совмещение, геопривязку космических снимков, геометрическую трансформацию в единую систему координат и помещение в банк данных в автоматическом режиме. Рассматриваемая задача актуальна при создании мультимасштабных и мультивременных покрытий территории (композитов) космическими снимками.

Введение

Задача анализа территорий на основе последовательностей космических снимков находит широкое применение в задачах оценки состояния посевов, природопользования и лесоустройства, предупреждения чрезвычайных ситуаций и т.д.[1,2]. При этом, как правило, используются снимки среднего разрешения (5-50 м) типа Spot 2/4, IRS 1C/1D P5/P6, LandSat 3-7, Монитор-Э, которые относительно недороги и могут поступать в режиме постоянного сброса на региональные станции приема (такая станция существует в Самарском государственном аэрокосмическом университете). Для совместного анализа снимков необходимо их привести к одному радиометрическому диапазону, к одной системе координат и, желательно, поместить в единое хранилище данных (региональный банк космических снимков - РБКС). Выделим особенности обработки таких снимков

– орбитальная привязка снимков весьма неточна (ошибка составляет до 500 метров), поэтому невозможно проводить геометрическую трансформацию только на основе этих данных;

– необходимость использования цифровых моделей рельефа (ЦМР) при геометрической трансформации;

– необходимость выделения полезной части снимка и отсечения облаков;

– большое количество снимков (несколько десятков в неделю объемом около 1 Гб на территорию области).

Все это ведет к большому времени ручного труда по приему, радиометрической коррекции и геопривязке снимков.

Цель работы – создать сквозную автоматизированную технологию предварительной обработки космических снимков и их помещения в РБКС, состоящую из следующих основных этапов:

1. Прием, нарезка на фрагменты, радиометрическая коррекция и предобработка данных.
2. Создание маски облачности
3. Автоматическая геопривязка и геометрическая трансформация
4. Создание метаданных и помещение изображения в РБКС
5. Яркостная коррекция изображений (приведение к динамическому диапазону исходного покрытия) – при необходимости.

Основной идеей является использование базового начального покрытия из космических снимков и геометрическая привязка к нему на основе корреляционного сопоставления с использованием информации о ЦМР.

В следующих параграфах каждый их этапов описывается более подробно.

Прием, нарезка на фрагменты, радиометрическая коррекция и предобработка данных

Исходные данные принимаются со спутника в виде полосы соответствующей пролету спутника над территорией. Для автоматической привязки данных SPOT необходимо представить данные в виде отдельных сцен, радиометрически-скорректированных и геометрически-трансформированных.

Этапы подготовки данных SPOT:

1) Получение данных со спутника в виде сырых данных файла пролета.
2) Обработки всех пролетов за день.



Обработка пролета состоит из следующих основных этапов:

a. Разрезание данных пролета на сегменты с разными режимами съемки.

b. Нарезка сегментов на сцены по технологии GRS и генерация метаданных для дальнейшей обработки

c. Генерация сцен с радиометрической нормализацией (по полученным метаданным).

d. Проверка принадлежности сцены заданной территории и сохранение сцен в специальную структуру каталогов для дальнейшей обработки.

e. Создание изображения сцены в натуральных цветах.

Для отдельных этапов технологии использовалось программное обеспечение, поставляемое вместе со станцией приема (http://www.scanex.ru). В работе была выполнена автоматизация этих этапов в виде единой технологической цепочки, включая резервное копирование и архивирование данных

При нарезке сегмента на сцены формируются файлы метаданных представляющими собой xml-файл с информацией относительно источника данных, параметров съемочной аппаратуры, предварительной привязки изображения и т.д.

Под радиометрической нормализацией данных понимается калибровка многозональных снимков в оптическом диапазоне:

$$B_\lambda^* = K_\lambda N + C_\lambda ,$$

где B_λ^* – энергетическая яркость для спектральной зоны λ, N – «сырые» значения яркости, K_λ – калибровочный коэффициент, C_λ – калибровочная константа, соответствующая минимальной величине регистрируемой яркости. Значения K_λ и C_λ распространяются вместе с цифровыми снимками в виде специального калибровочного файла.

Для создания изображения в натуральных цветах используются панхроматическая и мультиспектральная сцены. При этом мультиспектральное изображение используется для преобразования в натуральные цвета, а панхроматическая сцена для повышения разрешения.

Мультиспектральные изображения со спутника SPOT не имеют синего канала. Поэтому необходимо синтезировать изображение, имеющее синий канал, путем

созданием композитного цветного изображения, каналы которого рассчитываются на основе линейной комбинации каналов исходного мультиспектрального изображения:

$$R = Xs_2\ ;\ G = \frac{3Xs_1 + Xs_3}{4}\ ;\ B = \frac{3Xs_1 - Xs_3}{4}$$

где R, G и B – натуральные цветовые каналы,

Xs_1, Xs_2, Xs_3– исходные зеленый, красный и инфракрасный каналы,

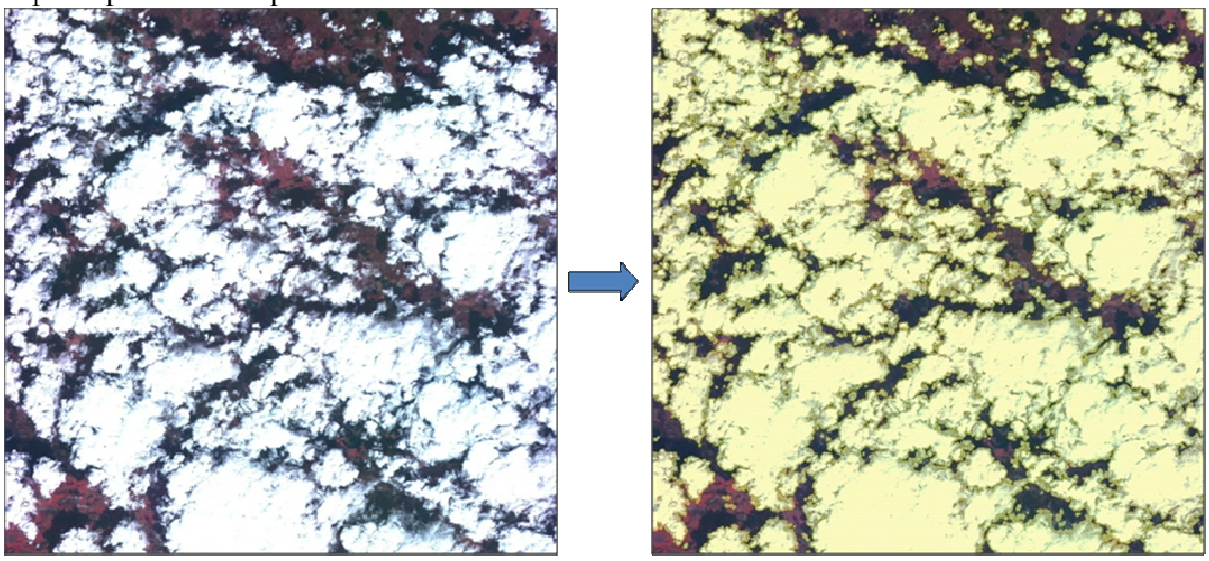
После получения изображения в натуральных цветах для повышения его разрешения используется HSI метод, состоящий в переводе исходного изображения в пространство HSI (Hue, Saturation, Intensity) и замене интенсивности на панхроматический канал [2,3] (Image Fusion). После проведения описанных операций производится обратное преобразование к исходным цветам.

Таким образом, в результате описанной процедуры получаем цветное изображение, в наиболее привычной человеческому глазу цветовой гамме (натуральных цветах) с разрешением соответствующем панхроматической сцене.



Создание маски облачности

На основе синтезированного изображения в натуральных цветах производится построение маски облачности. Для этого осуществляется пороговая обработка в цветовом пространстве HSI, с автоматическим выбором порога исходя из яркостных характеристик изображения.



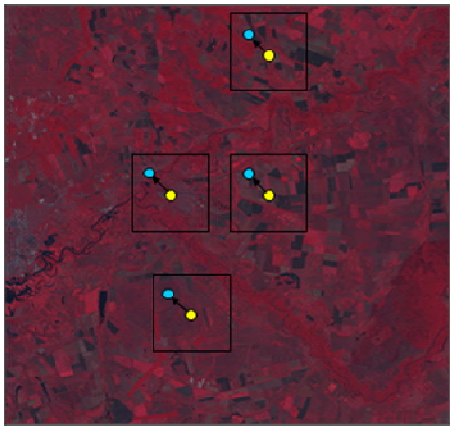
Автоматическая привязка сцен

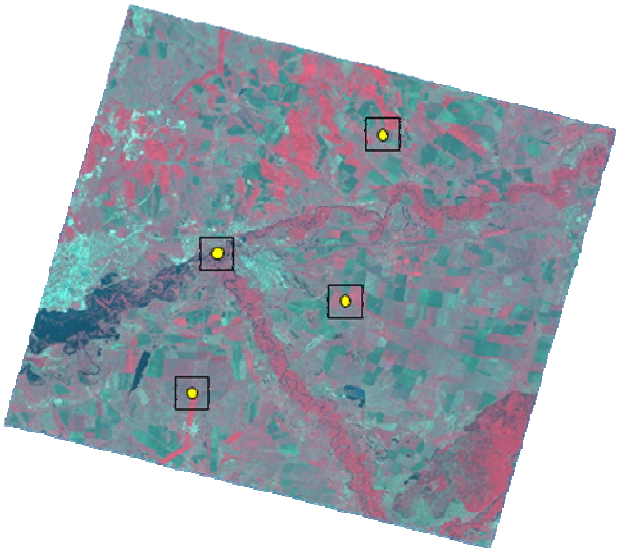
Автоматическая привязка осуществляется на основе обнаружения на привязываемом изображении опорных точек, хранимых в базе данных. Определение положения опорной точки производится путем использования корреляционного алгоритма обнаружения шаблона с центром в данной опорной точке, вырезаемого с базового покрытия [1]. При вырезании может потребоваться дополнительное преобразование между системами координат. Следует отметить, что точки, попавшие в

маску облачности, исключаются из рассмотрения. В качестве меры сходства используется нормированный коэффициент корреляции:

$$R_s(m,n) = \frac{\sum_j \sum_k G_1(j,k) G_2(j-m,k-n)}{\sqrt{\sum_j \sum_k G_1^2(j,k) \cdot \sum_j \sum_k G_2^2(j-m,k-n)}},$$

где $G_1(j,k)$ - центрированные отсчеты зоны поиска на привязываемом изображении $G_2(j,k)$ - центрированные отсчеты шаблона поиска.





Далее на основе найденных точек соответствия производится геометрическое преобразование изображения. Настройка преобразования выполняется с итеративным отбросом точек по критерию максимальной погрешности. При геометрической трансформации также необходимо использовать цифровую модель рельефа (ЦМР), так как преобразования без нее учета не дают нужную точность.

Создание метаданных и помещение изображения в РБКС

Для помещения привязанной сцены в РБКС необходимо создать файл метаданных[4], содержащий параметры помещения, привязки, описание сцены, характеристики съемочной аппаратуры и т.д. Этапы помещения данных SPOT в РБКС:

1) Создание файла метаданных формата xml с предварительными данными соответствующими определенной сцене (данные о файлах маски, контура, дополнительные файлы и т.д.)

2) Добавление в xml-файл метаданных сцены (размеры, границы, кол-во каналов, битность и т.д.)

3) Получение и добавление в файл метаданных параметров съемки (азимут, угол солнца, тип продукта, дата съемки и т.д.)

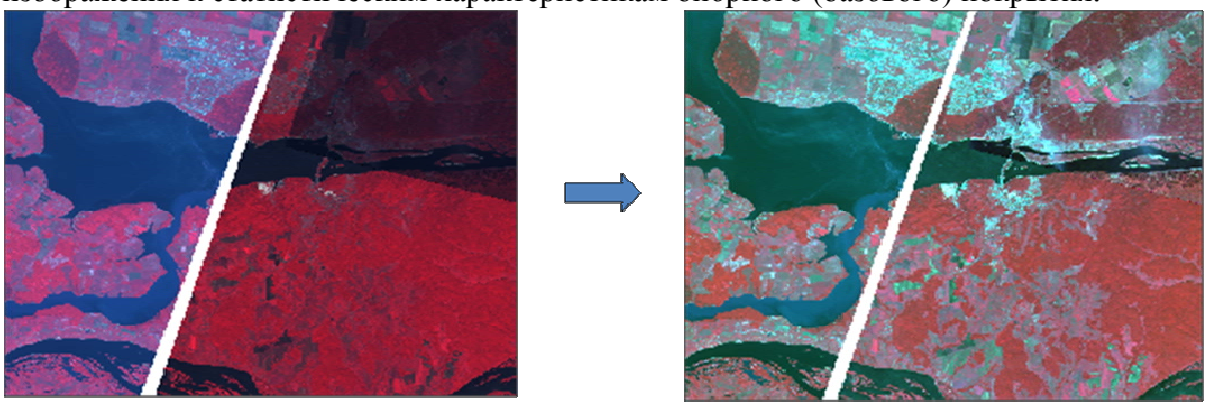
4) Добавление точек привязки и параметров привязки с проверкой возможности помещения сцены (облачность, точность привязки и т.д.)

5) Добавление следующих растровых изображений в РБКС по каждому космическому снимку: панхроматическая сцена, сцены в исходных цветах, и натуральных цветах, а также, при необходимости, производные продукты. Добавление информации из файла метаданных в базу метаданных.

Более подробная информация о структуре каталогов и формате РБКС представлена в статье «Региональный банк космических снимков самарской области»[5].

Яркостная коррекция изображений

Поступающие снимки могут иметь яркостные характеристики, отличающиеся от их аналогов для базового покрытия. Возникает задача улучшения их визуального восприятия и создания бесшовных композитов с помощью дополнительного приведения к одному яркостному диапазону. Для этого используется преобразование яркости и контраста, приводящие статистические характеристики каналов исходного изображения к статистическим характеристикам опорного (базового) покрытия.



Заключение

Разработана технология, позволяющая производить автоматическую предварительную обработку получаемых данных со спутника SPOT для дальнейшего их тематического совместного анализа. Предложенный подход с некоторыми модификациями может быть применен для обработки снимков с других космических аппаратов.

Информационная технология внедрена в Самарском государственном аэрокосмическом университете, НП «Поволжский центр космической геоинформатики» и используется при автоматическом наполнении регионального банка данных космических снимков Самарской области.

Литература

1. W. Pratt, Digital Image Processing // Wiley. – 2001.
2. R. Gonzelez, R. Woods. Digital image processing // Pearson Education. – 2002.
3. C. Pohl, J.L. Van Genderen. Multisensor image fusion in remote sensing: Concepts, methods and application // International Journal of Remote Sensing, Vol. 19, No. 5, pp. 823-854
4. B. Zitova, J. Flusser. Image registration: a survey // Image and Vision Computing. – 2003. pp. 977–1000
5. Chernov A.V., Gashnikov M.V., Glumov N.I., Sergeyev V.V. Regional bank of samara region satellite images //Proceedings of 9-th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-9-2008), Russian Federation, Nizhni Novgorod, September 15—19, 2008, Vol.1, pp. 162-165/

Анализ изображений на наличие локальных искусственных изменений со свойствами JPEG сжатия

А.В. Кузнецов
Институт систем обработки изображений РАН, Самара

В статье решается задача обнаружения локальных искусственных изменений [1] (фальсификаций) со свойствами JPEG сжатия. Известные методы обнаружения таких изменений [2, 3, 4] описывают только отличительные свойства изображений, подвергавшихся сжатию алгоритмом JPEG, от изображений без компрессии. В данной работе был разработаны алгоритм обнаружения на изображениях локальных встраиваний со свойствами сжатия и алгоритм определения смещения встроенных JPEG блоков относительно координат встраивания, кратных 8, а также получена зависимость периода пиков на спектре гистограммы коэффициентов ДКП от коэффициента качества алгоритма сжатия JPEG. В ходе исследований были получены численные результаты качества верного и ложного обнаружения разработанного алгоритма обнаружения встраиваний.

Введение

Алгоритм JPEG является стандартной процедурой сжатия изображений с потерями, основанной на применении дискретного косинусного преобразования (ДКП) [5]. Алгоритм сжатия состоит из трех последовательных шагов:

1. ДКП.

Изображение разбивается на блоки $f(x,y)$ с размерами 8×8 значения переводятся из диапазона $[0,255]$ в диапазон $[-128,127]$, затем вычисляется ДКП.

$$F(\omega_x,\omega_y)=\frac{1}{4}c(\omega_x)c(\omega_y)\sum_{x=0}^{7}\sum_{y=0}^{7}f(x,y)\cos\frac{(2x+1)\omega_x\pi}{16}\cos\frac{(2y+1)\omega_y\pi}{16},$$

$$\omega_x,\omega_y=0,..,7,$$

$$c(\omega)=\begin{cases}\frac{1}{\sqrt{2}},\omega=0\\ 1,\omega=1\end{cases}.$$

2. Квантование.

Квантованные значения получаются делением на шаг квантования и округлением до ближайшего целого числа:

$$F_q(\omega_x,\omega_y)=\left\lfloor\frac{F(\omega_x,\omega_y)}{s(\omega_x,\omega_y)}+\frac{1}{2}\right\rfloor,$$

$$\omega_x,\omega_y=0,..,7,$$

$$s(\omega_x,\omega_y)=\begin{cases}\max\left(\left\lfloor\frac{200-2Q}{100}C(\omega_x,\omega_y)+\frac{1}{2}\right\rfloor,1\right),Q\in[50,100]\\ \frac{50}{Q}C(\omega_x,\omega_y)+\frac{1}{2},Q\in(0,50)\end{cases}.$$

Матрица квантования $C(\omega_x,\omega_y)$, определенная экспериментально при разработке стандарта JPEG, является фиксированной.

3. Кодирование (например, алгоритм Хаффмана).

Алгоритм восстановления изображения, сжатого алгоритмом JPEG, заключается в следующем:

1. Декодирование сигнала.
2. Деквантование.
3. Обратное ДКП.

При сжатии цветных изображений RGB-компоненты переводятся в цветовое пространство YCrCb, для которых и применяется вышеописанная процедура.

Постановка задачи

Пусть $x[t]$ – одномерный дискретный сигнал. Тогда квантование будет описано следующей формулой:

$$q_a(u) = \left[\frac{u}{a}\right]$$

где a – шаг квантования, а u – значение выборки $x[t]$. Деквантование вычисляется по следующей формуле: $q_a^{-1}(u) = au$. Следует иметь в виду, что функция $q_a(u)$ необратима, а операция деквантования не является обратной для операции квантования. Операция двойного квантования может быть представлена в виде:

$$q_{ab}(u) = \left[\left[\frac{u}{b}\right]\frac{b}{a}\right]$$

где a, b – шаги квантования. Таким образом, двойное квантование может быть сделано за 3 операции: квантование с шагом b, деквантование с тем же шагом b и квантование с шагом a.

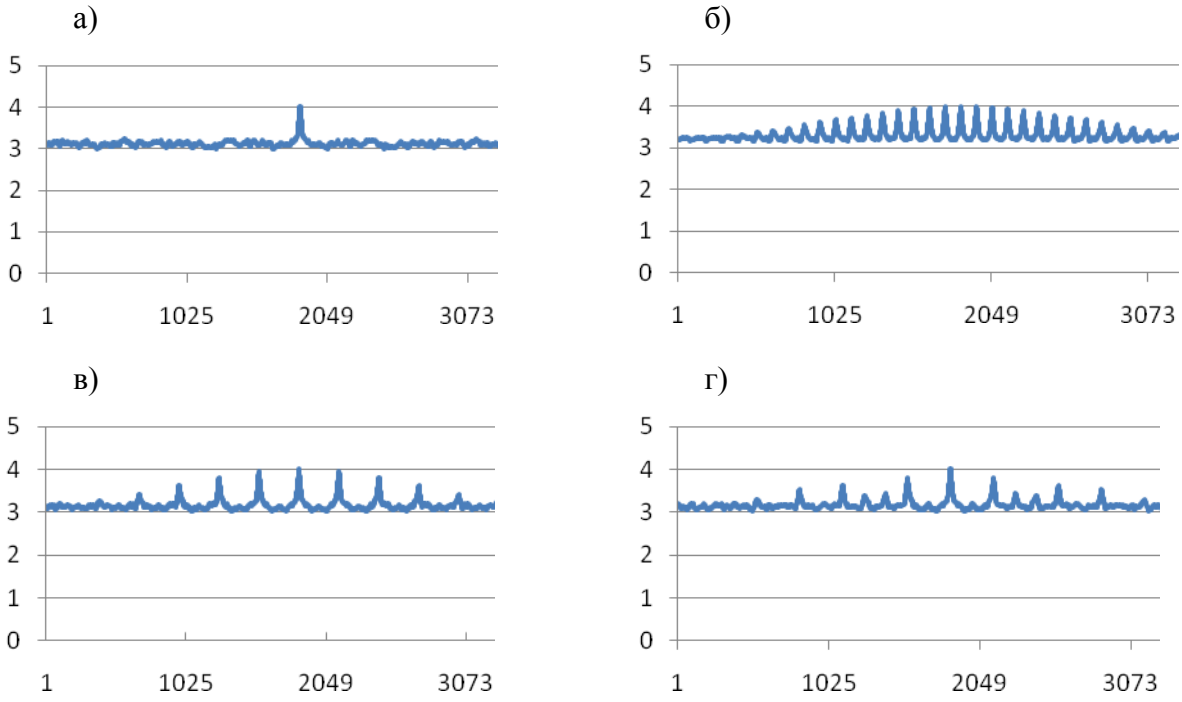


Рисунок 1 – Спектры гистограмм компоненты (0,7) ДКП: а) для исходного изображения; б) для сжатого изображения при Q =30; в) для сжатого изображения при Q =70; г) для дважды сжатого изображения при Q =70 и 90

Артефакты, проявляющиеся после квантования (однократного и двукратного) в алгоритме JPEG сжатия, визуально видны на спектрах гистограмм выбранной компоненты ДКП сигнала для разных случаев (рисунок 1). На рисунке 1а показан исходный сигнал без сжатия, поэтому на спектре один максимальный пик в центре. На

рисунке 1б и 1в представлены спектры гистограмм областей, подвергнутых JPEG сжатию с различными коэффициентами сжатия, о чем свидетельствуют многочисленные периодичные пики на спектрах. В зависимости от качества сжатия период пиков также меняется. На рисунке 1г показан спектр гистограммы области, двукратно сжатой алгоритмом JPEG, что видно по наличию второстепенных пиков, нарушающих монотонность убывания значений всех пиков на спектре.

Предлагаемое решение

Анализируя спектры гистограмм коэффициентов ДКП можно определить факт применения JPEG сжатия для встроенной части изображения. При этом возможны ситуации – JPEG сжатие не производилось, производилось однократно, производилось двукратно с различными параметрами качества. Был разработан следующий алгоритм обнаружения локальных встраиваний со свойствами JPEG сжатия:

1. Для цветных изображений производится преобразование в полутоновое, вычисляется Y компонента (YCrCb) и дальнейшая работа производится только над ней.

2. Вычисление основного сдвига матрицы блоков 8×8 Данная операция производится для того, чтобы обеспечить обнаружение встроенных JPEG блоков со смещением относительно исходной сетки.

3. Предварительная кластеризация блоков с учетом основного сдвига. На данном этапе осуществляется отделение блоков фонового изображения от блоков, обладающих какими-либо свойствами JPEG сжатия. В результате этой процедуры формируются два кластера: в первом находятся координаты блоков, не обладающих свойствами периодичности и монотонности, во втором, соответственно, все остальные, которые на последующем этапе кластеризации будут разделяться по качеству и кратности сжатий JPEG.

4. Кластеризация блоков изображения при найденном основном сдвиге. Она представляет собой итерационную процедуру, которая повторяется, пока не произойдет разделения блоков по кластерам или число кластеров не достигнет максимально допустимого.

Исследование алгоритма

В таблице 1 и на рисунке 2 приведены результаты исследований работы алгоритма обнаружения встраиваний с различными JPEG свойствами. Размеры встроенных областей менялись от 64×64 до 512×512 при размерах входных космических изображений от 1024×1024 до 4096×4096 пикселов.

Таблица 1 – Точность обнаружения двойного сжатия JPEG при разных комбинациях параметров качества

Q2 \ Q1	0%	0%	0%	0%	0%
50 %					.9
60 %					.95
70 %					.95
80 %			.95		.8
90 %	.95	.95	.95	.75	

Представленные данные в таблице и на рисунках являются средними по выборке из 40 космических изображений (Spot, Eros, Quickbird, Cartosat). По ним видно, что качество обнаружения очень близко к 1 в большинстве случаев. Это говорит о высокой точности разработанного алгоритма.

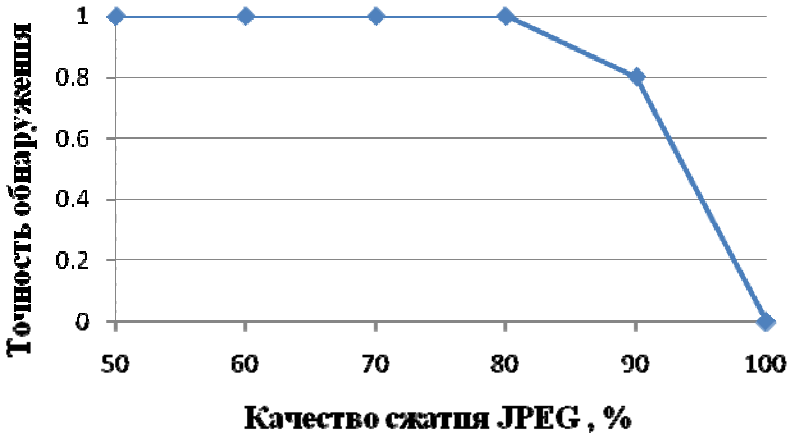


Рисунок 2 – Точность обнаружения однократного сжатия JPEG

Также в ходе исследований была получена зависимость между качеством однократного сжатия JPEG от значений периодов пиков на спектре коэффициентов ДКП. Результаты представлены на рисунке 3. При помощи данного графика, вычислив период по спектру гистограммы коэффициента ДКП, можно получить исходное значение качества сжатия JPEG.

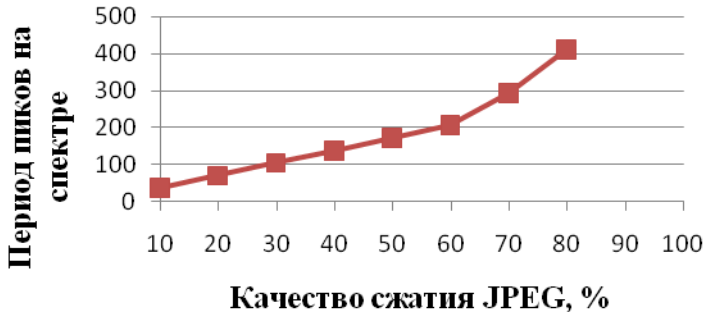


Рисунок 3 – Зависимость качества сжатия JPEG от значений периодов пиков на спектре коэффициентов ДКП

Выводы

В результате данного исследования был разработан алгоритм для обнаружения локальных искусственных изменений, отличающихся свойствами JPEG сжатия. Результаты экспериментов по оценке точности разработанного алгоритма показали возможность его применения на практике, для анализа крупноразмерных космических изображений земной поверхности. Также была выявлена зависимость между значениями периодов пиков на спектре гистограммы коэффициентов ДКП и качеством сжатия алгоритма JPEG.

Литература

1. Kuznetsov A.V, Glumov N.I. Local artificial image's changes detection. – Proceedings of the IASTED International Conference on Automation, Control, and Information Technology: Information and Communication Technology (ACIT-ICT 2010), June 15 – 18, 2010, Novosibirsk, Russia. 2010, P. 40-45.

2. A. C. Popescu. Statistical Tools for Digital Image Forensics. – PhD thesis, Dartmouth College, Department of Computer Science, Hanover, USA. – 2005. – P. 11-102.

3. J. Fridrich, J. Lukas. Estimation of primary quantization matrix in double compressed JPEG images. – Digital Forensic Research Workshop. – 2003. – P. 2-5.

4. Zhigang Fan, Ricardo L. de Queiroz. Identification of Bitmap Compression History: JPEG Detection and Quantizer Estimation. – IEEE Transactions On Image Processing, vol. 12, № 2 – 2003. – P. 230-235.

5. Методы компьютерной обработки изображений. Под ред. В.А. Сойфера. М.: ФИЗМАТЛИТ, 2003. 784 с.

Организация распределённых вычислений

И.И. Меркульев, М.В. Кузнецов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Организация распределенных вычислений является одним из самых востребованных направлений в сфере повышения быстродействия работы различных систем. Критичность быстродействия отельных задач особенно остро проявляется на маломощных пользовательских машинах, которые, согласно статистике, составляют более половины машинного парка средней компании. Согласно нашим исследованиям, основную нагрузку на персональные компьютеры накладывают системы безопасности, которые потребляют чрезмерно много вычислительных ресурсов, что приводит к сбоям в работе компьютера и, как следствие, к незапланированным губительным простоям в работе персонала. Именно поэтому, мы видим решение данной проблемы в том, что система безопасности должна использовать все преимущества распределенных вычислений, что позволит значительно уменьшить нагрузку на отдельные машины.

Опишем методологию организации распределенных вычислений. Функция организации распределённых вычислений в сети принадлежит центральному серверу. Для поддержки организации таких вычислений оснастка каждого конечного узла сети должна предоставлять весь необходимый функционал. Каждая корпоративная сеть, использующая технологию распределенных вычислений, становится так называемым «виртуальным суперкомпьютером». Или, иначе говоря, речь идет о GRID вычислениях.

Детализируя данную идею, необходимо отметить, что процесс организации должен основываться на собранной статистической информации о функционировании сети в течение некоторого периода, продолжительность которого зависит от количества компьютеров в сети и её топологии. Для сбора статистики в системе безопасности присутствует весь необходимый инструментарий, то есть, выполняя свои первичные функции защиты сети, система параллельно занимается сбором информации. В ходе работы получено, что длительность промежутка времени, в течение которого производится сбор данных, не имеет верхней границы критичности, то есть механизм сбора статистики будет функционировать постоянно, что позволит лучше распределять нагрузку между конечными узлами сети. В число собираемых характеристик должны входить: средний объём анализируемого трафика, необходимый конечному узлу в день, и количество запросов в день на распределённую проверку. Центром сбора и хранения такой информации будет являться центральный сервер, а инструментами сбора - механизм опроса оснасток. Цель сбора статистики – достижение оптимального значения быстродействия при решении поставленных задач.

Система безопасности располагает следующими методами сбора статистики:

1) **Сбор статистики в течение всего времени работы**. Преимущества – необходимо меньше времени на достижения оптимальных значений или значений, близких к оптимальным. Недостатки – оптимальные значение для данного метода гораздо ниже, т.к. оно не учитывает загруженности сети в разные интервалы времени.

2) **Сбор статистики в течение интервалов времени**. Является «обратным» методом к изложенному выше. Преимущества – учитывает загруженность сети в разные интервалы времени. Недостатки – необходимо собирать гораздо больше информации, объем которой растет с уменьшением длины временного интервала.

3) **Непостоянный сбор статистики**. При действующем механизме интервального сбора работает механизм пересчета значения быстродействия. Данный

механизм пересчета строит график изменения таких значений для каждого из интервалов и определяет степень улучшения этих значений. Если улучшение на каком-то из интервалов ниже определенного значения, то механизм справки предлагает администратору отключить сбор статистики на данном интервале ввиду малой эффективности результата. Например, это актуально в ночное время. Также, администратор может совершить аналогичные действия и без уведомления системы. Преимущества – частично исправляется главный недостаток интервального метода со сбором большого количества информации. Недостатки – при автоматическом переходе необходим работающий изначально интервальный метод. При ручном переходе все зависит от человеческого фактора.

4) **Отключенный сбор информации**. Так называемый, нулевой метод. Здесь также функционирует механизм пересчета значений быстродействия, который по истечении определенного промежутка времени определит, что оптимальное значение достигнуто и дальнейший пересчет отнимет больше ресурсов, чем поможет сэкономить. Тогда механизм справки информирует об этом администратора и предлагает отключить сбор статистики. Однако, администратор в праве отключить и включить сбор статистики в любое время, не дожидаясь оповещений системы.

Помимо верстки ранговой таблицы загруженности конечных узлов, организация распределенных вычислений предполагает создание аналогичной таблицы рангов для выполняемых задач. То есть, каждый тип задачи имеет свой приоритет, заданный в системе. Новые задачи с более высоким приоритетом будут находиться в очереди на выполнения выше, чем старые задачи с более низким приоритетом. Задачи с равным приоритетом будут ранжироваться по времени поступления запроса. В любой момент времени системный администратор может открыть текущую очередь задач, просмотреть процесс выполнения отдельных задач, удалить из очереди не актуальные, по его мнению, задачи, добавить на любое место в очереди новую задачу или переместить (читай, повысить/понизить приоритет) какую-либо задачу.

В ходе данной работы выяснено, что наиболее эффективно можно повысить быстродействие тех задач, которые возникают не спонтанно, а известны заранее. К таким задачам относятся, в первую очередь, некие запланированные события. А потому, система будет рекомендовать администратору использовать интервальный метод сбора статистики как минимум для выполнения таких задач. То есть, механизм сбора статистики будет использовать только ту информацию, которая была собрана в небольшом временном интервале, включающим в себя время выполнения запланированной задачи. Таким образом, каждая запланированная задача будет использовать свою собственную таблицу ранжирования, которая наиболее точно отражает загруженность сети во время выполнения запланированной задачи.

Таким образом, предложенная организация распределённых вычислений в системе безопасности представляет собой самообучаемый механизм, который с течением времени повышает эффективность работы всей системы. Предусмотрено множество методов сбора статистики, каждый из которых подходит отдельному типу корпоративной сети. Реализован прогрессивный механизм составления таблицы рангов, который адекватно отражает загруженность сети в данный момент времени (при выборе определенных методов сбора статистики). Однако, несмотря на мощный аппарат сбора и обработки информации о сети, администратор имеет достаточно полномочий, чтобы внести свои корректировки в процесс повышения быстродействия системы.

Разработка метода снижения размерности пространства признаков на основе опорных узлов

Е.В. Мясников
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара
Институт систем обработки изображений РАН, Самара

Введение

Нелинейные методы снижения размерности многомерных данных можно рассматривать не только как средство уменьшения размерности исходных данных с целью получения более компактного представления, но и как средство индикации многомерных данных в пространствах низкой размерности для анализа [3]. Снижение размерности данных до размерности 2 или 3 (построение двух- или трехмерных отображений), позволяющее получать представление о структуре многомерных данных, в контексте коллекций цифровых изображений приобретает еще большую практическую ценность, позволяя строить системы навигации по коллекциям цифровых изображений на основе признаковой информации [3,7].

Одним из наиболее распространенных и признанных методов нелинейного снижения размерности в области распознавания образов является метод Сэммона [3,22]. Данный метод позволяет минимизировать ошибку представления многомерных данных, выражаемую в виде:

$$\varepsilon = \frac{1}{\sum\limits_{\substack{o_i,o_j \in O \\ i<j}} d(o_i,o_j)} \cdot \sum_{\substack{o_i,o_j \in O \\ i<j}} \frac{\left(d(o_i,o_j) - d^*(o_i,o_j)\right)^2}{d(o_i,o_j)}. \qquad (1)$$

Здесь $d(o_i,o_j)$ и $d^*(o_i,o_j)$ - расстояние между объектами o_i и o_j, соответственно, в многомерном пространстве и пространстве малой размерности, O - все множество объектов.

Алгоритм, используемый в методе Сэммона для минимизации ошибки (1), носит итеративный характер. Он базируется на следующем рекуррентном соотношении для координат $\mathbf{y}_i$ в пространстве малой размерности:

$$\mathbf{y}_i(t+1) = \mathbf{y}_i(t) + m \cdot \boldsymbol{\sigma}_i(t), \qquad (2)$$

$$m = \frac{2 \cdot \alpha}{\sum\limits_{\substack{o_i,o_j \in O \\ i<j}} d(o_i,o_j)}, \quad \boldsymbol{\sigma}_i(t) = \sum_{\substack{o_j \in O \\ j \neq i}} \frac{d(o_i,o_j) - d^*(o_i,o_j)}{d(o_i,o_j) \cdot d^*(o_i,o_j)} \cdot \left(\mathbf{y}_i(t) - \mathbf{y}_j(t)\right). \qquad (3)$$

Настраиваемый параметр α влияет на скорость работы и сходимость алгоритма.

Этот алгоритм нелинейного снижения размерности хорошо работает на небольших объемах данных, однако на больших объемах возможность применения алгоритма ограничивает его высокая вычислительная сложность. Вследствие этого исследователями предпринимаются попытки построения методов, имеющих пониженную вычислительную сложность. Среди таких методов можно выделить: метод триангуляции [5, 6], линейного преобразования [6], стохастической аппроксимации [1,4], а также комбинированные методы, основанные на иерархической кластеризации [1,2].

Предлагаемый в работе метод снижения размерности пространства признаков основан на аппроксимации приращений координат точек. В отличие от предложенных

ранее методов, такую аппроксимацию предлагается выполнять с использованием списков опорных узлов, создаваемых на основе результатов иерархической кластеризации данных в исходном многомерном пространстве.

Описание предлагаемого метода снижения размерности

Разработанный метод снижения размерности состоит из четырех описанных ниже шагов:

1. Построение иерархии кластеров на основе координат в исходном пространстве.

Под иерархией кластеров будем понимать древовидную структуру данных, корнем которой является кластер верхнего уровня C^0, а каждая вершина-кластер C_i^l на уровне $l > 0$ содержит либо подкластеры $\left\{C_{i_1}^{l+1}, C_{i_2}^{l+1}, ..., C_{i_{|C_i^l|}}^{l+1}\right\}$, либо непосредственно объекты $\left\{o_{i_1}, o_{i_2}, ..., o_{i_{|C_i^l|}}\right\}$ исходного множества.

Для построения иерархии вложенных кластеров могут быть применены как алгоритмы иерархической агломеративной кластеризации, так и неиерархические алгоритмы кластеризации, запускаемые в рекурсивном порядке (например, алгоритм k-внутригрупповых средних, нейросетевые алгоритмы и т. д.).

2. Инициализация координат объектов в целевом пространстве

Нелинейный метод снижения размерности Сэммона, принятый в настоящей работе в качестве базового, позволяет строить начальное приближение различными способами. В частности, в литературе встречаются способы инициализации случайными значениями, результатами работы метода PCA [6]. При этом отмечено [3], что инициализация с использованием метода PCA позволяет обеспечить в среднем снижение размерности с более высоким качеством.

3. Построение списков опорных узлов для каждого из объектов

Под опорным узлом объекта o в настоящей работе будем понимать некоторый объект $o_i \neq o$ или группу объектов $\{o_i\}_{i=1..N}$, обладающих в исходном многомерном пространстве близкими характеристиками и рассматриваемых как единое целое. Под списком опорных узлов S объекта o будем понимать упорядоченное множество опорных узлов объекта o.

Формирование списка опорных узлов выполняется для каждого из объектов один раз до запуска процесса оптимизации. При построении списка опорных узлов для каждого из объектов принимается во внимание тот факт, что при построении аппроксимации приращений координат объекты, расположенные вблизи рассматриваемого, следует учитывать по-отдельности, а группы объектов, расположенные в удалении от рассматриваемого, возможно учитывать совместно.

Пусть o - некоторый объект, для которого требуется сформировать список S опорных узлов, C - произвольный кластер иерархии. Тогда формирование списка осуществляется с использованием следующего рекурсивного алгоритма.

1. Если кластер C близок к рассматриваемому объекту $r(C,o) \le T$ и в C содержатся подкластеры $C_i \in C, i = 1..N$ то применить этот рекурсивной алгоритм для каждого из подкластеров $C_1, C_2, .., C_N$.

2. Если кластер C близок к рассматриваемому объекту $r(C,o) \le T$ и в C содержатся объекты $o_i \in C, i = 1..N$, то

а) добавить в список опорных узлов S все объекты, расположенные близко к рассматриваемому: $S = S \cup \{o_i | d(o_i, o) \le T\}$;

б) выделить множество объектов V, расположенных в удалении от рассматриваемого в неполный кластер: $V = \{o_i | d(o_i, o) > T\}$;

в) в том случае, если множество V непустое, добавить его в список опорных узлов: $S = S \cup V$.

3. Если кластер C расположен в удалении от рассматриваемого объекта $r(C, o) > T$, то добавить его в список опорных узлов: $S = S \cup C$.

В рассмотренном алгоритме T - порог, определяющий, находится ли рассматриваемый объект близко к источнику, а $r(C, o)$ - функция, определяющая удаленность объекта от кластера.

4. Выполнение итеративной процедуры оптимизации

Процедура оптимизации в настоящей работе строится на основе аппроксимации вычисления рекуррентного соотношения (2). Наибольший интерес представляет приближенное вычисление для произвольного объекта $o \in O$ суммы $\boldsymbol{\sigma}_i(t)$ в (3), так как расчет множителя m для (2) может быть сделан заранее и не представляет затруднений. С использованием сформированного на предыдущем шаге работы метода списка опорных узлов S для объекта $o \in O$ приближенное значение суммы на шаге *t* может быть записано в виде

$$\tilde{\boldsymbol{\sigma}}_i = \sum_{s_j \in S} \boldsymbol{\varsigma}(o_i, s_j),$$

где функция $\boldsymbol{\varsigma}(o_i, s_j)$ принимает различный вид в зависимости от того, является ли s_j объектом, кластером или неполным кластером.

1. Если s_j является объектом o_j, то

$$\boldsymbol{\varsigma}(o_i, s_j) = \frac{d(o_i, o_j) - d^*(o_i, o_j)}{d(o_i, o_j) \cdot d^*(o_i, o_j)} \cdot (\mathbf{y}_i - \mathbf{y}_j).$$

2. Если s_j является кластером C_j (или неполным кластером V_j), то

$$\boldsymbol{\varsigma}(o_i, s_j) = |C_j| \cdot \frac{d(o_i, c_j) - d^*(o_i, c_j)}{d(o_i, c_j) \cdot d^*(o_i, c_j)} \cdot (\mathbf{y}_i - \mathbf{y}_{C_j}),$$

где $d(o_i, c_j)$ – расстояние от объекта до центра кластера (неполного кластера) в исходном пространстве, $d^*(o_i, c_j)$ – расстояние от объекта до центра кластера (неполного кластера) в результирующем пространстве, $d^*(o_i, c_j)$ – координаты центра кластера (неполного кластера) в результирующем пространстве.

Следует отметить, что центры кластеров и неполных кластеров в многомерном пространстве могут быть рассчитаны заранее. Центры кластеров в результирующем пространстве должны уточняться в процессе оптимизации, причем центры неполных кластеров могут быть рассчитаны через центры соответствующих полных кластеров и позиции исключенных объектов.

Экспериментальные исследования

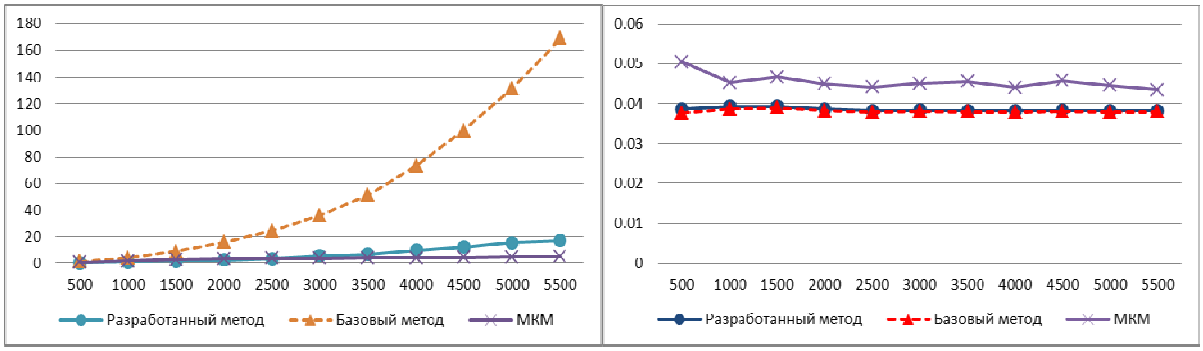
В рамках настоящей работы предложенный метод был реализован программно на языке C++ в интегрированной среде разработки Borland Turbo C++ 2006 Explorer. Были проведены экспериментальные исследования предложенного метода для ряда систем признаков, а также для коллекций изображений различного объема. При

проведении экспериментов использовался ПК на базе Intel(R)Core(TM)2 QuadCPU 2.5 ГГц.

Для проведения экспериментальных исследований качества работы разработанного метода для различных систем признаков использовался набор фрагментов космических снимков, полученных со спутника SPOT-4 с пространственным разрешением 10 м (1 канал, 8 бит). По изображениям рассчитывались статистические яркостные характеристики (гистограммы) и признаки на основе фильтров Габора. Результаты проведенных исследований приведены в таблице 1 (по причине ограниченного объема публикации в таблице представлены результаты только для набора из 3000 изображений). На рис. 1 показаны результаты экспериментов для наборов данных различного объема (в качестве признаков использовались стат. характеристики полутоновых изображений).

Таблица 1. Результаты экспериментальных исследований (для космических снимков)

Метод	**Среднее значение ошибки**	**Среднее время работы, сек**
Стат. характеристики полутоновых изображений		
МКМ	0,045	3,69
Разработанный метод	0,038	5,36
Базовый метод	0,038	35,83
Признаки на основе фильтров Габора		
МКМ	0,040	3,51
Разработанный метод	0,036	5,32
Базовый метод	0,035	47,81



а) *б)*

Рис. 1 – Работа методов на наборах данных различного объема (количество изображений показано по оси абсцисс): а) время работы методов (по оси ординат, в секундах); б) значение ошибки (по оси ординат)

В рамках другой группы экспериментов использовались цифровые фотографии различной тематики (люди, животные, техника, искусство, природа и т.п.) и искусственные 3D сцены. Во время проведения эксперимента из исходной коллекции изображений случайным образом генерировались наборы по 500, 1000 и 2000 изображений, и каждый из методов применялся к сгенерированным наборам. Результаты экспериментов представлены в таблице 2. В качестве системы признаков использовались хорошо зарекомендовавшие себя гистограммы цветов в цветовом пространстве CIE L*a*b*.

Как видно из результатов проведенных экспериментов, полученные значения ошибки для всех рассмотренных систем признаков не превышают 0.1, что свидетельствует об адекватном отображении структуры многомерных данных [8]. При этом значения ошибки, получаемые с использованием разработанного метода, практически не отличаются от значений, получаемых с использованием базового метода Сэммона, а время функционирования предложенного метода заметно меньше, чем у базового метода.

Следует отметить, что для работы разработанных методов, также как и для метода МКМ требуется выполнение иерархической кластеризации. В качестве метода кластеризации в работе использовался вызываемый в рекурсивном порядке нейросетевой алгоритм WTA, функционирующий на базе нейронной сети Кохонена. При этом среднее время, затрачиваемое на кластеризацию данных объемом 500, 1000 и 2000 изображений, составляет 0,1 сек., 0,2 сек. и 0,3 сек., соответственно. Максимальный размер кластеров был выбран равным 32 объектам в кластере.

Таблица 2. Результаты экспериментальных исследований (для цифровых фотографий)

Метод	**Среднее значение ошибки**			**СКО ошибки**			**Среднее время работы, сек**		
	Количество изображений в наборе								
	500	**1000**	**2000**	**500**	**1000**	**2000**	**500**	**1000**	**2000**
Базовый метод	0,026	0,026	0,024	0,005	0,004	0,001	7,0	91,3	178,3
Стохастическая аппрокс.*	0,038	0,033	0,031	0,009	0,007	0,012	-	-	-
МКМ	0,034	0,032	0,030	0,004	0,003	0,002	0,7	2,7	3,9
Разработанный метод	0,024	0,024	0,024	0,002	0,002	0,001	1,0	6,4	10,9

*Сведения о методе приведены здесь с целью сравнения и взяты из работы автора [1] (сведения о времени работы для указанного метода намеренно не приводятся, т.к. эксперименты проводились на другом аппаратном обеспечении). В той же работе можно найти аналогичные сведения для методов триангуляции и линейного преобразования.

Заключение

В рамках работы изложен разработанный автором нелинейный метод снижения размерности пространства признаков на основе опорных узлов. Проведен анализ качества работы предложенного метода для ряда систем признаков цифровых изображений, показавший, что разработанный метод можно рекомендовать к практическому применению.

Благодарности

Работа выполнена при поддержке Российского фонда фундаментальных исследований: проекты № 10-07-90703-моб_ст, 09-01-00434-а.

Литература

1. Мясников, Е.В. Анализ методов снижения размерности в задаче представления коллекций цифровых изображений // Компьютерная оптика. – 2008. - Том 32, №3. - стр. 296-301.

2. Мясников, Е.В. Разработка метода навигации по коллекциям цифровых изображений /Е.В. Мясников //Труды 9-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL’2007. - Переславль-Залесский, Россия. - 2007. - С. 185-194.

3. Фукунага, К. Введение в статистическую теорию распознавания образов /К. Фукунага.- М.Наука.- 1979. – 368 с.

4. Chalmers, M. A Linear Iteration Time Layout Algorithm for Visualizing High–Dimensional Data / M. Chalmers //Proc. IEEE Visualization `96. - San Francisco. – 1996. - P.127-132.

5. Lee, R.C.T. A Triangulation Method for the Sequential Mapping of Points from N-Space to Two-Space. /R.C.T. Lee, J.R. Slagle, H. Blum. //IEEE Transactions on Computers. – 1977. - V. 26, №3. - P. 288-292.

6. P"ekalska, E. A new method of generalizing Sammon mapping with application to algorithm speed-up. /E. P"ekalska, D. de Ridder, R.P.W. Duin, M.A. Kraaijveld //Proc. ASCI'99, 5th Annual Conf. of the Advanced School for Computing and Imaging.- Heijen, The Netherlands. – 1999. - June 15-17. - P. 221-228.

7. Rubner, Y. Perceptual Metrics for Image Database Navigation / Y. Rubner, C. Tomasi. – Boston. Kluwer Academic Publishers. – 2001. – 143 p.

8. Sammon, J.W., Jr. A nonlinear mapping for data structure analysis / IEEE Transactions on Computers. – 1969. - V. C-18, No.5. - P.401-409.

Модификация алгоритма «Split – Merge – Pixelwise refinement» для сегментации спутниковых изображений

А.Ю. Петрова, В.А. Федосеев
Институт систем обработки изображений РАН, Самара
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Аннотация

Работа посвящена сегментации спутниковых изображений. В качестве базового алгоритма сегментации используется метод “Split – Merge – Pixelwise refinement”. Выбраны признаки, подходящие для источника данных рассматриваемого типа. Выявлены алгоритмические особенности рассматриваемого метода, негативным образом влияющие на результаты сегментации. Предложены модификации алгоритма, ставящие целью скомпенсировать эти особенности.

Введение

Спутниковые изображения высокого пространственного разрешения стали важным источником информации для геопространственных приложений. Автоматическая сегментация таких изображений необходима в целом ряде аналитических задач, к числу которых можно отнести отыскание изменений антропогенного характера на серии снимков разного времени на одном участке местности, контроль сельскохозяйственной деятельности, точная привязка и оцифровка объектов, отделение наземных объектов от отбрасываемых ими теней и многие другие.

Важно отметить, что в задаче сегментации многоканальных спутниковых снимков высокого разрешения адекватные результаты могут быть достигнуты лишь при совместном использовании нескольких признаков. В связи с этим необходимо иметь общую методику сегментации, инвариантную к количеству и виду используемых признаков, отличающих одни сегменты от других, а также выбрать пространство признаков, подходящих для источника данных рассматриваемого типа.

Основная категория методов сегментации при работе с данными дистанционного зондирования основана на попиксельной классификации, которая включает как статистическую, так и кластерную классификацию [1]. Одним из наиболее часто используемых алгоритмов для сегментации изображений является кластеризация пространства признаков. Этот метод имеет два недостатка:

1) он не берет во внимание пространственную информацию;
2) количество кластеров не может быть получено напрямую и автоматически.

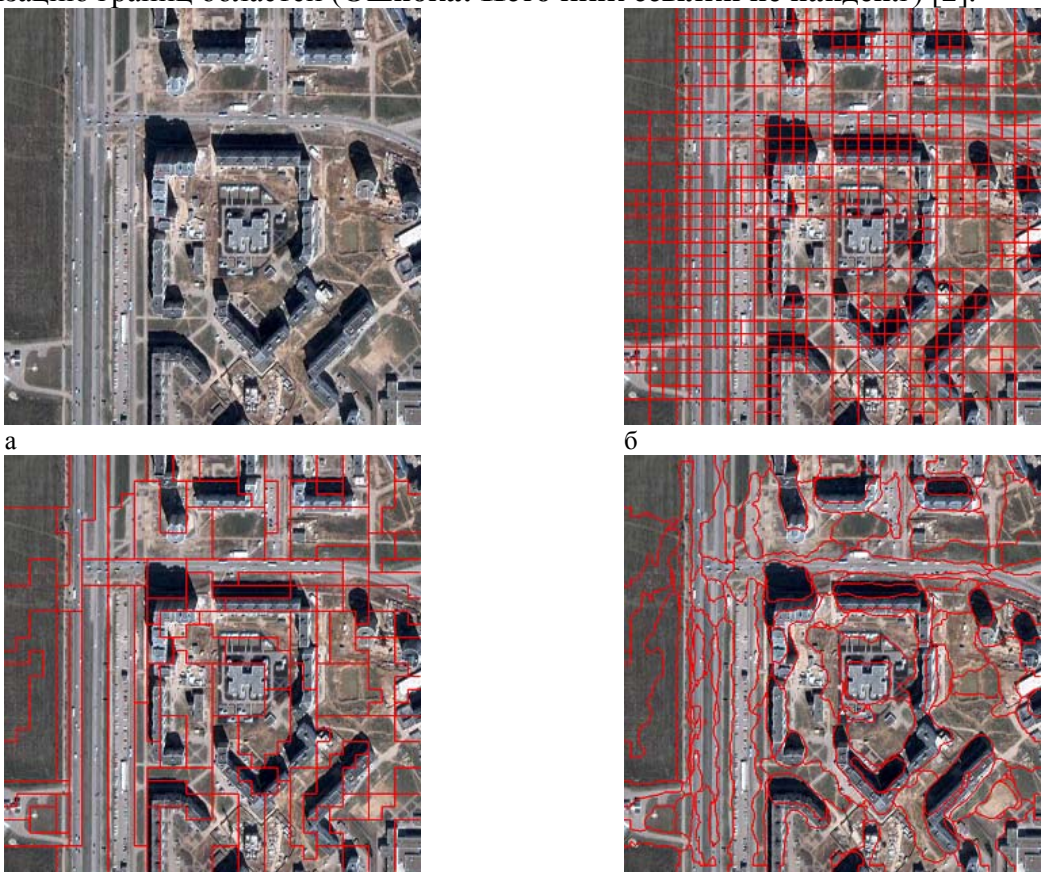
Методы, основанные на разбиении на области (т.е. выращивание, расщепление, слияние областей и их комбинации), используют статистическую однородность областей, следовательно, стремятся снизить чувствительность к шуму. Эти методы зачастую дают результаты, лучшие, нежели при использовании кластеризации пространства признаков или при пороговом подходе, т.к. в них одновременно рассматриваются как пространство признаков, так и пространственное взаимодействие между пикселями [1].

Руководствуясь данными соображениями, в качестве базового алгоритма в задаче сегментации спутниковых изображений был выбран метод под названием “Split – Merge – Pixelwise refinement”[2], который можно перевести как «расщепление – слияние – попиксельный допоиск». Важным его достоинством также является отделение собственно алгоритма сегментации от пространства признаков, используемого для сегментации.

Краткое описание метода

Метод сегментации [2] состоит из трех последовательных шагов: иерархическое расщепление, агломерационное слияние и попиксельный допоиск.

На этапе иерархического расщепления изображение разбивается на однородные сегменты (**Ошибка! Источник ссылки не найден.**б). Это не требует априорной информации о сценах на изображении. Различные сегменты разделяются в соответствии с их признаками (текстура, интенсивность, цвет и т.д.). Затем агломерационное слияние объединяет схожие смежные сегменты, пока не сработает критерий остановки (**Ошибка! Источник ссылки не найден.**в). В результате получается грубая оценка областей с различными признаками. Сегментация завершается использованием попиксельного допоиска для того, чтобы улучшить локализацию границ областей (**Ошибка! Источник ссылки не найден.**г) [2].



а б в г

Рисунок 1 – Основные шаги метода: а) исходное изображение б) результат иерархического расщепления, в) результат агломерационного слияния, г) результат попиксельного допоиска.

Иерархическое расщепление

Алгоритм иерархического расщепления рекурсивно разбивает исходное изображение на блоки разного размера. Прежде всего, происходит разбиение изображения на квадратные сегменты размера $S_{\max}$. Решение «разбить сегмент на четыре подсегмента» принимается на основе проверки на однородность. Эта процедура рекурсивно применяется ко всем подсегментам, пока не будет достигнут заранее заданный размер сегмента $S_{\min}$. Проверка на однородность заключается в оценке шести попарных G - расстояний **Ошибка! Источник ссылки не найден.** между четырьмя подсегментами. Сегмент считается неоднородным и разбивается, если относительное различие R в нем больше порога.

$$\tilde{S}_i = \left\{ S_i^{lu}, S_i^{ld}, S_i^{ru}, S_i^{rd} \right\},$$
$$G_{\min} = \min_{s,\sigma \in \tilde{S}_i; s \neq \sigma} G(s,\sigma) \quad G_{\max} = \max_{s,\sigma \in \tilde{S}_i; s \neq \sigma} G(s,\sigma),$$
$$R = \frac{G_{\max}}{G_{\min}} > X, \qquad (1)$$

где $\tilde{S}_i$ - множество из четырех подсегментов.

Агломерационное слияние

На данном этапе происходит слияние смежных сегментов до тех пор, пока не выполнится условие остановки. На отдельном шаге слияния происходит объединение той пары смежных сегментов, которая имеет наименьшее значение «важности слияния» MI. MI определяется как:

$$MI\left(S_i, S_j\right) = \min\left(\left|S_i\right|, \left|S_j\right|\right) \times G\left(S_i, S_j\right),$$
$$MI : \square\ [0,1] \to \square\ [0, V \cdot H - 1], \qquad (2)$$

где G - это расстояние, определяемое по формуле **Ошибка! Источник ссылки не найден.**.

Другими словами, на каждом шаге процедура слияния выбирает то слияние из всех возможных слияний, которое приведет к наименьшему изменению в сегментируемом изображении. Пусть N - количество сегментов после этапа расщепления, k - номер итерации при слиянии (он же количество уже объединенных сегментов). $S_i^{(k)}$ - i-ый сегмент на k-ой итерации. Тогда

$$MI_{\min}^{(k)} = \min_{0 \le i,j \le N-k, i \neq j} MI\left(S_i^{(k)}, S_j^{(k)}\right) \qquad (3)$$

- минимальное значение MI на k-ой итерации. $MI_{\min}^{(-1)} \square\ 0$.

$$MI_{\max}^{(k)} = \max_{0 \le i,j \le N-k, i \neq j} MI\left(S_i^{(k)}, S_j^{(k)}\right) \qquad (4)$$

- максимальное значение MI на k-ой итерации.

На каждом шаге объединяются два сегмента, обеспечивающие $\arg MI_{\min}^{(k)}$. Критерий остановки процедуры слияния MIR:

$$MIR^{(k)} = \begin{cases} MI_{\min}^{(k)} \Big/ \max\limits_{m=0..k-1} MI_{\min}^{(m)}, \max\limits_{m=0..k-1} MI_{\min}^{(m)} \neq 0 \\ 0, иначе \end{cases}; \qquad (5)$$
$$MIR > Y,$$

Попиксельный допоиск

Для лучшей локализации границ используется простой алгоритм попиксельного допоиска. Если иерархическое расщепление и агломерационное слияние прошли удачно, то в результате получается вполне реальная оценка границ областей разных признаков. Если пиксель принадлежит границе двух или более сегментов, то в него помещается центр диска радиуса r. Затем вычисляется G-расстояние между диском и соседними сегментами. Пиксель переразмечается, если метка ближайшего сегмента отличается от текущей метки пикселя. При следующем проходе по изображению проверяются только соседи тех пикселей, которые были переразмечены на предыдущем шаге. Процесс попиксельного улучшения продолжается до тех пор, пока не останется переразмеченных пикселей или число «чисток» не превзойдет максимум.

Описание признаков и меры подобия

Как уже было сказано выше, многоканальный космический снимок, как правило, имеет достаточно сложную структуру, и для адекватной его сегментации необходимо

использовать вектор из нескольких признаков. По мнению авторов работы [1], в состав этого вектора должны быть включены текстурные, яркостные, а также цветовые признаки.

В качестве способа агрегирования оценок близости двух сегментов по каждому из признаков можно использовать простейшую линейную модель:

$$G\left(\vec{f}\right)=\sum_i C_i\rho_i, \qquad 6)$$

где $\vec{f}$ – вектор типов признаков; $\vec{\rho}$ – вектор оценок близости по каждой из компонент вектора признаков; $\vec{C}$ – вектор весов оценок.

Результаты сегментации, представленные на рисунке 1 получены с использованием следующих признаков: LBP/C-распределение для оценки текстуры [3], распределение «насыщенность/тон» [1] для цветовых характеристик, а так же интенсивность.

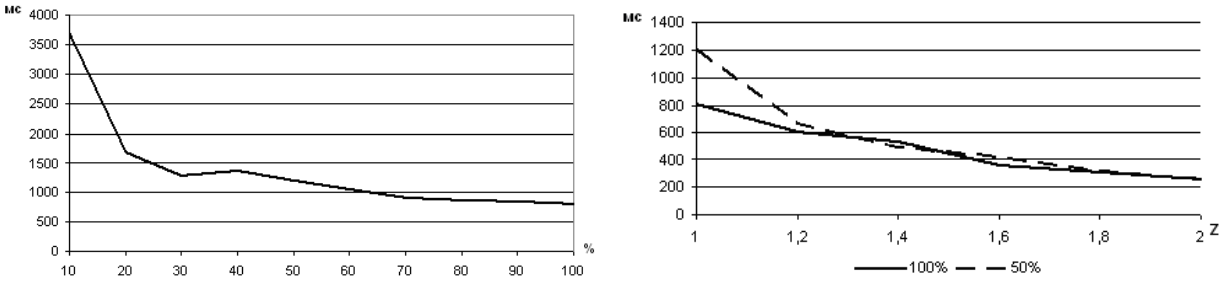
Модификация попиксельного допоиска

Несмотря на визуально удовлетворительные результаты сегментации на многих снимках, примером чему служит изображение на рисунке 1, в ходе исследований выявлены некоторые негативные особенности рассматриваемого алгоритма. Одна из них заключается в том, что при попиксельном допоиске изменяется сегментная принадлежность всех граничных пикселей, характеристики которых ближе к соседнему сегменту, нежели к текущему, то есть независимо от того, насколько характеристики пикселя ближе к соседнему сегменту. Ввиду неточного определения границы сегментов после этапа слияния такой недостаток может «направить» границы при попиксельной коррекции по неправильному пути.

Нами предложена следующая модификация, призванная скомпенсировать эту особенность. На каждом проходе переразмечаются не все «намеченные» пиксели, а только «наилучшие». Для каждого граничного пикселя вычисляется G-расстояние между диском и соседними сегментами. Если отношение расстояния до текущего сегмента к наименьшему из расстояний до соседних сегментов T превосходит порог Z, то пиксель становится кандидатом на переразметку. Среди всех таких кандидатов переразмечаются n процентов пикселей с наибольшим значением T.

$$S_b=\left\{S_b^i : i=2,3,...\right\},$$
$$G_{cur}=G\left(S_{cur},S_d\right)\quad G_{\max}=\max_{S_b^i\in S_b} G\left(S_b^i,S_d\right), \qquad 7)$$
$$T=\frac{G_{cur}}{G_{\max}}>Z,$$

где S_b - множество сегментов соседей для текущего пикселя, G_{cur} - расстояние до сегмента, которому принадлежит текущий пиксель, S_d - диск.

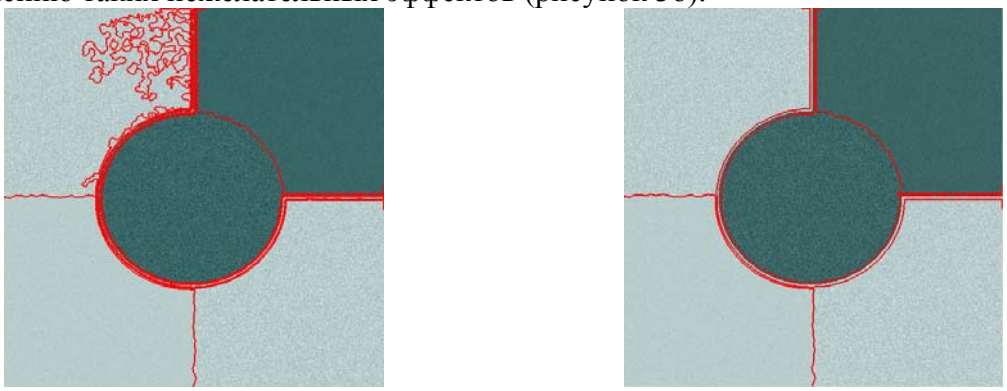


а б

Рисунок 2 – а) зависимость времени работы программы от количества точек при ПД; б) зависимость времени работы программы от порога при ПД

На рисунке 2 приведены графики зависимости работы программы от порога Z и от процентного количества точек для переразметки. Они показывают, что выбор порога равным 1.2 и отбор 50-70% точек не только существенно не увеличивает время выполнения процедуры, но и напротив, может выполняться быстрее, чем каноническая версия алгоритма.

Ещё один недостаток канонического алгоритма заключается в том, что в процессе попиксельного допоиска могут стать граничащими друг с другом сегменты, которые не граничили на этапе слияния. Такие сегменты необходимо также проверять на возможность слияния друг с другом, то есть выборочное слияние должно происходить и на этапе попиксельного допоиска. Иначе при попытке сдвига границы между ними ввиду возможной близости характеристик двух сегментов итоговая граница может принимать весьма причудливый характер, как показывает рисунок 3а. Применение выборочного слияния на этапе попиксельного допоиска приводит устранению таких нежелательных эффектов (рисунок 3б).



а б

Рисунок 3 – Сравнение результатов: а) исходный метод б) после модификации

Заключение

В работе делается попытка использования метода “Split – Merge – Pixelwise refinement” для сегментации спутниковых изображений. Пробное применение его дало визуально неплохие результаты при использовании в качестве вектора признаков совокупности текстурных, яркостных и цветовых признаков. Также выявлены алгоритмические особенности рассматриваемого метода, негативным образом влияющие на результаты сегментации, и предложены подходы, позволяющие скомпенсировать эти недостатки. В дальнейшем предполагается произвести более глубокие эксперименты по исследованию работоспособности используемого алгоритма сегментации в различных его модификациях и на различных векторах признаков с использованием численных показателей качества результатов сегментации.

Литература

1. Hu, X. Automatic Segmentation of High-resolution Satellite Imagery by Integrating Texture, Intensity, and Color Feature / X. Hu, C.V. Tao, B. Prenzel // Photogrammetric Engeneering & Remote Sensing. – 2005. – Vol. 71, № 12. – P. 1399-1406.

2. Ojala, T. Unsupervised texture segmentation using feature distributions / T. Ojala, M. Pietikäinen // Pattern Recognition. – 1999. – Vol. 32. – P. 477-486.

3. Ojala, T. Textural discrimination with multidimentional distributions of signed gray-level differences / T. Ojala, K. Valkealahti, E. Oja, M. Pietikäinen // Pattern Recognition. – 2001. – Vol. 34. – P. 727-739.

Построение текстурных признаков на основе регрессионных моделей

А.И. Пластинин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Многие естественные микро и макро изображения представляют собой комбинацию текстурных примитивов, например аэрофотоснимки земной поверхности, изображения нано-структур материалов, биомедицинские изображения. Поэтому во многих прикладных задачах обработки изображений (сегментация, кодирование, имитационное моделирование и т.д.) можно применить методы обработки текстур.

Можно выделить два основных подхода к анализу и синтезу текстур: структурный и стохастический [1]. В структурном подходе текстура рассматривается как совокупность повторяющихся базовых примитивов, взаимное расположение которых, определяется в соответствии некоторым правилам. В статистических методах текстура определяется как реализация некоторого двумерного случайного процесса, и характеризуется набором статистик. В данной статье рассматривается стохастический подход.

Следует отметить, что распространенные методы текстурного анализа, например, такие как матрицы взаимного распределения [2], спектральные дескрипторы [3], признаки Тамура и т.д., дают общие признаки для всех типов изображений, они не учитывают специфики задачи. Однако на практике возникает необходимость вычислять признаки для классификации ограниченного набора типов текстур (например, изображения клеток крови). Также к недостаткам этих методов следует отнести то, что многие методы ориентированы на работу с полутоновыми изображениями, то есть они не применимы для анализа цветных и многоспектральных текстурных изображений.

Методы, в основе которых лежит модель изображения, как реализации марковского случайного поля, успешно применяются во многих прикладных задачах. Эти методы оказались эффективными в задачах сегментации изображений, имитационного моделирования и т.д.[4-7].

В данной работе рассматривается возможность построения признаков, учитывающих особенности предметной области.

1. Задача распознавания текстурных изображений

Обозначим $\Im$ как множество всех возможных изображений, которое разделено на K классов $C_1,\ldots,C_K$, т.е. $\bigcup_{k=0}^{K} C_k = \Im$ и $C_i \bigcap C_j = \varnothing, i \neq j$, здесь C_0.– множество объектов, не относящихся ни к одному из классов. Разделение на классы изначально не известно.

Пусть задано множество изображений $\mathrm{I} = \{I_1,\ldots,I_N\} \subset \Im$, для каждого изображения указан класс, к которому оно относится, т.е. определено отображение $u : \mathrm{I} \to \mathbb{L}$, где $\mathbb{L} = \{1,\ldots,K\}$ – множество меток объектов. Обучающая выборка определяется как множество пар изображение-класс: $T = \mathrm{I} \times \mathbb{L}$.

Необходимо разработать алгоритм, который по данному изображению I отнесет его к одному из классов.

Для того чтобы воспользоваться методами распознавания образов [8, 9], необходимо поставить каждому объекту вектор признаков $f \in \mathbb{F} = \mathbb{R}^n$. Т.е. должно быть задано отображение $F : \Im \to \mathbb{F}$, где $\mathbb{F}$ пространство признаков. Существует

неизвестная зависимость $u: F \to L$, значения которой известны только на объектах конечной обучающей выборки $T = \{(f_i, l_i)\}_{i=1}^{N}$. Требуется построить алгоритм $\tilde{u}: F \to L$, который приближал бы неизвестную целевую зависимость, как на элементах выборки, так и на всем множестве F.

Многие существующие методы вычисления признаков, определены одинаково для любо типа задач, т.е. они не учитывают информацию о предметной области, определяемую множеством примеров и обучающим множеством.

Для того что бы учесть информацию будем строить параметрический алгоритм вычисления признаков F_A, где A – набор параметров алгоритма. Таким образом, на этапе обучения, по множеству примеров I, вычисляются конкретные параметры алгоритма вычисления признаков $a = P(\mathrm{I})$, а признаки для конкретного изображения оцениваются алгоритмом с использованием вычисленного вектора параметров $f = F_a(I)$.

Общая схема построения разработанной системы классификации текстурных изображений представлена на рис. 1.



Рис. 1. Схема классификации

На первом этапе по обучающему множеству I вычисляются параметры алгоритма вычисления признаков $c_0, \ldots, c_K$. Далее полученные параметры используются для вычисления множества обучающих признаков F, которые на следующем шаге используются для обучения классификатора.

2. Модель изображения

Для построения алгоритма вычисления признаков в случае анализа текстурных изображений можно использовать модель марковского случайного поля, которая эффективно используется в задачах анализа текстур [7].

Здесь под изображением будем понимать функцию:

$$I: L_x \times L_y \mapsto G, \qquad (1)$$

где $L_x = \{1, 2, \ldots, K_x\}$ и $L_y = \{1, 2, \ldots, K_y\}$ – пространственные координаты, а G множество цветов, например, в случае полутонового изображения $G = [0;1]$, для цветного RGB изображения $G = [0;1]^3$.

Пусть S – конечное множество узлов (в случае изображений это решетка $L_x \times L_y$). Пусть $Z = \{Z_1, \ldots, Z_m\}$ – семейство случайных переменных определенных на множестве S, каждая случайная переменная Z_i принимает значение $z_i \in G$. Семейство Z будем называть случайным полем. Будем использовать обозначение $Z_i = z_i$ для

события, что Z_i принимает значение z_i, а для совместного события $(Z_1 = z_1, \ldots, Z_m = z_m)$ будем использовать обозначение $Z = z$.

$P(Z_i = z_i)$ – вероятность того что случайная переменная Z_i принимает значение z_i.

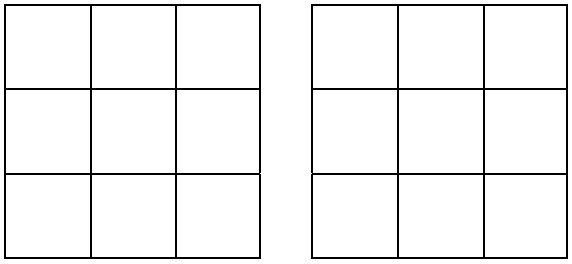
Связь узлов друг с другом определяется системой окрестностей, которая определяется как $\mathrm{N} = \{\mathrm{N}_s \mid s \in S\}$, где N_s множество узлов соседей узла s. Окрестности обладают следующим свойством $s \notin \mathrm{N}_s$. Окрестность называется симметричной, если выполнено условие: $s \in \mathrm{N}_t$, тогда и только тогда, когда $t \in \mathrm{N}_s$.

Случайное поле Z называется Марковским случайным полем (МСП), по отношению к системе окрестности N, если для любого $x \in X$, выполнено:

$$P(z_i \mid z_{S/i}) = P(z_i \mid z_{\mathrm{N}_i}). \quad (2)$$

Далее будем использовать следующие обозначения: N_p^k – k-ый элемент окрестности; $I_{\mathrm{N}_p} = \left(I_{\mathrm{N}_p^1}, \ldots, I_{\mathrm{N}_p^K}\right)$ – вектор значений отсчетов изображения в окрестности пикселя.

На рис. 2 приведен пример возможной нумерации пикселей в двумерных окрестностях.



а) *б)*

Рис. 2. Порядок пикселей в окрестности: а) некаузальная окрестность; б) каузальная окрестность

3. Метод построения признаков

Поскольку существует статистическая зависимость значения центрального пикселя x_0, от значений пикселей в окрестности, введем в рассмотрение функцию аппроксимации, т.е. функцию, аргументами которой будут являться значения пикселей в окрестности, а значением – оценка центрального пикселя. Такую функцию можно определить несколькими способами:

$$\hat{I}_p = f(I_{\mathrm{N}_p}) = \arg\max_x P(x \mid I_{\mathrm{N}_p}), \quad (3)$$

Или используя уравнение регрессии:

$$\hat{I}_p = f(I_{\mathrm{N}_p}) = \mathrm{M}[x \mid \mathrm{N}_p]. \quad (4)$$

Рассмотрим метод решения задачи регрессии (4). $D = \{(x_i, y_i)\} = \{(I_{\mathrm{N}_p}, I_p) \mid p \in L_x \times L_y\} \in \square^K \times \square$ – обучающее множество задачи регрессии. Будем использовать метод ridge-регрессии на основе ядер [10, 11]. Согласно теореме представления (representer theorem) [10] будем искать решение в виде:

$$f(x) = \sum_{i=1}^{l} \alpha_i k(x, x_i). \quad (5)$$

Метод ridge-регрессии заключается в минимизации функционала:

$$\sum_{i=1}^{l}\left(y_i-\alpha_i k(x,x_i)\right)^2+\lambda\sum_{i=1}^{l}\alpha_i^2\ . \tag{6}$$

Решение имеет вид:

$$\alpha=\left(K+\lambda I\right)^{-1}y\text{, где } K_{i,j}=K(x_i,x_j)\,. \tag{7}$$

Таким образом, каждому изображению будет соответствовать свой набор параметров α. Более того, поскольку изображения одного класса отличаются незначительно, то и параметры у них буду схожие.

Для построения признаков разделяющих несколько классов необходимо строить функцию регрессии с использованием обучающего множества построенного по всем изображением из обучающего множества исходной задачи.

Однако с этим связана другая проблема – это приведет к увеличению размера обучающего множества задачи регрессии. Что бы избежать этого роста необходимо по каждому изображению строить не все возможные вектора окрестностей, а выбирать только часть, поскольку вектора будут совпадать в силу свойств текстуры.

Далее из всего полученного множества необходимо выбрать только линейно независимое (в пространстве ядра K) подмножество.

4. Процедура построения линейно независимого подмножества

Пусть необходимо выбрать M линейно независимых векторов. Воспользуемся следующей итеративной процедурой [12]. На каждом шаге t будем выкидывать вектор x_t, для которого следующий функционал минимален:

$$\delta(x_t)=\min\left\|\sum_{j=1}^{m_{t-1}}a_j\phi(\tilde{x}_j)-\phi(x_t)\right\|^2\ , \tag{8}$$

где m_t – количество векторов на шаге t, $\{\tilde{x}_j\}$ – множество векторов на шаге t, $\phi(x)$ – отображение исходного пространства в гильбертово пространство с воспроизводящим ядром K, таким что $K(x,y)=\langle\phi(x),\phi(y)\rangle$ [10].

Обобщая сказанное выше, получим, что набор параметров A для каждой конкретной задачи представляет собой множество линейно независимых векторов окрестностей, построенных по всем изображениям обучающего множества, а вектор признаков изображения – коэффициенты разложения функции аппроксимации изображения по базису параметров A.

Заключение

В работе предлагается общий подход к построению текстурных признаков использующих информацию об исходном обучающем множестве.

Рассмотрен один из способов формализации алгоритма вычисления признаков на основе построения функции аппроксимации зависимости значения пикселя от пикселей в окрестности, для изображений удовлетворяющих модели марковского случайного поля.

Предложенный алгоритм обладает рядом преимуществ:

1. При использовании модели регрессии для вычисления признаков, можно использовать информации о характере изображений из обучающего множества, что позволяет улучшить качество классификации.
2. Вектора признаков объектов разных классов линейно независимы.
3. Признаки обладают устойчивостью относительно аддитивного шума на изображении.

Литература

1. R.M. Haralick. Statistical and structural approaches to texture. Proceedings of the IEEE In Proceedings of the IEEE, Vol. 67, No. 5. (1979), pp. 786-804.
2. R.M. Haralick, K. Shanmugam, I. Dinstein (1973). "Textural Features for Image Classification". IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6): 610–621.
3. William K. Pratt. Digital Image Processing 4th ed., John Wiley & Sons, Inc. 2007.
4. Winkler G. Image Analysis, Random Fields and Dynamic Monte Carlo Methods. – Springer-Verlag, 1995. – 324 p.
5. Paget R. Nonparametric Markov Random Field Models for Natural Textures Images, Ph.D. thesis, University of Queensland, St Lucia, QLD Australia, December 1999.
6. Kaulgud N., Desai U.B. Efficient color image restoration using Markov random field. TENCON 98. 1998 IEEE Region 10 International Conference on Global Connectivity in Energy, Computer, Communication and Control. Volume 1, 1998, pp. 41 - 44.
7. Stan Z. Li. Markov Random Field in Image Analysis.
8. Ту Дж. Принципы распознавания образов / Дж. Ту, Р. Гонсалес. – М.: Мир, 1978. – 412 с.
9. Vapnik V. Statistical Learning Theory. Wiley, 1998.
10. Bernhard Schölkopf and Alexander J. Smola. Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, 2001. 644 pp.
11. Hastie Trevor, Tibshirani Robert, Friedman Jerome. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. 2nd ed. Springer, 2009.
12. Yaakov Engel, Shie Mannor, Ron Meir. The Kernel Recursive Least-Squares Algorithm. IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 52, NO. 8, AUGUST 2004.

К вопросу обеспечения безопасности распределённых сетей и технической поддержки

М.С. Воронов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В настоящее время наблюдается увеличение количества корпоративных сетей, которое будет продолжаться и в будущем. Огромное количество компаний используют локальные сети уровня организации для передачи или обработки конфиденциальной информации. И каждый год придумывается большое количество методик совершения НСД с целью получения такой информации. В связи с этим важная роль в процессе функционирования интранета должна уделяться его безопасности и защите от НСД. Причём безопасность корпоративных сетей должна рассматриваться со всех аспектов, т.к. бесполезно ставить дорогой замок на картонную дверь.

В данной работе рассматривается некоторые частные аспекты нового комплексного подхода к обеспечению безопасности корпоративных сетей. К числу рассматриваемых аспектов относится техническая поддержка созданных сетей на базе Аналитического центра, а также предназначение и функции оснастки.

Работа Аналитического Центра

Основное предназначение Аналитического Центра в системе обеспечения безопасности корпоративных сетей – это техническая поддержка всех установленных систем защиты, которая, в первую очередь, должна решить проблему обновлений. Причём в число обновлений будет входить: базы сигнатур, необходимые для антивируса, IPS и спам-анализатора; векторы атак для системы «патчинга на лету»; обновления/исправления модулей и список «подозрительных» IP адресов.

Другая задача центра – сбор информации о функционировании всех систем защиты корпоративных сетей. К собираемой с системы информации должна относиться и статистика функционирования всех систем защиты, и результат работы анализатора аномалий.

Каждая установленная система, в зависимости от выставленных политик безопасности, может обращаться в центр за помощью в принятии решения о блокировании определённого трафика. Обращение должно происходить только в действительно «сложных» случаях. В таком случае человек может в реальном времени принять решение о разрешении или блокировании трафика, вызвавшего обращение. Такой прецедент может дать дополнительную информацию для системы обучения как IPS отдельных систем, так и системы принятия решений в самом центре.

Также предполагается наличие системы принятий решений на базе самого центра, призванной облегчить труд его сотрудников. Данная система должна быть обучаемой. Также она должна при принятии решения руководствоваться не только информацией с одной системы, но и со всех доступных систем. Это позволит при появлении однотипной аномалии сразу на нескольких системах быстро её пресечь, сконфигурировав таблицы правил файрволов остальных систем.

Работа системы с Аналитическим центром возможна только в случае наличия в системе модуля «Сервер обновлений» и согласия на эту работу со стороны пользователя системы. В план развития продукта должно быть предусмотрено увеличение количества таких центров в случае необходимости (можно сказать, что каждый такой центр должен быть ответственен за определённый территориальный участок). При всём этом, в образовавшейся структуре должен непременно находиться ведущий Аналитический центр, который может корректировать работу остальны.

Создание такого центра позволит не только в реальном времени наблюдать за распространением вирусной инфекции в любом уголке мира (к нему можно также подключить и уже установленные антивирусы о при желании их владельцев), но и препятствовать дальнейшему заражению, конфигурируя сенсоры межсетевых экранов.

В целом, работа такого Аналитического центра позволит в значительной степени усилить систему безопасности корпоративных сетей: защита отдельной интранетовской сети будет зависеть не только от принятия решения на уровне этой сети, но и от вычислительных способностей центра, квалификации работающих там специалистов, а также обширной аналитики, собранной с других подобных сетей. Это позволит эффективно противостоять угрозам не только всем существующим угрозам, но и многим угрозам завтрашнего дня.

Оснастка

Оснастка представляет собой основной модуль управления защитой на конечном компьютере сети. Она должна быть установлена на каждом компьютере сети, включая центральный сервер. Основная задача оснастки – управление модулями системы защиты на данном узле и предоставление удобного интерфейса для управления не только центральным сервером, но и администраторами сети.

Помимо этого, функциональность оснастки определяется следующими требованиями:

- оснастка должна иметь **клиентскую часть** для поддержания работы транспортной подсистемы системы защиты**;**
- оснастка должна обладать **кроссплатформенностью**, которая заключается в том, что вся оснастка должна работать как на Windows, так и на *nix-подобных системах;
- оснастка должна иметь удобного **интерфейс для установки/ изменения/ удаления модулей** и компонентов защиты на отдельных узлах сети;
- оснастка должна предоставлять **возможность импорта/экспорта настроек системы защиты** на отдельном узле в файл;
- оснастка должна иметь возможность **поддерживать систему распределённых вычислений;**
- оснастка должна иметь возможность **сбора обширной статистической информации** о функционировании системы защиты на отдельном хосте;
- оснастка должна обладать **возможностью** широкой интеграции в существующие ОС.

Таким образом, в данной работе разработана концепция функционирования распределённой системы комплексной безопасности на основе использования кроссплатформенных оснасток, предназначенных для защиты отдельных узлов сети и управления ими. А также предложена модель технической поддержки в виде создания Аналитического центра для принятия решений на уровне всех установленных систем защиты сетей.

Определение динамических характеристик капли жидкости, нанесенной на поверхность подложки, в задаче оценки степени ее чистоты

М.С. Глянько, П.Ю. Изотов
Институт систем обработки изображений РАН, Самара
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Оценка степени чистоты материала подложки является важным этапом в технологическом процессе формирования микрорельефа дифракционного оптического элемента. Для контроля степени чистоты поверхности подложек могут быть применены различные способы и устройства, но в основном они являются неэффективными или дорогостоящими. Наиболее перспективными и применяемыми на практике являются методы, базирующиеся на использовании закономерностей, имеющих место при смачивании поверхности подложки каплей жидкости. Методам, относящимся к последней категории посвящено большое количество работ, однако недостатком данных методов является тот факт, что оценка степени чистоты поверхности осуществляется по состоянию капли, находящейся в стационарном состоянии, что требует значительных временных затрат.

В статье [Бородин С.А., Волков А.В., Казанский Н.Л. Автоматизированное устройство для оценки степени чистоты подложки по динамическому состоянию капли жидкости, наносимой на ее поверхность // Компьютерная оптика. 2005. В. 28. С. 70–75.] рассмотрен метод оценки степени чистоты поверхности подложки, основанный на измерении динамических характеристик капли жидкости в процессе ее растекания по поверхности. На рисунке 1 приведено схематичное изображение измерительной системы.

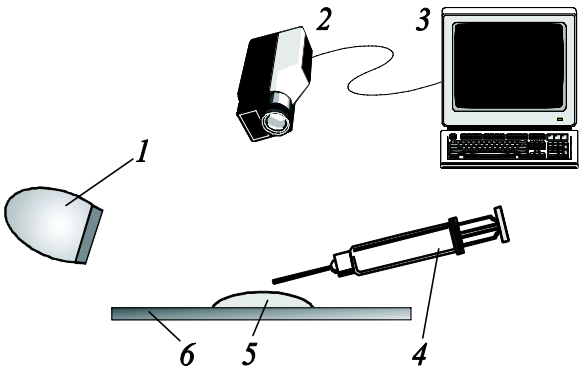


Рисунок 1 – Структурная схема устройства

Для проведения измерений используется установка, состоящая из основания, на котором установлены камера VS-FAST/CG6, дозатор жидкости, системы освещения капли жидкости, нанесенной на поверхность подложки и персонального компьютера, осуществляющего анализ изображений.

В развитие данного способа измерений нами была предложена методика оценки радиуса капли жидкости, в основе которого положены алгоритмы удаления фона, выделения границ капли, определения радиуса капли.

При проведении измерений характеристик подложек, изготовленных из разных материалов, необходимо учитывать особенности светоотражающей способности материала и то, как они влияют на яркостные характеристики изображения капли, растекающейся по поверхности подложки.

Процедура проведения эксперимента, в данном случае, независимо от материала подложки включает в себя несколько этапов:

- получение фона изображения (изображение поверхности подложки до нанесения на ее поверхность капли) как статической части изображения;
- нанесение на поверхность подложки капли жидкости и съемка процесса ее растекания;
- обработка полученной последовательности изображений – для каждого кадра в отснятой камерой последовательности произвести следующие действия:
  - преобразовать изображение кадра из цветного в градации серого;
  - увеличить значение яркости отдельных пикселей на одну и ту же постоянную величину K_1 (эмпирически было определено, что значение этой величины, равное 128 единицам, дает изображение, пригодное для дальнейшей обработке);
  - произвести поэлементное вычитание фонового изображения из полученного;
  - произвести пороговую обработку в интервале [0;255];
  - произвести выделение границ капли по алгоритму, индивидуальному для каждого вида подложки.

Ниже рассмотрены алгоритмы обработки изображений капли на поверхности стеклянной, ситалловой и хромированной подложек.

Алгоритм обработки изображения капли на стеклянной подложке

В результате проведения опытов на стеклянной подложке было установлено, что очертания капли на поздней стадии растекания достаточно хорошо определяются ее тенью (см. рисунок 2).

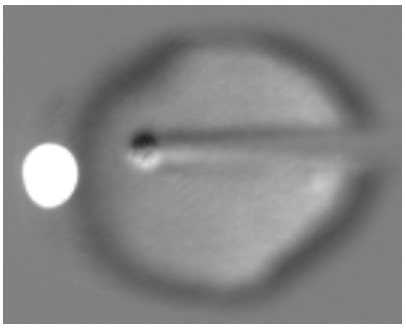
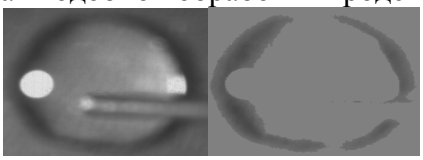


Рисунок 2 – Изображение капли на поздней стадии растекания по стеклянной подложке

Поэтому с помощью пороговой обработки в интервале [0;110], границы которого получены опытным путем, удаляются внешняя и внутренняя область капли, остается только ее тень. Результат подобной обработки продемонстрирован на рис. 3.



а) б)

Рисунок 3 – Изображение капли до и после пороговой обработки

В качестве размера капли выбирается максимум из высоты и ширины прямоугольника, описанного вокруг тени капли:

$$r = \max\left(\left|x_{\max} - x_{\min}\right|, \left|y_{\max} - y_{\min}\right|\right)$$

где $x_{\min}$ – абсцисса крайней левой, $x_{\max}$ – абсцисса крайней правой, $y_{\min}$ – ордината крайней нижней, а $y_{\max}$ – ордината крайней верхней точки тени.

Алгоритм обработки изображения капли на ситалловой подложке

В отличие от стеклянной подложки, в случае, изготовленной из ситалла, визуально отличить удается лишь границу, расположенную по противоположную сторону капли относительно источника освещения. Это можно сделать по ее тени. На этапе выделения границ капли процедуры определения размеров капли значения пикселей с яркостью в интервале [124;140] заменялись на K_1 что позволило устранить шумы матрицы камеры. После чего при поиске границ капли применялся тот же метод, что и в случае стеклянной подложки, но теперь рассматривались пиксели, имеющие яркость, отличную от K_1.

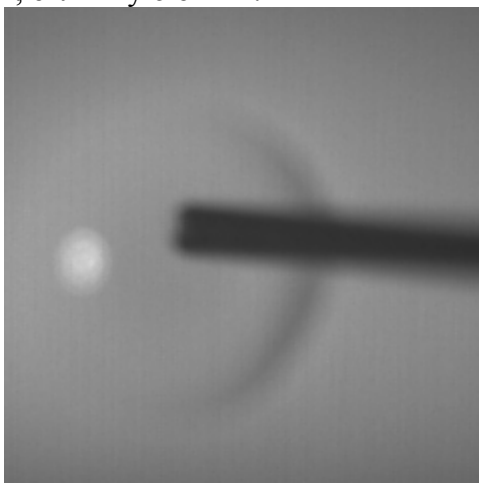


Рисунок 4 – Изображение капли на ситалловой подложке

Чтобы сделать левую границу капли более различимой, в конструкцию устройства был введен оптический барьер – непрозрачная пластинка, расположенная над проверяемой поверхностью перпендикулярно ей. Тень от пластинки накрывает левую часть капли, поэтому свет, испытывая отражения от границы раздела «капля-воздух» и «капля-поверхность», подсвечивает левую границу капли. В результате введения в систему оптического барьера некоторых случаях удается снизить требуемую скорость съемки для фиксации изображения до 50 кадров в секунду.

В процессе исследований было замечено, что в отдельных случаях игла дозатора может давать тень и блики на изображении, полученном в результате вычитания фона вследствие смещения в процессе съемки. Поэтому в процессе определения радиуса из обработки исключались пиксели, лежащие в прямоугольнике, покрывающем изображение иглы.

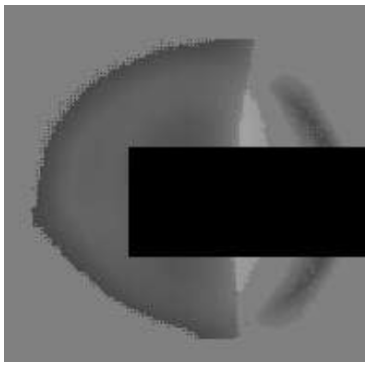


Рисунок 5 – Обработанное изображение капли
(вырезанный участок содержит изображение иглы)

Алгоритм обработки изображения капли на хромированной подложке

В случае хромированной подложки лучи, составляющие основную часть интенсивности освещения, попадающего в объектив камеры, отражаются от поверхности капли и соответствуют бликам (рисунок 6).



Рисунок 6 – Капля, растекающаяся по поверхности хромированной подложки

На этом этапе задача выделения границ капли переходит в разряд трудноreализуемых. Поэтому в качестве динамической характеристики капли, используемой для оценки шероховатости используемой подложки, было предложено использовать расстояние между бликами, расположенными на противоположных концах капли. Для этого вместо выделения границ капли и измерения ее радиуса была предложена следующая последовательность шагов:

- для избавления от шумов матрицы видеокамеры яркости пикселей, лежащих в интервале $[K_1 - K_2; K_1 + K_2]$, заменяются на K_1 (значение K_2 было эмпирическим путем выбрано равным 6);
- определяется смещение крайней правой точки правого блика относительно крайней левой точки левого блика.

Рассчитанная указанным способом характеристика имеет тот же смысл, что и радиус капли в случае стеклянной или ситалловой подложек.

Заключение

Разработано программное обеспечение, реализующее обработку изображений капли, получаемых в режиме реального времени непосредственно со скоростной видеокамеры VS-FAST/CG6.

Предложенная методика оценки параметров динамического состояния капли жидкости позволяет оценить степень чистоты подложек из различных материалов.

Разработка алгоритмов локализации инородного тела на рентгенографических изображениях с использованием искусственных нейронных сетей

Е.С. Залеткин, А.В. Куприянов, Н.Ю. Ильясова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара
Институт систем обработки изображений РАН

Введение

В современной медицине часто ставится задача обнаружения и локализации внутриглазных инородных объектов. Для хирургического вмешательства необходимо точно определить пространственное положение инородного объекта внутри окулярной области. Диагностика внутриглазных инородных тел (ВИТ) проводится по рентгенограммам, выполненным в 2-х проекциях: фронтальной и боковой. Расчет положения ВИТ относительно оси глаза, расстояние его от наружной капсулы глаза осуществляется с помощью специальных расчетных сеток с миллиметровыми делениями. Визуальное оценивание параметров и положений исследуемых объектов носит субъективный характер и позволяет дать только качественные оценки. Этого недостаточно для применения в медицине[1].

Перед проведением исследований на глазное яблоко пациента накладывается протез-индикатор со свинцовыми метками, которые служат ориентирами при расчете на снимке положения внутриглазного инородного тела и его размеров. Так как свинец – материал, имеющий большую плотность, то метки индикатора, полученные на изображении, имеют большую яркость и хорошо заметны. Кроме того, эти метки, будучи геометрически расположенными на двух взаимно перпендикулярных прямых, образуют подобие системы координат, с помощью которой возможно определить в плоскости изображения координаты залегания ВИТ.

Исходными данными являются изображения рентгеновских снимков черепной коробки с нанесенным протезом – индикатором на около окулярную область. Пример рентгенограммы представлен на рисунке 1.



Рисунок 1 – рентгенограмма черепной коробки, содержащей протез – индикатор и инородное тело

Применение искусственных нейронных сетей

Среди различных методов оптического распознавания образов применяют такие методы, как метод перебора каждого объекта под различными углами, масштабами, смещениями; исследование свойств контура объектов; использование искусственных

нейронных сетей и т.д. В данной работе продемонстрированы возможности использования нейронных сетей.

Искусственные нейронные сети (ИНС) не программируются в привычном смысле этого слова, они обучаются. Возможность обучения — одно из главных преимуществ нейронных сетей перед традиционными алгоритмами. Технически обучение заключается в нахождении коэффициентов связей между нейронами. В данной работе в качестве ИНС используется персептрон с входными и двумя активными слоями. Элементы входного слоя не выполняют никакой работы, а лишь передают сигналы на скрытый (ассоциативный) слой. Нейроны скрытого и выходных слоев содержат весовые коэффициенты соответствующие выходам предыдущего слоя. Скрытый слой проводит взвешенное суммирование и рассчитывается выходное значение нейрона с использованием функции активации. В данном методе мы использовали смещенную сигмоиду, т.к. смещение позволяет уменьшить количество итераций на обучении на 30-50% [2]. Результаты вычислений скрытого слоя передаются на последний(выходной) слой. Далее, происходит процедура, аналогичная скрытому слою и рассчитываются выходы последнего слоя, которые мы и принимаем за ответы.

Важной частью правильного функционирования нейронной сети является выбор обучающего алгоритма. Для корректного обучения персептрона требуется учитель, т.е. на этапе обучения к множеству тестовых входных данных необходимо также добавить еще и соответствующее им множество требуемых результатов. В многослойных сетях оптимальные выходные значения нейронов всех слоев, кроме последнего, как правило, не известны, и двух- или более слойный персептрон уже невозможно обучить, руководствуясь только величинами ошибок на выходах ИНС. Наиболее приемлемый вариант — распространение сигналов ошибки от выходов ИНС к ее входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы. Этот алгоритм обучения ИНС получил название процедуры обратного распространения[3].

Реализация алгоритма

Входными данными для нейронной сети послужили компоненты массива размером 121. Сам массив получается из окна поиска размером 11 на 11 пикселей. Элементы имеют тип double и характеризуют яркость пикселей. Перед вводом данных массив был нормирован до получения нулевого математического ожидания, чтобы уменьшить зависимость компонент входного массива от интенсивности освещения[4].

$$\beta = \frac{1}{N}\sum_{i,j} p_{ij},$$

где N – число отсчетов в массиве.

Реализация алгоритма локализации объекта проводилась в три этапа. На первом этапе исследуемое изображение обрабатывалось путем его полного построчного и постолбцового сканирования окном поиска. Строилось изображение, состоящее только из элементов, при обработке которых нейронная сеть давала выход, лежащий в установленном диапазоне. На втором этапе, для того, чтобы можно было применить геометрический анализ расположения полученных выше областей потребовалось провести кластеризацию. Для решения этой задачи был применен один из алгоритмов кластеризации, называемый методом ближайшего соседа. Создавался двумерный список. Каждый из его элементов содержал номер области. В область, в свою очередь, записывались координаты всех точек, принадлежащих ей. Первая точка из списка обозначалась как первая область. Далее, с ней сравнивались координаты остальных точек списка. Если расстояние между их координатами оказывалась больше заданного

порога, сравниваемая точка записывалась в новую область и координаты проверяемых точек сравнивались также и с этой областью. Далее, поочередно сравнивались все области на наличие точек, лежащих на расстоянии меньше порогового. В случае если такие точки были обнаружены, их области объединялись в одну. Проверка происходила до момента, когда не оставалось близлежащих областей. На третьем этапе выполнялся алгоритм, оценивающий геометрическое положение точек относительно друг друга. Искомый маркер образует некое подобие виртуального креста, состоящего из четырех точек. Пиксельное расстояние между точками известно. Сравнивались горизонтальные и вертикальные длинны отрезков, образуемые центрами областей. В том случае, если находилась пара, состоящая из горизонтального и вертикального отрезка, выполнялась проверка на принадлежность серединных точек отрезков области их пересечения.

Экспериментальные исследования

Для проверки работоспособности реализованного алгоритма были проведены серии экспериментальных измерений на наличие ошибок локализации. К таким ошибкам относятся: 1) количество областей, принадлежащих объекту на исходном изображении, но не выделенных алгоритмом; 2) количество областей, принадлежащих фону на исходном изображении, но выделенных алгоритмом, как искомые области.

Нейронная сеть обучалась на пустых снимках, содержащих только маркер и фон.

В качестве обучающих примеров, с положительными значениями вектора – учителя, использовались снимки с наиболее чистым изображением метки на протезе. Также в обучающие примеры входили области, с объектами, которые могут быть восприняты нейронной сетью как искомые маркеры. Вектор учитель в данном случае имел отрицательные значения. Это было выполнено с целью уменьшить количество ошибок, связанных с ложным распознаванием объектов, которые не являются частью протеза – индикатора. Обучение проводилось до момента, когда выходы нейронной сети не стали удовлетворять заданной погрешности. После получения удовлетворительных результатов на этапе обучения, были проведены исследования на реальных рентгенографических снимках черепной коробки, содержащей внутриглазную область и протез – индикатор, а также были исследованы воздействия шумов и размывки на устойчивость алгоритма. Графики представлены ниже на рисунках 2 и 3.

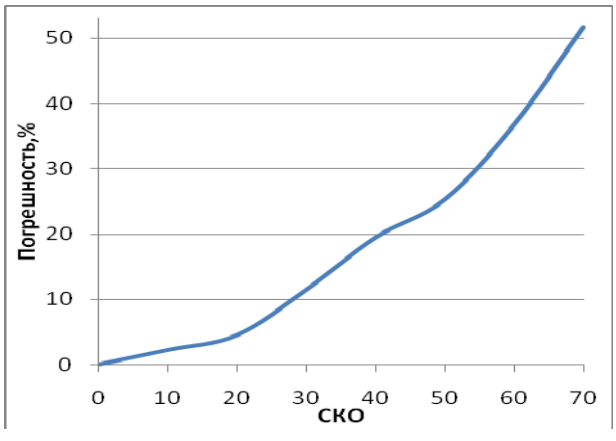


Рисунок 2 – График зависимости погрешности от СКО

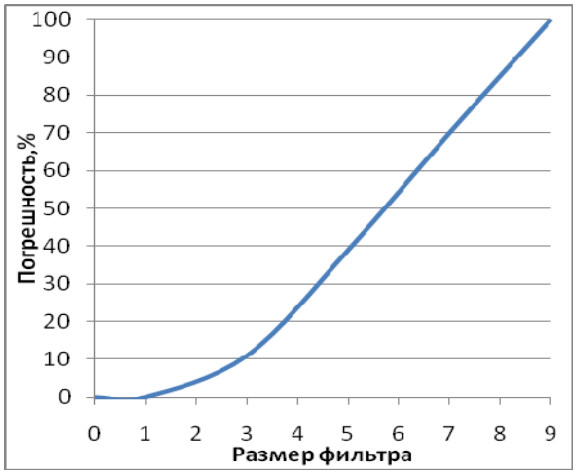


Рисунок 3 – График зависимости погрешности от порядка фильтра

Графики, характеризующие эти опыты, наглядно демонстрируют, что данный метод способен адекватно решать задачу локализации при уровне шума с СКО не более 30 и размеров смазывающего фильтра не более 3 и обеспечивает погрешность менее 10%.

Заключение

В результате исследований разработаны методы и алгоритмы локализации протеза – индикатора на рентгенографических снимках черепной коробки. Данные методы были основаны на применении искусственных нейронных сетей для задач распознавания и классификации объектов. Была спроектирована и реализована компьютерная система, предназначенная проведения сравнительного анализа различных систем классификации.

Были проведены серии вычислительных экспериментов, исследующих работоспособность разработанного метода:

1. Реализованный алгоритм был протестирован на пустых снимках индикатора с различными распределениями яркости на фоне.
2. Были проведены эксперименты по устойчивости указанного алгоритма к добавлению аддитивных шумов и размывки. Получены экспериментальные графики и таблицы ошибок в зависимости от уровня добавленного шума и размера сглаживающего фильтра.

На основании проведенных экспериментов можно сделать вывод, что разработанные методы, могут быть использованы при локализации инородных объектов на реальных рентгенографических снимках для автоматизации медицинских диагностик.

Литература

1. Волков В.В., Горбань А.И. Клиническое исследование глаза с помощью приборов. Санкт-Петербург, 1991. 320с.

2. H. Drucker, R. Schapire, P. Simard. Boosting Performance in Neural Networks, 1993. 703p.

3. В.Заенцев. Нейронные сети: Основные модели. Волгоград: Издательство ВолГУ, 2004. 225с.

4. И.С. Грузман, В.С. Киричук. Цифровая обработка изображений в информационных системах. Новосибирск: Изд-во НГТУ, 2002. 352с.

Исследование свойств внутриглазного инородного тела на основе анализа рентгенографических изображений черепа

Н.Ю.Ильясова, А.В. Куприянов, С.Г. Засканов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара
Институт систем обработки изображений РАН, Самара

Введение

Актуальность работы обоснована необходимостью автоматизации процессов диагностирования и локализации внутриглазного инородного тела. Одной из основных причин неудачного или несвоевременного извлечения внутриглазного инородного тела (ВИТ) является некачественная диагностика из-за несовершенства диагностической техники, ошибок в проведении диагностических манипуляций, наличия мелких (менее 1,0 мм) осколков, не улавливающихся ультразвуком. Среди методов, с помощью которых можно определить локализацию ВИТ, наиболее широкое распространение получило рентгеновское исследование, в результате которого наблюдаются изображения рентгеновских проекций и по ним проводится оценивание различных характеристик и параметров, позволяющих сделать предположение об уточненном диагнозе. Достоинствами метода является высокая информативность о взаимоотношении ВИТ и внутриглазных структур, быстрота и безболезненность исследования, возможность проведения его в ранние сроки после травмы, выявление материала инородных тел (металл, стекло, пластмасса, дерево и др.).

Данная работа осуществляется совместно с Самарской клинической офтальмологической больницей им. Т.И. Ерошевского, где применяемый в настоящее время метод рентгеновского исследования, основанный на использовании стандартно-схематической модели глазного яблока, не позволяет использовать весь спектр возможностей высококачественного современного диагностического оборудования.

Согласно методике, используемой в Самарской клинической офтальмологической больнице им. Т.И. Ерошевского, диагностика ВИТ проводится по рентгенограммам, выполненным в двух проекциях: фронтальной и боковой. Расчёт положения ВИТ относительно оси глаза, расстояние его от наружной капсулы глаза осуществляется с помощью специальных расчетных сеток [1] с миллиметровыми делениями. Перед приведением R-исследований на глазное яблоко пациента накладывается протез-индикатор со свинцовыми метками, которые служат ориентирами при расчете на R-грамме положения ВИТ и его размеров. Эти метки, будучи геометрически расположенными на двух взаимно перпендикулярных прямых, образуют подобие системы координат, с помощью которой возможно определить в плоскости изображения координаты залегания ВИТ. Итогом совместной с врачами работы является создание пространственной компьютерной модели внутриглазного инородного тела и параметрической модели глазного яблока с механизмом отслеживания временных изменений параметров на основе рентгенографических данных. Это позволяет на дооперационном этапе проводить обнаружение ВИТ, определять его взаимоотношения с окружающими тканями глаза, корректировать объем хирургического вмешательства.

Задачей данной работы является определение зависимости соотношений значений функции яркости опорных маркерных точек протеза-индикатора и объекта на рентгенографическом изображении от набора начальных данных, определяющих интенсивность излучения. В состав данных входят: задаваемое в рентгеновском аппарате напряжение, сила тока на рентгеновской трубке, экспозиция (время

излучения). Таким образом, может быть установлена непосредственная зависимость значения функции яркости отдельного элемента изображения ВИТ от интенсивности излучения.

Взаимодействие рентгеновских лучей и вещества

Поглощение рентгеновских лучей. Рассмотрим суммарное поглощение, вызываемое фотоэлектрическим эффектом и рассеянием [2][3]. Последняя причина поглощения к тому же незначительна по сравнению с первой, за исключением тех случаев, когда лучи с короткими длинами волн (но эти случаи остаются за пределами рассматриваемой нами области) проходят через лёгкие элементы.

Будем считать, что экран очень тонкий и однородный; пучок монохроматических лучей с сечением, равным единице, пересекая его, теряет энергию dI, пропорциональную пересеченной массе экрана на 1 $см^2$ dp и первичной интенсивности пучка I. Запишем это в виде: $dI = -\mu I dp$, где μ - коэффициент массового поглощения вещества экрана. Для экрана конечной толщины, интегрируя, находим, что отношение интенсивности падающего пучка к интенсивности прошедшего пучка зависит от массы на единицу поверхности p, пересеченной пучком, и определяется формулой $\frac{I}{I_0} = \exp(-\mu p)$, что равносильно $\frac{I}{I_0} = \exp(-\mu\delta x)$, где x - толщина слоя, а δ - плотность поглощающего вещества. Коэффициент массового вещества μ - постоянен для данного вещества, в каком бы физическом состоянии оно не находилось. Коэффициент линейного поглощения $\mu\delta$ зависит от плотности экрана. Поглощение рентгеновских лучей веществом – явление атомное: для сложного вещества оно равно сумме поглощений всеми составляющими элементами и в первом приближении не зависит от характера связи между ними [4]. Причина этой зависимости в том, что рентгеновские фотоны, обладая высокой энергией, взаимодействуют с электронами внутренних оболочек, тогда как для видимого света имеют значение внешние электроны, состояние которых уже существенно зависит от характера межатомных связей.

Зависимость поглощения от длины волны. Показатель поглощения элемента является одновременно функцией рассматриваемой длины волны и атомного номера элемента. Увеличение μ происходит скачками, неравномерно: поглощение резко возрастает, когда длина волны, уменьшаясь, проходит некоторые постоянные величины, называемые порогами поглощения. Эти разрывы объясняются механизмом фотоэлектрического эффекта, т. е. поглощения фотонов атомами. Как только λ становится меньше λ_K (предельная длина волны оболочки), поглощение слоем резко увеличивается, а затем постепенно убывает при дальнейшем уменьшении λ. Внутри одной области длин волн коэффициент поглощения приблизительно пропорционален λ^3.

Зависимость поглощения от природы элемента. Будем считать постоянной длину волны. Для данного монохроматического излучения поглощение, вообще говоря, возрастает с атомным номером Z. Легкие элементы (органические вещества, алюминий) поглощают достаточно слабо, тогда как тяжёлые элементы поглощают весьма сильно; хорошо известно, что для защиты от рентгеновских лучей используется свинец. Но увеличение поглощения по мере возрастания атомного номера не непрерывно, а происходит скачками; скачки эти вызываются теми же причинами, что и в предыдущем случае. Для области, заключенной между двумя скачками, коэффициент поглощения также приблизительно пропорционален Z^3.

Определение коэффициента поглощения. Используем эмпирические законы, связывающие μ с λ и Z, для интерполяционного расчета значений μ для различных элементов и длин волн. Будем считать, что коэффициент поглощения пропорционален и произведению $\lambda^3 \cdot Z^3$: $\mu = C \cdot \lambda^3 \cdot Z^3$, где C - неизвестная константа. Её мы определим, взяв из таблицы несколько пробных значений длин волн и атомных номеров интересующих нас элементов – алюминия и свинца. В результате получим приближенное значение: $C \approx 0.00619052 \cdot 10^{30}\, м^{-4} = 6.19052 \cdot 10^{-13} \left(\overset{\circ}{A} \right)^{-4}$.

С увеличением как тока, так и напряжения интенсивность излучения увеличивается, причём в первом случае оно линейное, во втором – почти квадратичное. Это подтверждает, в частности представление о том, что суммарная энергия излучения пропорциональна квадрату напряжения.

2. Экспериментальные исследования

Были обработаны рентгенографические изображения, представленные в медицинском формате DICOM – это снимки рассматриваемого алюминиевого протеза. Снимки получены в результате непосредственного эксперимента: ряд фотографий сделан при варьировании в некоторых пределах параметров установки – тока, напряжения и экспозиции. Обработка изображения (оцифрованного снимка) заключалась: а) в нахождении значения функции яркости $F(m,n)$ отдельных отсчётов (каждого снимка-изображения) в виде массива, путем считывания этой информации из dcm-файла, б) в подсчете средней яркости по конкретным областям, а именно: по маркерным точкам и непосредственно по «кольцу» протеза-индикатора. Также необходимо найти величину, при которой спектр (распределение по частотам) интенсивности оставался бы неизменным, и которая определялась бы поглощением излучения. Согласно закону поглощения – закон Бугера: $\ln\left(\frac{I}{I_0}\right) = -\left(\mu_{Al} x_{Al} + \mu_{Pb} x_{Pb}\right)$. Таким образом, этой величиной является логарифм отношения интенсивностей после и до поглощения, т.к. он пропорционален коэффициенту поглощения, т.е. может определять поглощающие свойства. В дальнейшем удобнее использовать в качестве оценки отношения интенсивностей отношение значений функций яркости: $k = \ln\left(\frac{\overline{F}}{\overline{F_0}}\right) = -\mu x$, т.к. между ними прямая монотонная зависимость, и подтвердим это экспериментально. В формуле обозначено: $\overline{F}$ - среднее значение функции яркости для кольца протеза, а $\overline{F_0}$ - среднее значение функции яркости маркера.

На рисунке 1 показан пример оцифрованного изображения типового снимка протеза-индикатора:

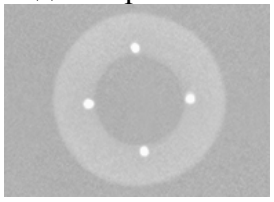


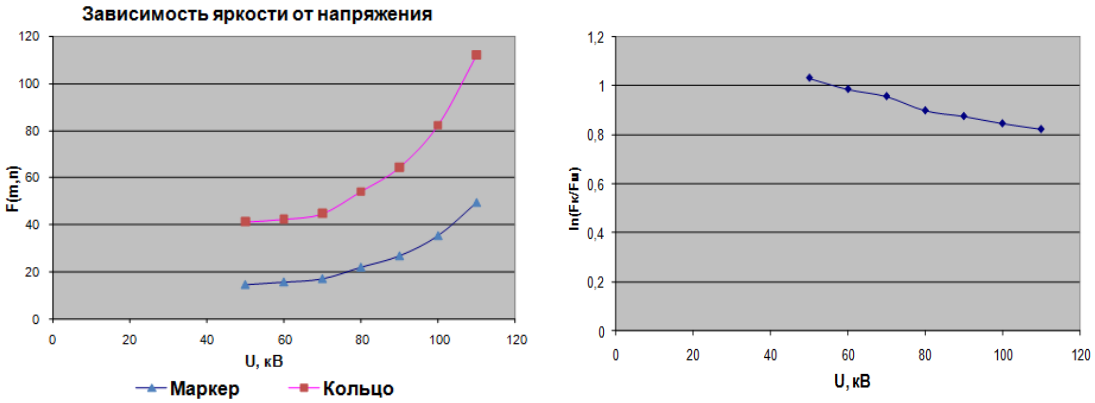
Рис. 1.

В результате эксперимента, заключавшегося в получении (при различных напряжениях) и дальнейшей оцифровке снимков протеза, были определены значения функции яркости отсчётов всего кольца и маркерных точек DICOM изображений. В последнем столбце таблицы 1 подсчитан логарифм отношения значений функций яркости кольца и маркера для соответствующего значения напряжения.

Здесь сила тока постоянна и равна i=320мА, параметр яркости – логарифм отношения яркостей кольца и маркера. Графики для этой таблицы приведены на рисунках 2а и 2б:

Таблица 1

№	U, кВ	Значение функции яркости кольца	Значение функции яркости маркера	Параметр k
1	50	41,15	14,66	1,032
3	70	44,66	17,15	0,957
5	90	64,40	26,84	0,875
7	110	112,41	49,39	0,822



а) *б)*

Рис. 2

В таблице 2 приведена зависимость яркостей тех же областей, что и в предыдущем случае, от значения силы электрического тока при постоянном напряжении U=110кВ:

Таблица 2

№	i, мА	Значение функции яркости кольца	Значение функции яркости маркера	Параметр k
1	320	111,00	49,39	0,819
2	250	95,55	42,15	0,820
3	200	82,62	36,38	0,820
4	160	74,45	32,73	0,821
5	100	69,08	30,39	0,821

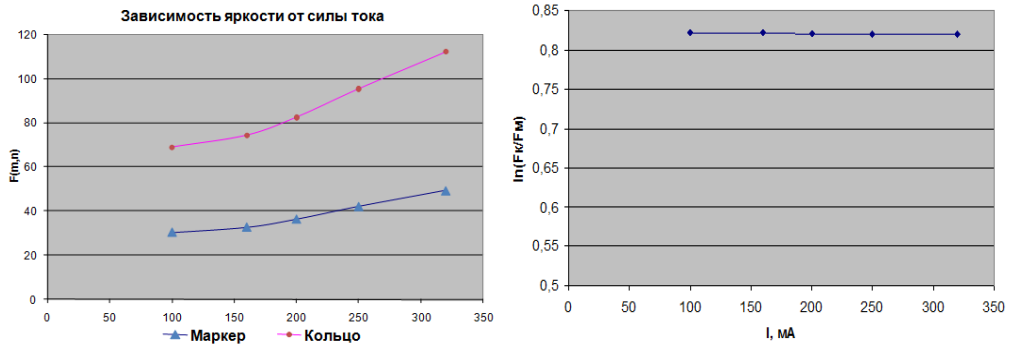
Графики для значений яркости и величины логарифма отношения яркостей приведены на рисунках 3а и 3б соответственно. Анализируя полученные графики, сделаем следующие выводы. График на рисунке 2а показывает, что с ростом напряжения абсолютное значение функции яркости возрастает; на рисунке 2б видно, что величина разности логарифмов яркостей кольца и маркера постепенно убывает, поскольку спадает поглощающая способность излучения и увеличивается его проникающая способность. Абсолютное значение функции яркости с ростом силы тока практически линейно возрастает (при неизменном напряжении), что отражено на рисунке 3а, а величина логарифма отношения яркостей остается почти постоянной (это видно на рисунке 3б), поскольку спектр (распределение) не меняется. Это говорит о том, что измерения, полученные на практике, хорошо согласуются с теорией, в данном случае – с законом поглощения.

Можно сформулировать алгоритм анализа свойств инородного тела (рис.4), который в дальнейшем можно использовать в задаче автоматического распознавания ВИТ на рентгеновских снимках, что является предметом дальнейшего исследования:

1. Используем изображение во фронтальной проекции, так как внутри области глаза фон можно считать однородным.

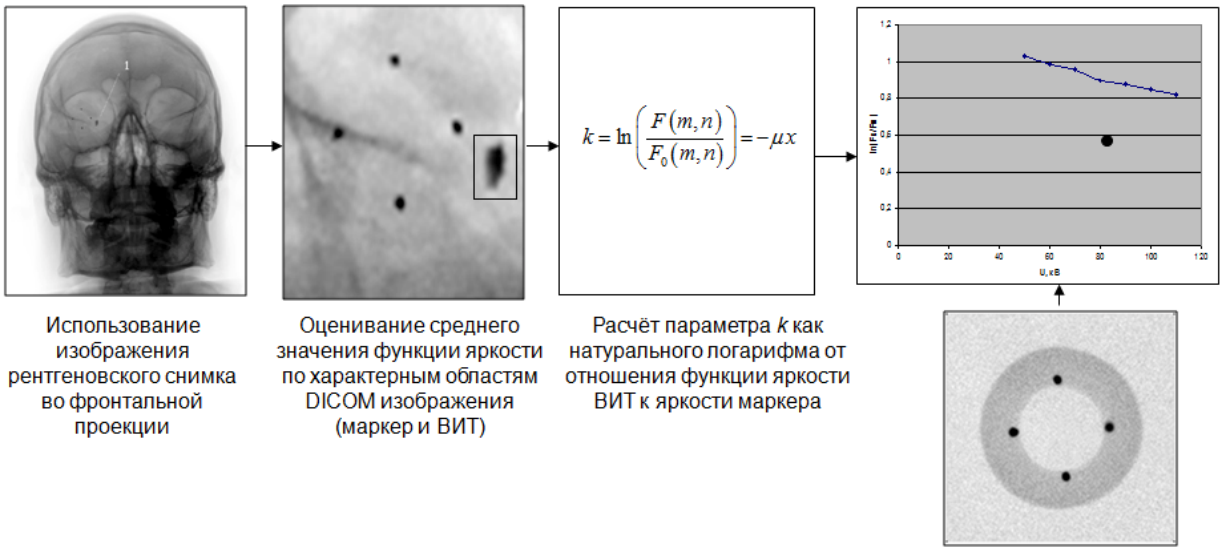
2. Определяем логарифм отношения функции яркости $F(m,n)$ маркеров и ВИТ (оценивание параметра k).

3. Анализируя полученное значение k и известные значения функции яркости маркерных точек при заданных значениях напряжений, силы тока, и времени экспозиции можно сделать выводы о физических свойствах инородного тела, сравнимого по толщине с толщиной маркерной точки.



а) *б)*

Рис. 3.



Использование изображения рентгеновского снимка во фронтальной проекции

Оценивание среднего значения функции яркости по характерным областям DICOM изображения (маркер и ВИТ)

Расчёт параметра k как натурального логарифма от отношения функции яркости ВИТ к яркости маркера

Рис. 4.

Заключение

В ходе работы было выяснено, что интенсивность рентгеновского излучения зависит практически линейно от силы тока установки и квадратично от напряжения, что соответствует верным энергетическим представлениям, поскольку энергия всего спектра пропорциональна квадрату напряжения, и физическим соображениям, так как подтверждается закон поглощения. Получен способ определения общих свойств внутриглазного инородного тела, основываясь на соотношениях яркостей отсчётов на рентгенограмме и на экспериментальных кривых зависимости яркости. Была выведена величина, определяемая поглощением (логарифм отношения яркостей отдельных областей), согласно которой определить свойства инородного тела, зная его значение функции яркости.

Литература

1 Поляк, Б. Л. Хирургическое лечение ранений с внутриглазными инородными телами [Текст]/Б. Л. Поляк. – М.: Л. Медгиз., 1957. – 388 с.

2 Гинье, А. Рентгенография металлов [Текст]/А. Гинье. – М.: Физматгиз, 1961. – 602 с.

3 Блохин, М.А. Физика рентгеновских лучей [Текст]/М.А. Блохин. – М.: Гос. изд-во технико-теор. лит., 1957. – 518 с.

Система комплексной информационной защиты распределённых сетей

М.В. Кузнецов, М.С. Воронов, А.А. Иванов, А.А. Солонщиков, Н.В. Студеникин, С.А. Бобров, М.В. Яшин, И.И. Меркульев,
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Информационная безопасность складывается из множества отдельных компонентов и, только применяя их в комплексе, можно организовать надёжную систему, обеспечивающую достоверность, конфиденциальность, точность и подлинность передаваемой в сети информации. При этом автоматизированная система безопасности должна противостоять действиям злоумышленников, пытающихся нарушить штатную работу, как отдельных элементов, так и сети в целом.

Всю систему безопасности можно условно разделить на две части: центральный сервер и оснастки. Центральным сервером может быть любой компьютер, на который необходимо всего лишь поставить соответствующее разработанное ПО. Оснастка предназначена для защиты и контроля конечных узлов. Она является кроссплатформенной и должна устанавливаться на каждый компьютер сети, включая серверы. А потом уже на оснастку посредством центрального сервера устанавливаются необходимые пользователю модули. Также можно выделить из второй своеобразную третью часть, которая, в целях повышения быстродействия, состоит из программно-аппаратного устройства (IPS+DLP+файрвол). Эта часть также поддерживает модульную систему, т.е. её функциональность напрямую зависит от установленных в системе модулей (Рис. 1). В целом, можно сказать, что система безопасности в некотором роде напоминает сетевую распределённую ОС.

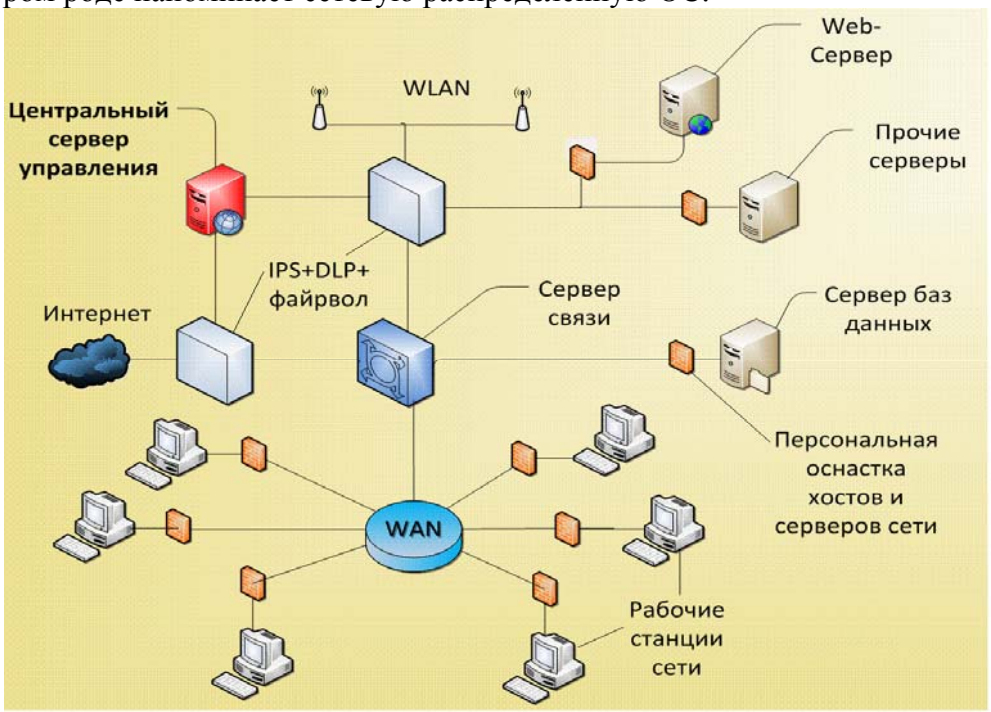


Рис.1 Структурная схема защиты сети.

Общесистемные требования к современной комплексной информационной безопасности распределённой сети

Система безопасности должна иметь централизованное управление. Для полноценного функционирования всех компонентов защиты предполагается наличие в

структуре системы безопасности выделенного узла с установленным центром управления защитой (ЦУЗ). Данный сервер должен помимо управления всеми остальными частями системы защиты, заниматься установкой обновлений, хранить собранную информацию о функционировании сети, контролировать параллельную обработку задач в рамках сети, а также периодически проводить аудит безопасности.

Такое централизованное управление позволит быстро обнаруживать и устранять уязвимости в системе безопасности.

Структура защиты сети должна иметь модульное строение. Основную (программную) часть системы безопасности можно разделить по функциональности на: модуль файрвола; модуль антивируса; модуль IPS/IDS; модуль DLP; модуль сервера обновлений; модуль системного журнала логов; модуль обеспечивающий защиту от спам-атак; модуль распараллеливания задач системы безопасности в рамках сети.

Все компоненты системы безопасности могут быть инсталлированы на центральный сервер, который затем произведёт реконфигурирование остальных хостов и установит, если требуется, программное сопровождение модуля.

Структура защиты должна строится на основе устройства, объединяющего функции IPS, файрвола, DLP и спам-анализатора. Это устройство – один из основных компонентов защиты от внешних угроз. Оно должно представлять собой программно-аппаратную реализацию и устанавливаться в точки соединения сети с внешним миром. А самый минимальный компонент модульной системы безопасности может включать только данный модуль.

Механизм защиты сетей должен использовать все преимущества распределённых вычислений, используя для этого ресурсы защищаемой сети.

Анализ и сбор информации по свободным вычислительным ресурсам сети – одна из функций центрального сервера. Она будет осуществляться следующим образом: сначала система на протяжении 1-2 недель собирает статистическую информацию о средней загруженности всех хостов сети, потом при необходимости использования распределённых ресурсов опрашивает хосты по возрастанию средней загруженности на предмет возможности использования их ресурсов. Преимущества распределённых вычислений будут использовать все критические узлы сети, например такие, как IPS+файрвол (для анализа заархивированного/зашифрованного трафика или просто при большой загруженности) и антивирусы на конечных узлах сети (для проверки файлов по таблице хешей).

Это позволит: уменьшить нагрузку как на отдельный хост сети, так и на систему IPS; увеличить объём пропускаемого трафика в единицу времени через IPS.

При разработке важно учесть то требование, что система параллельной обработки является всего лишь устанавливаемым модулем. Это необходимо, например, для того, чтобы обеспечивать защиту тех сетей, в которых уже организована параллельная обработка информации (например, популярные GRID-сети).

Для поддержания всех созданных систем безопасности корпоративных систем должен быть создан Аналитический центр. В число его задач будет входить: выпуск обновлений для систем безопасности. К этому относится не только обновление списка сигнатур и выпуск патчей, но и поддержка системы виртуального патчинга. После выпуска очередного пакета обновлений, центр «сообщает» об этом подключенным серверам сетей, которые их автоматически скачивают и устанавливают; сбор информации о функционировании всех систем безопасности сетей, созданных ЛК и подключенных к данному центру. К сбору информации также относится пополнение базы сигнатур вирусов и атак на основе логов IPS; принятия решения о блокировании трафика на уровне сервера центра в случае обнаружения схожей аномалии в других сетях.

Это позволит: в реальном времени наблюдать за распространением вирусной инфекции в любом уголке мира; препятствовать дальнейшему заражению, конфигурируя сенсоры межсетевых экранов таким образом, чтобы избежать попадания вредоносного трафика внутрь сети.

Работа такого Аналитического центра позволит в значительной степени усилить систему безопасности корпоративных сетей: защита отдельной интранетовской сети будет зависеть не только от принятия решения на уровне этой сети, но и от вычислительных способностей центра, квалификации работающих там специалистов, а также обширной аналитики, собранной с других подобных сетей. Это позволит эффективно противостоять угрозам не только всем существующим угрозам, но и многим угрозам завтрашнего дня.

Защита конечных узлов сети в составе защищаемой сети будет осуществляться за счёт установки на них оснастки. Оснастка будет включать в себя функции файрвола и DLP. Главное предназначение оснастки: осуществление политик безопасности; антивирусная защита; взаимодействие с центральным сервером для обеспечения распараллеливания задач; обновление журнала логов на сервере; обеспечение защиты от атак внутри сети; взаимодействие с центральным сервером. Основное требование к оснастке – высокая степень интеграции в существующие и разрабатываемые ОС.

Масштабируемость. Система должна обеспечивать работу в широком диапазоне различных количественных характеристик сети.

Система должна быть кроссплатформенной. Одним из основных преимуществ системы безопасности должна быть кроссплатфоменность, то есть модули могут быть развернуты на любой аппаратной архитектуре (RISC, Intel 32-bit, 64-bit) и под распространенными операционными серверными системами (AIX, Solaris, GNU/Linux, Windows NT), что позволит эффективнее использовать вычислительные мощности. Кроссплатформенность всей системы, в первую очередь обеспечивается за счёт кросспалтформенности оснастки и модулей, которые будут устанавливаться через неё в систему.

Интеграция с Active Directory. Немаловажным требованием, предъявляемым к оснастке, является интеграция её со службой каталогов Windows. Интеграция будет происходить на уровне учётных записей пользователей, которые будет контролировать система DLP, и на уровне оснастки, информация о каждом модуле которой для управление им также должна храниться в распределённой базе каталогов.

Система должна предоставлять удобный интерфейс для конфигурирования

Центральный сервер должен быть оснащён удобной программной системой управления сетью, которая должна включать в себя удобное управление всеми возможными настройками.

Система должна брать на себя основные обязанности по обеспечению информационной безопасности. Это позволит устанавливать её в те компании, где нет штатных специалистов по обеспечению информационной безопасности, а также снизить влияние человеческого фактора.

Поддержка большого числа протоколов связи. Для связи центрального сервера с оснастками может использоваться как протокол TCP/IP, так и IPX/SPX/IPV6, что позволяет развернуть систему безопасности, не меняя исторически сложившейся инфраструктуры сети.

Таким образом, в данной статье представлен новый вариант построения комплексной защиты распределённой сети, который позволяет обеспечить полную защиту корпоративной сети, как от всех существующих угроз, так и от угроз завтрашнего дня.

Требования к антивирусной защите в современных корпоративных сетях

А.А.Солонщиков, М.В.Кузнецов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Централизованная антивирусная защита компьютерных сетей – намного более сложная задача, нежели защита обычного домашнего компьютера, которую в большинстве случаев сможет обеспечить хорошее антивирусное программное обеспечение.

Корпоративные сети подвергаются опасности намного чаще, и зачастую компании рискуют крайне важной информацией (бух.отчетность, закрытая документация, пароли и пр.). Кроме того, вредоносное ПО, попадая в ИТ-инфраструктуру, гарантирует постоянные сбои в работе системы и неминуемо приводит к убыткам.

Существуют и на сегодняшний день с успехом применяются наиболее известные антивирусные продукты, которые не всегда успешно справляются с предъявляемыми к ним требованиям.

Антивирусная защита — программное средство, предназначенное для борьбы с вирусами, основными задачами которого являются:

- Препятствование проникновению вирусов в компьютерную систему
- Обнаружение наличия вирусов в компьютерной системе
- Устранение вирусов из компьютерной системы без нанесения повреждений другим объектам системы
- Минимизация ущерба от действий вирусов.

Сигнатурный анализ — метод обнаружения вирусов, заключающийся в проверке наличия в файлах сигнатур вирусов.

Это метод, при котором антивирусная программа, анализируя файл, обращается к антивирусным базам, составленным производителем программы-антивируса. В случае соответствия какого-либо участка кода просматриваемого файла (сигнатуре) вируса в базах, программа-антивирус может удалить инфицированный файл, заблокировать доступ к инфицированному файлу, отправить файл в карантин, попытаться «вылечить» файл, удалив тело вируса из файла.

Эвристический анализ — технология, основанная на вероятностных алгоритмах, результатом работы которых является выявление подозрительных объектов.

В процессе эвристического анализа проверяется структура файла, его соответствие вирусным шаблонам. Наиболее популярной эвристической технологией является проверка содержимого файла на предмет наличия модификаций уже известных сигнатур вирусов и их комбинаций. Это помогает определять гибриды и новые версии ранее известных вирусов без дополнительного обновления антивирусной базы.

Поведенческий анализ — технология, в которой решение о характере проверяемого объекта принимается на основе анализа выполняемых им операций.

Антивирусы, использующие метод обнаружения подозрительного поведения программ не пытаются идентифицировать известные вирусы, вместо этого они прослеживают поведение всех программ. Если программа пытается выполнить какие-либо подозрительные с точки зрения антивирусной программы действия, то такая

активность будет заблокирована, или же антивирус может предупредить пользователя о потенциально опасных действиях такой программы.

Метод обнаружения при помощи эмуляции. Некоторые программы-антивирусы пытаются имитировать начало выполнения кода каждой новой вызываемой на исполнение программы перед тем как передать ей управление. Если программа использует самоизменяющийся код или проявляет вирусную активность, такая программа будет считаться вредоносной, способной заразить другие файлы. Однако этот метод тоже изобилует большим количеством ошибочных предупреждений.

HIPS — система мониторинга всех приложений, работающих в системе, с чётким разделением прав для разных приложений. Таким образом, HIPS может предотвратить деструктивную деятельность вируса, не дав ему необходимых прав. Приложения делятся на группы, начиная от «Доверенных», права которых не ограничены, заканчивая «Заблокированными», которым HIPS не даст прав даже на запуск.

Последние результаты тестирования наиболее известных антивирусных продуктов представлены на Рис.1.

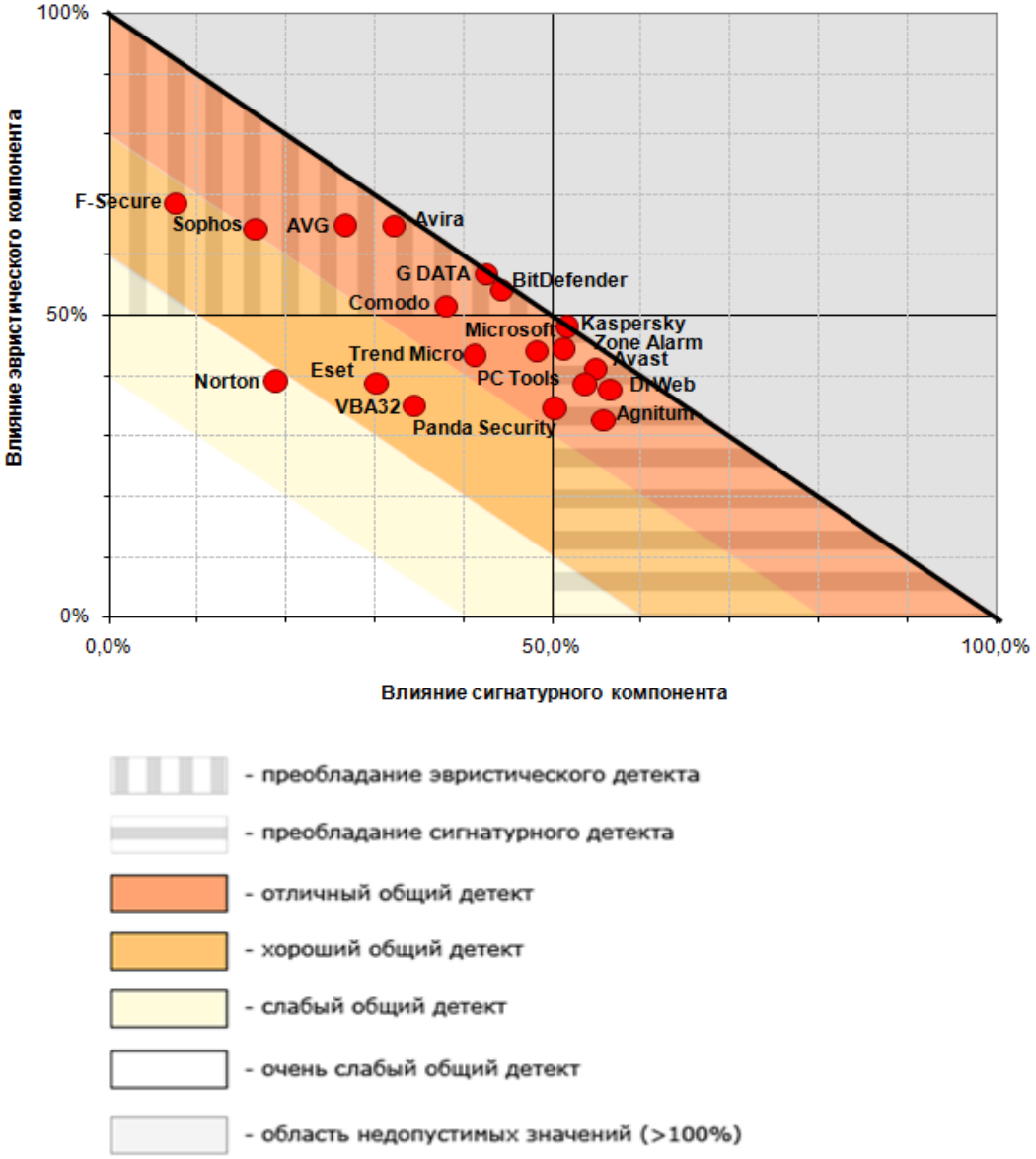


Рис.1. Влияние различных компонент антивирусной защиты на общий уровень обнаружения вредоносных программ

На основе проведенного анализа можно сформулировать требования для эффективной реализации антивирусной защиты в современных корпоративных сетях.

Основные требования:

1. Проверка в реальном режиме времени всей файловой системы (Microsoft Windows, Novell Netware, Unix/Linux)
2. Проверка любых объектов файловой системы по требованию
3. Обнаружение вирусов в составных файлах — архивах, файлах почтовых форматов и т. п.
4. Наличие эффективной и сбалансированной сигнатурной и проактивной защиты
5. Проверка сменных и сетевых дисков
6. Возможность выбора различных действий при обнаружении вируса, штатно

Требования к управлению:

1. Поддержка единой системы удаленного и централизованного управления
2. Возможность планирования запуска задач и выполнения действий
3. Возможность гибко исключать из проверки файлы, области, процессы

Требования к обновлению:

1. Быстрая реакция производителя на появление новых вредоносных программ — высокая частота выпуска обновлений
2. Поддержка различных источников обновления — HTTP- или FTP-ресурс, локальная или сетевая папка, желательно — централизованная система обновления
3. Возможность выполнения обновления вручную по запросу или автоматически по расписанию
4. Возможность выполнить откат обновлений антивирусных баз

Требования к диагностике:

1. Ведение журналов работы
2. Уведомление администратора обо всех важных событиях
3. Отображение статистики и счетчиков производительности в реальном режиме времени

Требования к производительности: Поддержка многопроцессорности.

Таким образом, централизованная антивирусная защита является одной из основных составляющих при построении защищенной сети. Такая защита предполагает комплексные решения и способна обеспечить необходимую безопасность, так как включает в себя множество функций, среди которых – применение технологий сигнатурного и вероятностного анализа, кроссплатформенность и наличие единой системы управления.

Анализ современных технологий защиты от спама

Н.В. Студеникин, М.В. Кузнецов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Проблема спама известна уже относительно давно, известны также потери, которые он нам приносит: лишний сетевой трафик, нагрузки на сервера, потери времени и концентрации сотрудников при разборе спама, а также потеря нормальной почты «в груде спама».

На сегодняшний день существуют следующие технологии борьбы со спамом:

- **«Черные списки» (DNSBL (DNS-based Blackhole Lists)) и «белые списки»**

Антиспам-технология, блокирующая почту, идущую с IP-серверов, перечисленных в списке. Для этого создана база RBL (Realtime Blackhole List), которая пополняется вручную IP-адресами серверов, на протяжении долгого времени рассылающих спам. В противоположность «черным», есть «белые списки», принцип работы которых в наличии определенного списка доверенных источников сообщений.

Недостатки:

- Не эффективная защита, которую они предоставляют.
- Белые списки также не надежны, ведь подделать адрес отправителя очень просто, так как используемый протокол позволяет любому человеку выдать себя за другого.

- **Методы контроля массовости**

Технология предполагает выявление в потоке почты массовых сообщений, которые абсолютно идентичны или различаются незначительно. Для построения работоспособного «массового» анализатора требуются огромные потоки почты, поэтому эту технологию предлагают крупные производители, обладающие значительными объемами почты, которую они могут подвергнуть анализу.

Недостатки:

- Можно ошибочно классифицировать как «спам» массовые рассылки вроде: Ozon.ru, Subscribe.ru;
- Спамеры используют специальное ПО для генерации разного контента для каждого письма

- **Контентная фильтрация**

Сюда относится фильтрация, как содержания письма, так и его заголовка, а также разновидность контентной фильтрации: по Байесу (самообучаемая, ее не нужно настраивать, необходимо только предварительное обучение по основным тематикам писем, типичных для данного конкретного пользователя). Спамерское сообщение проверяется на наличие специфических для спама слов, фрагментов текста, картинок и других характерных спамерских черт. Контентная фильтрация начиналась с анализа темы сообщения и тех его частей, которые содержали текст, но сейчас спам-фильтры проверяют все части, включая графические вложения.

Недостатки:

- Созданием контентых фильтров занимаются специально обученные люди, или даже антиспам-лаборатории, что очень затратно и отражается на стоимости продукта.
- Спамерские технологии развиваются в основном в сторону маскировки идентичности сообщений (зашумливание изображений и их разбиение на несколько частей, интеллектуальные генераторы, формирующие различающиеся тексты с одним содержанием и др.)

- Фильтрация по Байесу имеет неплохие результаты для индивидуального потока почты, для большой компании результат будет намного хуже, и может дать большое число ложных срабатываний.

Отдельно можно выделить еще 2 способа борьбы со спамом: защита на сетевом уровне с помощью репутационных механизмов и метод, использующий коэффициент риска.

- **Репутационный механизм**

Защита на сетевом уровне до момента приема тела сообщения и соответственно до его анализа. Отсюда вытекает необходимость введение так называемых репутационных механизмов. (Если у отпечатка плохая репутация, то письмо отвергается. Репутация письма формируется по количеству полученных писем с одинаковыми отпечатками).

Недостатки:

- Есть вероятность ошибки, в результате которой получателю не будет доставлено сообщение.

- **Метод, вычисляющий коэффициент риска**

Конкретно для компании или пользователя, можно ввести возможность указания всех возможных тематик сообщений, соответственно использовать коэффициент риска составляющийся из репутации письма, и информации полученной из его анализа и сравнения с уже имеющимися в базах сигнатурами.

Недостатки:

- Также как и для контентной фильтрации по Байесу, этот способ больше всего полезен в личном потоке почты, для крупной компании результат будет хуже, так как возможности точно определить коэффициент после анализа содержимого по тематике не будет.

Ниже представлены последние результаты тестирования наиболее известных антиспам-продуктов (рисунок 1).

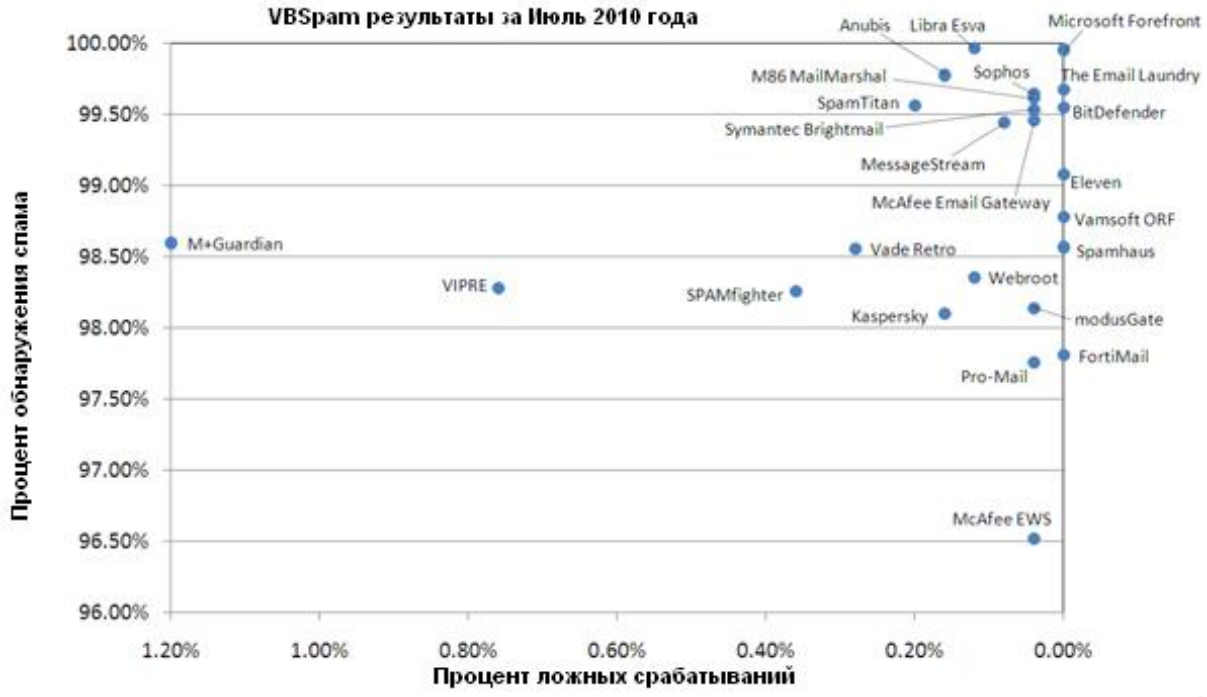


Рис. 1 Результаты тестов антиспам-продуктов от лаборатории Virus Bulletin

Всех разработчиков антиспам-продуктов можно разделить на 2 группы. Наиболее многочисленную среди них образуют антивирусные компании, для которых

спам-фильтр является дополнительным средством защиты от вредоносных программ. Такие производители, как правило, уже поддерживают базу данных подобных программ и сообщают сведения, содержащиеся в ней, на клиентские рабочие места. Вторую группу образуют компании, предлагающие специализированные спам-фильтры и получающие прибыль именно от продажи баз, встраивая фильтры в другие средства защиты.

Согласно тестам PC Magazine за 2010 год лучший антиспам у McAfee и ZoneAlarm (рисунок 2).

Продукт	Общая оценка	Фаервол	Антивирус	Антишпион	Производительность	Антиспам
avast! Internet Security 5.0	3.5	4.5	3.5	3.5	3.0	**4.0**
AVG Internet Security 9.0	3.5	3.5	4.0	4.0	2.5	**4.0**
BitDefender Internet Security 2010	4.0	4.0	4.0	4.0	2.5	**4.0**
CA Internet Security Suite Plus 2010	2.5	2.0	3.0	3.0	2.0	**1.5**
eScan Inernet Security Suite 10	2.0	1.5	3.0	2.5	3.5	**1.0**
FortiClient Endpoint Security Standart Edition 4.1	1.5	3.0	1.0	1.0	4.0	**4.0**
F-Secure Internet Security 2010	3.0	4.0	3.5	3.0	3.0	**3.0**
K7 TotalSecurity Version 10.0	2.5	2.0	3.0	3.5	4.0	**3.0**
Kaspersky Internet Security 2010	3.5	4.5	4.0	3.5	2.5	**3.0**
McAfee Total Protection 2010	4.0	3.5	4.0	4.0	3.5	**4.5**
Norton 360 v 4.0	4.5	5.0	4.5	4.5	3.5	**4.0**
Norton Internet Security 2010	4.5	5.0	4.5	4.5	4.0	**4.0**
Panda Internet Security 2010	3.0	3.5	3.5	3.0	4.0	**4.0**
PC Tools Internet Security 2010	3.5	3.0	3.5	4.5	3.0	**3.5**
Trend Micro Internet Security Pro (version 3)	3.5	3.0	3.5	3.5	2.5	**4.0**
ZoneAlarm Extreme Security 2010	4.0	4.5	4.0	3.5	2.5	**4.5**

Рис. 2. Результаты тестирования антивирусных и антиспам-решений от PC Magazine

Вывод

Таким образом, при оценке качественных показателей необходимо сформировать общие требования, предъявляемые к системе защиты информации от спама:

- поддержка всех существующих почтовых протоколов;
- обеспечение возможности пополнения базы сигнатур спама. Это подразумевает запись спам-сообщений и передача их в Аналитический центр;
- поддержка блокировки трафика на основе списка запрещённых IP-адресов. (100% отсекает почту из подозрительного источника);
- поддержка блокировки трафика на основе сигнатурного анализа. (Гибкость, возможность быстрой «тонкой» настройки. Системы, работающие на такой технологии, легко подстраиваются под новые виды спама и редко ошибаются с разграничением спама и нормальной почты);
- обеспечение возможности применения репутационных механизмов (экономия сетевого трафика и ресурсов почтового сервера, оперативность реагирования).

Применение локального корректирующего фильтра для устранения «блочности» на изображениях, сжатых методом JPEG

Е.И. Тимбай
Институт систем обработки изображений РАН, Самара

Введение

Наиболее заметным артефактом, характерным для изображений, сжатых методом JPEG с высоким значением коэффициента компрессии, является «блочность» - изображение рассыпается на блоки размером 8×8 пикселей (рис. 1). Этот эффект особенно заметен на областях изображения с плавными изменениями яркости. Причина возникновения такого типа искажений заключается в следующем. В процессе сжатия изображение разбивается на блоки 8×8 пикселей, к каждому блоку применяется дискретное косинусное преобразование (DCT), выполняется квантование и кодирование коэффициентов DCT. Артефакты «блочности» на восстановленном изображении появляются из-за независимой обработки блоков на этапе компрессии и особенно заметны при высоких степенях сжатия из-за грубого квантования высокочастотных коэффициентов DCT



Рис.1 – Изображение «Lena», восстановленное после сжатия JPEG с $K_c = 24.38$ и $PSNR = 32.9478$

На сегодняшний день существует много алгоритмов для устранения «блочности». Их можно разделить на две категории.

Часть методов работает внутри декодера [1-4], то есть обработка заключается в модификации метода декодирования сжатых изображений. В этом случае при выполнении деблокинга известна полная информация о параметрах квантования изображения на этапе компрессии.

Другая группа методов не учитывает структуру кодека и применяется как самостоятельная постобработка восстановленных после сжатия изображений. Здесь можно выделить методы, использующие в своей работе вейвлет-преобразования [5, 6], а также методы проекций на ограниченные множества [7]. Поскольку применение независимой постобработки не требует вмешательства в существующие стандарты кодеков, то реализация и разработка таких методов носит более практический характер.

В качестве третьего подхода к устранению «блочности» можно предложить использование локального корректирующего фильтра, описанного в работах [8, 9]. Исходными данными для построения фильтра являются два изображения – исходное и восстановленное после компрессии методом JPEG. По восстановленному изображению вычисляются некоторые признаки, на основе которых строится многомерная функция восстановления, параметры, описывающие которую, должны добавлены к архиву. Эта функция применяется для постобработки изображения на этапе декомпрессии для повышения качества его восстановления. Такой подход также не учитывает структуру кодека, поскольку построение и применение фильтра заключатся лишь в «достраивании» стандартной процедуры компрессии и декомпрессии изображения, однако требует передачи дополнительного объема данных, пусть и незначительного по сравнению с размером самого архива.

Локальный корректирующий фильтр

Построение и применение корректирующего фильтра заключатся в «достраивании» стандартной процедуры компрессии и декомпрессии изображения.

Модификация этапа компрессии заключается в следующем. Сразу же после сжатия исходного изображения X методом JPEG осуществляется восстановление изображения Y и вычисляется поле ошибок восстановления:

$E = X - Y$.

Далее выполняется «обучение» фильтра, которое заключается в построении функции φ, аппроксимирующей поле ошибок восстановления E, зависящей от некоторых выбранных признаков, вычисленных в «скользящем окне» обработки на основании пикселей изображения Y:

$\hat{E} = \varphi(\nu_0, ..., \nu_{K-1}, \alpha_0, ..., \alpha_K)$,

где $\hat{E}$ - аппроксимированная ошибка восстановления, $\{\nu_k(n_1, n_2)\}_0^{K-1}$, $n_1 = \overline{0, N_1 - 1}$, $n_2 = \overline{0, N_2 - 1}$ - локальные признаки, вычисляемые по изображению Y; N_1, N_2 - размеры изображения; $\alpha_0, ..., \alpha_K$ - коэффициенты аппроксимации (параметры фильтра, полученные при его обучении) определяются из условия минимума среднеквадратичной ошибки аппроксимации; φ - функция аппроксимации.

Результаты обучения, т.е. параметры фильтра, должны быть сохранены в архив вместе с компрессированными данными исходного изображения.

На стадии восстановления изображения сначала происходит разделение компрессированных данных на основную и дополнительную части. Из основной части восстанавливается изображение Y, то есть выполняется стандартная процедура декомпрессии. Восстановленные из второй части параметры фильтра используются для приближенного описания поля ошибок $\hat{E}$, компенсации ошибки восстановления и получения окончательно восстановленного изображения $\hat{X}$:

$\hat{X} = Y + \hat{E}$.

Для построения конкретной реализации фильтра были выбраны:

1) система из $K = 7$ признаков, вычисление которых основано на линейных сглаживающих фильтрах. Нулевой признак – значения пикселей восстановленного изображения Y. Признаки с номерами 1-6 – отсчеты восстановленного изображения Y, сглаженного прямоугольными окнами 3×3, 1×3, 3×1, 7×7, 5×3, 3×5, из которых вычтено восстановленное изображение Y.

2) кусочно-линейная функция аппроксимации. В процессе построения функции признаковое пространство, представляющее собой K-мерный гиперкуб, последовательно разбивается по осям (признакам) и порождает в памяти ЭВМ древовидную структуру. В каждой из областей, полученных в результате разбиения, выполняется линейная аппроксимация. Области с малой ошибкой аппроксимации принимаются за терминальные вершины дерева. Те области, в которых ошибка велика, подвергаются дальнейшему разбиению. Процесс завершается при выполнении заданного условия.

После завершения построения кусочно-линейной функции аппроксимации сформировано дерево, полностью описывающее этот процесс. В каждой терминальной вершине дерева хранятся коэффициенты $\alpha_0, ..., \alpha_K$. Чем меньше объем данных, необходимый для хранения одного коэффициента аппроксимации, тем меньше объем дополнительной информации, сохраняемой в архиве, и тем эффективней работа фильтра-корректора. Задача уменьшения объема дополнительной информации, описывающей построение фильтра, решается с помощью равномерного квантования

коэффициентов аппроксимации. Количество бит l, отводимое для хранения одного коэффициента, вычисляется по формуле:

$$l = \left\lceil \frac{1}{2} \cdot \log_2 \left(\frac{(\alpha_{\max} - \alpha_{\min})^2}{12\mu\varepsilon^2} \sum_{k=0}^{K} \sum_{i=0}^{N-1} \sum_{(n_1,n_2)\in V^{(i)}} \nu_k^2(n_1,n_2) \right) \right\rceil + 1,$$

где ε^2 - суммарная квадратичная погрешность аппроксимации ошибки восстановления, вычисленная для всего изображения; $\mu << 1$ - параметр, задающий допустимое превышение дисперсии ошибки, вызванной квантованием коэффициентов аппроксимации, над величиной ε^2; $\alpha_{\min}, \alpha_{\max}$ - минимальный и максимальный коэффициенты аппроксимации; $V^{(i)}$ - области признакового пространства, соответствующие N терминальным вершинам дерева:

Таким образом, объем дополнительной информации I_E, описывающей фильтр-корректор с кусочно-линейной функцией аппроксимации, составит величину

$I_E = N \cdot K \cdot l + M \cdot b$,

где M - количество нетерминальных вершин дерева, b - количество бит, которое отводится для хранения номера признака, по которому было выполнено разбиение данной вершины.

Параметры корректирующего фильтра, то есть $\alpha_0, ..., \alpha_K$, в каждой подобласти признакового пространства определяются из условия минимума среднеквадратичной ошибки аппроксимации, поэтому применение фильтра при декомпрессии позволяет повысить качество восстановленного изображения по показателю *PSNR* [10].

Исследование метода

Наличие большого количества различных методов деблокинга и отсутствие открытых кодов для перечисленных во введении алгоритмов, значительно затрудняют проведение полного сравнительного анализа эффективности устранения «блочности» этими методами и корректирующим фильтром. Поэтому для исследований был выбран свободно распространяемый программный продукт Jpeg Enhancer 1.8 [11], являющийся простой, но эффективной программой, предназначенной для устранения «блочности» на изображениях, искаженных при сжатии методом JPEG. В отличие от многих алгоритмов устранения артефактов, методы, применяемые в Jpeg Enhancer 1.8, не просто размывают изображения, но восстанавливают их, улучшая качество по показателю *PSNR*.

Для проведения экспериментальных исследований в качестве тестовых изображений были использованы наборы полутоновых и полноцветных изображений из коллекции Waterloo [12] и базы изображений USC-SIPI [13]. Метод JPEG был взят в реализации IrfanView 4.25 [14].

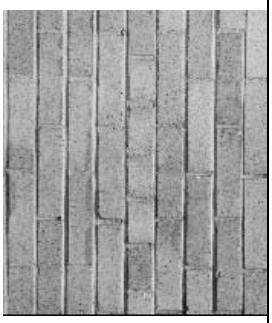
На рисунке 2 представлены примеры устранения «блочности» на изображении «Lena» с помощью программы Jpeg Enhancer 1.8 (рис. 2в) и корректирующего фильтра с кусочно-линейной аппроксимирующей функцией, построенного для системы признаков, основанной на линейных сглаживающих фильтрах (рис. 2г).

Согласно результатам проведенных экспериментов, обработка фильтром-корректором позволяет получить изображение визуально более близкое к оригиналу за счет меньшего размытия и сохранения большего количества деталей.

На рисунке 3 приведены примеры графиков зависимостей $PSNR(K_c)$, построенных для изображений «Lena», «Washsat» и «Brick wall 1» (табл. 1), восстановленных после базовой компрессии методом JPEG и обработанных

Jpeg Enhancer 1.8 и фильтром-корректором. Из графиков видно, что корректирующий фильтр обеспечивает лучшее качество восстановления изображений по показателю *PSNR* по сравнению с методами устранения «блочности», применяемыми в Jpeg Enhancer 1.8 (рис. 3).

Таблица 1. Примеры тестовых полутоновых изображений

Lena	Washsat	Brick wall
		

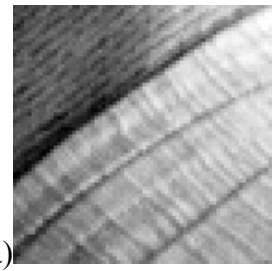
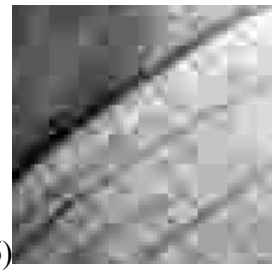
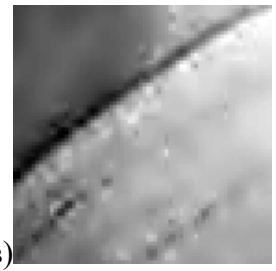
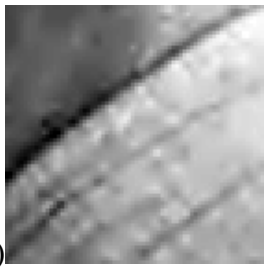
а)  б)  в)  г) 

Рис. 2 – Фрагмент исходного изображения «Lena» (а), фрагмент изображения «Lena», восстановленного после сжатия JPEG с $K_c = 24.38$ и $PSNR = 32.9478$ (б), результат устранения «блочности» с помощью Jpeg Enhancer 1.8 (в), результат работы корректирующего фильтра (г)

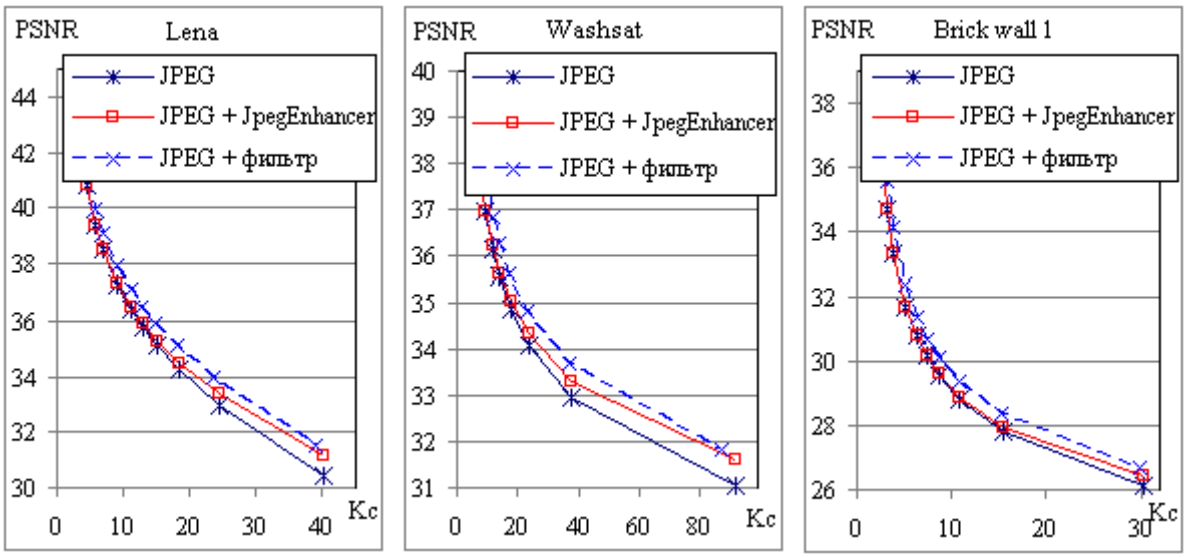


Рис. 2 - Зависимости $PSNR(K_c)$, построенные для изображений «Lena», «Washsat», «Brick wall 1», восстановленных после компрессии JPEG, обработанных Jpeg Enhancer 1.8 и корректирующим фильтром

Результаты проведенных экспериментальных исследований позволяют сделать следующий вывод: предложенный корректирующий фильтр может быть эффективно использован для улучшения качества изображений, сжатых методом JPEG. Кроме того,

предложенный фильтр, позволяет получить лучшее качество восстановления изображений, чем методы устранения «блочности», применяемые в Jpeg Enhancer 1.8.

Литература

1. Farrelle P., Jain A. Recursive blocking-coding-A new approach to transform coding // IEEE Trans Commun, 1986, pp. 161-179.
2. Lynch W.E., Reibman A.R., Liu B. Edge compensated transform coding // Proc IEEE ICIP Image Process (1994), 105-109.
3. Malvar H.S., Staelin D.H. The LOT: Transform coding without blocking effects // IEEE Trans Acoust Speech Signal Process, 1989, pp. 553-559.
4. Pearson D., Whybray M. Transform coding of images using interleaved blocks // Proc IEEE Radar Signal Process, 1984, pp. 466-472.
5. Xiong Z., Orchard M. T., Zhang Y. Q. A Deblocking Algorithm for JPEG Compressed Images Us-ing Overcomplete Wavelet Representation //. IEEE Trans-actions on Circuits and Systems for Video Technology, No. 7, pp. 433-437.
6. R.A.Gopinath, M.Lang, H.Guo, J.E.Odegam. Wavelet-Based Post-processing of Low Bitrate Transform Coded Images // In Proc. of Intl. Conf. in Image Processing, Austin, TX, USA, November 13-16, 1994.
7. Yang Y., Galatsanos N., Katsaggelos A. Projection-Based Spatially-Adaptive Reconstraction of Block-Transform Compressed Image // IEEE Trans. on Image Processing, Vol. 4, No. 7, pp. 896-908, July 1995.
8. Сергеев В. В., Тимбай Е. И., Чернов А. В. Адаптивные методы повышения качества изображения после компрессии // Сборник трудов всероссийского семинара по моделированию, дифракционной оптике и обработке изображений, Самара, 3-7 июля 2006, с. 38-41.
9. Sergeev V.V., Timbay E.I. The investigation of image local feature sets in a problem of image enhancement after compression // In Proc. of 9-th Intl. Conf. on Pattern Recognition and Image Analysis: New Information Technologies, Russian Federation, Nizhni Novgorod, September 15—19, 2008, Vol.2, pp. 155-158.
10. Сэломон, Д. Сжатие данных, изображений, звука. – М.: Техносфера, 2004. -368 с.
11. Jpeg Enhancer – Overview. // http://www.vicman.net/jpegenhancer/index.htm
12. The Waterloo Fractal Coding and Analysis Group. Image Repository. // http://links.uwaterloo.ca/Repository.html
13. The USC-SIPI Image Database // http://sipi.usc.edu/database/
14. IrfanWiew //http://www.irfanview.com

Алгоритм обработки изображений при фотограмметрических измерениях

П.П. Чураков
Пензенский государственный университет, Пенза

Рассмотрены специфика и варианты обработки изображений фотограмметрических измерений при съемки камерой мобильного телефона.

Наиболее часто при проведении фотограмметрических измерений используют стереоскопические фотоаппараты. При этом делается два снимка одинаковыми объективами выставленными в определенных направлениях.

Для вычисления результатов измерения в трехмерных координатах полагают следующие допущения [1];

- точки размещения фотокамеры по горизонтальной оси симметричны (x_1= 0,5b; y = 0; z =0 и x_2 = -0,5b; y=0; z=0; b - расстояние между центрами осей фотоаппаратов при съемке);
- оптическая ось фотоаппарата параллельна вертикальной оси z измерительной системы координат;
- оси x и y систем координат снимка и измерительной системы совпадают;
- отсутствует искажения геометрии изображения (дисторсия);

Координаты точки объекта x, y, z определяются из подобия треугольников [1] путем нахождения координат отображений *u* и *v* на первом и втором снимках.

$$u_1 = -f\frac{x - b/2}{z}, \qquad v_1 = -f\frac{y}{z}.$$

1)

$$u_2 = -f\frac{x + b/2}{z}, \quad v_2 = -f\frac{y}{z}.$$

Решая систему уравнений (1) находим координаты точки объекта

$$z = \frac{b \cdot f}{u_2 - u_1}$$

$$x = -\frac{1}{2}(u_1 + u_2)\frac{z}{f} = \frac{1}{2}(u_1 + u_2)\frac{b}{u_2 - u_1} \qquad 2)$$

$$y = -v \cdot \frac{z}{f} = v\frac{b}{u_2 - u_1}$$

Разность $u_2 - u_1$ называется параллаксом точки; *f* фокусное расстояние от центра объектива до плоскости изображения

При использовании обычного фотоаппарата (камера мобильного телефона) необходимо выполнить несколько снимков с разных положений и учесть координаты (X_C, Y_C, Z_C) и ориентацию (углы Эйлера) Ω, K, Φ фотоаппарата в пространстве. Трехмерные координаты и система уравнений (1) трансформируется в

систему координат (СК), связанную с положением фотоаппарата, номером снимка k и номером точки i

$$u_{ik} = -f\frac{(\mathrm{X}^i - \mathrm{X}_C^k)m_{xx}^k + (\mathrm{Y}^i - \mathrm{Y}_C^k)m_{xy}^k + (\mathrm{Z}^i - \mathrm{Z}_C^k)m_{xz}^k}{(\mathrm{X}^i - \mathrm{X}_C^k)m_{zx}^k + (\mathrm{Y}^i - \mathrm{Y}_C^k)m_{zy}^k + (\mathrm{Z}^i - \mathrm{Z}_C^k)m_{zz}^k}$$

$$v_{ik} = -f\frac{(\mathrm{X}^i - \mathrm{X}_C^k)m_{yx}^k + (\mathrm{Y}^i - \mathrm{Y}_C^k)m_{yy}^k + (\mathrm{Z}^i - \mathrm{Z}_C^k)m_{yz}^k}{(\mathrm{X}^i - \mathrm{X}_C^k)m_{zx}^k + (\mathrm{Y}^i - \mathrm{Y}_C^k)m_{zy}^k + (\mathrm{Z}^i - \mathrm{Z}_C^k)m_{zz}^k} \qquad 3)$$

Здесь коэффициенты $m_{xx}^k, m_{xy}^k, m_{xz}^k, m_{yx}^k, m_{yy}^k, m_{yz}^k, m_{zx}^k, m_{zy}^k, m_{zz}^k$ – компоненты матрицы поворота M [2], определяемые углами $\Omega^k, \mathrm{K}^k, \Phi^k$:

$$M = \begin{vmatrix} \cos\phi\cdot\cos\kappa & \cos\omega\cdot\sin\kappa + \sin\omega\cdot\sin\phi\cdot\cos\kappa & \sin\kappa\cdot\sin\omega - \cos\omega\cdot\sin\phi\cdot\cos\kappa \\ -\cos\phi\cdot\sin\kappa & \cos\kappa\cdot\cos\omega - \sin\omega\cdot\sin\phi\cdot\sin\kappa & \cos\kappa\cdot\sin\omega + \cos\omega\cdot\sin\phi\cdot\sin\kappa \\ \sin\phi & -\sin\omega\cdot\cos\phi & \cos\phi\cdot\cos\omega \end{vmatrix}$$

Для определения координат различных точек на различных снимках необходимо осуществить преобразование СК. При преобразовании необходимо учитывать вид преобразования и измеренных параметров. В зависимости от этого различают следующие варианты.

1. Преобразование одной СК в другую с использованием сочетаний различных шести параметров:

– - начало координат (X_C, Y_C, Z_C) и углы Эйлера Ω, K, Φ Ориентации фотоаппарата в пространстве;

– вектор начала координат X и базисные вектора E_x E_z по осям x и z.

– углы ориентации Ω, K, Φ в пространстве, определяющие матрицу поворота M [2].

2. Преобразование координат из внешней СК осуществляется [2] согласно системе уравнений:

$$x' = (x - \mathrm{X}_C)m_{xx} + (y - \mathrm{Y}_C)m_{xy} + (z - \mathrm{Z}_C)m_{xz}$$
$$y' = (x - \mathrm{X}_C)m_{yx} + (y - \mathrm{Y}_C)m_{yy} + (z - \mathrm{Z}_C)m_{yz} \qquad 4)$$
$$z' = (x - \mathrm{X}_C)m_{zx} + (y - \mathrm{Y}_C)m_{zy} + (z - \mathrm{Z}_C)m_{zz}$$

Величины $m_{xx}, m_{xy}, ..., m_{zz}$ – компоненты матрицы поворота M [2]. В векторном виде: $x' = \mathrm{M}(x - \mathrm{X}_C)$. Обратное преобразование: $x = X_C + \mathrm{M}^T x'$.

3. Если две СК (СК1 и СК2) определены во внешней СК:

(СК1: $\mathrm{X}_{C1}, \mathrm{M}_1$ и СК2: $\mathrm{X}_{C2}, \mathrm{M}_2$). То преобразование из СК1 в СК2 будет иметь вид $x'' = \mathrm{M}_2(X_{C1} + \mathrm{M}_1^T x' - X_{C2})$. Взаимосвязь СК2 и СК1 определяется: $\mathrm{M}_{21} = \mathrm{M}_2\mathrm{M}_1^T$, $X_{C21} = M_2(X_{C2} - X_{C1})$. По компонентам матрицы

M_{21} определяются угловые параметры ориентирования $\Omega_{21}, K_{21}, \Phi_{21}$ при переходе из СК2 в СК1.

Измерения размеров объектов сводятся к определению координат изображений одних и тех же точек объекта на разных снимках и решению системы уравнений (3). Определяются трехмерные координаты точек объекта и положения фотоаппарата в момент съемки. Важным моментом при обработки информации является наличие необходимого числа опорных точек с известными трехмерными координатами.

Рассмотрим возможные варианты решения системы уравнений (3) с известным f:

1. Изображения одной и той же точки предмета, измеренные на двух снимках с известными параметрами ориентирования $X_C, Y_C, Z_C, \Omega, K, \Phi$ дают 4 уравнения, по которым можно определить трехмерные координаты точки (X, Y, Z).

2. Три точки с известными трехмерными координатами, измеренные на одном снимке, дают 6 уравнений. Можно определить параметры ориентирования снимка $X_C, Y_C, Z_C, \Omega, K, \Phi$ (*Внешнее ориентирование снимка*).

3. Для пары снимков можно выбрать общую внешнюю систему координат, в которой уравнения имеют простой вид (1) и (2) (*Базисная система координат*); сами изображения приводятся в эту систему координат проективным преобразованием плоскости. Преобразованные изображения образуют стереопару, по ней удобно выполнять измерения, можно наблюдать стереоизображение. *Базисную систему координат* можно определить по известным параметрам внешнего ориентирования снимков $X_C, Y_C, Z_C, \Omega, K, \Phi$.

4. Пять точек с неизвестными трехмерными координатами, измеренные на двух снимках, дают 5 (5*2*2-5*3=5) уравнений. Можно определить угловые положения одного снимка относительно другого, и следовательно, положения снимков относительно *базисной системы координат* этой пары снимков с точностью до расстояния между точками съемки. (*Взаимное ориентирование пары снимков*). Положение базисной системы координат относительно СК объекта остаётся неизвестным. По связанной таким образом паре снимков можно выполнить трехмерные измерения с точностью до масштаба

При использовании бытовой фотокамеры требуется учесть её внутренние параметры:

– положение главной точки u_M, v_M, т.е. проекцию оптического центра объектива на плоскость изображения (она не обязательно совпадает с центром кадра);

– реальные фокусные расстояния по вертикали и горизонтали кадра (неквадратные пиксели) f_U, f_V;

– нелинейные геометрические искажения, которые можно представить в виде разложения поправок по степеням измеренных на снимке координат [3,6]:

$$u' = u + D_{U\_UUU} u^3 + D_{U\_UVV} uv^2 + ...$$
$$v' = v + D_{V\_VVV} v^3 + D_{V\_VUU} vu^2 + ... \qquad 5)$$

Где, D_{U_UUU}, D_{U_UVV} … коэффициенты дисторсии

В данном случае необходимо решение системы уравнений:

$$u_{ik} - u_M + D_{U\_UUU}u_{ik}^{\ 3} + D_{U\_UVV}u_{ik}v_{ik}^{\ 2} + ... =$$
$$= -f_U \frac{(\mathrm{X}^i - \mathrm{X}_C^k)m_{xx}^k + (\mathrm{Y}^i - \mathrm{Y}_C^k)m_{xy}^k + (\mathrm{Z}^i - \mathrm{Z}_C^k)m_{xz}^k}{(\mathrm{X}^i - \mathrm{X}_C^k)m_{zx}^k + (\mathrm{Y}^i - \mathrm{Y}_C^k)m_{zy}^k + (\mathrm{Z}^i - \mathrm{Z}_C^k)m_{zz}^k}$$
$$v_{ik} - v_M + D_{V\_VVV}v_{ik}^{\ 3} + D_{V\_VUU}v_{ik}u_{ik}^{\ 2} + ... = \qquad 6)$$
$$= -f_V \frac{(\mathrm{X}^i - \mathrm{X}_C^k)m_{yx}^k + (\mathrm{Y}^i - \mathrm{Y}_C^k)m_{yy}^k + (\mathrm{Z}^i - \mathrm{Z}_C^k)m_{yz}^k}{(\mathrm{X}^i - \mathrm{X}_C^k)m_{zx}^k + (\mathrm{Y}^i - \mathrm{Y}_C^k)m_{zy}^k + (\mathrm{Z}^i - \mathrm{Z}_C^k)m_{zz}^k}$$

Структура уравнений (6) может быть представлена в следующем виде (рис. 1.2).

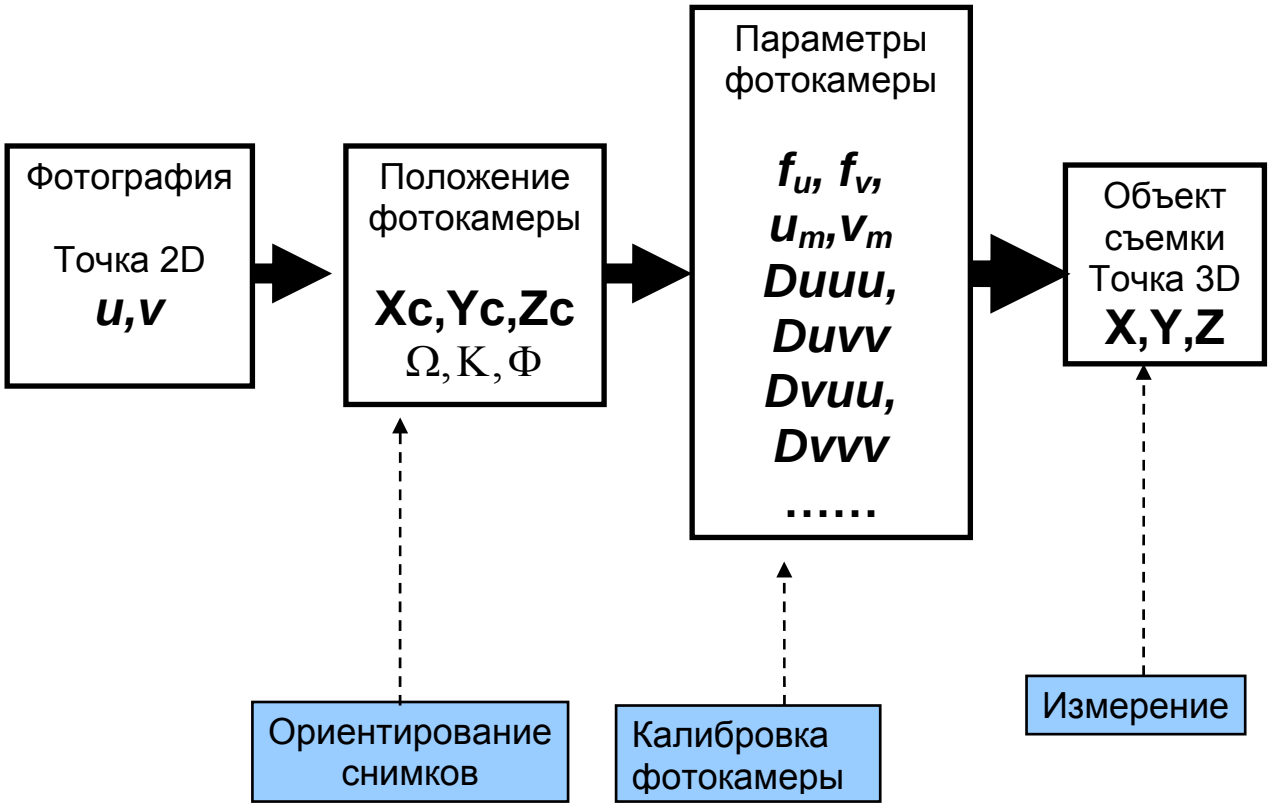


Рис. 1.2 – Структура уравнений.

Таким образом, предлагается решать систему уравнений (6) по частям:

1. Калибровка фотокамеры. Требуемые параметры камеры определяются заранее по снимкам калибровочных объектов с известной геометрией, предполагается, что параметры камеры сохраняются во время рабочих измерений. Возможно уточнение параметров по рабочим снимкам.

2. Ориентирование снимков. Выполняется связывание снимков по известным опорным точкам и возможно, по некоторым точкам измеряемого предмета. Используются хорошо опознаваемые и надежно измеряемые точки. Оценивается точность полученной стерео модели.

3. Измерение точек предмета. Для удобства измерений могут формироваться стереопары.

В ходе работы были проведены измерения концевых мер способом описанным выше. Концевые меры – это метрологические устройства, предназначенные для точного измерения геометрических размеров объектов, их метрологический параметр, в нашем случае толщина, нормирован и подвергается периодической аттестации. Для проведения испытаний были взяты четыре типа мер с толщинами 6, 7, 8 и 9 мм. Тестовые объекты были сфотографированы в приспособлении, их фотографии подвергнуты обработке с помощью специального программного обеспечения рисунок 1. Контрольные точки устанавливались как на переднем плане фотоснимков, так и на заднем .Результаты вычислений представлены в таблице 1.



Рисунок 1- Измерение концевых мер.

Таблица 1

	Мера «6»	Мера «7»	Мера «8»	Мера «9»
Размеры меры, мм	5,996	6,997	7,999	8,999
Результаты измерений.				
мм	6,005	7,080	7,958	9,092
Погрешность				
мм	0,009	0,083	0,041	0,093

Как видно из таблицы 1 погрешность измерения не превышает 100 мкм что дает возможность использования подобных устройств для измерения размеров трехмерных объектов в оперативных условиях (например при фиксации ДТП или летучем контроле деталей на производстве).

Литература

1. E.Mikhail, J.Bethel, and J.C.McGlone. Introduction to Modern Photogrammetry. Wiley&Sons Inc, 2001
2. Первунинских Д.В. Перспективы использования триангуляционного метода определения размеров трехмерных объектов Журнал Современные технологии безопасности 2007, стр. 35.
3. Назаров А.С. Фотограмметрия. Учебное пособие для студентов вузов,- М; ТетраСистемс, 2006, 360 с.: ил.

4. Форсайт Д. А., Понс Ж. Компьютерное зрение. Современный подход.: Пер. с англ. - М.: Издательство Вильямс, 2004.
5. Книжников Ю.Ф. Цифровая стереоскопическая модель местности, - М.: Научный мир 2004, 244с.
6. Русинов М.М. Инженерная фотограмметрия –М; Недра, 1966, 248с.

Анализ способов применения централизированного управления в системах комплексной информационной защиты распределённых сетей

М.В. Яшин
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В настоящее время вопросы, связанные с информационной безопасностью, выходят на новый уровень. Это связано с увеличением масштабов систем обработки информации с отдельного компьютера до огромных по количеству машин и распределенных территориально сетей. Возможностей обычных средств защиты (например, персональных антивирусов) не хватает для решения задач в подобных сетях. Поэтому силы всех крупнейших компаний в области разработки нацелены сейчас на работу над собственными системами комплексной информационной защиты распределенных сетей.

Основным отличием системы безопасности распределенной сети от системы защиты отдельного компьютера является разделение на клиентскую и серверную части. В связи с этим возникает проблема взаимодействия удаленных компонентов, синхронизации, обмена информацией между ними, построения иерархии управления. Эти и многие другие функции выполняет подсистема централизированного управления.

Для изучения подсистемы централизированного управления были выбраны некоторые продукты ведущих компаний в области информационной безопасности:

- Dr.Web Enterprise Suite.
- Kaspersky Administration Kit
- McAfee ePolicy Orchestrator 2.5
- Avira Security Management Center
- ESET Remote Administrator
- Symantec Control Compliance Suite

Ни одна программа не оказалась похожей на другую, все они имеют различную реализацию, построение и функционал. Это позволило провести разностороннее изучение и сравнение всех особенностей как систем безопасности в целом, так и подсистемы безопасности в частности. На основании проделанного исследования был сделан вывод, что для полноценного функционирования всех компонентов защиты необходимо наличие в структуре сети узла с установленным центром управления защитой (ЦУЗ). Данный сервер должен помимо управления всеми остальными частями системы защиты, заниматься установкой обновлений, хранить собранную информацию о функционировании сети, контролировать параллельную обработку задач в рамках сети, а также периодически проводить аудит безопасности.

Преимуществами использования такого сервера будут:

- возможность производить удаленно установку новых и обновление существующих модулей системы безопасности на компьютерах сети;
- возможность решить проблему установки обновлений (ведь именно это по статистике – основная причина взлома);
- создание политик безопасности для групп пользователей, что облегчит настройку отдельных компьютеров для специальных целей;
- доступность быстрого конфигурирования политик безопасности сети;
- контролирование сервером распараллеливания анализа трафика и файлов

позволит эффективно использовать свободные ресурсы других компьютеров сети;

- ведение логов о событиях в сети позволит провести быстрый аудит безопасности специально предусмотренными для этого штатными средствами, что приведёт к заметному сокращению времени поиска уязвимостей в сети;
- контроль выполнения задач на компьютерах, которые по каким-то причинам были недоступны в момент активации группового задания;
- быстрое устранение обнаруженных уязвимостей или последствия атак.

Таким образом, в данной работе произведен анализ элементов существующих систем комплексной информационной защиты распределенных сетей, рассмотрено их функционирование и взаимодействие. Было выявлено, что подсистема централизованного управления выполняет важнейшую роль интеграции отдельных модулей в целостную систему безопасности. Был определен наиболее оптимальный состав, функционал и представление централизованного управления защитой в собственной системе комплексной безопасности, а также создана модель собственной подсистемы централизированного управления.

Системы защиты от утечек данных (DLP-системы)

С.А. Бобров, И.С. Кузнецова
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Эффективность бизнеса во многих случаях зависит от сохранения конфиденциальности, целостности и доступности информации. В настоящее время одной из наиболее актуальных угроз в области информационной безопасности является утечка конфиденциальных данных от несанкционированных действий пользователей.

Это обусловлено тем, что большая часть традиционных средств защиты, таких как антивирусы, межсетевые экраны и системы аутентификации не способны обеспечить эффективную защиту от внутренних нарушителей. Целью такого рода нарушителей является передача информации за пределы компании с целью её последующего несанкционированного использования.

DLP (Data Loss/Leak Prevention) - технологии, предназначенные для обнаружения и предотвращения несанкционированного использования и передачи конфиденциальной информации.

Следует начать с того, что методы распознавания конфиденциальной информации можно разделить на два основных типа:

- **Контейнерный анализ**

Этот метод анализирует свойства файла или другого контейнера, в котором находится информация. Каждый контейнер содержит некую метку, которая однозначно определяет тип содержащегося внутри контейнера контента. Этот метод практически не требует вычислительных ресурсов для анализа перемещаемой информации, поскольку метка полностью описывает права пользователя на перемещение контента по любому маршруту.

Метки обычно подразделяют на атрибутные, форматные и внешние. Как следует из названия, первые размещаются в атрибутах файлов, вторые — в полях самого файла и третьи — ассоциируются с ним внешними программами.

Недостатком является необходимость разрабатывать процедуру расстановки меток на новые и входящие документы, а также систему противодействия переносу информации из помеченного контейнера в непомеченный посредством операций с буфером, файловых операций, копирования информации из временных файлов и т.д.

- **Контентный анализ**

Задача этого метода — извлечь значащий контент из контейнера или перехватить передачу по каналу связи и проанализировать информацию на наличие запрещенного содержимого. Основными технологиями в определении конфиденциальной информации в контейнерах являются контроль сигнатур, контроль на основе хеш-функций и лингвистические методы.

- **Сигнатуры**

Самый простой метод контроля — поиск в потоке данных некоторой последовательности символов. В подавляющем большинстве случаев сигнатурные системы настроены на поиск нескольких слов и частоту встречаемости терминов. К достоинствам этого метода можно отнести независимость от языка и простоту пополнения словаря запрещенных терминов.

- **Маски**

Расширением функционала поиска сигнатур является поиск их масок. Он представляет собой поиск такого содержания, которое невозможно точно указать в базе, но можно указать его элемент или структуру. К такой информации следует

отнести любые коды, характеризующие персону или предприятие: ИНН, номера счетов, документов и т.д.

- **Хеш-функции**

Суть этого метода заключается в наборе базы «отпечатков» с образцов конфиденциальных документов, то есть из документа извлекается значимое содержимое, которое приводится к некоторому нормальному виду и затем снимаются хеши всего содержимого и его частей. Перехваченный документ обрабатывается точно так же, и при совпадении отпечатков он считается конфиденциальным.

Важным плюсом такой технологии является ее нацеленность на работу с произвольными последовательностями символов. Из этого следует независимость от языка текста и возможность снятия отпечатков с нетекстовой информации.

- **Лингвистические методы**

Лингвистика как наука состоит из многих дисциплин — от морфологии до семантики. Поэтому лингвистические методы анализа тоже различаются между собой. Есть методы, использующие лишь стоп-слова, только вводящиеся на уровне корней, а сама система уже составляет полный словарь; есть базирующиеся на расставлении весов встречающихся в тексте терминов; есть и свои отпечатки, базирующиеся на статистике.

Достоинством метода является полная нечувствительность к количеству документов. База контентной фильтрации (набор ключевых словарных классов и правил) не меняется в размере от появления новых документов или процессов в компании. Из недостатков можно отметить зависимость от языка и высокий процент ошибок второго рода.

Структура системы и ее функции

- **Центральный сервер управления**

- объединение всех остальных компонентов решения в единую систему
- определение данных, содержащих конфиденциальную информацию
- создание, редактирование и распространение политик работы с конфиденциальными данными
- сбор, хранение и обработку инцидентов, создание и рассылку отчетов
- предоставление ролевого доступа к управлению системой сотрудникам службы информационной безопасности

- **Модули мониторинга и блокировки**

- анализ трафика выходящего за пределы защищаемой информационной системы
- пресечение передачи данных в случае обнаружении конфиденциальной информации

- **Агенты для рабочих станций и серверов**

- контроль перемещения конфиденциальных данных на сменные носители информации
- контроль помещения данных в буфер обмена
- контроль функции снятия снимка с экрана
- контроль отправки файлов на печать

Критерии оценки программных продуктов

- **Многоканальность**

Решение DLP не должно быть сосредоточено только на одном канале утечек. Это должно быть комплексное решение, охватывающее максимальное количество каналов.

- **Унифицированный менеджмент**

Система должна обладать унифицированными средствами управления всех компонентов, которые она в себя включает.

- **Активная защита**

Система должна не только фиксировать утечку конфиденциальной информации, но и давать возможность ее блокировать.

- **Классификация информации**

Классификация информации с учетом, как содержимого, так и контекста. Утечки конфиденциальной информации должны базироваться не только на содержимом пересылаемой информации, но и на контексте, в котором она происходит.

Исходя из всего вышесказанного, можно сделать вывод о том, что любой разработанный вариант DLP-системы позволяет структурировать и систематизировать данные, а так же контролировать процессы их передачи; снизить риски связанные с потерей, кражей и уничтожением важной информации; сохранять и архивировать все действия, связанные с перемещением данных внутри информационной системы. То есть является важной составляющей комплекса информационной безопасности любой организации.

Секция 5

КОМПЬЮТЕРНАЯ ОПТИКА И НАНОФОТОНИКА

Решение задач компьютерной оптики на графических вычислительных устройствах

Д.Л. Головашкин, Н.Л. Казанский
Институт систем обработки изображений РАН, Самара
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

На протяжении достаточно длительного времени магистральное направление развития вычислительной техники было связано с увеличением тактовой частоты центрального процессора. Так, с 1978 года тактовая частота процессоров архитектуры x86 возросла в 700 раз (с 4,77 МГц у Intel 8086 до 3,33 ГГц у Intel Core i7) [1]. Однако последние годы отмеченный рост существенно замедлился, что связывают [2] с “кремниевым тупиком” – физическими ограничениями, присущими технологии изготовления процессоров. Выход из положения ведущие производители вычислительной техники нашли в размещении некоторого количества ядер на одном кристалле и нескольких процессоров на одной плате (например, 8 для персонального компьютера).

Основным недостатком этого экстенсивного пути, по мнению авторов настоящей работы, является возросшая сложность архитектуры таких систем. В одном вычислительном устройстве предусмотрены обычные скалярные вычисления, суперскалярные (начиная с Intel Pentium), векторные (технология SSE, начиная с Intel Pentium III) и параллельные (в силу многоядерности и многопроцессорности). Универсализация обычно отрицательно сказывается на производительности. Для производства векторных вычислений (SIMD по классификации Флинна [3]), которым посвящен данный доклад, предпочтительнее оказываются специализированные графические процессоры (Graphics Processing Unit, GPU), отличающиеся лучшим быстродействием и распространенностью[1,2,4].

Следуя принятой в [5] классификации будем различать пять поколений таковых. Процессоры первого поколения (например, 3Dfx Voodoo I, 1996 г.) аппаратно переводили трехмерную сцену в двумерное ее изображение. Второго поколения (NVidia GeForce 256, 1999г.) – дополнительно рассчитывали освещенность проецируемых объектов. Возможность программирования графических процессоров появилась в третьем поколении (NVidia GeForce 3, 2001 г.). В четвертом (ATI Radeon X850 XT Platinum, 2004 г.) программы для GPU (шейдеры) уже могли содержать операции ветвления и циклов, записываясь на языках высокого уровня (GLSL, Cg). С этих пор начинаются вычисления общего назначения на GPU (GPGPU - General-Purpose computation on GPU). Графические процессоры пятого поколения (например NVidia GeForce GTX 295, 2009 г.) отличаются универсальной шейдерной архитектурой:

конкретные вычислительные блоки более не специализированы на исполнение шейдеров определенного типа (пиксельных, вершинных либо геометрических).

Первая публикации по применению GPU в оптических вычислениях относятся ко времени появления процессоров четвертого поколения [6]. В то время доступ к ресурсам GPU осуществлялся через специализированный интерфейс (API - Application Programming Interface), ориентированный на работу с графикой (OpenGL, Direct3D) [7], что накладывало определенные неудобства. Программа писалась на двух языках: традиционном (для CPU) и шейдерном (для GPU). Входные данные представлялись в виде текстур, подлежащих визуализации. Взаимодействие между параллельно обрабатываемыми пикселами не предусматривалось.

Попыткой решить указанные проблемы являлся язык BrookGPU (2004 г.) [8], представляющий собой расширение языка программирования C, позволяющее абстрагироваться от графического интерфейса при написании программ (однако фактически вычисления производились через него). Настоящим прорывом в области GPGPU оказалось появление в 2007 году технологии CUDA [9] (Compute Unified Device Architecture) для графических процессоров фирмы NVidia пятого поколения, представленной этой же фирмой. Теперь связь с процессором осуществляется через драйвер видеокарты, минуя графический API. При этом в распоряжении исследователя оказывается C-подобный язык высокого уровня с развитыми библиотеками. Аналогичный проект AMD FireStream [10] фирмы AMD, развиваемый с 2006 года, не добился аналогичной популярности в силу большей закрытости на начальной стадии разработки. В другом современном проекте – OpenCL (Open Computing Language) [11] – создается аппаратно-независимый стандарт, включающий программы для GPGPU и пока еще несколько отстающий по производительности от CUDA и AMD FireStream [12].

Представленное совершенствование аппаратно-программной базы дало импульс к разработке алгоритмов и программных комплексов решения задач компьютерной оптики [13-14] на GPU, которым и посвящено дальнейшее изложение.

1. Трассировка лучей

Метод трассировки лучей (Ray Tracing) традиционно используется [14-15] для расчета световых полей в рамках геометрической оптики. Он основан на формировании изображения лучами, исходящими от источников света (прямая трассировка) либо наблюдателя (обратная трассировка), претерпевающими отражения и преломления от изучаемых объектов. Основным недостатком метода является высокая вычислительная сложность, связанная с необходимостью рассчитывать ход большого числа лучей для построения изображения объектов с хорошим качеством.

Первый программный инструмент, осуществляющий расчет трассировки лучей на графическом вычислительном устройстве, появился в 2007 году и был разработан группой исследователей из Стэндфордского университета [16] с использованием языка BrookGPU [8]. Авторы проекта сосредоточились на графическом приложении метода трассировки, рассматривая задачу построения проекции освещенных трехмерных сцен на двумерный экран. Современные программные продукты (OptiX фирмы NVidia [17] и RTE Лаборатории компьютерной графики и мультимедиа ф-та ВМК МГУ [18]), реализующие трассировку лучей на GPU по технологии CUDA, также специализированы для решения задачи графической трассировки.

Мы обнаружили лишь одну некоммерческую разработку, посвященную приложению метода в оптике. Программа LINZIK [19] предназначена для трассировки лучей через сферические и асферические оптические поверхности. Она может быть использована для расчета оптических схем (в качестве тестового примера приводится схема телескопа). При организации вычислений данные оптических поверхностей,

список длин волн и данные решетки лучей передаются с центрального процессора в раздел памяти constant [20] GPU, каждому лучу соответствует своя нить (node [20]), которая оставляет в глобальной памяти координату прохождения луча через фокальную плоскость. В зависимости от типа графического процессора допускается применение технологий CUDA и AMD FireStream. Проведенные автором программы тесты на графической карте NVIDIA GeForce 8800 GT (112 потоковых процессоров) продемонстрировали ускорение расчётов от 2 до 10 раз по сравнению с вычислениями на центральном процессоре Intel Pentium E2180 (два ядра, 2 ГГц). Выигрыш от использования GPU возрастал с увеличением количества оптических поверхностей в схеме. Представленное в [21] сравнение для NVIDIA GeForce 9800 GTX (G92, 1,69 ГГц, 128 ALU) и Intel Core 2 Duo E6600 (2,4 ГГц) характеризовалось шестикратным ускорением вычислений.

Независимо от области применения метода трассировки (графической или оптической) следует признать его структуру весьма удачной для реализации на GPU: ставя в соответствие одному лучу одну нить можно одновременно задействовать все потоковые процессоры GPU в силу многочисленности лучей. Вместе с тем, в ряде источников [1, 22] указывается на недостаточность числа регистров современных графических процессоров для эффективной реализации трассировки.

Отметим, что векторные алгоритмы трассировки помимо графических процессоров разрабатывают и для CPU, поддерживающих инструкции SSE [23]. Следует ожидать появления программных пакетов трассировки, написанных на OpenCL, что позволило бы объединить векторные ресурсы центрального и нескольких графических процессоров на одной плате.

2. Метод FDTD

Метод конечных разностей для решения уравнений Максвелла [13, 24] (FDTD, Finite-Difference Time-Domain) получил исключительную популярность благодаря широкой области применения. Расчеты по нему производятся в рамках строгой теории дифракции при изучении распространения импульсов и гармонических электромагнитных полей в линейных и нелинейных следах, характеризующихся произвольными диэлектрической и магнитной проницаемостями.

Следствием учета всех компонент электромагнитного поля и их связей является высокая вычислительная сложность метода, обусловившая его первую реализацию на графическом процессоре сразу вслед за появлением такой возможности [6]. В силу необходимости использовать в то время графический интерфейс, электромагнитные поля хранились в виде текстур, при этом узлы сеточной области соответствовали пикселям.

В работе [25] произведен расчет по FDTD методу дифракционной картины на идеально проводящем цилиндре (двумерный случай, 1300 узлов по осям x и y, 20000 по времени) и кольцевом резонаторе (трехмерный случай, 300, 480, 40 – число узлов по x,y,z, 8000 по времени) с использованием интерфейса OpenGL и графических процессоров NVidia 6800 GT и 7800 GTX. Сравнивая два приведенных процессора, авторы отмечают пятикратный скачек производительности при переходе к более современной карте в двумерном случае и трехкратный в трехмерном. Показано десятикратное превосходство в быстродействии NVidia 7800 GTX над специализированными (программируемая пользователем вентильная матрица) под FDTD микросхемами из [26].

Сравнение реализаций метода конечных разностей для решения уравнения Максвелла на GPU и неспециализированных CPU проведено в работе [27]. Если взять за основу производительность процессора Opteron 270 (2 GHz), то при реализации FDTD метода на центральных процессорах Opteron 890 (2,8 GHz), Core 2 Duo T7600 и

графических процессорах GForce 7400 GO, GForce 7400 GS, GForce 8800 GTX ускорение вычислений составит 2,2; 2,16 и 12,4; 67,82; 429,2 раза. Обсуждается зависимость быстродействия от количества узлов сеточной области: для небольших сеток использование CPU оказывается выгоднее в силу простоя большого числа потоковых процессоров GPU. При постановке вычислительных экспериментов для связи с видеокартой использовался интерфейс OpenGL, программы записывались на языках C и Cg.

С появлением технологии CUDA реализация FDTD на GPU значительно упростилась. В настоящее время оптику-исследователю не только нет нужды знакомиться с API OpenGL и Direct3D (что требует специализированных знаний в области разработки графических приложений), но даже не обязательно изучать расширенный C-подобный язык CUD'ы. Компания NVidia разработала плагины для языков MATLAB, Mathematica и LabView [28], позволяющие задействовать возможности графических процессоров. К сожалению, такая поддержка появилась совсем недавно - Atef Elsherbeni и Veysel Demir, авторы известной монографии 2008 года издания [29], посвященной реализации FDTD на MatLab'е, в главе об использовании графических процессоров (GPU Acceleration of FDTD) еще обращаются к инструментарию языка Brook [8].

В 2007 году NVidia начала сотрудничество с компанией Acceleware с целью разработки библиотеки CUDA FDTD Library. В настоящее время библиотека позволяет [30] моделировать распространение электромагнитного излучения через различные среды (диэлектрики, проводники и анизотропные среды), ограничивая область вычислительного эксперимента PML слоями, электрической/магнитной стенкой, условиями Mur'a или Higdon'a [13, 31-32]. При исследовании резонатора в виде кубической каверны в идеально проводящем материале была достигнута следующая производительность (определяется в количестве значений сеточных функций, рассчитанных за одну секунду [миллион в секунду]): 800 для GPU Tesla C1060 (при объеме сеточной области в 100 миллионов узлов), 1600 для вычислительной системы из 2 процессоров Quadro Plex D2 (200 миллионов узлов) и 2800 для 4-х процессорной системы, объединяющей GPU Tesla S1070 (500 миллионов узлов).

Разработка Acceleware используется в коммерческих пакетах W2405 Agilent FDTD Simulation Element (компания Agilent) [33], Concerto (Cobham Technical Sevices) [34], SEMCAD X OPTICS (Speag) [35] и Xfdtd (RemCom) [36].

Свободно распространяемым является пакет FastFDTD [37]. По свидетельству разработчиков он позволяет получить на видеокартах NVIDIA серии GeForce 6000 и выше 30-50 кратные ускорения вычислений. Независимый источник [38] подтверждает 5-кратное ускорение на GPU GeForce 6800XT и 25-кратное на GeForce 7800 GTX по сравнению с процессором Pentium 4 3,2 ГГц.

Комментируя прогресс в представленной области необходимо отметить очевидность векторизации метода FDTD - вектор сеточных функций посредством операций сложения и вычитания выражается через вектора на предыдущих слоях. Вместе с тем, отображению FDTD на архитектуру GPU свойственна весьма существенная проблема. При моделировании распространения излучения через трехмерные оптические элементы на наиболее распространенных графических процессорах серии GeForce ограниченный объем их оперативной памяти вынуждает организовывать многочисленные обмены с глобальной памятью, что сразу сводит на нет преимущества GPU. Преодоление указанного недостатка авторы настоящей работы видят в использовании декомпозиции вычислительной области [39-40], позволяющей за счет принципа суперпозиции электромагнитных волн многократно снизить количество обменов между глобальной памятью и памятью видеокарты.

Заключение

Появление и бурное развитие вычислений на графических процессорах позволило за несколько последних лет на порядок снизить длительность расчетов на персональном компьютере по актуальным методам компьютерной оптики, что открывает широкие перспективы использования данных методов не только в оптике (трассировка лучей), но и в дифракционной нанофотонике (решение уравнений Максвелла) [41-42].

Благодарности

Работа выполнена при поддержке грантов РФФИ № 10-07-00553, 09-07-12147, 10-01-00453, 09-07-92421 и гранта Президента РФ №НШ-7414.2010.9.

Литература

1. БоресковА.В.,Харламов А.А.Основы работы с технологией CUDA.-М.:ДМК Пресс,2010.-232с.

2. Фролов В. В. Введение в технологию CUDA http://cgm.computergraphics.ru/issues/issue16/cuda

3. Michael J. Flynn. Computer Architecture: Pipelined and Parallel Processor Design.- Jones and Bartlett, Boston, 1995

4. Берилло А. NVIDIA CUDA — неграфические вычисления на графических процессорах. Часть 1 http://www.ixbt.com/video3/cuda-1.shtml

5. Скрябин В. История появления графических процессоров http://cgm.computergraphics.ru/issues

6. S.E. Krakiwsky, L.E. Turner and M.M. Okoniewski Graphics Processor Unit (GPU) Acceleration of Finite-Difference Time-Domain (FDTD) Algorithm//Microwave Symposium Digest, pp. 1033-1036, June 2004.

7. Чеканов Д. NVidia CUDA: вычисления на видеокарте или смерть CPU? http://www.thg.ru/graphic/nvidia_cuda/print.html

8. Brook and BrookGPU research projects at the Stanford University Graphics Lab http://graphics.stanford.edu/projects/brookgpu/

9. Технология CUDA компании NVidia http://www.nvidia.ru/object/cuda_home_new_ru.html

10.Технология ATIStream компании ATI http://developer.amd.com/gpu/ATIStreamSDK/Pages/default.aspx

11. Проект OpenCL http://www.khronos.org/opencl/

12. Сравнение OpenCL с CUDA, GLSL и OpenMP http://habrahabr.ru/blogs/hi/96122/

13. Дифракционная компьютерная оптика. Под редакцией В.А. Сойфера. – М.: Физматлит, 2007

14. Казанский Н.Л. Исследовательский комплекс для решения задач компьютерной оптики // Компьютерная оптика, 2006, № 29, с. 58-77.

15. Казанский Н.Л. Математическое моделирование оптических систем. – Самара: СГАУ, 2005

16. Daniel Horn, Jeremy Sugerman, Mike Houston, Pat Hanrahan Interactive k-D Tree GPU Raytracing http://graphics.stanford.edu/papers/i3dkdtree/

17. Пакет OptiX http://developer.nvidia.com/object/optix-examples.html

18. Пакет RTE http://www.ray-tracing.ru/

19. Пакет LINZIK linzik.com

20. CUDA, Published by NVIDIA Corporation – 2701 San Tomas Expressway Santa Clara, CA 95050, http:/www.nvidia.com/object/cuda_home_new.html

21. Берилло А. NVIDIA CUDA — неграфические вычисления на графических процессорах. Часть 2 http://www.ixbt.com/video3/cuda-2.shtml

22. David Luebke and Steven Parker Interactive Ray Tracing with CUDA http://developer.nvidia.com/object/nvision08-IRT.html

23. Интерактивная трассировка лучей с использованием SIMD инструкций http://software.intel.com/ru-ru/articles/interactive-ray-tracing/

24. Taflove, A. Computational Electrodynamics: The Finite-Difference Time-Domain Method: 2nd. ed. / A. Taflove, S. Hagness – Boston:Arthech House Publishers, 2000.- 852 p.

25. D.K. Price, J.R. Humphrey and E.J. Kelmelis GPU-based accelerated 2D and 3d FDTD solvers // Phisics and Simulation of Optoelectronic Devices XV, vol. 6468 of Proceedings of SPIE, San Jose, Calif, USA, January 2007

26. J.P. Durbano FPGA-based Acceleration of the Three-Dimensional Finite-Difference Time-Domain Method for Electromagnetic Calculations// Global Signal Processing Expo & Conference (GSPx), 2004

27. S. Adams, J. Payne, R. Boppana Finite Difference Time Domain (FDTD Simulations Using Graphics Processors // Proceedings of the 2007 DoD High Performance Computing Modernization Program Users Group Conference, 2007 – p. 334-338

28. Программное обеспечение для MATLAB, использующее графические процессоры с архитектурой CUDA http://www.nvidia.ru/object/matlab_acceleration_ru.html

Расчет пластмассово-линзовых микрообъективов для массовых видеокамер наблюдения

Г.И. Грейсух, Е.Г. Ежов, И.А. Левин, С.А. Степанов
Пензенский государственный университет архитектуры и строительства, г. Пенза

Введение

Расширение круга задач, решаемых с помощью систем видеонаблюдения и, в частности, видеокамер, устанавливаемых в салонах пассажирских самолетов и местах контроля пассажиропотоков в аэропортах, обусловило рост требований, предъявляемых к качеству формируемого видеокамерами изображения. Качество изображения, получаемого с помощью видеокамеры, определяется характеристиками ее ПЗС-матрицы и микрообъектива. Хроматические и монохроматические аберрации последнего являются существенным фактором, ограничивающим разрешение системы в целом и, следовательно, должны быть скорректированы. Сочетание требований низкой стоимости и весьма высоких оптических характеристик микрообъективов обуславливает целесообразность изготовления элементов их оптических схем из пластмасс, так как современные методы формообразования на основе прецизионной штамповки позволяют легко тиражировать пластмассовые линзы с асферическими преломляющими поверхностями и, кроме того, при необходимости штамповать эти поверхности с дифракционным микрорельефом [1-3].

В работах [4-8] показано, что при ограниченном выборе оптических материалов коррекция хроматизма первого порядка рефракционно-линзовых оптических систем может выполняться за счет использования дифракционных линз (ДЛ). Методика ахро- и апохроматизации оптической системы, чей хроматизм не удалось снизить до требуемого уровня на предшествующих этапах расчета, путем введения в нее дифракционно-рефракционного корректора, состоящего из ДЛ и одной либо двух рефракционных линз (РЛ), описана в работе [9].

Анализ серийно производимых различными фирмами микрообъективов для видеокамер наблюдения позволил в качестве базовых принять нижеследующие значения их основных параметров: заднее фокусное расстояние f'=3,7 мм; относительное отверстие 1:2,4; угловое поле в пространстве предметов 2ω=60° [10]. Кроме того, следует принять во внимание, что выпускаемые видеокамеры наблюдения могут быть предназначены для работы в режиме «день» или в режиме «день-ночь». В первом случае камера работает с видимым излучением и рассчитывается на интервал, ограниченный синей F– и красной C–линиями водорода ($\lambda_{min} = \lambda_F = 0{,}48613$ мкм и $\lambda_{max} = \lambda_C = 0{,}65626$ мкм). Во втором случае, кроме видимого используется и ближнее инфракрасное излучение, в результате камера рассчитывается на спектральный интервал от λ_{min}=0,4 мкм до λ_{max}=0,9 мкм.

Компоновка схем и результаты оптимизации ахроматов

При расчете микрообъектива камеры наблюдения, предназначенной для работы в режиме «день», в силу относительной узости спектрального диапазона представляется возможным ограничиться ахроматизацией [11], обеспечивающей равенство задних отрезков системы на крайних длинах волн выбранного спектрального диапазона.

При принятых выше характеристиках микрообъективов вполне удовлетворительной стартовой схемой для ахромата может служить схема, содержащая четыре РЛ, выполненные из одного и того же кроноподобного пластика и имеющие асферические поверхности. В качестве кроноподобного оптического материала нами

выбран полиметилметакрилат (ПММА) ($n_d = 1{,}491756$; $\nu_d = 57{,}4408$). Показатели преломления и числа Аббе, также как и все оптические характеристики, в данной работе приведены на желтой d–линии гелия ($\lambda_d = 0{,}58756$ мкм).

Начальные значения конструктивных параметров схемы (радиусов преломляющих поверхностей, толщин линз и воздушных промежутков) могут быть найдены с использованием любой из известных методик габаритного и аберрационного расчета. Одна из таких методик применительно к малогабаритным объективам для мобильных телефонов изложена в работе [12]. Однако авторами использовалась псевдолучевая методика, предполагающая прослеживание через систему хода лучей в приближении различных порядков малости [13].

В соответствии с методикой подавления хроматизма, описанной в [9], для ахроматической коррекции исходной оптической системы, все РЛ которой выполнены из ПММА, оказалось вполне достаточным включение в схему одиночной ДЛ [11]. На рис. 1 приведена оптическая схема такого пятилинзового ахромата.

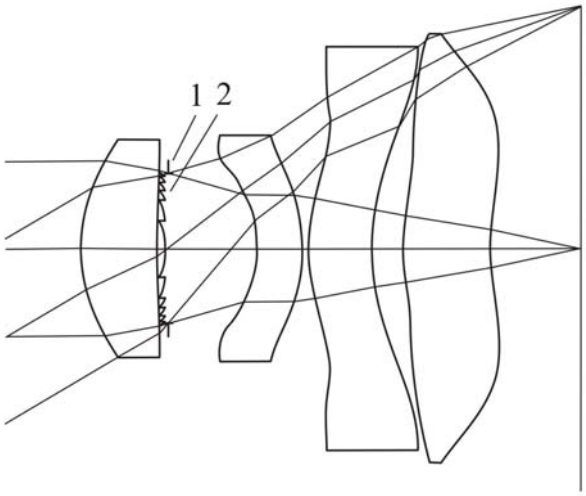


Рис. 1. Принципиальная оптическая схема пятилинзового ахромата:
1– апертурная диафрагма; 2 –ДЛ

Анализ аберрационных кривых микрообъектива показал, что у него действительно удалось совместить ахроматизацию с существенным сокращением вторичного спектра, ограничив изменение заднего фокального отрезка во всем видимом спектральном диапазоне величиной $\Delta s'_F \le 6{,}6$ мкм. При этом удалось также существенно снизить монохроматические аберрации и сферохроматизм. По всему угловому полю $2\omega = 60^\circ$ разрешение составляет 120 мм$^{-1}$ при контрасте не ниже 0,5, а модуль дисторсии не превышает 1%.

Компоновка схем и результаты оптимизации апохроматов

При расчете микрообъектива камеры наблюдения, предназначенной для работы в режиме «день-ночь», требования к коррекции хроматизма ужесточаются, что приводит, как правило, к необходимости апохроматизации, т.е. достижения равенства задних фокальных отрезков на трех длинах волн заданного спектрального диапазона. Исследования показали, что в этом случае исходная оптическая схема микрообъектива должна наряду с положительными линзами, выполненными из кроноподобного оптического материала (например, ПММА), включать, по крайней мере, одну отрицательную линзу, выполненную из флинтоподобного материала [например, поликарбоната (ПК) ($n_d = 1{,}585470$; $\nu_d = 29{,}9092$)]. На одной из поверхностей этой отрицательной линзы или вблизи нее и будет размещена ДЛ. Кроме того, еще до включения в схему ДЛ исходная схема должна обеспечить возможность ограничения

перепада значений фокусного расстояния в пределах выбранного спектрального интервала величиной, равной протяженности вдоль оптической оси главного максимума дифракционного распределения интенсивности в фокальной области [14]:

$$\Delta f' = \left| f'(\lambda_{\max}) - f'(\lambda_{\min}) \right| \le 16\lambda_d K^2, \quad (1)$$

где K - диафрагменное число. При этом модуль радиуса поверхности Петцваля $|R|$, на которой апохроматизированная оптическая система смогла бы формировать стигматическое изображение [14], должен в случае $2\omega=60^\circ$ удовлетворять условию:

$$|R| \ge 2{,}5 f'. \quad (2)$$

Требования (1) и (2) в сочетании с необходимостью ограничения комы и астигматизма при кривизнах поверхностей РЛ, приемлемых с точки зрения высот, на которых они будут пересекаться лучами при выбранных выше относительном отверстии и угловом поле в пространстве предметов, могут быть удовлетворены все той же схемой, содержащей четыре РЛ, выполненные из одного и того же кроноподобного пластика и имеющие асферические поверхности. Апохроматизация же, как указывалось выше, достигается заменой материала отрицательной линзы на ПК и включением в схему ДЛ.

После размещения ДЛ на фронтальной поверхности отрицательной РЛ и оптимизации, выполненной с помощью компьютерной программы оптического проектирования ZEMAX [15], была получена окончательная схема пятилинзового микрообъектива, представленного на рис. 2.

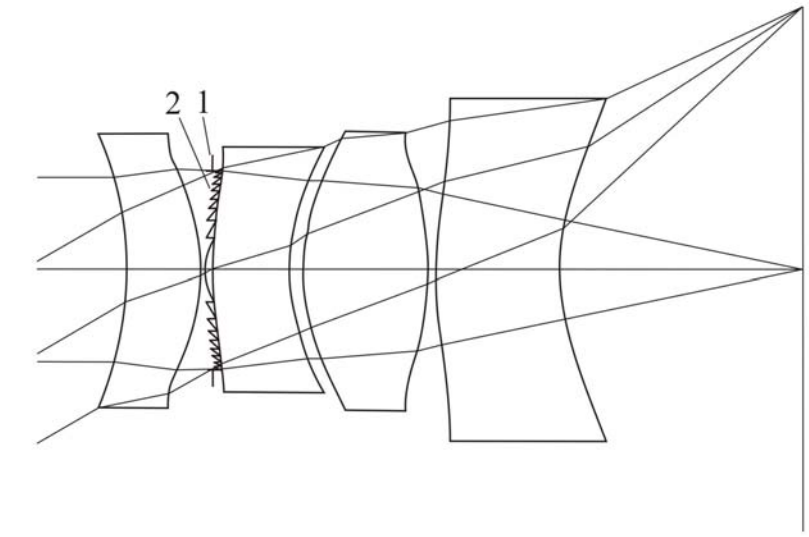


Рис. 2. Принципиальная оптическая схема пятилинзового апохромата:
1– апертурная диафрагма; 2 – ДЛ

Полученный дисторсирующий план-апохромат при относительном отверстии 1:2,4 обеспечивает в плоскости наилучшей установки s'=1,99 мм разрешение 100 мм$^{-1}$ при контрасте не ниже 0,5 в пределах полевого угла $2\omega \le 60^\circ$. Остаточный хроматизм положения в диапазоне от $\lambda_{\min} = 0{,}4$ мкм до $\lambda_{\max} = 0{,}9$ мкм не превышает 6,3 мкм, а модуль дисторсии не превышает 0,35%.

Наиболее ощутимым недостатком данного объектива является весьма широкий диапазон полевых углов в пространстве изображений, приводящий к неоднородности освещения по плоскости ПЗС-матрицы, так как на ее периферийные элементы главные лучи падают под углами порядка 30°.

Диапазон полевых углов в пространстве изображений можно существенно (более чем в 3,5 раза) снизить, тем самым приблизив ход лучей к телецентрическому,

если перейти к восьмилинзовой схеме, включающей три двухлинзовых склеенных компонента. Оптимизация данной оптической системы, представленной на рис. 3, позволила получить план-апохромат с фокусным расстоянием f'=3,7 мм. Микрообъектив при относительном отверстии 1:2,4 обеспечивает в плоскости наилучшей установки s'=2,175 мм разрешение 96 мм$^{-1}$ при контрасте не ниже 0,5 в пределах полевого угла $2\omega \le 60^{\circ}$. Остаточный хроматизм положения в диапазоне от $\lambda_{min} = 0{,}4$ мкм до $\lambda_{max} = 0{,}9$ мкм не превышает 6,9 мкм, а модуль дисторсии не превышает 1%.

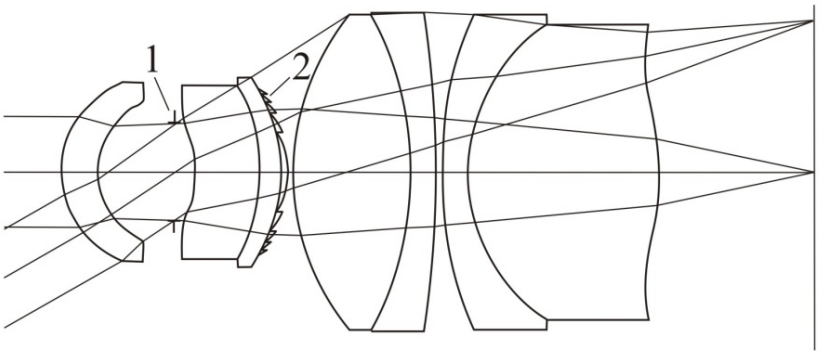


Рис. 3. Принципиальная оптическая схема восьмилинзового апохромата с ходом лучей в пространстве изображений, близким к телецентрическому:
1– апертурная диафрагма; 2 – ДЛ

Особо отметим, что структуры ДЛ всех вышепредставленных микрообъективов - низкочастотные и содержат не более десяти кольцевых зон. Низкочастотность структур позволяет выполнить их с пилообразным профилем штриха и более того двухслойными, для чего могут быть использованы те же самые оптические пластмассы, т.е. ПММА и ПК. Это обеспечит дифракционную эффективность, близкую к 100% в пределах всего заданного спектрального диапазона [16].

Заключение

Обобщая вышеприведенные результаты, можно сделать следующие основные выводы:

- включение в схему пластмассово-линзового микрообъектива одиночной ДЛ позволяет произвести коррекцию хроматизма как в видимом, так и в широком спектральном интервале, охватывающем видимый и ближний ИК диапазоны;
- рекомендации авторов дают возможность скомпоновать исходную схему объектива и получить конструктивные параметры, необходимые для последующей оптимизации;
- достижимые оптические характеристики пластмассово-линзовых микрообъективов удовлетворяют требованиям, предъявляемым к объективам массовых видеокамер наблюдения.

Благодарности

Работа выполнена при поддержке министерства образования и науки РФ в рамках федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы по лоту «Проведение научных исследований научными группами под руководством докторов наук (шифр «2010-1.2.1-102-018») по теме: «Фокусирующая и изображающая дифракционная оптика полихроматического излучения» (шифр заявки «2010-1.2.1-102-018-033»).

Литература

1. Потелов В. В., Сенник Б. Н. Асферизация высокоточных оптических элементов методом вакуумного напыления // Оптический Журнал. - 2004. - Т. 71, № 12. - С. 14-19.
2. Radtke D., Zeitner U.D. Lazer-lithography on non-planar surfaces // Optics Express. 2007. V. 15, No 3. P. 1167-1174.
3. G-S plastic optics [Электронный ресурс]. – Режим доступа: http://www.gsoptics.com
4. Ган М.А. Теория и методы расчета голограммных и киноформных оптических элементов. Л.: ГОИ, 1984. 140 с.
5. Stone T., George N. Hybrid diffractive-refractive lenses and achromats // Appl. Opt. 1988. V. 27, No 14. P. 2960-2971.
6. Gan M.A. Optical systems with holographic and kinoform elements // Proc. SPIE. 1989. V. 1136. P. 150.
7. Gan M. A. Kinoforms long focal objectives for astronomy // Adaptive optics and optical structures / Proc. of the Meeting, European Congress on Optics. 1990. P. 330–338.
8. Gan M., Potyemin I., Perveev A. High-speed apo-lens with kinoform element // Proc. SPIE. 1991. V. 1574. P. 243–249.
9. Greisukh G.I., Ezhov E.G., Stepanov S.A. Diffractive-refractive hybrid corrector for achro- and apochromatic corrections of optical systems // Appl. Opt. 2006. V. 45, No 24. P. 6137-6141.
10. http://www.ukaoptics.com/ccd.html.
11. Greisukh G.I., Ezhov E.G., Levin I.A., Stepanov S.A. Design of achromatic and apochromatic plastic micro-objectives // Appl. Opt. 2006. V. 49, No 23. P. 4379-4384.
12. Бронштейн И.Г., Зверев В.А., Лившиц И.Л., Kim Young-Gi, Kim Tae-Young, Young Phil-Ho. Выбор оптической схемы и расчет малогабаритных объективов для мобильных телефонов // Оптический Журнал. - 2009. - Т. 76, № 5. - С. 25-31.
13. Greisukh G.I., Bobrov S.T., Stepanov S.A. Optics of diffractive and gradient-index elements and systems. Bellingham: SPIE Press, 1997. 414 p.
14. Борн М., Вольф Э. Основы оптики. М.: Наука, 1973. 720 с.
15. ZEMAX Optical Design Program [Электронный ресурс]. – Режим доступа: http://www.focus-software.com.
16. Грейсух Г.И., Безус Е.А., Быков Д.А., Ежов Е.Г., Степанов С.А. Подавление спектральной селективности двухслойных рельефно-фазовых дифракционных структур // Опт. и спектр. 2009. Т. 106, вып. 4. С. 694-699.

Коррекция хроматизма оптических систем, содержащих дифракционные элементы

Г.И. Грейсух, Е.Г. Ежов, С.В. Казин, С.А. Степанов
Пензенский государственный университет архитектуры и строительства, г. Пенза

Введение

Совершенствование оптических систем, освоение новых нетрадиционных для оптики, спектральных областей требует и нетрадиционных подходов к коррекции хроматизма. Действительно, в системах, предназначенных для видимого и инфракрасного (ИК) диапазонов, основным приемом исправления хроматизма является использование материалов с различной дисперсией [1]. Это, однако, существенно, ограничивает возможности коррекции хроматизма, либо требует использования необычных оптических материалов, являющихся весьма дорогими и имеющими, как правило, невысокие эксплуатационные характеристики [1-3]. С другой стороны, включение в оптическую систему наряду с рефракционными линзами (РЛ) дифракционных оптических элементов (ДОЭ) и, в частности, оптически слабых дифракционных линз (ДЛ) позволяет, как упростить систему, так и отказаться от использования необычных оптических материалов [4-6].

Проблема построения оптических систем на базе РЛ еще более усугубляется при переходе в коротковолновую часть электромагнитного спектра [7]: ультрафиолетовую (УФ), вакуумную ультрафиолетовую (ВУФ) и рентгеновскую области. Уже в ВУФ-области каталог используемых материалов ограничивается несколькими единицами, что же касается рентгеновской области, то в силу особых свойств материалов в этой области РЛ должны иметь радиусы кривизн в несколько микрометров, что, помимо технологических проблем создания таких линз еще и существенно ограничивает их световые диаметры [8]. Последние же успехи в лазерной и электронной литографии позволяют изготавливать ДОЭ, как для всех поддиапазонов оптического (включая ИК и УФ), так и рентгеновского излучения [9, 10]. Однако оптические системы, включающие ДОЭ с высокочастотной пространственной структурой, обладают значительным хроматизмом и в области его коррекции, как в нашей стране, так и за рубежом сделаны лишь первые шаги.

Коррекция хроматизма гибридных систем оптического диапазона

В оптическом диапазоне использование оптически слабых ДЛ наиболее эффективно для достижения апохроматизации [2-6] и суперахроматизации [11]. Апохроматизация или исправление вторичного спектра предполагает достижение равенства задних фокальных отрезков на трех длинах волн заданного спектрального диапазона, а суперахроматизация, т.е. исправление третичного спектра предполагает устранение хроматизма положения уже для четырех длин волн заданного спектрального диапазона и направлена на достижение равенства задних отрезков на этих длинах волн.

На рис. 1 приведена принципиальная оптическая схема суперахромата, полученного из классического триплета Кука в результате замены его внутренней отрицательной линзы, расположенной вблизи апертурной диафрагмы, на дифракционно-рефракционный корректор (ДРК), состоящего из одной ДЛ и двух РЛ. Выбор триплета обусловлен, в частности, тем, что именно этот объектив, использовался в работе [6] для демонстрации возможности апохроматической коррекции с помощью ДРК, состоящего из одной ДЛ и одной РЛ.

Суперахроматизация осуществлялась в видимом диапазоне, ограниченном синей F– и красной C–линиями водорода ($\lambda_{min} = \lambda_F = 0{,}48613$ мкм; $\lambda_{max} = \lambda_C = 0{,}65626$ мкм). В качестве центральной длины волны выбиралась желтая d–линия гелия ($\bar{\lambda} = \lambda_d = 0{,}58756$ мкм), а промежуточные длины волн вычислялись по формулам $\lambda_1 = (\lambda_C + 2\lambda_F)/3$; $\lambda_2 = (\lambda_F + 2\lambda_C)/3$. Эффективное фокусное расстояние объектива с ДРК на центральной длине волны $f' = 240$ мм, а задний фокальный отрезок s'_F=224,03 мм.

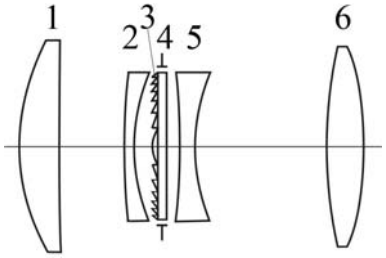


Рис. 1. Оптическая схема объектива с ДРК: 1; 2; 5; 6 – РЛ; 3 – ДЛ; 4 – апертурная диафрагма

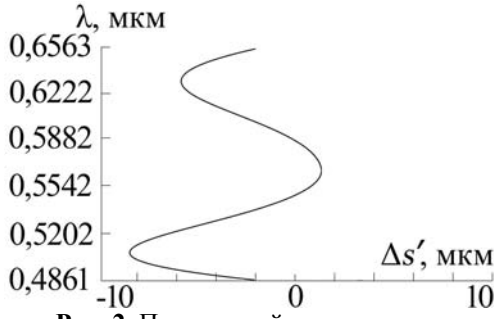


Рис. 2. Продольный хроматизм объектива с ДРК

Несмотря на то, что все стекла, используемые в объективе, имеют обычный ход частной дисперсии, достигнута, как видно из рис. 2, строгая суперахроматизация. В результате изменение величины заднего фокального отрезка в пределах всего выбранного спектрального интервала уменьшено от $\Delta s'_F$=606 мкм у исходного триплета до $\Delta s'_F$=9,8 мкм у объектива с ДРК. При этом, благодаря оптимальному сочетанию прогиба РЛ и значений коэффициентов разложения функции пространственной частоты структуры ДЛ, удалось существенно снизить как сферическую аберрацию, так и сферохроматизм. Наконец, абсолютная величина радиуса поверхности Петцваля благодаря ДРК увеличилась с $|R| = 776$ мм до $|R| = 1190$ мм. В результате полихроматическое разрешение объектива, рассчитанное при относительном отверстии 1:6 и при контрасте T=0,2 на оси ($\omega = 0$) повысилось с $N = 13$ мм$^{-1}$ до $N = 120$ мм$^{-1}$, а на краю углового поля ($\omega = 16°$) - с $N = 13$ мм$^{-1}$ до $N = 32{,}5$ мм$^{-1}$.

Ахроматизация дифракционных оптических систем мягкого рентгеновского диапазона

Проблеме ахроматизации, т.е. достижению равенства задних фокальных отрезков на двух длинах волн заданного спектрального диапазона у чисто дифракционных оптических систем посвящен ряд работ (см., например, [12, 13]). В частности, в [13] найдены четыре схемы дифракционных дублетов-ахроматов. Общим для всех этих схем является то, что в отличие, например, от зонных пластинок Френеля, используемые в схемах ДОЭ имеют на оси ненулевую пространственную частоту, т.е. являются аксиконами с дополнительными фокусирующими свойствами (рис. 3).

В цитированной работе [13] возможности дифракционных ахроматизированных дублетов, а также сопоставительный анализ фокусирующих свойств систем различных типов приведены для ИК-диапазона (длина волны порядка 10 мкм), когда отношение светового диаметра к длине волны составляет величину порядка 10^4. В рентгеновском же диапазоне это отношение как минимум на два порядка больше. Данное отношение входит в волновую сферохроматическую аберрацию, выраженную в долях длины

волны, и определяет качество фокусировки. Поэтому, чтобы оценить возможности использования дифракционного дублета для фокусировки полихроматического излучения рентгеновского диапазона требуется провести дополнительные исследования.

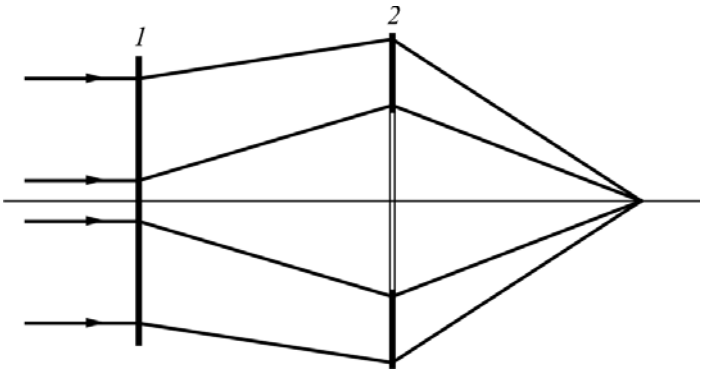


Рис. 3. Оптическая схема дифракционного дублета, скоррегированного на две длины волны: 1, 2 – ДОЭ

Авторами настоящей работы предложена методика получения функций распределения пространственных частот структур ДОЭ, обеспечивающих ахроматизацию дублета, отличающаяся от известных тем, что позволяет задавать величину минимального периода в структурах элементов и исследовать влияние этой величины на основные параметры дублета [14]. Такая особенность методики особенно важна при расчете дублета для рентгеновского излучения из-за имеющихся в этом случае технологических ограничений на минимальный период используемых ДОЭ.

Учитывая, что ДОЭ дублета представляют собой аксиконы с дополнительными фокусирующими свойствами, а также с целью минимизации числа параметров, определяемых итерационно, предлагается при выборе функции распределения пространственной частоты фронтального ДОЭ, рассматривать его как «обратную» осевую асферизованную голограмму точки, у которой пространственная частота растет не от оси к краю апертуры, а наоборот. Выражение, описывающее пространственную частоту этой синтезированной голограммы, имеет вид

$$\Omega_1(\rho_1)=\frac{1}{\lambda_1}\left\{C-\rho_1\left[\Omega_{\text{H}}(\rho_1)-\sum_{i=1}^{I}B_i\rho_1^{i-2}\right]\right\}. \qquad (1)$$

Здесь

$$\Omega_{\text{H}}(\rho_1)=-\frac{1/z_1}{\sqrt{1+(\rho_1/z_1)^2}}+\frac{1/z_2}{\sqrt{1+(\rho_1/z_2)^2}}; \qquad (2)$$

$$C=B_0\rho_{1,\max}\left(\Omega_{\text{H}}(\rho_{1,\max})-\sum_{i=1}^{I}B_i\rho_{1,\max}^{i-2}\right); \qquad (3)$$

λ_1 – длина волны записи голограммы и одновременно одна из двух крайних длин волн спектрального диапазона, в пределах которого производится ахроматизация; ρ_1 - расстояние от оси структуры в плоскости голограммы, $\rho_{1,\max}$ - радиус апертуры голограммы; z_1, z_2 - расстояния от плоскости регистрации до соответствующих источников записи, т.е. отрезки записи голограммы, определяющие ее оптическую силу. Входящий в формулу (3) коэффициент B_0 определяет расстояние от центра симметрии структуры голограммы, до окружности, на которой пространственная частота структуры обращается в нуль и, как показали исследования, является весьма эффективным коррекционным параметром, влияющим на минимальный период в

структуре обоих элементов дублета и его хроматические свойства. При расчете этот коэффициент задается, а коэффициенты B_1 , …, B_I , обеспечивающие асферизацию формируемого фронтальным ДОЭ волнового фронта, находятся итерационно. Итерационный процесс направлен на минимизацию диаграммы рассеяния лучей, формируемой в задней фокальной плоскости дублета при его освещении осевым параллельным пучком на второй из двух крайних длин волн выбранного диапазона $-\lambda_2$.

Второй ДОЭ дублета представляет собой голограмму, моделируемую на каждом шаге итерационного процесса. Один из фронтов записи этой голограммы формируется фронтальным ДОЭ при его нормальном освещении плоским фронтом электромагнитного излучения с длиной волны λ_1, а второй фронт записи сферический сходящийся в задний фокус дублета. То есть функция распределения пространственной частоты структуры второго ДОЭ определяется на каждом шаге итерационного процесса из условия идеальной фокусировки излучения на длине волны λ_1.

С использованием вышеописанной методики исследовалась зависимость хроматизма дублета от его конструктивных параметров, включая минимальный период в структурах ДОЭ. Исследования проводились, в частности, при λ_1=5 нм, λ_2=6 нм, световом диаметре первого ДОЭ D_1=3,5 мм и при относительном отверстии дублета, под которым в данном случае можно понимать отношение светового диаметра D_1 к заднему фокальному отрезку s'_F, равном 0,1.

Для сравнения возможностей фокусировки полихроматического излучения дублетом-ахроматом и одиночной ДЛ рассматривался дублет, с минимальным периодом в структурах $\Lambda_{\min}$=54 нм. Предельная ширина спектрального диапазона высококачественной фокусировки определялась из условия равенства среднеквадратических радиусов диаграмм рассеяния лучей монохроматического излучения в центре диапазона и на его краях. Для дублета с выбранными параметрами этот диапазон был ограничен длинами волн 4,85 нм и 6,15 нм, а среднеквадратический радиус диаграммы рассеяния лучей полихроматического излучения в пределах этого диапазона δ_{RMS}=24 мкм. При фокусировке же полихроматического излучения указанного спектрального диапазона одиночной ДЛ с D=3,5 мм среднеквадратический радиус диаграммы рассеяния δ_{RMS}=99 мкм, т.е. при фокусировке полихроматического излучения данный дублет-ахромат дает выигрыш по сравнению с одиночной ДЛ более чем в 4 раза.

Заключение

Дифракционные линзы с небольшой оптической силой являются эффективными коррекционными элементами, позволяющими достигать высокой степени коррекции хроматизма у оптических систем, с рефракционными силовыми элементами. В то же время хроматизм оптических систем, состоящих только из дифракционных линз со структурами подобными структуре зонной пластинке Френеля, неустраним. Ахроматизация чисто дифракционной системы возможна лишь в случае, если она компонуется из аксиконов с дополнительными фокусирующими свойствами. Такие системы позволяют произвести качественную фокусировку практически только осевого пучка. Поэтому их использование наиболее целесообразно в мягком рентгеновском диапазоне, где они являются практически единственной альтернативой зеркальной оптике скользящего падения, также работающей в очень узком диапазоне полевых углов.

Благодарности

Работа выполнена при поддержке министерства образования и науки РФ в рамках федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы по лоту «Проведение научных исследований научными группами под руководством докторов наук (шифр «2010-1.2.1-102-018») по теме: «Фокусирующая и изображающая дифракционная оптика полихроматического излучения» (шифр заявки «2010-1.2.1-102-018-033»).

Литература

1. Слюсарев Г. Г. Методы расчета оптических систем. - Л.: Машиностроение, 1969. – 672 с.
2. Rayces J.L., Rosete-Agilar M. Selection of glasses for achromatic doublets with reduced secondary spectrum. I. Tolerance conditions for secondary spectrum, spherochromatism, and fifth-order spherical aberration // Applied Optics. 2001. Vol. 40, No 31. P. 5663-5676.
3. Rosete-Agilar M., Rayces J.L. Selection of glasses for achromatic doublets with reduced secondary spectrum. II. Application of the method for selecting pairs of glasses with reduced secondary spectrum // Aplied Optics. 2001. Vol. 40, No 31. P. 5677- 5692.
4. Ган М.А. Теория и методы расчета голограммных и киноформных оптических элементов. Л.: ГОИ, 1984. 140 с.
5. Gan M., Potyemin I., Perveev A. High-speed apo-lens with kinoform element // Proceedings of SPIE. 1991. V. 1574. P. 243–249.
6. Greisukh G.I., Ezhov E.G., Stepanov S.A. Diffractive-refractive hybrid corrector for achro- and apochromatic corrections of optical systems // Applied Optics. 2006. V. 45, No 24. P. 6137-6141.
7. Бельский А.Б., Ган М.А., Миронов И.А. Создание и аттестация флюорита. Перспективы разработки оптических систем для УФ фотолитографии. // Международный оптический конгресс «Оптика-XXI век»: сб. трудов. СПб.: СПбГУ ИТМО, 2006 - С. 9-17.
8. Аристов В. В., Шабельников Л.Г. Современные достижения рентгеновской оптики преломления // Успехи физических наук. 2008. Т 178, № 1. С. 61-83.
9. Полещук А.Г., Корольков В.П. Методы изготовления и контроля дифракционных оптических элементов // Сборник трудов научно-практической конференции «Голография в России и за рубежом Наука и практика». М.: ООО «ГОЛОГРАФИЯ-СЕРВИС». 2007. С. 11-14.
10. Аристов В. В. Современная рентгеновская оптика высокого разрешения // Вестник Российской академии наук. 2002. Т. 72, № 11. С. 963-968.
11. *Грейсух Г.И., Ежов Е.Г., Казин С.В., Степанов С.А.* Дифракционно-рефракционный корректор третичного спектра // Оптический Журнал. - 2010. - Т. 77, № 9. - С. 22-29.
12. Bennett S.J. Achromatic combinations of hologram optical elements // Applied Optics. - 1976. Vol. 15, No 2. - P. 542-545.
13. Farn M.W., Goodman J.W. Diffractive doublets corrected at two wavelengths // J. Opt. Soc. Am. A. - 1991. - Vol. 8, No 6. - P. 860-867.
14. Грейсух Г.И., Ежов Е.Г., Казин С.В., Степанов С.А. Анализ возможностей ахроматизации оптических систем, состоящих из дифракционных элементов // Компьютерная оптика. – 2010. - Том 34, № 2. - С. 87-193.

Возможности суперахроматизации пластмассово-линзовых микрообъективов

Г.И. Грейсух, Е.Г. Ежов, И.А. Левин, С.А. Степанов
Пензенский государственный университет архитектуры и строительства, г. Пенза

Введение

Суперахроматическая коррекция (исправление третичного спектра) предполагает устранение хроматизма положения для четырех длин волн заданного спектрального диапазона и направлена на достижение равенства задних отрезков на этих длинах волн [1].

При проектировании оптических систем на основе рефракционных линз (РЛ) для исправления третичного спектра используют оптические материалы с близкими по значению относительными частными дисперсиями, но с различными коэффициентами дисперсий [2-6]. К таким материалам относятся специальные марки стекол, объединяемые в группы особых флинтов и кронов, а также флюорит.

Уникальные дисперсионные свойства дифракционных линз (ДЛ) и несомненные успехи в области технологии их изготовления обусловили резко возросший в последние годы интерес к использованию таких элементов для коррекции хроматизма [7–11].

Методика расчета дифракционно-рефракционного корректора (ДРК)

Приращение заднего отрезка оптической системы при переходе с одной длины волны на другую практически пропорционально первой хроматической сумме этой системы S_{ch}, вычисленной для данной пары длин волн. Отсюда следует, что суперахроматизация требует обращения в нуль первых хроматических сумм, вычисленных для четырех длин волн, ограничивающих три примыкающих спектральных диапазона.

Полагая ДРК бесконечно тонким, и учитывая, что его оптическая сила Φ на центральной длине волны $\overline{\lambda}$ должна иметь фиксированное значение, условие суперахроматизации оптической системы, в которую вводится этот корректор, можно записать в виде системы четырех уравнений

$$\left.\begin{aligned}
&\varphi_{\text{D}} + \sum_{i=1}^{I} \varphi_{\text{R}i} = \Phi \\
&\left(\frac{\varphi_{\text{D}}}{\nu_{\text{D}}} + \sum_{i=1}^{I} \frac{\varphi_{\text{R}i}}{\nu_{\text{R}i}} \right) h_{\text{A}}^2 - S_{\text{ch}}^{(\lambda_{\min}, \lambda_{\max})} = 0 \\
&\left(\frac{\varphi_{\text{D}}}{\nu_{\text{D}}} \gamma_{1\text{D}} + \sum_{i=1}^{I} \gamma_{1\text{R}i} \frac{\varphi_{\text{R}i}}{\nu_{\text{R}i}} \right) h_{\text{A}}^2 - S_{\text{ch}}^{(\lambda_1, \lambda_{\max})} = 0 \\
&\left(\frac{\varphi_{\text{D}}}{\nu_{\text{D}}} \gamma_{2\text{D}} + \sum_{i=1}^{I} \gamma_{2\text{R}i} \frac{\varphi_{\text{R}i}}{\nu_{\text{R}i}} \right) h_{\text{A}}^2 - S_{\text{ch}}^{(\lambda_{\min}, \lambda_2)} = 0
\end{aligned}\right\}. \qquad (1)$$

Здесь φ_{D} и φ_{R} – оптические силы ДЛ и РЛ корректора на длине волны $\overline{\lambda}$; h_{A} – высота апертурного нулевого луча в плоскости ДРК, вычисленная на той же длине волны; I – количество РЛ в корректоре; $S_{\text{ch}}^{(\lambda_{\min}, \lambda_{\max})}$, $S_{\text{ch}}^{(\lambda_1, \lambda_{\max})}$ и $S_{\text{ch}}^{(\lambda_{\min}, \lambda_2)}$ – вклады в первую хроматическую сумму всех элементов исходной оптической системы (за исключением заменяемого компонента), вычисленные для соответствующих пар длин волн.

Коэффициенты дисперсии ν и относительные частные дисперсии γ материала РЛ и ДЛ, входящие в систему уравнений (1), имеют вид

$$\nu_R = \left(n_{\bar{\lambda}} - 1\right) / \left(n_{\lambda_{min}} - n_{\lambda_{max}}\right), \tag{2}$$

$$\gamma_{1R} = \left(n_{\lambda_1} - n_{\lambda_{max}}\right) / \left(n_{\lambda_{min}} - n_{\lambda_{max}}\right), \tag{3}$$

$$\gamma_{2R} = \left(n_{\lambda_{min}} - n_{\lambda_2}\right) / \left(n_{\lambda_{min}} - n_{\lambda_{max}}\right), \tag{4}$$

$$\nu_D = \bar{\lambda} / \left(\lambda_{min} - \lambda_{max}\right), \tag{5}$$

$$\gamma_{1D} = \left(\lambda_1 - \lambda_{max}\right) / \left(\lambda_{min} - \lambda_{max}\right), \tag{6}$$

$$\gamma_{2D} = \left(\lambda_{min} - \lambda_2\right) / \left(\lambda_{min} - \lambda_{max}\right). \tag{7}$$

Выбранные длины волн спектрального диапазона удовлетворяют условию $\lambda_{min} < \lambda_1 < \bar{\lambda} < \lambda_2 < \lambda_{max}$.

Система уравнений (1) является линейной относительно оптических сил элементов ДРК. Если ДРК наряду с ДЛ содержит три РЛ, эта система практически всегда имеет приемлемые (с точки зрения оптических сил элементов корректора) решения. Более того, эти решения существуют для весьма широкого ряда наборов стекол РЛ «с нормальным ходом» частной дисперсии. В случае же ДРК, включающего наряду с ДЛ две РЛ, рассматриваемая система линейных уравнений совместна только при выполнении условия [12]:

$$\begin{vmatrix} 1 & 1 & 1 & -\Phi \\ 1/\nu_D & 1/\nu_{R_1} & 1/\nu_{R_2} & -S_{ch}^{(\lambda_{min},\,\lambda_{max})}/h_A^2 \\ \gamma_{1D}/\nu_D & \gamma_{1R_1}/\nu_{R_1} & \gamma_{1R_2}/\nu_{R_2} & -S_{ch}^{(\lambda_1,\,\lambda_{max})}/h_A^2 \\ \gamma_{2D}/\nu_D & \gamma_{2R_1}/\nu_{R_1} & \gamma_{2R_2}/\nu_{R_2} & -S_{ch}^{(\lambda_{min},\,\lambda_2)}/h_A^2 \end{vmatrix} = 0. \tag{8}$$

Если ДРК подобно исходному объективу компонуется РЛ, выполненными из реальных оптических стекол, то вероятность точного решения уравнения (8) на априорно выбранных длинах волн близка к нулю. В то же время требование минимизации модуля левой части уравнения (8) может быть положено в основу поиска оптимальной пары стекол РЛ трехэлементного ДРК, обеспечивающей возможность решения системы уравнений (1) с наилучшим приближением.

В силу нелинейности показателя преломления стекла как функции длины волны, условия равенства нулю первой хроматической суммы и отсутствия хроматизма положения для данной пары длин волн, строго говоря, не совпадают. Поэтому оптические силы элементов ДРК, полученные при решении системы уравнений (1) рассматриваются в качестве нулевого приближения и затем уточняются путем итерационного решения системы уравнений, одно из которых обеспечивает заданное фокусное расстояние f' оптической системы с установленным ДРК на длине волны $\bar{\lambda}$, а три других - равенство задних отрезков этой системы на четырех длинах волн выбранного спектрального диапазона:

$$\left.\begin{aligned} f'_{\bar{\lambda}} &= f' \\ s'_{\lambda_{min}} &= s'_{\lambda_{max}} \\ s'_{\lambda_1} &= s'_{\lambda_{max}} \\ s'_{\lambda_{min}} &= s'_{\lambda_2} \end{aligned}\right\}. \tag{9}$$

Переход от системы уравнений (1) к системе (9) позволяет, кроме того, ввести толщины элементов и воздушных промежутков и оперировать с «реальными стеклами», описывая их не основанными на линеаризации зависимости показателя от

длины волны параметрами ν_R и γ_R, а соответствующими дисперсионными формулами.

Суперахроматизация микрообъектива

Анализ серийно производимых различными фирмами микрообъективов для видеокамер наблюдения позволил в качестве базовых принять нижеследующие значения их основных параметров: заднее фокусное расстояние f'=3,71 мм; относительное отверстие 1:2,4; угловое поле в пространстве предметов 2ω=54° [13].

Перечисленные требования в сочетании с необходимостью ограничения комы и астигматизма при кривизнах поверхностей линз, приемлемых с точки зрения высот, на которых они будут пересекаться лучами при выбранных выше относительном отверстии и угловом поле в пространстве предметов, могут быть удовлетворены уже триплетом, подобным описанному в работе [14]. Это обусловлено, в частности, и тем, что у исходной оптической схемы коррекция сферической аберрации не требуется, так как изначально предполагается введение в нее ДЛ и возможность асферизации преломляющих поверхностей. После включения в исходный триплет трехэлементного ДРК, состоящего из одной ДЛ и двух РЛ, призванного обеспечить суперахроматизацию микрообъектива в целом, приходим к пятилинзовой схеме, в которой на одну из поверхностей РЛ, расположенной вблизи апертурной диафрагмы, нанесена структура ДЛ.

Начальные значения конструктивных параметров схемы (радиусов преломляющих поверхностей, толщин линз и воздушных промежутков) могут быть найдены с использованием любой из известных методик габаритного и аберрационного расчета. Одна из таких методик применительно к малогабаритным объективам для мобильных телефонов изложена в работе [15]. Однако авторами использовалась псевдолучевая методика, предполагающая прослеживание через систему хода лучей в приближении различных порядков малости [16].

Суперахроматизация осуществлялась в спектральном интервале от λ_{min}=0,4 мкм до λ_{max}=0,9 мкм. В качестве центральной длины волны выбиралась желтая d–линия гелия ($\overline{\lambda} = \lambda_d = 0{,}58756$ мкм), а промежуточные длины волн вычислялись по формулам $\lambda_1 = (\lambda_{max}+2\lambda_{min})/3$; $\lambda_2 = (\lambda_{min}+2\lambda_{max})/3$.

В процессе поиска оптимальной пары оптических пластмасс для РЛ корректора на этапе минимизации модуля левой части уравнения (8), итерационного решения системы уравнений (9) и оценки кривизны Петцваля, были выбраны марки акрил (кроноподобный пластик) и поликарбонат (флинтоподобный пластик), которые наряду с суперахроматизацией обеспечивают возможность формирования стигматического изображения оптической системой на поверхности с минимальной кривизной.

Оптимизация рассчитанной оптической системы, представленной на рисунке, позволила получить суперахромат, с фокусным расстоянием f'=3,71 мм, который при относительном отверстии 1:2,4 обеспечивает в плоскости наилучшей установки s'=2,412 мм разрешение 110 мм$^{-1}$ при контрасте не ниже 0,5 в пределах полевого угла $2\omega \le 54^{\circ}$. Остаточный хроматизм положения в диапазоне от $\lambda_{min} = 0{,}4$ мкм до $\lambda_{max} = 0{,}9$ мкм не превышает 3,9 мкм, а модуль дисторсии не превышает 0,1%.

Также отметим, что структура ДЛ представленного микрообъектива низкочастотная и содержат не более десяти кольцевых зон. Низкочастотность структуры позволяет выполнить её с пилообразным профилем штриха и более того двухслойной, для чего могут быть использованы те же самые оптические пластмассы, т.е. акрил и поликарбонат. Это обеспечит дифракционную эффективность, близкую к 100% в пределах всего заданного спектрального диапазона [17].

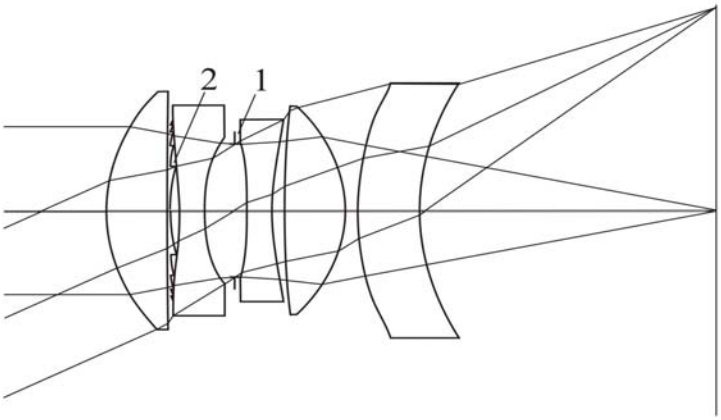


Рис. Принципиальная оптическая схема пятилинзового микрообъектива:
1– апертурная диафрагма; 2 – ДЛ

Существенным недостатком данного объектива является широкий диапазон полевых углов в пространстве изображений, приводящих к неоднородности освещения по плоскости многоэлементного фотоприемника. Исправление данного недостатка с сохранением суперахроматизации только за счет РЛ объектива приводит к существенному увеличению дисторсии, поэтому необходимо введение в схему дополнительно рефракционно-линзового корректора, что повлечет рост габаритов и усложнение исходной системы.

Заключение

Обобщая результаты проведенной работы, можно сделать следующие основные выводы:

− включение в схему пластмассово-линзового микрообъектива одиночной ДЛ позволяет произвести суперахроматизацию в широком интервале длин волн, охватывающем видимый и ближний ИК диапазоны;

− рекомендации авторов дают возможность скомпоновать исходную схему объектива и получить конструктивные параметры, необходимые для последующей оптимизации;

− достижимые оптические характеристики пластмассово-линзовых микрообъективов удовлетворяют требованиям, предъявляемым к объективам массовых видеокамер наблюдения.

Благодарности

Работа выполнена при поддержке министерства образования и науки РФ в рамках федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы по лоту «Проведение научных исследований научными группами под руководством докторов наук (шифр «2010-1.2.1-102-018») по теме: «Фокусирующая и изображающая дифракционная оптика полихроматического излучения» (шифр заявки «2010-1.2.1-102-018-033»).

Литература

1. Herzberger M. Color correction in optical systems and a new dispersion formula // Optica Acta. 1959. V. 6. P. 197-215.

2. Чуриловский В.Н. Теория хроматизма и аберраций третьего порядка. Л.: Машиностроение, 1968. 383 с.

3. Вычислительная оптика: Справочник / Русинов М.М., Грамматин А.П., Иванов П.Д. и др. / Под общ. ред. М.М. Русинова. Л.: Машиностроение, 1984. 423 с.

4. Русинов М.М. Композиция оптических систем. Л.: Машиностроение, 1989. 383 с.

5. Herzberger M, Mc Clure N.R. The design of superachromatic lenses // Applied Optics. 1963. V. 2. P. 553-560.

6. Maxwell J. Tertiary spectrum manipulation in apochromats // **Proceedings** of SPIE. 1990. V. 1354. P. 408–416.

7. Ган М.А. Теория и методы расчета голограммных и киноформных оптических элементов. Л.: ГОИ, 1984. 140 с.

8. Stone T., George N. Hybrid diffractive-refractive lenses and achromats // Applied Optics. 1988. V. 27, No 14. P. 2960-2971.

9. Gan M.A. Optical systems with holographic and kinoform elements // **Proceedings** of SPIE. 1989. V. 1136. P. 150.

10. Gan M.A. Kinoforms long focal objectives for astronomy // Adaptive optics and optical structures / **Proceedings of the Meeting**, **European Congress on Optics**. 1990. P. 330–338.

11. Gan M.A, Potyemin I., Perveev A. High-speed apo-lens with kinoform element // **Proceedings** of SPIE. 1991. V. 1574. P. 243–249.

12. Грейсух Г.И., Ежов Е.Г., Казин С.В., Степанов С.А. Дифракционно-рефракционный корректор третичного спектра // Оптический Журнал. - 2010. - Т. 77, № 9. - С. 22-29.

13. http://www.ukaoptics.com/ccd.html.

14. Ежов, Е. Г. Высококачественный массовый микрообъектив для CCTV-систем // Современные технологии безопасности. – 2006. – № 3, 4. - С. 12-15.

15. Бронштейн И.Г., Зверев В.А., Лившиц И.Л., Kim Young-Gi, Kim Tae-Young, Young Phil-Ho. Выбор оптической схемы и расчет малогабаритных объективов для мобильных телефонов // Оптический Журнал. - 2009. - Т. 76, № 5. - С. 25-31.

16. Greisukh G.I., Bobrov S.T., Stepanov S.A. Optics of diffractive and gradient-index elements and systems. Bellingham: SPIE Press, 1997. 414 p.

17. Грейсух Г.И., Безус Е.А., Быков Д.А., Ежов Е.Г., Степанов С.А. Подавление спектральной селективности двухслойных рельефно-фазовых дифракционных структур // Опт. и спектр. 2009. Т. 106, вып. 4. С. 694-699.

Методика формирования падающей волны при разностном решении уравнения Даламбера

М.А. Астраханцева
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Дифракционная компьютерная оптика развивается более 25 лет. За прошедшие годы решены фундаментальные задачи фокусировки и селекции мод лазерного излучения [1], формирования бездифракционных пучков [1] и другие.

До недавнего времени, при моделировании распространения света через оптические элементы, исследователи ограничивались преимущественно геометрическим приближением [2]. Достижения технологии производства оптических элементов привели к такой степени их миниатюризации, что приближение геометрической оптики утратило адекватность. Моделирование работы таких элементов традиционно проводится в рамках строгой теории дифракции с привлечением аппарата вычислительной электродинамики.

Численные методы решения основных уравнений электродинамики известны достаточно давно [3]. Их можно разделить на три наиболее распространенных в компьютерной оптике: аналитические, RCWA и FDTD. Особое внимание в литературе уделяется последнему – методу решения уравнений Максвелла посредством конечных разностей во временной области, как наиболее общему. Появившись в середине прошлого века [3], численный метод решения уравнений Максвелла прошел несколько стадий развития.

Ранее всего были записаны разностные уравнения, обладающие высокими порядками аппроксимации исходной дифференциальной задачи по времени и пространству [4]. Одновременно с FDTD развивался и разностный метод решения волнового уравнения, основанный на аналогичных разностных схемах [5].

В настоящее время известно несколько эффективных граничных условий поглощения используемых вместе с методом FDTD. К ним можно отнести системы идеально согласованных слоев (PML) [6] и граничные условия Мура [7]. Для волнового уравнения в фундаментальной работе [8] использовались последние. Граничные условия PLM для волнового уравнения не применялись в силу того, что ко времени их создания (94-95гг.) интерес к разностному решению волнового уравнения применительно к задачам оптики был потерян.

Задача моделирования источника падающей волны, поставленная еще Yee в работе [4], решается с разной степенью точности во многих работах вплоть до настоящего времени. Первый способ задания падающей волны, позволяющий ограничить вычислительную область исследуемым объектом и его ближайшей окрестностью, был сформулирован в работе [9]. Более точный метод опубликован в 1980 году [10] в рамках TF/SF методики (Total-Field/Scattering-Field technique). К сожалению, автором настоящей работы было выяснено, что к волновому уравнению эта методика (точнее метод TF/RF как частный случай TF/SF представления) не применима. Видимо это и предопределило отсутствие интереса к разностному решению волнового уравнения. Настоящая работа посвящена преодолению указанной проблемы.

Численная форма методики TF/SF для волнового уравнения ранее не применялась, вследствие упомянутой приостановки исследований в данной области. Целью настоящей работы является перенос современных методик задания падающей

волны в методе разностного решения уравнений Максвелла на математической аппарат решения волнового уравнения.

Сравнивая разностное решение волнового уравнения и уравнений Максвелла отметим преимущества и недостатки первого подхода. К недостаткам можно отнести более узкую область применения разностного решения волнового уравнения. В частности оно актуально для решения задач дифракции ТЕ-волны на цилиндрическом оптическом элементе. В других случаях (случай ТМ-волны, либо сферического оптического элемента) имеет смысл использования FDTD метода. Вместе с тем оперируя одной сеточной функцией в одном уравнении (а не тремя функциями в трех уравнениях для двумерной задачи как в случае системы уравнений Максвелла) разностное решение волнового уравнения позволяет экономить оперативную память ЭВМ и характеризуется меньшей вычислительной сложностью.

Аналитическое задание падающей волны в рамках TF/RF – методики

В силу того, что разработанные способы формирования падающей волны являются частью методики и способствуют развитию TR/SF представления, автор сочла необходимым предварить основное изложение описанием этого представления. TF/RF – методика основывается на разделении исследуемой области на зоны результирующего (Total Field) и отраженного (Reflected Field) полей (рис. 1). Падающее поле считается известным и задается либо аналитически (как в настоящем параграфе) либо численно (как в следующем).

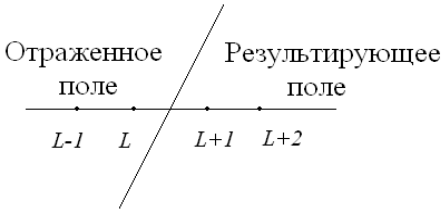


Рис.1. Фрагмент сеточной области

Для расчета полей во всей области традиционно применяется схема [8]:

$$\frac{E_i^{n+1}-2E_i^n+E_i^{n-1}}{h_t^2}=\frac{c^2}{\varepsilon_i\mu}\frac{E_{i-1}^n-2E_i^n+E_{i+1}^n}{h_y^2} \quad . \tag{1}$$

В одномерном случае на область вычислительного эксперимента D ($0\le y\le L_x$, $0<t\le T$) налагается сеточная структура Dh, в узлах которой $\{(t_m,y_i)$: $t_n=nh_t$, n=-1, 0, 1, .., $N=T/h_t$, $y_i=ih_y$, i=0, .., $I=L_y/h_y\}$ определены значения электрической компоненты E электромагнитного поля в случае распространения ТЕМ-волны. Среда и оптический элемент характеризуются диэлектрической и магнитной проницаемостями ε_i и μ, величина c соответствует скорости света в свободном пространстве.

Начальное условие соответствует отсутствию поля в области к моменту t=0. На границах располагаются электрические стенки.

Разделение результирующего и рассеянного полей в узле L (узел, делящий отраженное и прошедшее поля) обеспечивается следующей модификацией разностной схемы:

$$\begin{aligned}\frac{E_L^{n+1}-2E_L^n+E_L^{n-1}}{h_t^2}&=\frac{c}{\varepsilon_i\mu}\frac{E_{L-1}^n-2E_L^n+(E_{L+1}^n-\widetilde{E}_{L+1}^n)}{h_x^2};\\ \frac{E_{L+1}^{n+1}-2E_{L+1}^n+E_{L+1}^{n-1}}{h_t^2}&=\frac{c}{\varepsilon_i\mu}\frac{(E_L^n+\widetilde{E}_L^n)-2E_{L+1}^n+E_{L+2}^n}{h_x^2},\end{aligned} \tag{2}$$

На практике формула (2) используется при выводе метода TF/RF, реализация которого связана с вычислениями по формулам (1), (3).

$$E_L^{n+1} = E_L^{n+1} - \frac{ch_t^2}{\varepsilon\mu}\widetilde{E}_L^n,$$
$$E_{L+1}^{n+1} = E_{L+1}^{n+1} + \frac{ch_t^2}{\varepsilon\mu}\widetilde{E}_L^n. \qquad (3)$$

Знаком “ ˜ ” обозначена электрическая напряженность падающего поля, которая задается в (1) и (3) аналитически.

Исследуя погрешность вычислений по аналитической реализации метода TF/RF, организуем тестовый вычислительный эксперимент, связанный с распространением TEM-волны в вакууме. Разместим узел L посередине сеточной области. Результирующее поле будет формироваться справа от узла L, рассеянное соответственно слева. Справа от узла L на расстоянии одной длины волны будем изучать равномерную погрешность разностного решения. В данном эксперименте излучается цуг падающей волны в 10λ на области L_y=20λ, за это время формирования поле в выбранной подобласти можно считать устоявшимся с высокой точностью. Переход в частотную область завершает расчет поля. Параметры дискретизации сеточной области при представлении результатов в таблице 1 заданы в виде пары чисел Q и Qt. Первое соответствует количеству узлов сеточной области по пространству, приходящееся на длину волны, второе – количеству узлов по времени, приходящееся на временной интервал, за который плоский волновой фронт в вакууме пройдет расстояние в одну длину волны.

На Рисунке 2 и в Таблице 1 представлены результаты упомянутого эксперимента.

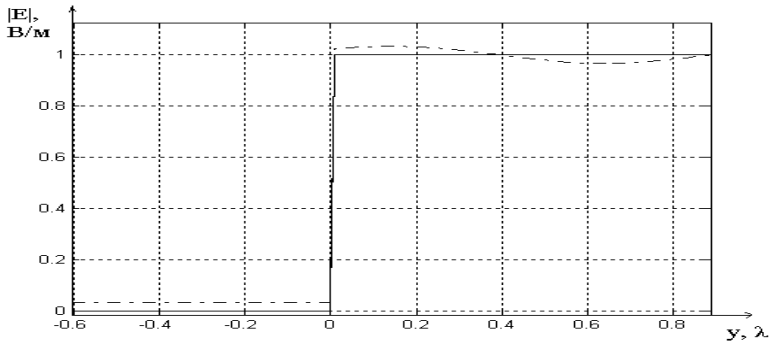


Рис.2. Графики модуля комплексной амплитуды при аналитическом (штрихпунктирная линия) и численном (сплошная линия) задании падающей волны методом TF/RF.

Показан фрагмент вычислительной области (-4λ< y <4λ) для дискретизации Q=100, Qt=200.

Таблица 1. Зависимость погрешностей (в %) вычислительных экспериментов от параметров дискретизации сеточной области

Q	Qt	TF/RF аналитический	TF/RF численный
10	24	29,290	0,202
10	52	27,026	0,143
10	100	24,851	0,240
10	200	23,454	0,253
25	52	13,734	0,051
25	100	11,257	0,040
25	200	10,223	0,037
50	100	6,803	0,012
50	200	5,747	0,013
100	200	3,369	0,005

Следует заключить, что аналитическое задание падающей волны методом TF/RF приводит к значительному росту погрешностей за счет аналитического задания разности фаз падающей волны в узлах L и L+1.

Численное задание падающей волны в рамках TF/RF – методики при распространении волны в однородной среде.

Исследуя погрешность вычислений по численной реализации метода TF/RF, организуем дополнительный вычислительный эксперимент, связанный с распространением TEM-волны в вакууме и “жестким” [8] источником излучения. Найденные в этой второй задаче значения электрического поля в узлах L, L+1 подставим в уравнения (3) первой задачи в качестве падающей волны.

Как видно из Таблицы 1 погрешность при использовании численной методики TF/RF формирования падающей волны снизилась на 2 (для разреженных сеток), 3 (для густых сеток) порядка. Более того во 2-ом случае погрешность разностного решения совпадала с таковой при использовании жесткого источника.

Аналогичное явление наблюдается и при сравнении указанных методик задания падающей волны применительно к FDTD методу [11]. В указанной работе повышение точности при переходе к численной методике задания связывается с компенсацией погрешностей разностного решения: замены производных разностными отношениями в основной и вспомогательной задачах. При аналитическом задании такой компенсации не происходит. Кроме того, в отличие от результатов из работы [11] аналитическое задание падающей волны при разностном решении волнового уравнения не представляет практического интереса в силу значительной величины погрешности даже на густых сеточных областях (3,37% для Q=100/Qt=200, и 0,091% для той же дискретизации в [11]).

Численное задание падающей волны в рамках TF/RF – методики при распространении волны в неоднородной среде.

Продолжая исследования численного задания падающей волны по методике TF/RF, автор поставила серию вычислительных экспериментов связанных с отражением электромагнитной волны от идеального зеркала и от границ раздела хлорид серебра/вакуум, вакуум/хлорид серебра (полагалось λ=1 мкм). Для случаев распространения волны при наличии зеркала и из вакуума в хлорид серебра, излучался цуг падающей волны в10λ, в случае вакуум/хлорид серебра - 20λ, (в данной среде волна имеет меньшую скорость распространения, чем вакууме и только за это время формирования поле в выбранной подобласти можно считать устоявшимся с высокой точностью). В Таблице 3 представлены результаты проведенных экспериментов.

Таблица 3. Зависимость погрешностей (в %) вычислительных экспериментов от параметров дискретизации сеточной области

Q	Qt	Зеркало	Прошедшая волна (вакуум / хлорид серебра)	Отраженная волна (вакуум / хлорид серебра)	Прошедшая волна (хлорид серебра / вакуум)	Отраженная волна (хлорид серебра / вакуум)
10	24	1,182	8,771	33,857	13,148	36,74
10	52	1,498	10,379	32,073	13,345	37,311
10	100	1,555	10,481	32,601	13,443	36,456
10	200	1,482	10,477	33,104	13,464	36,063
25	52	0,316	1,5754	4,537	1,726	4,605
25	100	0,336	1,2376	4,638	1,739	4,754
25	200	0,317	1,345	4,674	1,745	4,552
50	100	0,131	0,354	1,385	0,427	1,1862
50	200	0,136	0,409	1,313	0,428	1,1917
100	200	0,052	0,119	0,392	0,109	0,3816

Таким образом, представленная методика обеспечивает сходимость разностного решения волнового уравнения во всех случаях и может успешно применяться при моделировании дифракционных оптических элементов.

Заключение

Разработанный способ формирования падающей волны в рамках численной реализации TF/RF – методики при разностном решении волнового уравнения значительно превосходит аналитическую реализацию этой же методики по точности. Будучи распространенной на двумерный случай, заявленная методика позволит значительно снизить вычислительную сложность моделирования дифракции на элементах компьютерной оптики.

Литература

1. Дифракционная компьютерная оптика: под ред. Сойфера В.А.- М. Физматлит, 2007.- 736с
2. Борн М., Вольф Э. Основы оптики: Пер. с англ.- М.:Наука, 1973.- 720 с.
3. Kron G. Equivalent circuit of the field equations of Maxwell-I//Proc.IRE.1944.32.P.289.
4. Yee K.S. Numerical solution of initial boundaryvalue problems involving Maxwell's equations inisotropic media // IEEE Trans. Antennas Propag. AP-14, 1966. P. 302-307.
5. Самарский А.А. Теория разностных схем.- М.:Наука, 1989.-614 с.
6. Berenger J.-P. A perfectly matched layer for the absorption of electromagnetic waves// Journ, Computat. Phys. 1994. N 114. P. 185.
7. Mur G. Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations // IEEE Trans. Electromagn.Compat., 1981. — Vol.EMC-23, No. 4. — P. 377–382.
8. Taflove A., Hagness S. Computational Electrodynamics: The Finite-Difference Time-Domain Method: 3nd. ed. – Boston:Arthech House Publishers, 2005.- 852 p.
9. Taflove A., Brodwin M. Computation of the electromagnetic fields and induced temperatures with a model of the microwave irradiated human eye // IEEE Transactions of microwave theory and techniques, 1975.- vol. mtt-23, № 11.- p. 888–896.
10. Taflove A. Application of the finite-difference time-domain method to sinusoidal steady-state electromagnetic-penetration problems // IEEE Transactions of microwave theory and techniques, 1980.- vol. mtt-22, № 3.- p. 191–202
11. Головашкин Д.Л., Казанский Н.Л. Методика формирования падающей волны при разностном решении уравнений Максвелла (одномерный случай) // Автометрия, 2007.- Том 43, № 6, с. 78-85.

Резонансное усиление затухающих порядков дифракции в дифракционных спектральных отражающих фильтрах для наноразмерной фотолитографии

Е.А. Безус, Л.Л. Досколович, Н.Л. Казанский
Институт систем обработки изображений РАН, Самара
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Быстрое развитие нанотехнологий, имеющее место в последние годы, делает актуальной разработку методов изготовления наноструктур с наноразмерными деталями [1–3]. В настоящее время одним из основных методов создания наноструктур является проекционная фотолитография на основе лазеров с длиной волны, лежащей в дальнем ультрафиолетовом диапазоне. Такие системы могут обеспечить разрешение менее 50 нм, однако их стоимость и сложность очень высоки. В качестве альтернативы уменьшению длины волны были предложены различные варианты фотолитографии в ближнем поле, основанной на регистрации интерференционных картин затухающих [4–7] или поверхностных [8–15] электромагнитных волн, позволяющие избежать указанных выше недостатков и формировать структуры с высоким разрешением. В настоящей работе исследуется возможность формирования высокочастотных интерференционных картин затухающих электромагнитных волн с помощью дифракционных спектральных фильтров (в англоязычной литературе — guided-mode resonance filters). Фильтрующие свойства указанных структур интенсивно исследовались на протяжении последних лет многими научными группами [16–18], но относительно немного внимания было уделено распределению ближнего поля в таких структурах при резонансных условиях [19, 20].

Рассматриваемая структура представляет собой бинарную диэлектрическую дифракционную решетку, окруженную двумя диэлектрическими средами (рис. 1). Численное моделирование дифракции света на структуре на основе строгого метода Фурье-мод [21–23] проводилось для следующих параметров: λ=453 нм, нормальное падение, d=250*n нм, где дифракционные порядки с номерами $\pm n$ возбуждают моды структуры, ε_{super}=2.15 (SiO_2), ε_{gr1}=4.41 (ZnO), ε_{gr2}=2.56 (иммерсионная жидкость или фоторезист), ε_{sub}=2.56 (фоторезист).

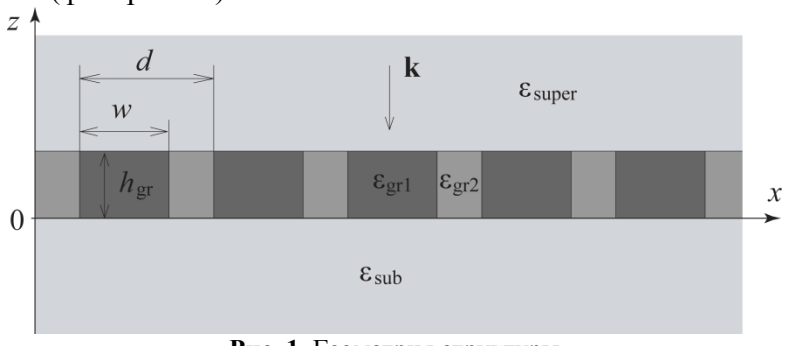


Рис. 1. Геометрия структуры

Значения геометрических параметров структуры w и h_{gr} были определены из оптимизационной процедуры. В качестве целевой функции была выбрана мера близости расчетной интерференционной картины к «идеальной» интерференционной картине, получающейся при интерференции двух затухающих волн. На рисунке 2(а) показано распределение электрического поля в структуре, мода в которой возбуждается при ТМ-поляризации падающей волны и обладающей следующими параметрами: n=3, d=750 нм, h_{gr}=141 нм, w=675 нм. Интенсивность падающей волны полагалась равной

единице. Интенсивность сформированной под структурой интерференционной картины на различных расстояниях от границы раздела решетка/фоторезист показана на рисунке 2(б). Период сформированной картины составляет 125 нм (в 6 раз меньше, чем период решетки), таким образом, может быть достигнут размер деталей порядка 50–60 нм (более чем в 7 раз меньше длины волны падающего света).

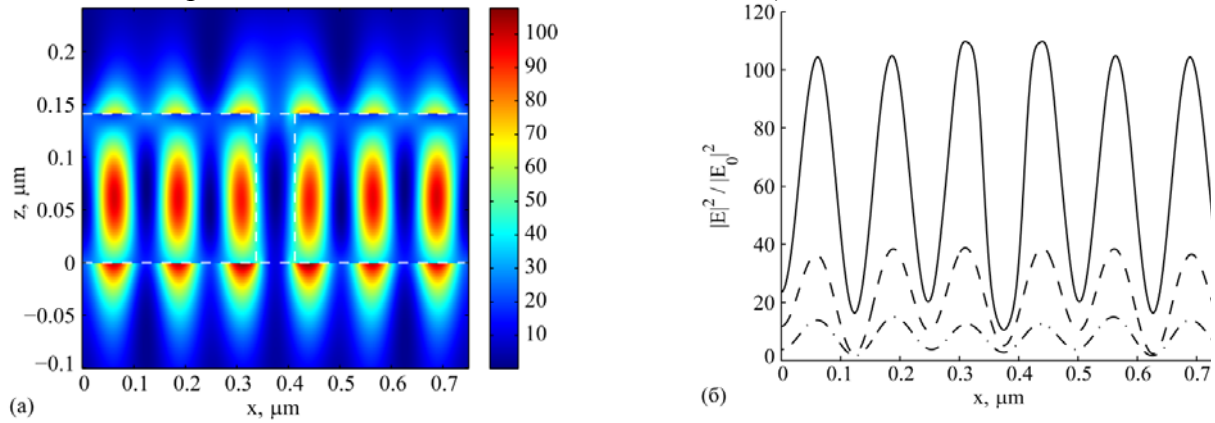


Рис. 2. Распределение интенсивности электрического поля в рассматриваемой структуре (1 период решетки, решетка показана пунктирными линиями) (а), интенсивность сформированной интерференционной картины на различных расстояниях от границы раздела на одном периоде решетки (0 нм — сплошная линия, 50 нм — пунктирная линия, 100 нм — штрих-пунктирная линия) (б)

Из рисунка 2 видно, что интенсивность сформированной под структурой интерференционной картины сравнима с интенсивностью, достигаемой в плазмонной интерференционной литографии [11, 12]. Более того, сформированная картина обладает высоким контрастом: при z=0 (на границе раздела решетка/подложка) контраст составляет 0.68 и медленно уменьшается при удалении от границы. При z=–100 нм минимальный локальный контраст составляет примерно 0.54, что существенно превышает минимальное значение, необходимое для литографии с использованием распространенных негативных фоторезистов (около 0.2).

Следует сделать несколько дополнительных замечаний, касающихся качества интерференционной картины. Если мода структуры возбуждается ±1 дифракционными порядками, пропускание может быть очень мало (менее 0.01), и контраст будет очень слабо зависеть от расстояния от границы. Наоборот, возможно использование высших порядков дифракции (например, порядков с номерами ±5), однако, однородность картины и глубина проникновения в резист будут при этом уменьшаться, что влечет за собой необходимость использования тонких слоев резиста и ограничивает высоту изготавливаемых структур.

В заключение, в настоящей работе с помощью моделирования в рамках строгой электромагнитной теории дифракции показано, что при резонансных условиях амплитуды затухающих прошедших дифракционных порядков с константой распространения, близкой к константе распространения моды рассматриваемой структуры, могут значительно усиливаться, при этом будет формироваться контрастная интерференционная картина с наноразмерными деталями. Показано, что для формирования картин могут быть использованы высшие порядки дифракции, период сформированной картины при этом в 6–10 раз меньше периода используемой дифракционной решетки. Предлагаемый подход может быть применен для создания периодических структур с наноразмерными деталями по методу контактной фотолитографии в ближнем поле.

Работа была выполнена при поддержке грантов РФФИ № 09-07-92421, 09-07-12147 и 10-07-00553, гранта Президента РФ поддержки ведущих научных школ НШ-7414.2010.9 и российско-американской программы «Фундаментальные исследования и высшее образование» (грант № PG08-014-1).

Литература

1. Bates A.K., Rothschild M., Bloomstein T.M., Fedynyshyn T.H., Kunz R.R., Liberman V., Switkes M., IBM J. Res. Dev. 45 (5) (2001) 605–614.
2. Gwyn C.W., Stulen R., Sweeney D., Attwood D., J. Vac. Sci. Technol. B 16 (6) (1998) 3142–3149.
3. Silverman J. P., Vac J.. Sci. Technol. B 16 (6) (1998) 3137–3141.
4. Blaikie R.J., McNab S.J., Appl. Opt. 40 (10) (2001) 1692–1698.
5. Martinez-Anton J.C., Opt J.. A.: Pure Appl. Opt. 8 (4) (2006) 213–218.
6. Chua J.K., Murukeshan V.M., Tan S.K., Lin Q.Y., Opt. Express 15 (6) (2007) 3437–3451.
7. Murukeshan V.M., Chua J.K., Tan S.K., Lin Q.Y., Opt. Express 16 (18) (2008) 13857–13870.
8. Luo X., Ishihara T., Appl. Phys. Lett. 84 (23) (2004) 4780–4782.
9. Srituravanich W., Fang N., Sun C., Luo Q., Zhang X., Nano Lett. 4 (6) (2004) 1085–1088.
10. Jiao X., Wang P., Zhang D., Tang L., Xie J., Ming H., Opt. Express 14 (11) (2006) 4850–4860.
11. Bezus E.A., Bykov D.A., Doskolovich L.L., Kadomin I.I., Opt J.. A: Pure Appl. Opt. 10 (9) (2008) 095204.
12. Bezus E.A., Doskolovich L.L., Opt. Commun. 283 (10) (2010) 2020–2025.
13. Xiong Y., Liu Z., Zhang X., Appl. Phys. Lett. 93 (11) (2008) 111116.
14. Yang X., Zeng B., Wang C., Luo X., Opt. Express 17 (24) (2009) 21560–21565.
15. Guo X., J. Du, Y. Guo, J. Yao, Opt. Lett. 31 (17) (2006) 2613–2615.
16. Brundrett D.L., E.N. Glytsis, T.K. Gaylord, J.M. Bendickson, J. Opt. Soc. Am. A 17 (7) (2000) 1221–1230.
17. Magnusson R., D. Shin, Z.S. Liu, Opt. Lett. 23 (8) (1998) 612–614.
18. Tamir T., S. Zhang, J. Opt. Soc. Am. A 14 (7) (1997) 1607–1616.
19. Wei C., S. Liu, D. Deng, J. Shen, J. Shao, Z. Fan, Opt. Lett. 31 (9) (2006) 1223–1225.
20. Sun T., Ma J., Wang J., Jin Y., He H., Shao J., Fan Z., J. Opt. A: Pure Appl. Opt. 10 (12) (2008) 125003.
21. Moharam M.G., Pommet D.A., Grann E.B., Gaylord T.K., J. Opt. Soc. Am. A. 12 (5) (1995) 1077–1086.
22. Moharam M. G., Grann E.B., Pommet D.A., Gaylord T.K., J. Opt. Soc. Am. A. 12 (5) (1995) 1068–1076.
23. Li L., J. Opt. Soc. Am. A. 13 (9) (1996) 1870–1876.

Расчет вторичной оптики светодиодного прожектора для формирования заданного распределения освещенности

А. А. Белоусов, М.А. Моисеев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара
Институт систем обработки изображений РАН, Самара

Введение

Использование светоизлучающих диодов (СИД) становится все более масштабным. Их применяют в автомобильной светотехнике, системах подсветки жидкокристаллических мониторов, системах освещения и т.д. Это связано с увеличением их светового потока и уменьшением стоимости. СИД компактны, устойчивы к вибрациям и механическим ударам, имеют большой срок службы и большую эффективность. К 2025 году использование СИД в США позволит сократить потребление электроэнергии на освещение на 29% и сэкономить 125 биллионов долларов. В то же время, развитие светотехнических устройств на основе СИД тормозит отсутствие эффективных методов расчета, позволяющих проектировать оптические элементы для формирования заданных распределений освещенности с высокой световой эффективностью.

Авторами был разработан метод, представленный в работе [1], позволяющий рассчитывать оптические поверхности для формирования заданных распределений освещенности с большими угловыми размерами. Метод решает задачу в общем непараксиальном приближении, с учетом френелевских потерь и протяженности источника излучения. Метод позволяет получать поверхности сложной формы (freeform в англоязычной литературе) формирующие заданные распределения освещенности с высокой точностью и эффективностью.

При помощи разработанного метода [1] в настоящей работе был проведен расчет вторичной оптики светодиодного прожектора. Данное светотехнической устройство позволяет равномерно освещать заданные области большой площади, при малом размере самого прожектора.

Постановка задачи

Геометрия задачи представлена на рис. 1. Источник излучения должен равномерно освещать прямоугольную область с размерами 30x14 м, находящуюся на расстоянии 12 м от источника под углом 15° к нормали поверхности источника. Источник смещен относительно центра требуемой области освещения на 3.5 м по осиY.

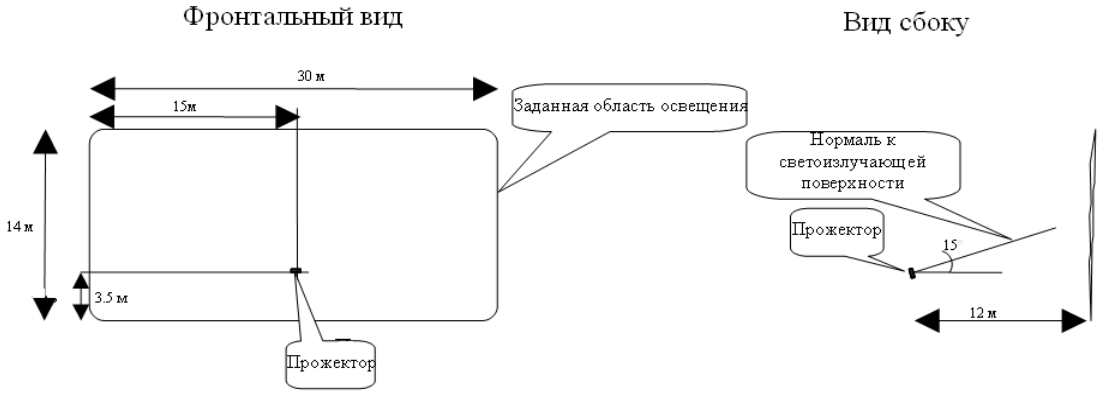


Рис.1. Схема расположения светодиодного прожектора и освещаемой области

В качестве источника излучения предлагается использовать светодиодный прожектор. Светодиодный планарный прожектор состоит из набора СИД установленных в одной плоскости. Каждый СИД снабжен преломляющим оптическим элементом (вторичной оптикой), который формирует требуемое распределение освещенности.

Необходимо рассчитать оптический преломляющий элемент из условия формирования требуемого распределения освещенности в заданной области от протяженного источника излучения. Форма этого элемент будет одинакова для всех СИД планарного прожектора.

Результаты расчета

Оптический элемент был получен при помощи метода расчета преломляющей поверхности для формирования заданного распределения освещенности [1]. Данный метод позволяет рассчитать оптический элемент, формирующий заданное распределение освещенности с большими угловыми размерами при протяженном источнике излучения. Метод основан на градиентной оптимизации преломляющей поверхности, представленной в виде бикубического сплайна в сферических координатах. Метод использует непараксиальное приближение и учитывает френелевские потери при преломлении на границе оптического элемента.

Данная оптическая поверхность была рассчитана для СИД Cree XLamp XP-G White [2] (рис. 3). Световой поток данного СИД составляет 92 люмен/Ватт, размер излучающего тела 1,5×1,5 мм. На рис. 4 представлена диаграмма направленности излучения такого источника. Расположение рассчитанного преломляющего элемента и СИД показано на рис. 5.

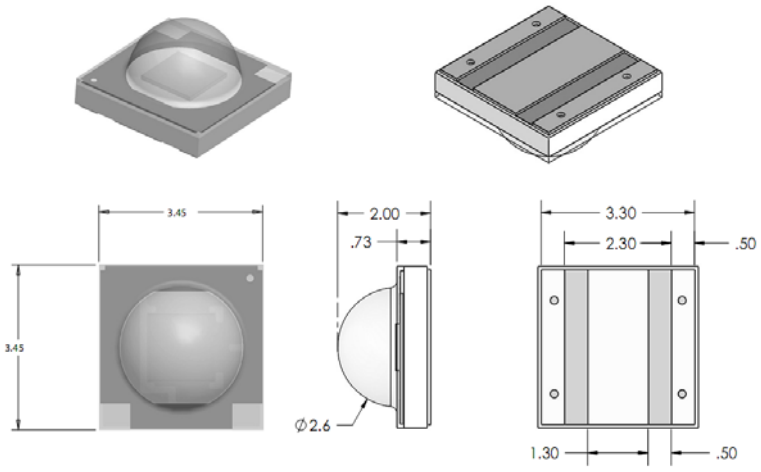


Рис. 3. СИД Cree XLamp XP-G White (размеры приведены в мм)

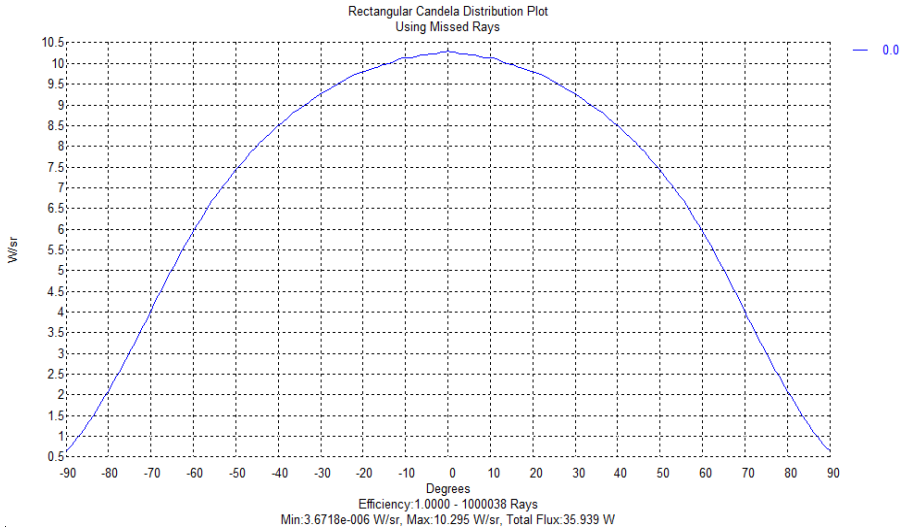


Рис. 4. Диаграмма направленности излучения СИД

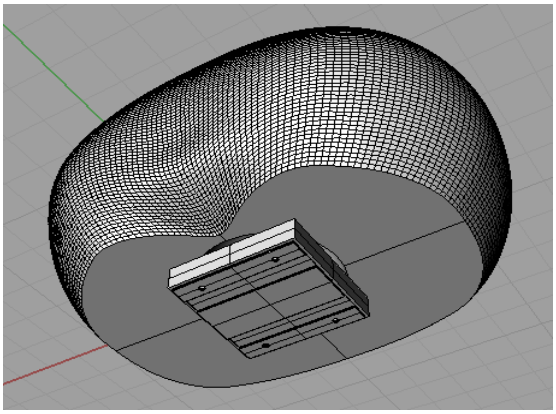


Рис. 5. Расположение СИД относительно рассчитанного оптического элемента

На рис. 6 изображено распределение освещенности, формируемое рассчитанным оптическим элементом на выходной плоскости. Энергетическая эффективность элемента составляет 75,4 %, а неравномерность освещенности – менее 9,9 %. Распределение освещенности на рис. 6 получено с помощью специализированного программного обеспечения TracePro® [3]. Необходимо отметить, что TracePro® не позволяет решать рассматриваемые задачи расчета оптических элементов и используется в работе только для моделирования формируемого распределения освещенности.

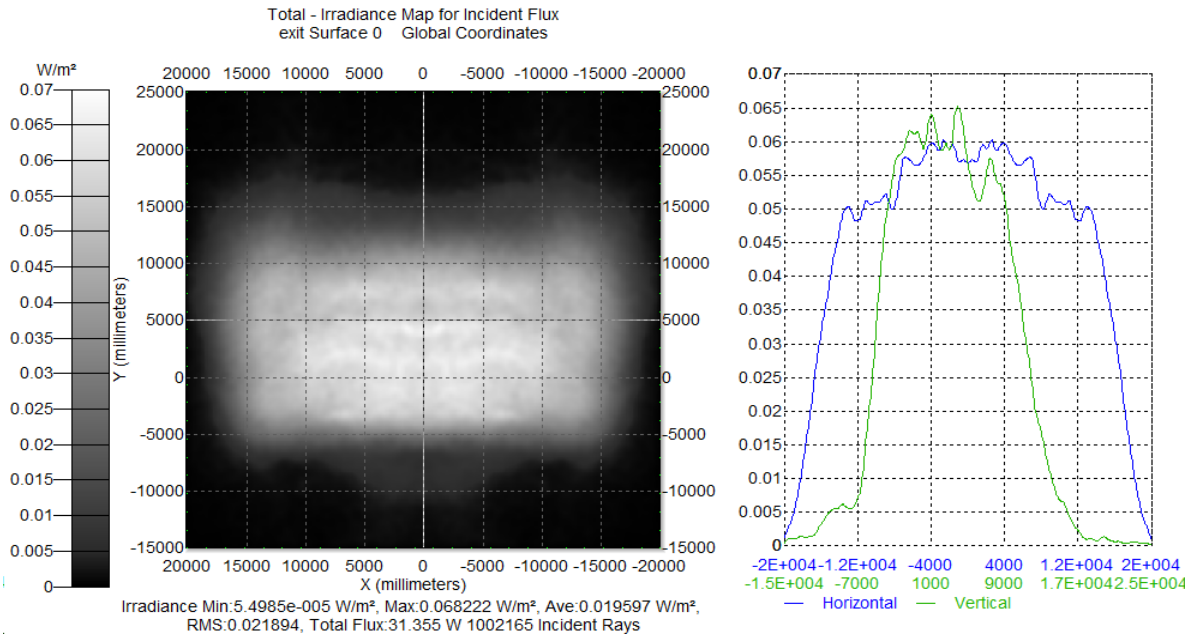


Рис. 6. Распределение освещенности формируемое рассчитанным оптическим элементом

Заключение

Метод расчета преломляющих поверхностей, представленный в работе [1], использован для расчета вторичной оптики светодиодного прожектора, формирующего равномерное распределение освещенности в прямоугольной области 30x14 м. В качестве источника излучения был использован светоизлучающий диод Cree XLamp XP-G White. Энергетическая эффективность рассчитанного оптического элемента превосходит 75 %, ОСКО формируемого распределения освещенности от заданного – менее 10 %.

Литература

Моисеев М.А. Расчет преломляющего оптического элемента для формирования заданного распределения освещенности при протяженном источнике излучения / М.А. Моисеев, Л.Л. Досколович // Компьютерная оптика, 2009. – Т. 34, № 2. – С. 194–202. – ISSN 0134-2452.

http://www.cree.com/products/xlamp_xpg.asp

http://lambdares.com/software_products/tracepro/

Изучение возможности применения хрома в качестве маскирующего слоя при глубоком травлении кремния

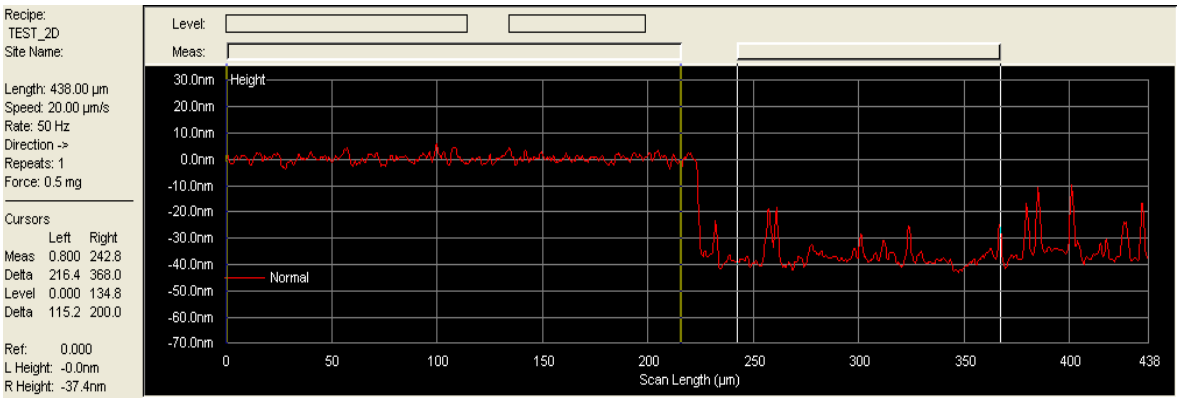
Н.А. Болобанов, А.В. Волков, Г.Ф. Костюк, С.Д. Полетаев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Современные технологии создания микро и нано структур часто требуют формирования элементов малой размерности. Для этой задачи хорошо подходит технология ионно-химического травления. Процесс ионно-химического травления в вакууме обеспечивает качественно высокий уровень решения многих технологических задач: создание масок в тонких плёнках, травление различных материалов с высокой разрешающей способностью. В данной работе исследуются возможности ионно-химического травления плёнок хрома. Этот материал выбран из-за высоких адгезивных свойств и отработанности технологии напыления плёнок хрома. Хром широко применяется в качестве материала для маскирующего слоя, поэтому данные о плазмостойкости плёнок хрома в средах различного состава позволяют рассчитать время за которое хромовая маска разрушится в тех или иных условиях. Маска, с помощью которой создается структура травления, является одним из основных элементов, определяющих производительность и качество процессов ионного травления. Основными требованиями, предъявляемыми к маскам, являются стойкость к тепловым нагрузкам, вызванными ионно-плазменной обработкой, низкая скорость травления. Особенно интересно исследовать возможность использования хрома в качестве маскирующего слоя для глубокого травления кремния в элегазе с целью создания ДОЭ для инфракрасного излучения. Целью данной работы является сравнительный анализ скоростей травления хрома в следующих средах: элегаз(SF_6), четырёххлористый углерод, хладон-14(CF_4).

Постановка эксперимента

Исследования проводились на установке плазмохимического травления УТП ЦДЭ-125-009 («Интра»). Плёнки напылялись на установке магратронного напыления «Каролина Д12-А». Плёнки хрома напылялись на ситаловые подложки в одну загрузку установки магратронного напыления «Каролина». Толщина напылённых плёнок



измерялась на профилографе-профилометре «KLA Tencor P-16+», она составила 38 нм.

Рис. 1 - Профилограмма границы плёнки Cr

Все эксперименты проводились в одинаковых режимах: ВЧ мощность, поступающая в реактор составляла 580 Вт, предварительный вакуум 0,01 Па,

напряжение смещения на столике с креплением для подложки, 350 Вольт, ток электромагнита 0,9 Ампер.

Время травления определялось следующим образом: первый раз определялось примерное время стравливания плёнки путём дотравливания и визуального контроля наличия плёнки через смотровое окно реактора, после этого время травления подбиралось так чтобы за один раз сошла вся плёнка хрома. Дело в том что процесс травления начинается не сразу, поэтому чтобы время установления стационарного процесса оказывало наименьшее влияние на результативное время необходимо проводить процесс ИХТ в один цикл. Также для предотвращения перегрева подложки столик имеет жидкостное охлаждение.

Полученные результаты

Измеренные скорости травления хрома в различных средах представлены в таблице:

среда	$d_{Cr}, \mathring{A}$	$t_{mp}, мин.$	$V_{mp}, \mathring{A}/мин$
CCl_4	380	5,07	75
CF_4 (хладон-14)	380	12,6	30
SF_6 (элегаз)	380	19	20

Для оценки возможности травления кремния в атмосфере элегаза был проведён эксперимент по определению скорости травления кремниевых пластин ЕТО 035124ТУ марки КДБ 4,5 ориентация кристаллической решётки (100). По результатам эксперимента скорость травления кремния составила $V_{mp}= 7200\,\mathring{A}/мин$, что означает что селективность при глубоком травлении кремния под хромовой маской будет составлять $N= 7200/20= 360$.

Вывод

Применение хрома в качестве маскирующего слоя при глубоком (10-15 мкм) травлении кремния в элегазе, позволяет добиться высокой плазмостойкости маски и как следствие получения структур высокого качества. Также перспективным является применение хромовых масок при травлении различных материалов в CF_4, ввиду высокой плазмостойкости хрома в этой среде. Такая ситуация с плазмостойкостью хрома в атмосфере содержащей фтор может быть объяснена тем что фторид хрома не летуч и удаления продуктов реакции происходит вследствие ионного распыления что замедляет разрушение плёнки хрома.

Литература

1. Ивановский Г.Ф., Петров В.И., «Ионно-плазменная обработка материалов», М:, «Радио и связь», 1986, 232 с.

2. Данилин Б.С., «Вакуумное нанесение тонких плёнок», М:, «Энергия», 1967, 312 с.

3. Блохин В. Г. Лекции «Технологии производства микроэлектронной аппаратуры», М:, РГТУ-МАТИ им. Циолковского, 2004-2005.

4. Boswell R.W. Very efficient plasma generation by whistler waves near lower hybrid frequency. – Plasma Phys. and Contr. Fusion, vol. 26, No. 10.

Магнитооптические свойства периодических дифракционных структур, намагниченных в плоскости. Эффект резонансного изменения фазы

Д.А. Быков, Л.Л. Досколович
Институт систем обработки изображений РАН, Самара
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Изучение взаимодействия света с периодическими дифракционными структурами является предметом интенсивных исследований. В таких структурах существует широкий круг экстраординарных (резонансных) оптических эффектов, включающих экстраординарное пропускание, полное поглощение падающего излучения, резонансное изменение спектров прохождения и отражения на определённых частотах, формирование областей с высокой степенью локализации энергии вследствие интерференции затухающих и плазмонных волн [1–2].

Особый интерес представляют дифракционные структуры, содержащие намагниченные материалы. В таких структурах возникают новые, магнитооптические эффекты, связанные с преобразованием поляризации света [3–5] и резонансным изменением спектров отражения и пропускания при изменении намагниченности [6–8]. В настоящей работе исследуются магнитооптические эффекты, заключающиеся в резонансном изменении фазы порядка дифракции. Величина соответствующего магнитооптического эффекта характеризуется величиной.

$$\Delta\varphi(g,\omega) = \arg A(g,\omega) - \arg A(0,\omega), \qquad (1)$$

где $A(g,\omega)$ — комплексная амплитуда прошедшей (отражённой) волны при частоте ω и намагниченности материала структуры, характеризуемой модулем вектора гирации среды g.

В данной работе фазовые МО эффекты (1) исследуются для периодической дифракционной структуры. Структура состоит из дифракционной решётки и однородного слоя (рис. 1).

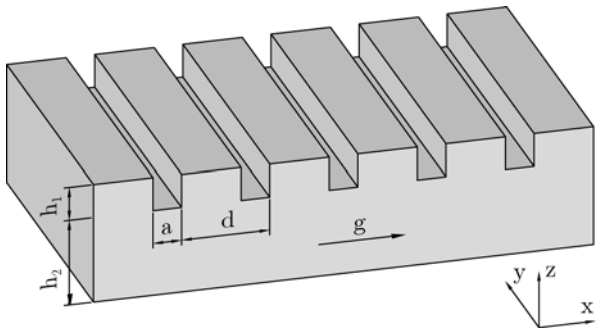


Рис. 1. Геометрия рассматриваемой структуры (период $d = 2200$ нм, высота решётки $h_1 = 700$ нм, ширина отверстия решётки $r = 1370$ нм, высота слоя $h_2 = 6430$ нм)

Вектор намагниченности материала структуры направлен перпендикулярно щелям решётки. В этом случае тензор диэлектрической проницаемости материала структуры имеет следующий вид [9]:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon & 0 & 0 \\ 0 & \varepsilon & ig \\ 0 & -ig & \varepsilon \end{bmatrix}, \qquad (2)$$

Структура предполагается выполненной из распространённого материала магнитооптики $Bi_{2.2}Dy_{0.8}Fe_5O_{12}$ (висмут-замещённый диспрозиевый феррит гранат) [9]. В области ближнего ИК (в окрестности $\lambda = 1200\,\text{нм}$, $\omega \approx 1{,}57 \cdot 10^{15}\,\text{Гц}$), параметры тензора (2) для данного материала принимают следующие значения: $\varepsilon = 5{,}06 + 4{,}3 \cdot 10^{-4} i$, $g = 0{,}015 - 3 \cdot 10^{-5} i$.

Теоретически показано, что в общем случае фаза порядков дифракции резонансно меняется одновременно с интенсивностью порядков. Однако, при надлежащем выборе параметров резонанса становится возможным управлять только фазой порядка дифракции при практически постоянной его интенсивности [10, 11].

На рис. 2*а* представлена зависимость интенсивности нулевого прошедшего порядка дифракции от величины намагниченности структуры. На рис. 2*б* представлена зависимость фазы порядка дифракции от величины намагниченности структуры.

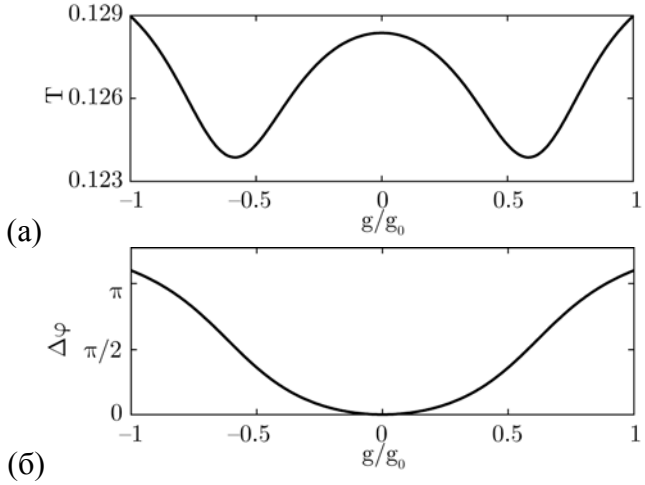


Рис. 2. Зависимость интенсивности (а) и фазы (б) нулевого прошедшего порядка дифракции от величины намагниченности при нормальном падении ТМ-поляризованной волны при частоте $\omega = 1.57 \cdot 10^{15}\,\text{Гц}$

Согласно рис. 2, фаза нулевого порядка изменяется в диапазоне $[0;\pi]$ при фактически постоянном пропускании $T \approx 0.126$ (величина изменения интенсивности 0-го порядка при намагничивании $\Delta T < 0.005$). Приведенные значения изменения фазы и постоянное пропускание были получены за счет оптимизации по значениям геометрических параметров структуры d, r, h_1, h_2 (см. рис. 1). Отметим, что полученная величина магнитооптического эффекта на рис. 2б для рассматриваемой дифракционной структуры более чем в 300 раз превышает аналогичный эффект для однородного намагниченного слоя [12].

Показано соответствие теоретического описания рассматриваемых резонансов с результатами моделирования в рамках электромагнитной теории. На основе расчёта резонансов структуры по методу матрицы рассеяния [13–15] показано, что резонансы в рассматриваемой структуре соответствуют возбуждению её собственных мод [10, 11].

Работа выполнена при поддержке грантов РФФИ № 09-07-12147, 10-02-01391, 09-07.92421 и 10-07-00553, гранта президента РФ поддержки ведущих научных школ НШ-7414.2010.9 и российско-американской программы «Фундаментальные исследования и высшее образование» (грант BRHE PG08-014-1).

Литература

1. Ebessen, T. W. Extraordinary optical transmission through sub-wavelength hole arrays / T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, P. A. Wolff // Nature. — 1998. — Vol. 391, N. 6668. — P. 667–669.

2. Bonod, N. Total absorption of light by lamellar metallic gratings / N. Bonod, G. Tayeb, D. Maystre, S. Enoch, E. Popov // Opt. Express. — 2008. — Vol. 16, N. 20. — P. 15431–38.
3. Belotelov, V. I. Magnetooptical effects in the metal-dielectric gratings / V. I. Belotelov, L. L. Doskolovich, V. A. Kotov, E. A. Bezus, D. A. Bykov, A. K. Zvezdin // Optics Communications. — 2007. — Vol. 278, N. 1. — P. 104–109.
4. Bai, B. Polarization conversion in resonant magneto-optic gratings / B. Bai, J. Tervo, J. Turunen // New Journal of Physics. — 2006. — Vol. 8, N. 9. — P. 205.
5. Khanikaev, A. B. Anomalous faraday effect of a system with extraordinary optical transmittance / A. B. Khanikaev, A. V. Baryshev, A. A. Fedyanin, A. B. Granovsky, M. Inoue // Opt. Express. — 2007. — Vol. 15, N. 11. — P. 6612–6622.
6. Белотелов, В. И. Оптические свойства перфорированных металлодиэлектрических гетероструктур, намагниченных в плоскости / В. И. Белотелов, Д. А. Быков, Л. Л. Досколович, А. Н. Калиш, А. К. Звездин // Физика твердого тела. — 2009. — Т. 51, № 8. — С. 1562–1567.
7. Belotelov, V. I. Giant magneto-optical orientational effect in plasmonic heterostructures / V. I. Belotelov, D. A. Bykov, L. L. Doskolovich, A. N. Kalish, V. A. Kotov, A. K. Zvezdin // Optics Letters. — 2009. — Vol. 34, N. 4. — P. 398–400.
8. Быков Д. А Исследование магнитооптических резонансов в диэлектрических дифракционных решётках на основе анализа дисперсии и структуры собственных мод / Д. А. Быков, Л. Л. Досколович, В. А. Сойфер // Компьютерная оптика. — 2009. — Т. 33, № 4, С. 384–392
9. Zvezdin A. K., Kotov V. A. Modern Magnetooptics and Magnetooptical Materials. Bristol and Philadelphia. IOP Publishing, 1997. 386 p.
10. Быков, Д. А. Экстраординарный магнитооптический эффект изменения фазы прошедшей волны в периодических дифракционных структурах / Д. А. Быков, Л. Л. Досколович // Письма в ЖТФ. — 2010. — Т. 36, № 13. — С. 22–30.
11. Быков Д. А. Экстарординарныый магнитооптический эффект изменения фазы дифракционных порядков в диэлектрических дифракционных решётках / Д. А. Быков, Л. Л. Досколович, В. А. Сойфер, Н. Л. Казанский // ЖЭТФ. — 2010, — принято к печати.
12. Penfold C. Transverse magneto-optical Kerr effect: the phase change of reflected light / C. Penfold, R. T. Collins, A. P. B. Tufaile, Y. Souche // JMMM. — 2002. — Vol. 242-245. — P. 964–966.
13. Tikhodeev, S. G. Quasiguided modes and optical properties of photonic crystal slabs / S. G. Tikhodeev, A. L. Yablonskii, E. A. Muljarov, N. A. Gippius, T. Ishihara // Phys. Rev. B. — 2002. — Jul. — Vol. 66, N. 4. — P. 045102.
14. Gippius, N. A. Optical properties of photonic crystal slabs with an asymmetrical unit cell / N. A. Gippius, S. G. Tikhodeev, T. Ishihara // Phys. Rev. B. — 2005. — Jul. — Vol. 72, N. 4. — P. 045138.
15. Гиппиус, Н. А. Применение метода матрицы рассеяния для расчёта оптических свойств метаматериалов / Н. А. Гиппиус, С. Г. Тиходеев // Успехи физических наук. — 2009. — Т. 179, № 9. — С. 1027–1030.
16. FDTD solver компании Acceleware http://www.acceleware.com/fdtd-solvers
17. Головашкин Д.Л., Казанский Н.Л. Методика формирования падающей волны при разностном решении уравнений Максвелла. Одномерный случай // Автометрия, 2006, том 42, № 6, с.78-85.
18. Головашкин Д.Л., Казанский Н.Л. Методика формирования падающей волны при разностном решении уравнений Максвелла. Двумерный случай // Автометрия, 2007, том 43, № 6, с.78-88.
19. Страница пакета W2405 Agilent FDTD Simulation Element компании Agilent http://www.home.agilent.com/agilent/product.jspx
20. Страница пакета Concerto компании Cobham Technical Sevices http://www.vectorfields.com/concerto.php
21. Страница пакета SEMCAD X OPTICS компании Speag http://www.speag.com/products/semcad/solutions/optics/
22. Страница пакета Xfdtd компании RemCom http://www.remcom.com/xf7
23. Страница пакета FastFDTD http://www.emphotonics.com/products/fastfdtd
24. СуперЭВМ и «Нано»// Нанотехнологическое сообщество “Нанометр” www.nanometer.ru/2007/09/12/super_4233.html
25. Головашкин Д.Л., Казанский Н.Л. Декомпозиция сеточной области при разностном решении уравнений Максвелла // Математическое моделирование, 2007.- Том 19, №2.- с. 48-58.

Моделирование распространения лазерных пучков с фазовой сингулярностью в оптических волноводах

С.А. Дегтярев, С.Н. Хонина
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Вихревым лазерным пучкам посвящены многочисленные исследования и публикации как российских учёных-оптиков, так и их зарубежных коллег. В настоящее время изучение свойств лазерных пучков с фазовой сингулярностью выделено в отдельный раздел оптики, называемый «сингулярная оптика».

Сингулярные особенности в световых полях могут появляться при их прохождении через случайно-неоднородные и нелинейные среды. Также возможно возбуждение вихревых полей в лазерных резонаторах и многомодовых волоконных световодах.

Распространение волн с различными начальными распределениями, имеющих фазовую сингулярность в планарном волноводе

Рассмотрим модель плоского волновода, описываемого двумя пространственными координатами x, y.

Планарный волновод (рисунок 1) представляет собой плоский слой, ограниченный двумя параллельными отражающими стенками.

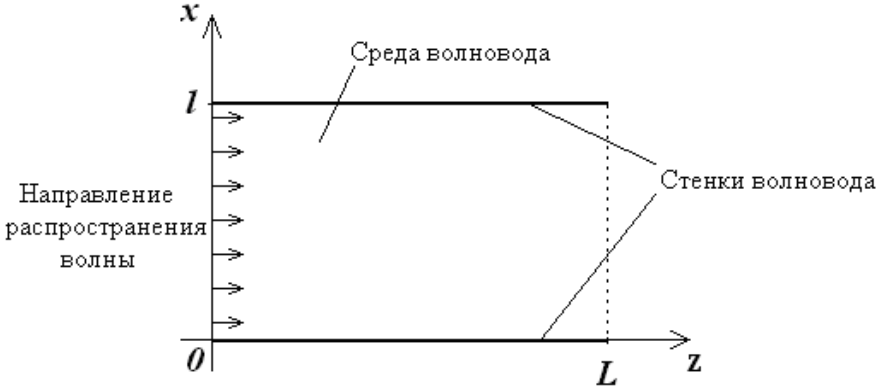


Рис. 1. Планарный волновод

Среда между стенками волновода оптически однородна.

Уравнение распространения световой волны в планарном волноводе имеет вид:

$$\frac{\partial u}{\partial z} - \frac{\mathrm{i}}{2kn}\frac{\partial^2 u}{\partial x^2} = 0, \qquad (2)$$

где $0 \le x \le l,\ 0 \le z \le L$,

значения параметров выберем следующими: $l = 30 мкм,\ L = 200 мкм,\ \lambda = 2 мкм,\ n = 1{,}5$

Для численного моделирования процесса распространения электромагнитных волн в волноводе был использован программный пакет BeamProp. Этот пакет использует неявную конечно-разностную схему Кранка-Николсона. Граничные условия аппроксимируют реальный процесс отражения от границы раздела двух сред и рассчитываются по эвристическому алгоритму.

Пусть начальное распределение имеет вид:

$$\mathrm{Re}\left(u\big|_{z=0}\right)=\sin\left(\frac{3\pi x}{l}\right),$$

$$\mathrm{Im}\left(u\big|_{z=0}\right)=0.$$

Такое начальное распределение соответствует третьей моде идеального планарного волновода с электрическими стенками.

График зависимости амплитуды и фазы напряженности электрического поля от координат волновода изображен на рисунке 2.

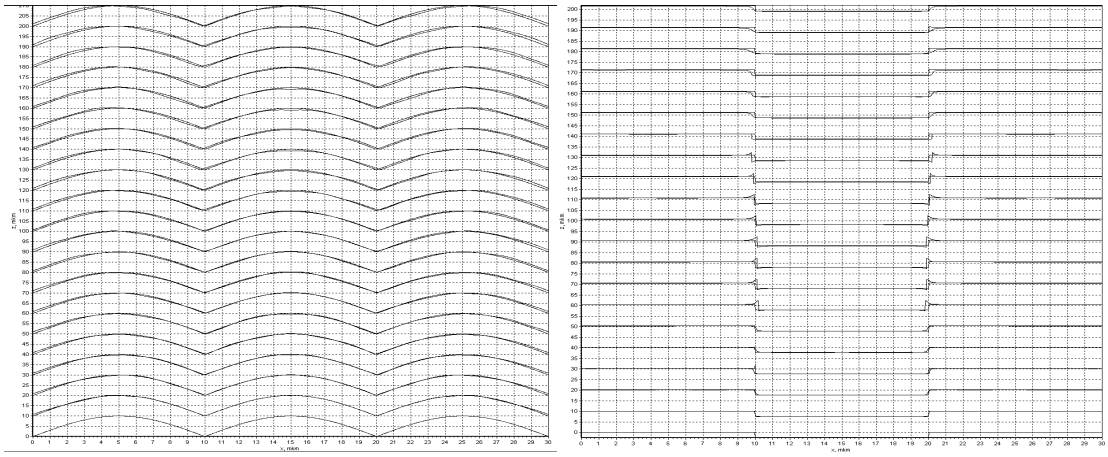


Рис. 2. – Амплитуда (слева) и фаза (справа) при начальном распределении – третьей моде волновода

На рисунке 2 видим, что с ростом z в поперечных сечениях волновода распределение амплитуды и фазы комплексной амплитуды напряженности почти не изменяются.

Пусть на входе в волновод имеется распределение напряженности электрического поля, соответствующее первой моде Гаусса-Эрмита:

$$\mathrm{Re}\left(u\big|_{z=0}\right)=\left(x-\frac{l}{2}\right)\exp\left(-\frac{\left(x-\frac{l}{2}\right)^2}{\left(\frac{l}{m}\right)^2}\right),\qquad \mathrm{Im}\left(u\big|_{z=0}\right)=0.$$

Амплитуда и фаза напряженности электрического поля приведена на рисунке 3

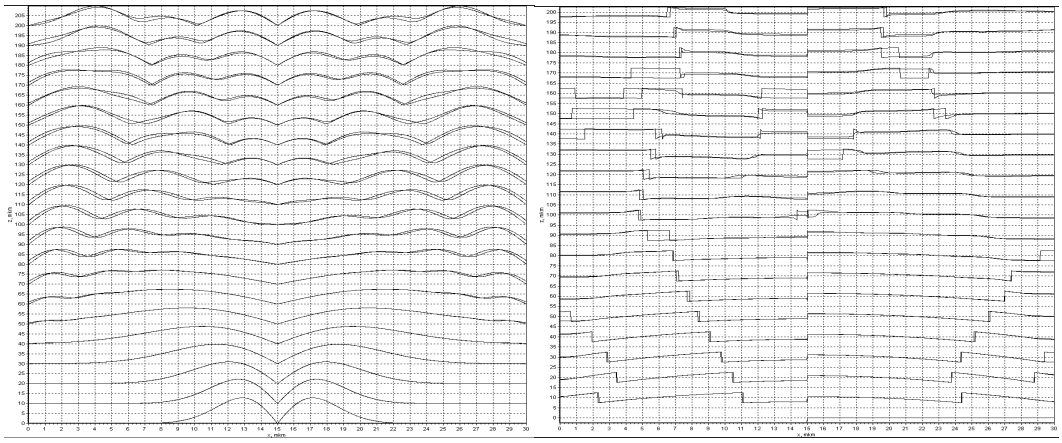


Рис. 3. – Амплитуда (слева) и фаза (справа) при начальном распределении – первой моде Гаусса – Эрмита

Из рисунка 3 следует, что в середине волновода наблюдается нулевое значение амплитуды, сингулярность фазы сохраняется с ростом z.

Следует отметить, что если в точке скачок фазы равен 2π, тогда эта точка не является сингулярной. В каждой точке к фазе можно добавить слагаемое, кратное 2π, что не изменит физическую картину процесса.

На расстоянии примерно 120 мкм от входа в волновод начинают образовываться дополнительные точки сингулярности.

Если начальное распределение электрического поля имеет фазовый скачок, расположенный не симметрично относительно стенок волновода, тогда сингулярность фазы исчезает с распространением поля в волноводе.

Распространение оптических вихрей в круглом оптическом волноводе

Круглый оптический волновод представляет собой цилиндрическую область пространства радиуса R, ограниченную отражающей стенкой. Распространение света в круглом волноводе описывается уравнением:

$$\frac{\partial u}{\partial z} - \frac{\mathrm{i}}{2kn}\left(\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \varphi}\right) = 0, \qquad (3)$$

где $u = u(r,\varphi)$ – комплексная амплитуда напряженности электрического поля;

(r,φ) – полярные координаты, описывающие поперечное сечение волновода;

z – координата, отсчитываемая вдоль оси симметрии волновода;

$k = \frac{2\pi}{\lambda}$ – волновое число (λ – длина электромагнитной волны);

$\mathrm{i} = \sqrt{-1}$;

n – показатель преломления среды волновода.

Оптический вихрь – это электромагнитное поле, имеющее следующий вид:

$$u_m = A(r,\varphi,z)\exp(im\varphi + \varphi_0), \qquad (4)$$

где $A(r,\varphi)$ – действительная часть амплитуды напряженности;

m – коэффициент, характеризующий степень закрутки вихря;

φ_0 – начальная фаза.

Точки $(r,\varphi) = (0,0)$ являются сингулярными точками фазы электрического поля. Амплитуда напряженности в этих точках должна быть равна 0. В противном случае амплитуда электрического поля будет иметь разрыв, что не позволит использовать схему Кранка-Николсона для численного моделирования процесса. Более того, физически трудно создать поле с ненулевой амплитудой в точке сингулярности фазы.

Приведенным выше условиям удовлетворяет первая мода Гаусса-Эрмита, распределение действительной части амплитуды напряженности имеет вид:

$$A(r,\varphi,0) = A_0 r \exp\left(\frac{r^2}{\sigma^2}\right) \qquad (5)$$

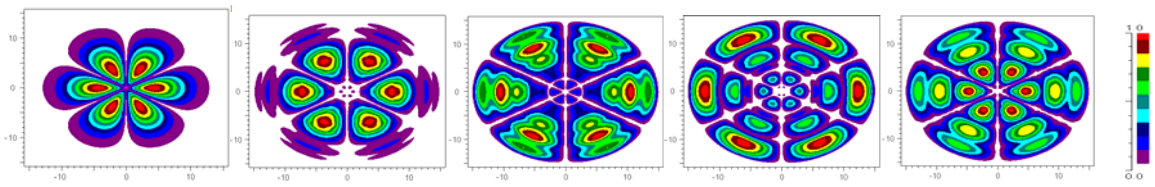
Пусть на входе в волновод начальное распределение соответствует суперпозиции двух оптических вихрей:

$$u = u_3 + u_{-3}, \qquad (6)$$

Для моделирования были использованы следующие значения параметров:

Радиус волновода $R = 30 мкм$, параметры моды Гаусса-Эрмита: $A_0 = 1, \quad \sigma = 5 мкм$.

На рисунке 4 представлены графики зависимости модуля амплитуды напряженности электрического поля от координат для различных сечений z.



$z = 0мкм$ $z = 100мкм$ $z = 200мкм$ $z = 300мкм$ $z = 400мкм$

Рис. 4. модуль амплитуды электрического поля при распространении суперпозиции $u = u_3 + u_{-3}$ для различных сечений z

На рисунке 4 видно, что структура поля не поворачивается с распространением в волноводе, что следует из компенсации поворотов вихрей u_3 и u_{-3}.

На рисунке 5 приведен график зависимости модуля амплитуды напряженности электрического поля для продольного сечения волновода. Длина волновода – $1000мкм$

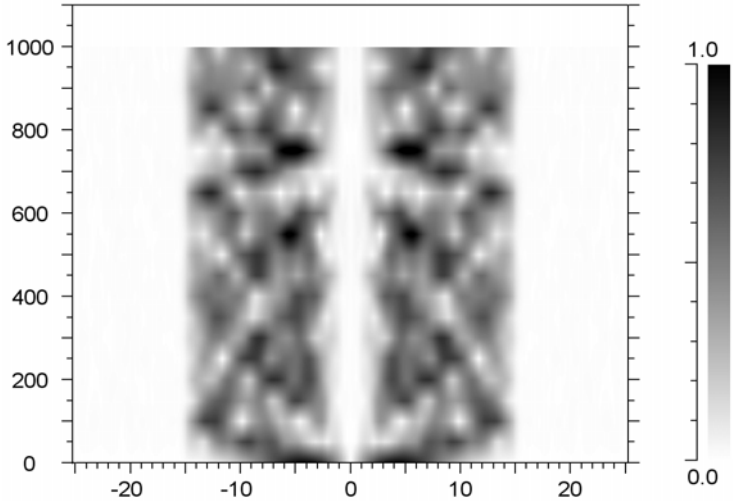


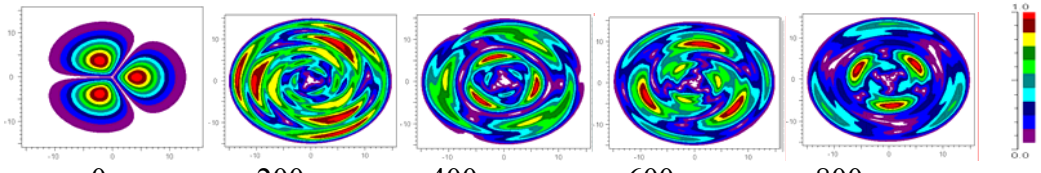
Рис. 5. модуль амплитуды электрического поля при распространении суперпозиции $u = u_3 + u_{-3}$ для продольного сечения волновода

Пусть на входе в волновод начальное распределение соответствует суперпозиции двух оптических вихрей:

$$u = u_5 + u_2, \qquad (6)$$

Параметры модели те же, что и для предыдущего случая.

На рисунке 6 представлены графики зависимости модуля амплитуды напряженности электрического поля от координат для различных сечений z.



$z = 0мкм$ $z = 200мкм$ $z = 400мкм$ $z = 600мкм$ $z = 800мкм$

Рис. 6. модуль амплитуды электрического поля при распространении суперпозиции $u = u_5 + u_2$ для различных сечений z

На рисунке 6 видно, что структура частично смазывается, хотя частично сохраняет свою форму. Также видно вращение структуры поля (против часовой стрелки).

На рисунке 7 приведен график зависимости модуля амплитуды напряженности электрического поля для продольного сечения волновода. Длина волновода – $1000мкм$

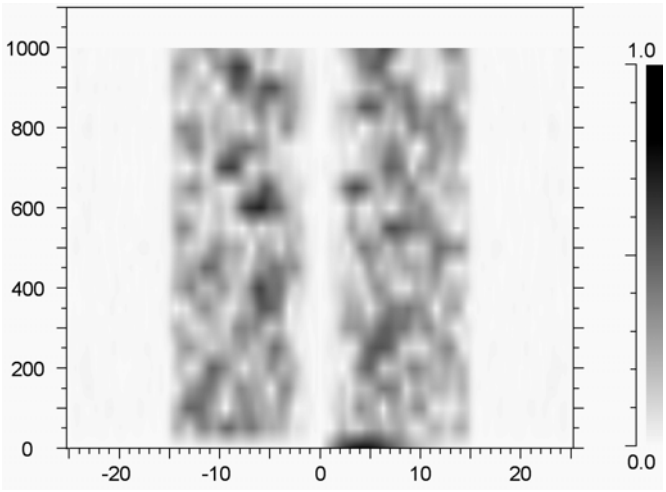


Рис. 7. модуль амплитуды электрического поля при распространении суперпозиции $u = u_5 + u_2$ для продольного сечения волновода

Литература

1. Хонина С.Н. Многомодовые лазерные поля в ступенчатых оптических волноводах: Метод.указания, Самар. гос. аэрокосм. ун-т. Самара, 2004. 30с.

2. Самарский А.А. Теория разностных схем. Учебное пособие. – М.: Наука, 1983. – 616 с.

3. RSoft Design Group Inc. BeamPROP 5.1. Manual. 2003. – 246 p.

4. Kishan Dholakia, Gabriel Spalding, Michael MacDonald. “Optical tweezers: the next generation”. Physics World. October 2002. pp. 31-35.

Расчет преломляющих поверхностей для формирования диаграммы направленности в виде линии

А.Ю. Дмитриев, Л.Л. Досколович
Институт систем обработки изображений РАН, Самара, Россия
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Формирование заданных диаграмм направленности (ДН) является актуальным для задач лазерной оптики, радиофизики и светотехники. Задача расчета оптического элемента для формирования заданной ДН состоит в определении формы преломляющей (или отражающей) поверхности элемента из условия, что преломленные (отраженные) лучи имеют заданные направления. Решение данной задачи сводится к решению нелинейного дифференциального уравнения типа уравнения Монже-Ампера [1,2]. В общем трехмерном случае решение данного уравнения является крайне сложным. Аналитическое решение возможно только в частном случае радиальной или цилиндрической симметрии.

Методы решения задач данного класса в приближении геометрической оптики хорошо разработаны для дифракционных оптических элементов [3-5]. В этом случае задача ставится как задача расчета эйконала светового поля на некоторой поверхности (обычно на плоскости), обеспечивающего заданные направления распространения лучей. Восстановление формы поверхности дифракционного элемента по функции эйконала предполагает использование приближенных соотношений типа приближения тонкого оптического элемента [6]. Использование приближенных соотношений для восстановления поверхности дифракционного микрорельефа позволяет рассчитывать оптические элементы только для формирования осевых ДН с малым угловым размером в 5-10 градусов [3,5].

Для формирования ДН в виде линии (ДН, представленных вектор-функцией одного аргумента) разработан ряд методов. В работах [7-9] приведено общее решение задачи расчета зеркала для формирования ДН в виде линии при точечном источнике излучения. Поверхность зеркала в [7-9] имеет вид огибающей семейства параболоидов вращения. Наибольшее практическое значение имеет задача формирования ДН в виде отрезка. Данная задача является эталонной при формировании сложных ДН, состоящих из набора отрезков. Законченное аналитическое решение этой задачи получено в [10,11]. Расчет преломляющих оптических элементов для формирования ДН в виде линии и отрезка в работах [7-11] не рассматривался.

В настоящее время широкое распространение получили светодиоды. В этой связи методы расчета преломляющих оптических элементов (головок светодиодов) имеют большую практическую значимость. В данной работе рассматривается расчет преломляющей поверхности для формирования ДН в виде линии. Преломляющую поверхность предлагается определить в виде огибающей семейства эллипсоидов или гиперболоидов вращения. С помощью предложенного подхода были рассчитаны преломляющие поверхности, формирующие диаграмму направленности в виде отрезка.

Расчет преломляющей поверхности

Геометрия задачи представлена на рисунке 1. Задача состоит в расчете формы преломляющей поверхности из условия, что направления преломленных лучей описываются заданной единичной векторной функцией одного аргумента.

$$\mathbf{p}(\sigma) = \left(p_x(\sigma), p_y(\sigma), p_z(\sigma) \right), \ |\mathbf{p}(\sigma)| = 1, \qquad (1)$$

где σ – некоторый параметр.

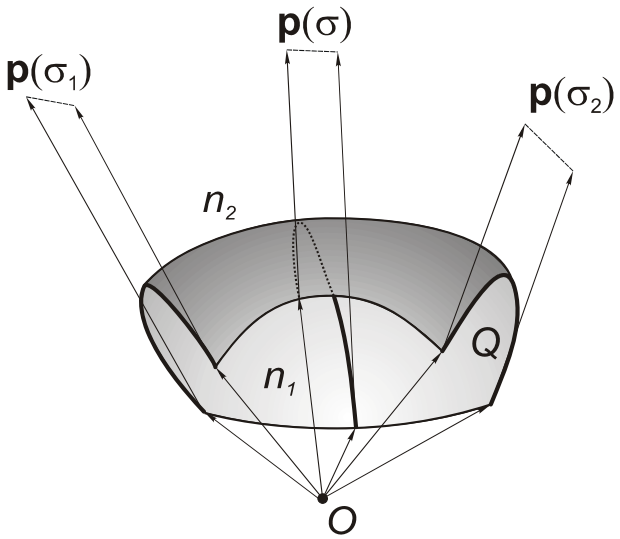


Рис. 1. Геометрия задачи формирования ДН

Будем рассчитывать преломляющую поверхность в виде огибающей семейства поверхностей $\mathbf{S}(u,v;\sigma)$, каждая из которых преобразует сферический пучок от источника в плоский с направлением $\mathbf{p}(\sigma)$. (u,v) – некоторые криволинейные координаты. Функция $\mathbf{S}(u,v;\sigma)$ является эллипсоидом или гиперболоидом вращения (в зависимости от показателей преломления сред n_1 и n_2 – до и после поверхности соответственно) с первым фокусом в начале координат и направлением большой оси $\mathbf{p}(\sigma)$, а огибающая в криволинейных координатах (u,v) может быть записана в виде

$$\begin{cases} \mathbf{S}(u,v;\sigma) = \mathbf{e}(u,v)\cdot\dfrac{\psi(\sigma)}{1-\varepsilon\left(\mathbf{e}(u,v),\mathbf{p}(\sigma)\right)}, \\ \left(\mathbf{e}(u,v),\dfrac{\mathbf{v}(\sigma)}{|\mathbf{v}(\sigma)|}\right)=\cos\left(\alpha(\sigma)\right), \end{cases} \qquad (2)$$

где

$$\mathbf{v}(\sigma)=\left(\mathbf{p}(\sigma)\frac{d\psi(\sigma)}{d\sigma}-\frac{d\mathbf{p}(\sigma)}{d\sigma}\psi(\sigma)\right),$$
$$\cos\left(\alpha(\sigma)\right)=\frac{1}{\varepsilon}\frac{d\psi(\sigma)}{d\sigma}/|\mathbf{v}(\sigma)|, \qquad (3)$$

где $\mathbf{e}(u,v)$ – единичный вектор направления луча, $\varepsilon=n_2/n_1$ – эксцентриситет эллипсоида, $\psi(\sigma)$ – фокальный параметр. В случае, когда $\cos(\alpha(\sigma))\le 1$, второе уравнение в (2) является уравнением кругового конуса с вершиной в начале координат. Тогда при фиксированном σ уравнения (2) определяют кривую $\mathbf{L}(u,\sigma)$, которая соответствует пересечению поверхности с конусом

$$\left(\mathbf{e}(u,v),\,\mathbf{a}(\sigma)\right)=\alpha(\sigma), \qquad (4)$$

где $\mathbf{a}(\sigma)=\mathbf{v}(\sigma)/|\mathbf{v}(\sigma)|$ – единичный вектор оси, а $\alpha(\sigma)$ - косинус угла при вершине конуса.

Неявное уравнение эллипсоида (гиперболоида) в декартовых координатах имеет вид

$$\begin{cases} |\mathbf{M}|-\varepsilon\left(\mathbf{p}(\sigma),\mathbf{M}\right)=\psi(\sigma), \\ \left(\dfrac{d\mathbf{p}(\sigma)}{d\sigma},\mathbf{M}\right)=-\dfrac{1}{\varepsilon}\dfrac{d\psi(\sigma)}{d\sigma}. \end{cases} \qquad (5)$$

Уравнения (5) показывают, что преломляющая поверхность состоит из кривых, которые соответствуют пересечениям поверхностей с плоскостями.

Функция $\psi(\sigma)$ в (2), (5) определяет распределение энергии вдоль ДН и может быть рассчитана из условия формирования заданной интенсивности на ДН из закона сохранения светового потока:

$$I(\sigma)\left|\frac{d\mathbf{p}(\sigma)}{d\sigma}\right|d\sigma = \int I_0(u,\sigma)d\Omega(u,\sigma), \qquad (6)$$

где $d\Omega(u,\sigma)$ – телесный угол между конусами, $I(\sigma)$ – распределение интенсивности вдоль диаграммы направленности, $I_0(u,\sigma)$ – интенсивность источника.

Рассмотрим задачу расчета преломляющей поверхности для формирования ДН в виде отрезка с заданным распределением интенсивности. Преломленные лучи в данном случае будут иметь направления

$$\mathbf{p}(\sigma) = \left(\sin(\beta(\sigma)), 0, \cos(\beta(\sigma))\right),\ \sigma \in \left[-\sigma_{\max}, \sigma_{\max}\right], \qquad (7)$$

где $\beta(\sigma)$ – угол, который образует преломленный луч с осью Oz. Предлагается расчет трехмерной преломляющей поверхности для формирования ДН (6) проводить через расчет цилиндрического профиля $r(\sigma)$, являющегося сечением поверхности плоскостью y=0. Функция $r(\sigma)$ определяется из дифференциального уравнения, разрешенного относительно производной

$$\frac{dr(\sigma)}{d\sigma} = -r(\sigma)\frac{\sin(\sigma-\beta(\sigma))}{n_1/n_2 - \cos(\sigma-\beta(\sigma))}. \qquad (8)$$

Функция $\psi(\sigma)$ вычисляется из условия, что каждый эллипсоид в семействе содержал точку $(r(\sigma)\sin(\sigma), 0, r(\sigma)\sin(\sigma))$ центрального профиля в виде:

$$\psi(\sigma) = r(\sigma)\cdot\left(1-\varepsilon\cos\left(\sigma-\beta(\sigma)\right)\right). \quad (9)$$

Функция $\beta(\sigma)$ определяет распределение энергии вдоль отрезка фокусировки и может быть вычислена из закона сохранения светового потока:

$$I(\beta)\frac{d\beta}{d\sigma} = \int I_0(\sigma,u)\left|\frac{\partial\mathbf{e}(u,\sigma)}{\partial u}\times\frac{\partial\mathbf{e}(u,\sigma)}{\partial\sigma}\right|du. \qquad (10)$$

Единичный вектор луча можно представить в виде:

$$\mathbf{e}(u,\sigma) = \begin{pmatrix} \sin(\alpha(\sigma))\sin(\tau(\sigma))\cos(u) + \cos(\alpha(\sigma))\cos(\tau(\sigma)) \\ \sin(u)\sin(\alpha(\sigma)) \\ -\sin(\alpha(\sigma))\cos(\tau(\sigma))\cos(u) + \cos(\alpha(\sigma))\sin(\tau(\sigma)) \end{pmatrix}, \quad (11)$$

где $\cos(\alpha(\sigma)) = \dfrac{\sin(\sigma-\beta(\sigma))}{\sqrt{1-2\varepsilon\cos(\sigma-\beta(\sigma))+\varepsilon^2}}$, $\tau(\sigma) = \alpha(\sigma)+\sigma$.

Тогда функция $\beta(\sigma)$ находится из следующего уравнения:

$$\frac{d\beta(\sigma)}{d\sigma} = \frac{K(\sigma,\beta(\sigma))\left[F_1(u,\sigma)-F_2(u,\sigma)\right]+F_2(u,\sigma)}{-K(\sigma,\beta(\sigma))\left[F_1(u,\sigma)-F_2(u,\sigma)\right]-I(\beta(\sigma))/\sin(\alpha(\sigma))}, \qquad (12)$$

где

$$K(\sigma,\beta(\sigma)) = \frac{\left(1-\varepsilon^2\cos(\sigma-\beta(\sigma))\right)}{1-2\varepsilon\cos(\sigma-\beta(\sigma))+\varepsilon^2},\ F_1(u,\sigma) = \int_0^{u_{\max}} I_0(u,\sigma)du,$$

$$F_2(u,\sigma) = \int_0^{u_{\max}} I_0(u,\sigma)\cos u\,du.$$

Формулы (2), (8), (10), (12) представляют законченное аналитическое решение задачи формирования ДН в виде отрезка. Таким образом, расчет преломляющей поверхности для формирования ДН в виде отрезка с заданным распределением энергии сведен к расчету дифференциального уравнения первого порядка, разрешенного относительно производной.

С использованием полученных формул была рассчитана преломляющая поверхность для формирования ДН в виде отрезка с угловым размером 150°. Показатель преломления материала поверхности n_1=1.5. Расчет производился для ламбертовского источника, излучающего в полусферу. Рассчитанная поверхность представлена на рисунке 2. Максимальные размеры поверхности вдоль осей координат Ox, Oy, Oz составляют 22.3мм, 10.4мм и 10мм, соответственно.

Для подтверждения правильности расчетов было выполнено моделирование работы оптического элемента на рисунке 2 с использованием специализированной программы по светотехнике TracePro [12]. На рисунке 3 приведено расчетное распределение интенсивности, формируемое при компактном ламбертовском источнике с диаметром 0.5 мм.

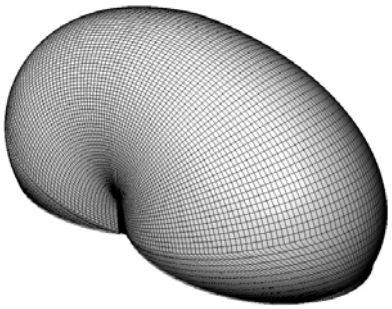


Рис. 2. Преломляющая поверхность для формирования ДН в виде отрезка с угловым размером 150°

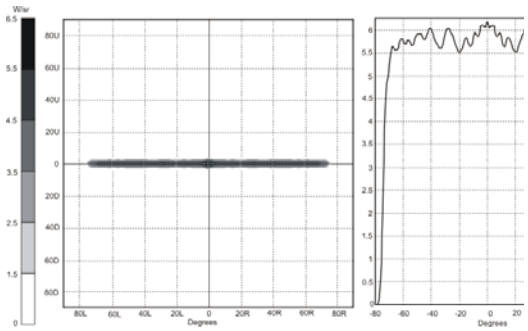


Рис. 3. Распределение интенсивности вдоль ДН, формируемой поверхностью

Результат моделирования показывает высокое качество ДН и подтверждает правильность и работоспособность предложенного метода расчета.

Литература

1. Muschaweck, J. Tailored freeform optical surface / J. Muschaweck, H. Ries // J. Opt. Soc. Am. A. 2002. V.19. №3. P. 590–595.
2. Guan, P. On a Monge-Ampere equation arising in geometric optics / P. Guan, X-J. Wang // J. Differential Geometry. 1998. V.48. №2. P. 205–223.
3. Дифракционная компьютерная оптика / под ред. Сойфера В.А. М.: Физматлит. 2007. 736 с.
4. Doskolovich, L.L. A DOE to form a line-shaped directivity diagram. / L.L. Doskolovich, N.L. Kazanskiy, V.A. Soifer, S.I. Khatritonov, P. Perlo // Journal of Modern Optics. 2004. V.51. №13. P. 1999.
5. Soifer, V., Kotlyar V., Doskolovich L. Iterative Methods for Diffractive Optical Elements Computation. London, Taylor & Francis Ltd., 1997. 256 c.
6. Борн, М. Основы оптики / М. Борн, Э. Вольф – М.: Наука, 1973.
7. Досколович, Л.Л. Расчет зеркала для формирования однопараметрической диаграммы направленности / Л.Л. Досколович, Н.Л. Казанский // Автометрия. 2004. Т.40. № 5. С. 104-111.
8. Doskolovich, L.L. Designing reflectors to generate a line–shaped directivity diagram / L.L. Doskolovich, N.L. Kazanskiy, S.I. Kharitonov, P. Perlo, S. Bernard // J. of Mod. Opt. 2005. V.52. №11. P. 1529–1536.
9. Досколович, Л.Л., Расчет формы поверхности зеркал для формирования изображения в виде линии / Л.Л. Досколович, С.И. Харитонов // Оптический журнал. 2005. Т.72. №4. С.34.
10. Doskolovich, L.L. Designing a mirror to form a line-shaped directivity diagram / L.L. Doskolovich, N.L. Kazanskiy, S. Bernard // Journal of Modern Optics. 2007. V.54. №4. P. 589–597.
11. Досколович, Л.Л. Расчет зеркала для формирования диаграммы направленности в виде отрезка / Л.Л. Досколович, Н.Л. Казанский, М.А. Тренина // Автометрия. 2006. Т.42. №4. С. 67–75.
12. http://www.lambdares.com/software_products/tracepro/

Синтез и исследование трехмерных металлодиэлектрических фотонных кристаллов

П.Н. Дьяченко[1,2], С.В. Карпеева[2], Е.В. Фесик[2], Ю.В. Микляев[3], В.С. Павельева[2], Г.Д. Мальчиков[2]

[1]*Институт систем обработки изображений РАН, Самара*

[2]*Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*

[3]*Южно-Уральский государственный университет, Челябинск*

Реализован синтез полимерных матриц фотонных кристаллов методом интерференционной литографии с последующим нанесением нанослоя золота. Запись решетки осуществлялась излучением гелий-кадмиевого лазера на длине волны 442нм в фоторезисте SU-8. Исследованы спектры отражения фотонных кристаллов в инфракрасном диапазоне.

Введение

Фотонные кристаллы представляют собой периодически структурированный материал, в котором достигается большая амплитуда модуляции диэлектрической проницаемости. Световые волны, распространяющиеся в средах с периодически распределенной диэлектрической проницаемостью, могут быть описаны в терминах фотонных энергетических зон с возможностью существования запрещенных зон, где распространение электромагнитных волн невозможно. Идея о контроле спонтанного излучения с помощью периодически изменяющегося показателя преломления впервые высказана В.П. Быковым в 1972г.[1]. После работ Э. Яблоновича и С. Джона [2,3] фотонные кристаллы стали одним из наиболее интенсивно исследуемых объектов в современной оптике.

К настоящему моменту предложено множество способов синтеза фотонных кристаллов [4-7]. Однако получение макроскопически однородных, бездефектных кристаллов до сих пор представляется весьма сложной технической задачей. Одним из наиболее перспективных методов синтеза в настоящее время представляется метод интерференционной литографии [8]. Данный метод состоит в получении трехмерной структуры из фотополимера за счет освещения трехмерной интерференционной картиной четырьмя (или более) когерентными пучками света. Преимуществами данного метода являются идеальная периодичность решетки, отсутствие дефектов структуры, возможность получения образцов большой площади и низкая стоимость. Интерес к данному методу особенно возрос после того, как в ряде работ [9,10] было показано, что с его помощью можно изготовить трехмерные фотонные кристаллы с запрещенной зоной при относительно небольших показателях преломления материала ($n > 2.0$).

Для формирования трехмерной решетки из фоторезиста методом интерференционной литографии использовалось трехкратное экспонирование пленки фоторезиста картиной интерференции двух волн, аналогично тому, как это было сделано в работе [11]. Схема эксперимента показана на рис.1. После каждого экспонирования образец поворачивается на 120 градусов вокруг вертикальной оси. Время каждой из экспозиций должно быть одинаковым. В результате экспонирования в объеме фоторезиста получается следующее распределение поглощенной энергии:

$$I(\vec{r}) = \sum_{i=1}^{N} I_i \cos^2(\vec{b}_i \vec{r} + \phi_i)$$

где $b_i = k1_i - k2_i$, $k1_i$, $k2_i$ – волновые вектора интерферирующих волн при i-ой экспозиции. Из (1) легко видеть, что вектора b_i являются базисными векторы обратной

решетки синтезируемого кристалла. На данном этапе нами реализован только случай, когда угол между интерферирующими пучками в каждой из трех экспозиций был одинаковым, т.е. базисные векторы обратной решетки имели одинаковую длину. Угол между базисными векторами также был одинаковым. Таким образом, нами реализовывался случай орторомбической решетки.

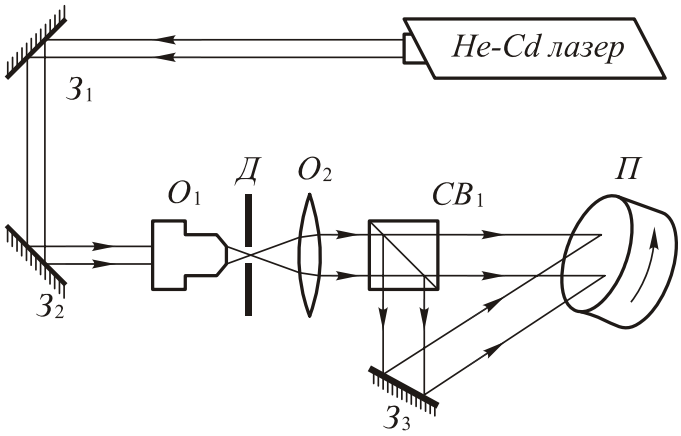


Рис.1. Буквами обозначены: З1, З2, З3, - зеркала, О1 , О2 – объективы, Д – диафрагма, СВ1, - делительный кубик , П - подложка из стекла с нанесенным на нее фоторезистом.

Для синтеза полимерных матриц фотонных кристаллов описанным методом была собрана экспериментальная установка в соответствие с оптической схемой рис. 1. Мы использовали гелий-кадмиевый лазер ГКЛ-60В(И) мощностью 80 мВт. Луч лазера расширялся примерно в 2,5 раза при помощи микрообъектива 8Х О1 , фильтра - диафрагмы Д диаметром около 50 мкм и объектива О2 с фокусным расстоянием 50 мм. Для поворота и разделения луча использовались зеркала З1, З2, З3 и светоделитель СВ1 с просветлением граней под данную длину волны. Фоторезист наносился на подложку методом центрифугирования в соответствие с рекомендациями производителя, то есть в два этапа. Предварительное распределение фоторезиста по поверхности образца при 500 об/мин в течение 10 сек формирование нужной толщины при скорости 3000об/мин в течение 30 сек. Описанная процедура давала толщину слоя при использовании SU-8-50 около 40-50мкм, в соответствии со спецификациями фоторезиста. Затем образец подвергался сушке также в два этапа – вначале 5 минут при температуре 60 градусов Цельсия, а затем 20 мин. при температуре 95 градусов для удаления растворителя. Экспонированный фоторезист запекался при температуре 95 градусов Цельсия в течении 6 минут. После запекания фоторезист помещался в проявитель PGMEA (2-(1-метокси) пропилацетат) на 5 – 7 минут, и затем промывался в изопропиловом спирте. Время экспозиций было одинаковым и составляло от 10 до 20 минут.

Далее проводились исследования режимов записи трехмерно-периодических решеток с различными периодами и орторомбическим типом симметрии. При этом для разных периодов подбиралось время экспозиции, запекания и проявления. Менялись и углы наклона образца для сохранения симметрии решеток. Указанные параметры оказывают комплексное влияние на качество решеток, и подбираться должны также в комплексе.

На рис. 2 представлена электронная фотография полученного фотонного кристалла. Период решетки составил 2.5 мкм. Представленные на рис. 2 решетки не позволяют существовать фотонной запрещенной зоне, поэтому в дальнейшем производилось нанесение нанослоя золота на данную решетку.

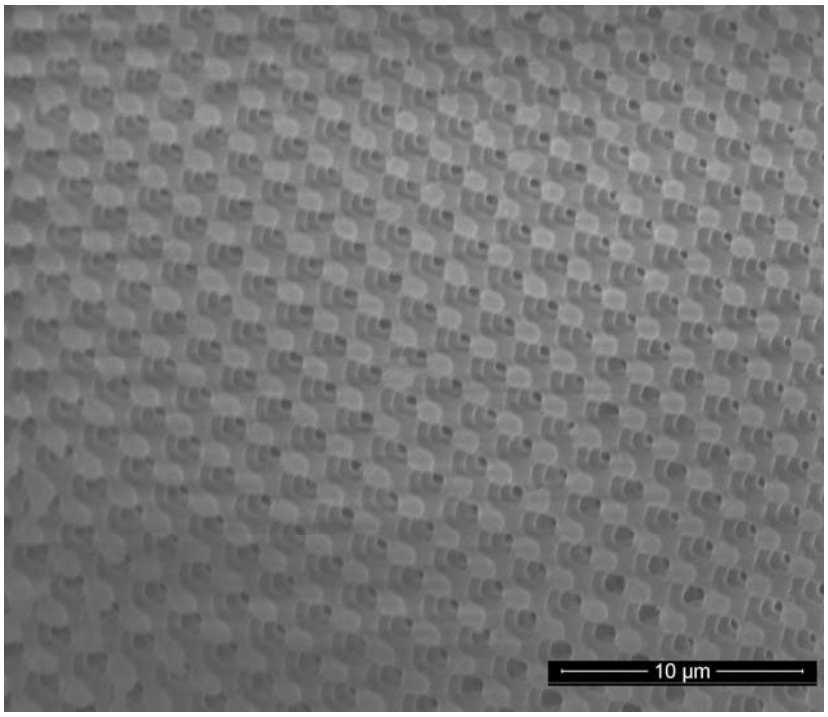


Рис. 2. Электронная фотография трехмерного фотонного кристалла.

Методом магнетронного распыления (SPI-Module Sputter Coater (США)) на фотонные кристаллы был нанесен слой золота толщиной 50 нм. Толщина слоя определялась при помощи кварцевого датчика толщины (SPI Quartz Crystal Thickness Monitor Model 12161). Электронные фотографии полученных образцов представлены на рис. 3.

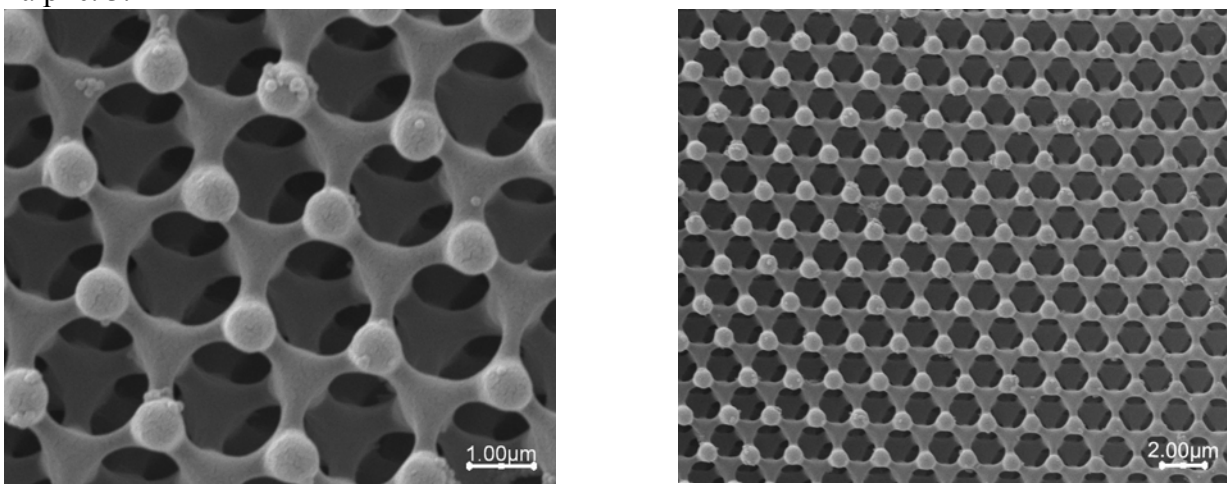


Рис. 3. Электронные фотографии фотонного кристалла с нанесенным на него слоем золота 50 нм.

Для исследования оптических свойств фотонных кристаллов использовался инфракрасный микроскоп Hyperion 1000 (Bruker Optics (Германия)) с фурье-спектрометром Tensor 27 (Bruker Optics (Германия)), при помощи которого измерялись коэффициенты отражения структур. Коэффициенты отражения фотонных кристаллов получались путем нормирования полученных спектров структур к спектрам золотого зеркала. Полученные спектры отражения представлены на рис. 4. Кривой черного цвета обозначен спектр отражения фотонного кристалла, полученный при помощи ИК-объектива Schwarzschild 15x. Кривой красного цвета обозначен спектр отражения фотонного кристалла полученный при помощи объектива скользящего падения 15x. Синим цветом отмечен спектр отражения пленки золота на фотополимере SU-8. Как можно заметить из графика, в спектре отражения фотонного кристалла присутствует

пик на длине волны 2,6-2,8 мкм, что равняется периоду решетки фотонного кристалла. Данный пик в спектре отражения указывает на появление фотонной запрещенной зоны с центром на длине волны 2,6-2,8 мкм.

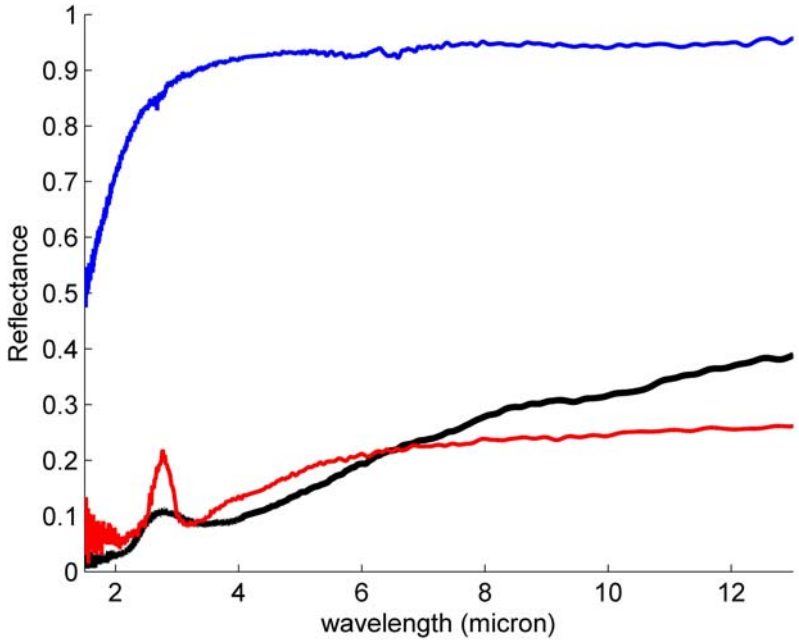


Рис. 4. Спектры отражения фотонного кристалла. Черным цветом обозначен спектр полученный от ИК-объектива Schwarzschild 15x. Красным цветом обозначен спектр от объектива скользящего падения 15x. Синим цветом отмечен спектр отражения пленки золота на фотополимере SU-8.

Литература

[1] Bykov, V. P.; “Spontaneous emission in a periodic structure,” Sov. Phys. JETP. 35, 269-273 (1972).

[2] Yablonovitch, Y.; “Inhibited spontaneous emission in solid-state physics and electronics,” Phys. Rev. Lett. 58, 2059-2062 (1987).

[3] John, S.; “Strong localization of photons in certain disordered dielectric superlattices,” Phys. Rev. Lett. 58, 2486-2489 (1987).

[4] Wijnhoven J.E.G.J.; Vos W.L.; “Preparation of photonic crystals made of air spheres in titania,” Science., 281, 802-804 (1998).

[5] Lin, S.Y.; Fleming, J.G.; Hetherington, D.L.; Smith, B.K; Biswas, R.B; Ho, K.M.; Sigalas, M.M.; Zubrzycki, W.; Kurtz, S.R.; Bur, J.; “A three-dimensional photonic crystal operating at infrared wavelengths,” Nature 394, 251-253 (1998).

[6] Ho, K.M.; Chan, C.T.; Soukoulis, C.M.; Biswas, R.; Sigalas, M.; “Photonic band gaps in three dimensions: New layer-by-layer periodic structures,” Solid State Communications 89, 413-416 (1994).

[7] Blanco, A.; Chomski, E.; Grabtchak, S.; Ibisate, M.; John, S.; Leonard, S.W.; Lopez, C.; Meseguer, F.; Miguez, H.; Mondla, J.P.; Ozin, G.A.; Toader, O.; Van Driel, H.M.; “Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1.5 micrometres,” Nature 405, 437-440 (2000).

[8] Campbell, M.; Sharp, D.N.; Harrison, M.T.; Denning, R.G.; Turberfield, A.J.; “Fabrication of photonic crystals for the visible spectrum by holographic lithography,” Nature 404, 53-56 (2000).

[9] Sharp, D.N.; Turberfield, A.J.; Denning, R.G.; “Holographic photonic crystals with diamond symmetry,” Phys. Rev. B 68, 205102 (2003).

[10] Ullal, C.K.; Maldovan, M.; Thomas, E.L.; Chen, G.; Han, Y.-J.; Yang, S.; “Photonic crystals through holographic lithography: Simple cubic, diamond-like, and gyroid-like structures,” Appl. Phys. Lett. 84, 5434-5436 (2004).

[11] Miklyaev, Yu. V., Karpeev, S. V., Dyachenko, P. N., Pavelyev, V. S. “Fabrication of three-dimensional photonics crystals by interference lithography with low light absorption,” Journal of Modern Optics 56, 1133–1136 (2009).

Астигматическое преобразование одномерных распределений, согласованных с функциями Эйри

О.В. Зотеева[1], С.Н. Хонина[1,2]
[1]*Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*
[2]*Институт систем обработки изображений РАН, Самара*

Введение

Вихревые лазерные пучки или пучки с вихревыми фазовыми особенностями (оптические воронки) успешно применяются для решения очень широкого спектра задач, включая оптическое манипулирование микро- и нанообъектами (захват и управление движением вплоть до отдельных атомов), высокоточную метрологию, уплотнение записи информации и даже астрологические исследования , что объясняет большое количество работ и разнообразие подходов для решения задачи генерации таких пучков.

В данной работе рассматривается формирование вихревых пучков с помощью средств астигматической оптики и цилиндрических непериодических дифракционных решеток, в том числе согласованных с функциями Эйри.

1. Преобразование одномерных распределений

В 1991 году Е. Абрамочкин и В. Волостников предложили получать моды Лагерра-Гаусса (ЛГ), имеющие вихревую сингулярную фазу из мод Эрмита-Гаусса (ЭГ), фаза которых имеет бинарную структуру, с помощью цилиндрических линз [1]. Было найдено интегральное преобразование мод ЭГ в моды ЛГ и представлены результаты натурных экспериментов. Преобразование одномерных распределений с помощью цилиндрической линзы было рассмотрено в работе [2] как частный случай астигматических модовых преобразований.

1.1. Астигматическое модовое преобразование

Астигматическое преобразование пучков ЭГ $H_{n,m}(x,y)$ в пучки ЛГ $L_{p,q}(x,y)$ выглядит следующим образом:

$$\iint_{R^2}\exp\left[-i\left(x\xi+y\eta\right)+\frac{2i\xi\eta}{\rho^2}\right]H_{n,m}\left(\frac{\xi}{\rho},\frac{\eta}{\rho}\right)d\xi d\eta=$$

$$=\frac{\pi\rho^2}{\sqrt{2}}(-1)^{n+m}\exp\left(-\frac{i\rho^2xy}{4}\right)\begin{cases}(2i)^n\,m!L_{m,n-m}\left(\frac{\rho x}{2\sqrt{2}},\frac{\rho y}{2\sqrt{2}}\right),n\ge m,\\(2i)^m\,n!L_{n,m-n}\left(\frac{\rho y}{2\sqrt{2}},\frac{\rho x}{2\sqrt{2}}\right),n\ge m.\end{cases}\qquad(1)$$

Его частный вариант:

$$\iint_{R^2}\exp\left[-i\left(x\xi+y\eta\right)+\frac{2i\xi\eta}{\rho^2}\right]H_{n,0}\left(\frac{\xi}{\rho},\frac{\eta}{\rho}\right)d\xi d\eta=\frac{\pi\rho^2}{\sqrt{2}}(-2i)^n\exp\left(-\frac{i\rho^2xy}{4}\right)L_{0,n}\left(\frac{\rho x}{2\sqrt{2}},\frac{\rho y}{2\sqrt{2}}\right)$$

дает возможность синтеза «спиральных пучков» $\exp(-z\bar{z})z^n$ и может быть обобщен следующим образом:

$$\iint_{R^2}\exp\left[-i\left(x\xi+y\eta\right)+\frac{2i\xi\eta}{\rho^2}-\frac{\eta^2}{\rho^2}\right]g\left(\frac{\xi}{\rho}\right)d\xi d\eta=\sqrt{\pi}\rho^2\exp\left(-\frac{i\rho^2xy}{4}\right)S\left(\rho\left(x+iy\right),\rho\left(x-iy\right)\right),\qquad(2)$$

где $g(x)$ – одномерное распределение.

Тогда спиральный пучок $S(z,\bar{z})$, полученный по формуле (2), имеет вид:

$$S(z,\bar{z}) = \exp\left(-\frac{1}{8}z\bar{z} + \frac{1}{8}z^2\right)\int_R \exp\left(-\xi^2 - iz\xi\right)g(\xi)d\xi\,. \quad (3)$$

Через декартовы координаты выражение (3) записывается следующим образом:

$$S(x,y) = \exp\left(-\frac{1}{4}y^2 + \frac{1}{4}ixy\right)\int_R \exp\left(-\xi^2 - ix\xi + y\xi\right)g(\xi)d\xi\,. \quad (4)$$

Как видно из выражения (4) выполнить такое астигматическое преобразование можно с помощью одной цилиндрической линзы.

1.2. Преобразование функции Эйри

Интерес к функциям Эйри, возникший в конце семидесятых годов прошлого века в контексте квантовой механики [3], был реанимирован в последнее время в связи с новыми возможностями, которые предлагает дифракционная оптика в физической реализации математических абстракций. Среди других сопротивляющихся дифракции мод пучки Эйри выделяются особыми «баллистическими» свойствами, а именно параболическое отклонение от прямолинейной траектории основного лепестка при распространении в свободном пространстве [4].

Аналогично модам ЭГ двумерные пучки Эйри синтезируются из произведения одномерных распределений, однако основная проблема связана с усечением [4-7] этих пространственно бесконечных решений параксиального волнового уравнения.

Основное приложение пучки Эйри получили в оптическом манипулировании [8, 9], но как и модам ЭГ, такие пучки не обладают орбитальным угловым моментом, связанным с наличием вихревых фазовых особенностей.

Создать на основе одномерных пучков Эйри комплексные распределения, содержащие оптические вихри, можно с помощью астигматического конвертора и бинарных ДОЭ, согласованных с функциями Эйри [9].

Функции Эйри являются решением одномерного дифференциального уравнения:

$$Ai''(x) - xAi(x) = 0\,, \quad (6)$$

и для действительных значений аргумента могут быть записаны в виде:

$$Ai(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(\frac{it^3}{3} + ixt\right)dt\,. \quad (7)$$

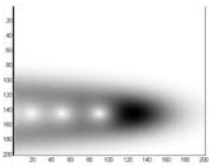
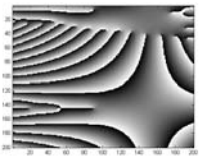
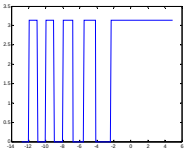
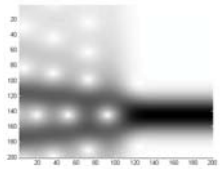
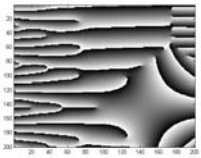
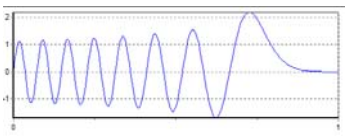
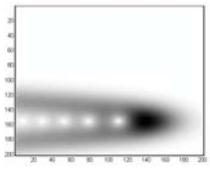
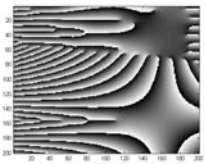
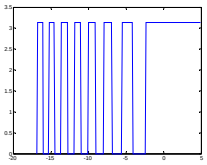
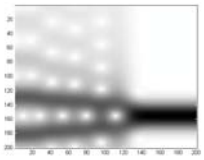
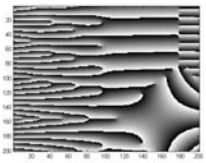
В случае простейшего бинарного фазового кодирования [9] комплексная функция пропускания ДОЭ, формирующего пучок Эйри, представляется следующим образом:

$$g(x) = \exp\left\{i\arg\left[Ai(x)\right]\right\}. \quad (8)$$

Численные эксперименты по астигматическому преобразованию как самой функции Эйри (7), так и фазовой функции, вычисляемой по формуле (8), приведены в таблице 3.

Усечение функций Эйри в (7) и (8) выполнялось как в работе [6] по количеству нулей.

Таблица 3. Астигматическое преобразование распределений, согласованных с функциями Эйри.

Число нулей	Входная функция	Выходное распределение	
		Амплитуда	Фаза
n=10	$Ai(x)$ 		
	$\arg\left[Ai(x)\right]$ 		
n=15	$Ai(x)$ 		
	$\arg\left[Ai(x)\right]$ 		

Из таблицы 3 видно, что имеется определенное сходство в получаемых на выходе распределениях. Однако в случае использования обычной усеченной функции Эйри формируется одномерная цепочка оптических воронок, а при подаче на вход согласованной бинарной фазовой функции – целая матрица оптических воронок. Проследить различие в положении вихревых фазовых сингулярных точек можно по фазовым картинам выходных распределений.

2. Применение астигматических конвертеров

На практике для получения результатов, аналогичных приведённым в пункте 1.3, часто используются астигматические конвертеры.

2.1. Параксиальные астигматические конвертеры

Конвертеры данного вида могут состоять из различного числа линз и иметь различные оптические характеристики. В работе рассматриваются преимущественно конвертеры из одной, двух и трёх линз.

Как уже говорилось ранее, на результат действия конвертера влияет соотношение параметров лазерного пучка и оптических параметров установки.

Рассмотрим конвертер, состоящий из трёх цилиндрических линз. На рисунке 1 изображена схема этого конвертера:

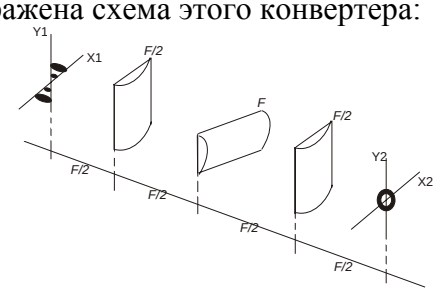


Рис. 1. Трёхлинзовый $\pi/2$ конвертер

2.2. Непараксиальные астигматические преобразования

В отличие от первого типа конвертеров, теоретически непараксиальные не используют принцип параксиального приближения, поэтому результаты могут отличаться от предыдущего случая. Но так как в оптике все расчеты, преобразования и схемы неизбежно используют принцип параксиальности, то необходимо соблюдение некоторых условий для параметров оптических схем и элементов.

Рассмотрим однолинзовую астигматическую установку, схема которой приведена на рисунке 2.

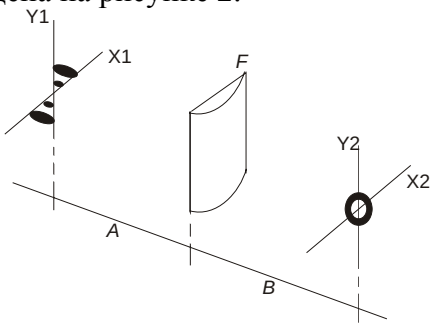


Рис. 2. Схема однолинзовой астигматической установки

На рисунке 2 f – фокусное расстояние линзы, a – расстояние, на котором находится линза от входного пучка, b – расстояние, на котором отслеживается результирующее распределение.

Далее будем рассматривать конвертеры второго типа, состоящие из одной цилиндрической линзы.

2.3. Результаты астигматического конвертирования

На примере функции Эйри, содержащей 10 нулей, рассмотрим результаты, получаемые с применением астигматического конвертера.

На вход будем подавать чисто фазовое распределение, соответствующее данной функции. Пример приведён на рисунке 3.

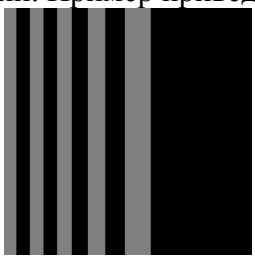
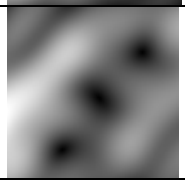
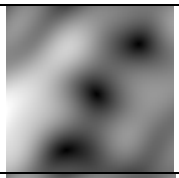
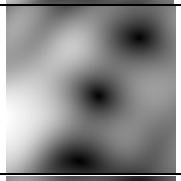
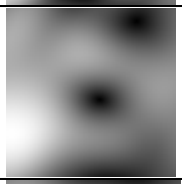
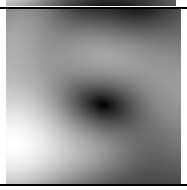
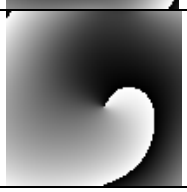


Рис. 3. Фаза, подаваемая на вход конвертера

Таблица 1. Результат преобразования входного распределения одной линзой

Параметры конвертера	**Амплитуда**	**Фаза**
Полуширина=2 мм a=b=1500 мм f=1000 мм		
Полуширина=2 мм a=b=1800 мм f=1000 мм		

Полуширина=2 мм a=b=1900 мм f=1000 мм		
Полуширина=2 мм a=b=2000 мм f=1000 мм		
Полуширина=2 мм a=b=2200 мм f=1000 мм		
Полуширина=2 мм a=b=2400 мм f=1000 мм		

Заключение

В работе рассмотрено преобразование одномерных распределений, в том числе согласованных с функциями Эйри, с помощью астигматических конвертеров. В результате таких преобразований можно формировать двумерные матрицы оптических вихрей различной конфигурации, что, в частности, расширяет возможности оптического микроманипулирования.

Литература

1. Abramochkin E., Volostnikov V. // Optics Communications. 1991. N 83, P. 123–135.
2. Абрамочкин Е.Г., Волостников В.Г. // Успехи физических наук. 2004. Вып. 12. С. 1273–1300.
3. Berry M.V., Balazs N.L., Nonspreding wave packets, Am. J. Phys. 47(3), 264-267 (1979)
4. Siviloglou G.A., Christodoulides D.N., Accelerating finite energy Airy beams, Opt. Letters 32(8), 979-981 (2007)
5. Banders M.A., Gutierrez-Vega J.C., Airy-Gauss beams and their transformation by paraxial optical systems, Opt. Express 15(25), 16719-16728 (2007).
6. Хонина С.Н., Волотовский С.Г., Ограниченные 1D пучки Эйри: лазерный веер, Компьютерная оптика. – 2008. –Т. 32, № 2. – С. 168-174.
7. Хонина С.Н., Волотовский С.Г., Зеркальные лазерные пучки Эйри, Компьютерная оптика, 34(2), 203-213 (2010)
8. . Jörg Baumgartl, Michael Mazilu and Kishan Dholakia, Optically mediated particle clearing using Airy wavepackets, Nature Photonics 2, 675 - 678 (2008)
9. Хонина С.Н., Скиданов Р.В., Моисеев О.Ю., Формирование лазерных пучков Эйри с помощью бинарно-кодированных дифракционных оптических элементов для манипулирования микрочастицами // Компьютерная оптика. – 2009. –Т. 33, № 2. – С. 138-146.

Анализ методов экспресс-контроля чистоты поверхности

Н.А. Ивлиев, В.А. Колпаков, С.В. Кричевский
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

При современном развитии микро-, наноэлектроники, дифракционной оптики и нанофотоники размеры загрязнений на поверхности подложек становятся соизмеримыми с размерами формируемых микро- и наноструктур, что приводит к необходимости жесткого контроля степени чистоты поверхности. Для анализа степени чистоты поверхности широкоформатных подложек широко применяются методы экспресс-контроля [1].

Основными методами, используемыми для экспресс-контроля чистоты поверхности подложек, являются методы спектроскопии многократно нарушенного полного внутреннего отражения (МНПВО); измерения контактной разности потенциалов (КРП); смачиваемости, подразделяемые на методы окунания, конденсации, смачиваемости поверхности подложки каплей жидкости; трибометрического контроля. Проведем краткий анализ вышеперечисленных методов с целью определения наиболее приемлемого метода, способного удовлетворить требования дифракционной оптики и нанофотоники к процессу контроля чистоты поверхности.

Метод спектроскопии многократно нарушенного полного внутреннего отражения

При осуществлении многократного отражения светового потока от структуры подложка – загрязнение на поверхности подложки происходит ослабление отраженного луча из-за проникновения части данного потока в оптически менее плотную среду и поглощения в ней. Причем исследуемая подложка либо прикладывается загрязненной поверхностью к элементу МНПВО, либо сама используется в качестве элемента МНПВО [2]. Из этого следует, что контролируемая подложка должна обеспечивать полное внутреннее отражение светового луча в обоих случаях, обеспечивая соответствующие оптические параметры. Выполнение данных условий приводит к ограничениям: чувствительности метода - 10^{-7} г/см^2 [2]; типов материала подложек и контролируемых загрязнений на их поверхности.

Кроме того, в процессе измерения осуществляется контакт поверхности контролируемой подложки с элементом МНПВО, поверхность последнего загрязняется, и при последующих измерениях требуется проводить операцию его очистки, что является существенным недостатком, препятствующим проведению экспресс-контроля чистоты поверхности подложек.

Таким образом, применение метода МНПВО ограничено оптическими свойствами материалов подложек и загрязнений, а также возможностью измерения чистоты поверхности подложек только в диапазоне до 10^{-7} г/см^2.

Метод измерения контактной разности потенциалов

Принцип измерения заключается в том, что над исследуемой поверхностью вибрирует с заданной частотой точечный зонд, емкостные параметры которого реагируют на загрязнения поверхности подложек. Численные значения емкости определяются как свойствами загрязнений, так и диэлектрика. Согласно [3] поверхность твердых тел обладает значительной концентрацией положительных и отрицательных зарядов, активно формирующих емкостные свойства вибрирующего зонда. Это приводит к существенным трудностям при интерпретации результатов

измерения. Из-за вибрации зонда возникает поток газа [4,5], частицы которого взаимодействуют с исследуемой поверхностью, изменяя ее свойства. Для исключения данного явления измерения КРП необходимо проводить или в инертной атмосфере, или в особо чистой среде.

Высокая чувствительность метода к наличию на поверхности заряда делают его крайне восприимчивым к изменению свойств поверхности исходных подложек. Поэтому при использовании новой партии подложек или состава газа в рабочей камере необходимо повторять калибровку параметров прибора. Следует вывод: метод КРП не является оптимальным для проведения экспресс-контроля чистоты поверхности подложек.

Методы контроля качества очистки по смачиваемости поверхности подложки

Наиболее простыми и доступными методами, дающими качественное представление о количестве загрязнений на поверхности подложек, являются способы окунания и конденсации [6, 7], чувствительность которых довольно высока: окунания - $1 \cdot 10^{-6} \div 1 \cdot 10^{-7}$ г/см^2, а конденсации - $1 \cdot 10^{-7} \div 1 \cdot 10^{-8}$ г/см^2. Однако, несмотря на простоту, данные методы обладают существенными недостатками: дают возможность контролировать только гидрофобные загрязнения, загрязняют поверхность контролируемой подложки остатками жидкости в процессе измерения.

Трибометрический метод контроля

Трибометрический метод контроля чистоты поверхности основан на зависимости коэффициента трения покоя от концентрации поверхностных загрязнений и позволяет определять величину данной концентрации до $1 \cdot 10^{-9}$ г/см^2 [1,8]. Схема конструкции стандартного прибора, реализующего данный метод, представлена на рис.1.

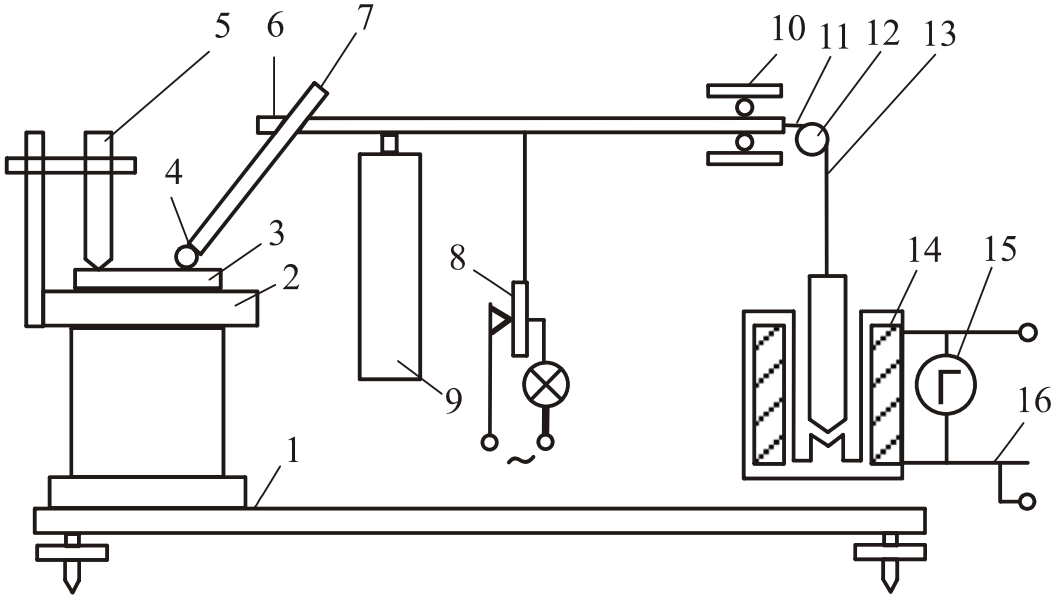


Рис. 1. Схема трибометра ИЧ-2: 1 – станина; 2 – предметный столик; 3 – измеряемая подложка; 4 – зонд; 5 – прижимное устройство; 6 – основная ось; 7 – ось индентора; 8 – контакты сигнальные; 9 – груз для создания нагрузки; 10 – подшипник шариковый; 11 – тяга; 12 – ролик; 13 – сердечник;14 – соленоид; 15 – гальванометр; 16 – потенциометр загрязнений на поверхности подложки

Измерение осуществляется путем определения значения величины тока, создающего в соленоиде в момент сдвига зонда-индентора силу тяги, равную силе трения покоя. Условные показания гальванометра соответствуют измеряемому

коэффициенту трения, а следовательно, и количеству загрязнений на поверхности подложки.

Анализ конструкции рассматриваемого прибора выявил его существенные недостатки, обусловленные плохой воспроизводимостью сопротивления сигнальных контактов из-за их загрязнения; наличием сил трения в механически движущихся узлах прибора, величина которых соизмерима с контролируемой силой трения зонда-индентора по исследуемой поверхности и суммируется с ней (см. рис.1). Перечисленное определяет существенную погрешность измерения величины коэффициента трения покоя, устранение которой представляет собой технически сложную задачу. Наряду с уже рассмотренными недостатками прибора и погрешностью измерения, вызванной пульсациями напряжения на соленоиде [9], имеется значительный недостаток и самого метода измерения, заключающийся в том, что при неподвижном контакте зонда-индентора с поверхностью исследуемой подложки происходит продавливание им слоя загрязнений [10]. Наличие данного явления приводит к взаимодействию на молекулярном уровне атомов поверхностей зонда-индентора и подложки. Атомы загрязнений в этом процессе участвуют лишь частично по периметру пятна трибометрического контакта. Такой механизм взаимодействия приводит к росту погрешности измерения.

Недостатком рассматриваемого прибора является также необходимость применения специальной технологии очистки поверхности зонда-индентора и эталонной (калибровочной) поверхности [8].

Таким образом, анализ существующих методов и приборов экспресс-контроля чистоты поверхности позволяет сделать вывод, что по точности измерения трибометрический метод на порядок превосходит все остальные. Однако для устранения его недостатков требуется модификация трибометра и разработка соответствующей методики измерения

Предлагаемый способ

В работе [11] предложен способ экспресс-контроля чистоты поверхности на основе трибометрического взаимодействия подложек, предусматривающий возможность определения коэффициента трения покоя, для оценки величины концентрации поверхностного загрязнения. Подложку, на поверхности которой необходимо измерить степень загрязнения, размещают под углом к горизонту (α), а подложку-зонд под углом к исследуемой подложке (β), формируя тем самым движущую силу и точечный контакт с определенными характеристиками.

Определение коэффициента трения покоя, путем плавного увеличения угла α до момента страгивания подложки-зонда, позволяет, зная табличный коэффициент трения покоя для данного материала, оценить относительную величину концентрации загрязнения.

Использование в качестве движущей силы - силы тяжести, для измерения коэффициента трения скольжения, позволяет упростить прибор, описанный в [1] и устранить погрешности механических узлов, задающих сдвиговое воздействие.

Литература

1. Полтавцев Ю.Г., Князев А.С. Технология обработки поверхности в микроэлектронике. – Киев: Тэхника, 1990. – 192 с.

2. Харрик Н. Спектроскопия внутреннего отражения / пер. с англ. Н. Харрик. – М.: Мир, 1970. – 335 .

3. Волькенштейн Ф.Ф. Электронные процессы на поверхности полупро-водников при хемосорбции. – М.: Наука, 1987. – 311 с.

4. Жарких Ю.С., Пастушенко А.М., Мисюра А.В., Тронько Т.В. Влияние химических обработок на гетерогенность поверхностного потенциала кремния // Полупроводниковая техника и микроэлектроника. – 1977. – № 25. – С. 40-44.

5. Жарких Ю.С., Евдокимов А.Д., Полтавцев Ю.Г., Левитская Р.О. Сравнение двух способов контроля предокислительных обработок кремниевых пластин // Оптоэлектроника и полупроводниковая техника. – 1983. – № 4. – С. 3-4.

6. Вестфаль О.Л., Мягков А.Т. Методы и средства контроля чистоты и качества поверхности полупроводников // Обзоры по электронной технике. – Сер. 6 «Материалы». – Вып. 9(406). – М.: ЦНИИ «Электроника», 1976. – 35 с.

7. Богатырев А.Е., Шушунова Л.И., Цыганов Г.М. Новые методы контроля чистоты и дефектности поверхности деталей // Обзоры по электронной технике. – 1980. – № 3 (707). – С. 19-27.

8. Перескокова А.П., Солодовникова Л.В., Акимова А.М. Применение трибометрического метода для контроля чистоты поверхности деталей и технологических сред // Электронная техника. Сер. 7. Технология, организация производства и оборудование. – 1979. – вып. 1. – С. 143-151.

9. Моисеев О.Ю. Исследование методов формирования микрорельефа дифракцирнных оптических элементов инфракрасного диапазона с использованием фоторезистов и фотополимерных композиций: дисс. канд. техн. наук: 01.00.01: защищена 07.06.00: утв. 26.12.00 / Моисеев Олег Юрьевич. – Самара, 2000. – 125 с. – 05245271765.

10. Хебда М. Справочник по триботехнике. В 3т. Т.1. Теоретические основы / под общ. ред. М. Хебды, А.В. Чичинадзе. – М.: Машиностроение, 1989. – 400 с.

11. Пат. 2307339 Российская Федерация, МПК[7] G 01 N 19/08. Способ измерения чистоты поверхности подложек / Казанский Н.Л., Колпаков В.А., Кричевский С.В., Ивлиев Н.А.; заявитель и патентообладатель ИСОИ РАН. – № 2005118279; заявл. 14.06.05; опубл. 27.09.07, Бюл. № 27. – 5 с.

Исследование формирования осевого отрезка с помощью оптимизированного бинарного ДОЭ

Д.Г. Качалов[1], В.С. Павельев[1,2], С.Н. Хонина[2], Р.В. Скиданов[2], О.Ю. Моисеев[2]
[1]*Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*
[2]*Институт систем обработки изображений РАН, Самара*

Аннотация

Проведено численное и экспериментальное исследование действия оптимизированного бинарного ДОЭ [Д.Г. Качалов, В.С. Павельев, С.Н. Хонина, Компьютерная оптика, 2009. Т. 33, № 4. С. 441-445], предназначенного для формирования продольного отрезка с равномерной интенсивностью, в условиях изменения параметров освещающего пучка.

Введение

Элементы дифракционной микрооптики, формирующие заданные продольные распределения интенсивности (соосные отрезки, наборы последовательных осевых фокусов и т.д.), находят применение во многих областях техники. Например, при выравнивании и совмещении конструкций на больших расстояниях [1]; при формировании изображений протяженных или движущихся объектов, что особенно актуально в медицине [2-4], а также для неразрушающего исследования материалов [5] и устройств [6], в метрологии для сканирующих [7] и интерференционных [8, 9] оптических систем.

Классическим рефракционным оптическим элементом для генерации осевой световой нити является аксикон [10].

Однако распределение интенсивности вдоль оси, формируемое аксиконом, очень не-равномерное – кроме существенных флуктуаций наблюдается линейный рост интенсивности при увеличении расстояния от оптического элемента.

Оптимизация фазовой функции дифракционного оптического элемента, фокусирующего лазерное излучение в продольный отрезок, в рамках геометрооптического подхода рассматривалась в работах [11-13]. Отметим, что геометрооптический расчет принципиально не позволяет учесть дифракционных эффектов.

Также известны итерационные методы расчета ДОЭ [14], основанные на алгоритме Герчберга-Сакстона [15], которые позволяют учитывать дифракционные эффекты, однако не обладают сходимостью (может быть показана лишь нерасходимость). Кроме того, необходимость применения обратного оператора на каждой итерации метода затрудняет получение квантованной фазы оптического элемента, в то время как значительная часть технологий (главным образом литографических) ориентирована как раз на получение квантованного микро-нанорельефа [16]. Поэтому в работе [17] для синтеза бинарного рельефа был применен метод последовательных приближений с бинаризацией фазовой функции на каждом шаге. Полученный ДОЭ имел вид кольцевой решетки из 186 колец, причем на периферии период был близок к 10 мкм при радиусе элемента 4 мм.

В случае, когда шаг поперечной дискретизации должен быть не меньше определенной величины, и соответственно, количество колец в структуре ДОЭ желательно свести к минимуму, предпочтительнее применение стохастических алгоритмов прямого поиска, так как поиск функции высоты микрорельефа можно организовать на множестве квантованных решений, к тому же алгоритмы прямого поиска обладают сходимостью.

В работах [18, 19] обсуждался расчет бинарных ДОЭ, формирующих множество фокусов, а также фокусирующих в заданные кривые в электромагнитном приближении. Однако при формировании осевых световых распределений, удовлетворяющих параксиальному приближению, применение скалярной теории дифракции является вполне корректным и ресурсосберегающим подходом.

В работе [20] с помощью стохастического алгоритма прямого поиска был рассчитан оптимизированный бинарный дифракционный оптический элемент (ДОЭ), предназначенный для формирования из гауссового пучка продольного отрезка с равномерной интенсивностью.

В данной работе проведено численное исследование действия данного оптического элемента в условиях изменения параметров освещающего пучка. Рассмотрено три типа освещающего пучка: TEM00, TEM10 и плоский, ограниченный круглой апертурой радиуса ДОЭ. Также проведено моделирование изменения длины волны в случае исполнения ДОЭ в виде фазового или амплитудного оптического элемента.

В данной работе также приведены экспериментальные результаты формирования осевого отрезка с помощью амплитудного ДОЭ при использовании лазерного излучения с различными длинами волн.

Литература

1. Wang, K. Influence of the incident wave-front on intensity distribution of the nondiffracting beam used in large-scale measurement / K. Wang, L. Zeng, and Ch. Yin // Opt. Commun. 2003. Vol. 216, P. 99-103.
2. Lu, Jian-yu Diffraction-limited beams and their applications for ultrasonic imaging and tissue characterization / Jian-yu Lu, J. F. Greenleaf // Proc. SPIE 1992. Vol. 1733, P. 92-119.
3. Leitgeb, R. A. Extended focus depth for Fourier domain optical coherence microscopy / R. A. Leitgeb, M. Villiger, A. H. Bachmann, L. Stein-mann, and T. Lasser // Opt. Lett. 2006. Vol. 31, № 16. P. 2450-2452.
4. Lee, Kye-Sung Bessel beam spectral-domain high-resolution optical coherence tomography with micro-optic axicon providing extended focusing range / Kye-Sung Lee, Jannick P. Rolland // Opt. Lett. 2008. Vol. 33, № 15. P. 1696-1698.
5. Lu, Jian-yu Producing deep depth of field and depth-independent resolution in NDE with limited diffraction beams / Jian-yu Lu, J. F. Greenleaf // Ultrason. Imag. 1993. Vol. 15, № 2. P. 134-149.
6. Lunazzi, José J. Photographing by means of a diffractive axicon / José J. Lunazzi, Daniel S. F. Magalhães // XXIX ENFMC - Annals of Optics. 2006. P. 1-4.
7. Arimoto, R. Imaging properties of axicon in a scanning optical system / R. Arimoto, C. Saloma, T. Tanaka, and S. Kawata // Appl. Opt. 1992. Vol.31, № 31. P. 6653-6657.
8. Fortin, M. Optical tests with Bessel beam interferometry / Mathieu Fortin, Michel Piché and Ermanno F. Borra // Optics Express 2004. Vol. 12, № 24. P. 5887-5895.
9. Reichelt, S. Self-calibration of wavefront testing interferometers by use of diffractive elements / Stephan Reichelt, Hans Tiziani, Hans Zappe // Proc. of SPIE 2006. Vol. 6292, P. 629205.1-629205.10.
10. McLeod, J.H. The axicon: a new type of optical element // J. Opt. Soc. Am. 1954. Vol. 44, P. 592–597.
11. Голуб, М.А. Фокусировка излучения в заданную область пространства / М.А. Голуб, С.В. Карпеев, А.М. Прохоров, И.Н. Сисакян, В.А. Сойфер // Письма в ЖТФ. 1981. Т. 7. № 10, С. 618-623.
12. Kazanskii, N.L. Correction of focuser phase function by computer-experimental methods // Computer Optics. Pergamon Press. 1989. Vol. 1, № 1. P.69-73.
13. Doskolovich L.L. Analysis of quasiperiodic and geometric optical solutions of the problem of focusing into an axial segment / L.L. Doskolovich, N.L. Kazanskiy, V.A. Soifer, A.Ye. Tzaregorodtzev // Optik. 1995. Vol 101, № 2. P.37-41.
14. Котляр В.В. Дифракционный расчет фокусаторов в продольный отрезок / В.В. Котляр, В.А. Сойфер, С.Н. Хонина // Письма в ЖТФ. 1991. Т. 17, № 24. С. 63-66.
15. Gerchberg, R. W. A practical algorithm for the determination of the phase from image and diffraction plane pictures / R. W. Gerchberg, W. O. Saxton // Optik 1972. Vol. 35, P. 237.
16. Методы компьютерной оптики (Издание второе, исправленное) / под ред. В.А. Сойфера М.: Физматлит, 2003. 688 с.
17. Волков А.В. Бинарный дифракционный оптический элемент для фокусировки гауссового пучка в продольный отрезок / Волков А.В., Котляр В.В., Моисеев О.Ю., Рыбаков О.Е., Скиданов Р.В., Сойфер В.А., Хонина С.Н. // Оптика и спектроскопия. 2000. Т. 89, № 2. С. 347-352.

18. Soifer, V.A. Multifocal diffractive elements / V.A. Soifer, L.L. Doskolovich, N.L. Kazanskiy // Optical Engineering. - 1994. - Vol.33, № 11. - P.3610-3615.

19. Сойфер, В.А. Синтез бинарного фокусатора в произвольную кривую в электромагнитном приближении / В.А. Сойфер, Н.Л. Казанский, С.И. Харитонов // Компьютерная оптика. 1996. № 16, С. 22-27.

20. Качалов, Д.Г. Оптимизация бинарных ДОЭ для формирования продольных распределений интенсивности / Д.Г. Качалов, В.С. Павельев, С.Н. Хонина // Компьютерная оптика. 2009. Т. 33, № 4. С. 441-445.

21. Хонина, С.Н. Сравнительный анализ распределений интенсивности, формируемых дифракционным аксиконом и дифракционным логарифмическим аксиконом / С.Н. Хонина, С.А. Балалаев // Компьютерная оптика Т. 33, № 2. 162-174.

Наноструктуры на базе пористого кремния для фотоэлектрических преобразователей

В.А. Китаева
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Свойства наноструктурированных сред (в том числе и пористых материалов) зачастую значительно отличаются от свойств, как объемного материала, так и свойств отдельных наночастиц, формирующих наноструктуру. Таким образом, нанокомпозитные среды являются той базой, на которой создаются новые материалы с заданными структурными, электрическими и оптическими свойствами, которые определяются размерами, формой и упорядоченностью составляющих их нанокомпонентов. В частности, возникновение искусственной анизотропии в подобных средах, усиление локального поля в микрочастицах и возможность заполнения пор веществами с высоким нелинейно-оптическим откликом открывает широкие возможности по созданию новых эффективных преобразователей частоты света, материалов с новыми оптическими и оптоэлектрическими свойствами.

В связи с этим детального исследования заслуживает влияние параметров структуры (рельеф поверхности, степень упорядоченности, форма и характерные размеры пор и нанокластеров, толщина наноструткурированного слоя) на характеристики конкретного нанокомпозитного материала.

Равным образом важным является изучение возможности управления свойствами наноструктурированных, в частности пористых, материалов, путем вариации режимов формирования. Установление зависимостей характеристик нанопористых слоев от параметров их получения – важная задача, решение которой позволит создать математические модели процесса порообразования и применять их, впоследствии, в производстве новых материалов с определенным, заведомо известным, набором свойств.

Методология формирования слоев пористого кремния (ПК)

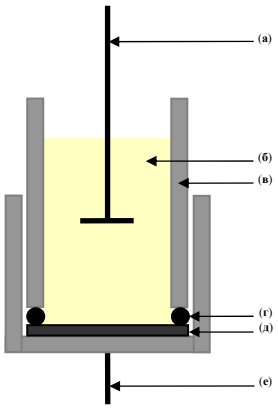


Рис. 1 – Электро-химическая ячейка

Лидирующее положение среди материалов для научных исследований занимает кремний в связи с рядом его уникальных свойств (он технологичен, стоек к высоким температурам и агрессивным химическим средам, имеет высокий коэффициент диффузии легирующих элементов и пр.).

Традиционным методом получения слоев ПК является электрохимическое анодное травление пластин монокристаллического кремния в растворах, содержащих плавиковую кислоту [1-8].

На рисунке 1 представлено схематическое изображение ячейки для электрохимического травления (а – электрод; б – электролит; в – стенки ячейки; г – кольцевая прокладка; д – подложка, на которой и формируется пористый слой; е – электрический контакт).

Параметрами, влияющими на характеристики пор, являются природа электролита, его концентрация и температура в процессе воздействия, а также плотность анодного тока.

Для формирования пористого слоя можно применять как гальваностатический, так и потенциостатический режимы, которые позволяют с высокой точностью контролировать глубину залегания пор по изменению напряжения или плотности тока соответственно.

На сегодняшний день существует два основных подхода для создания пористых материалов:

- саморегулируемое формирование зародышей для роста макропор, последующее травление с образованием пор на месте зародышей [1,2];
- предварительное создание V-образных затравок и последующее электрохимическое травление кремния [2-5].

Образцы и методика эксперимента

В эксперименте использовались кремниевые подложки ориентации 100, с удельным поверхностным сопротивлением 45 Ом/см^2, р-типа проводимости.

На первом этапе эксперимента на поверхности подложек путем травления в горячем насыщенном растворе NaOH был сформирован микрорельеф, представленный на рисунке 2.

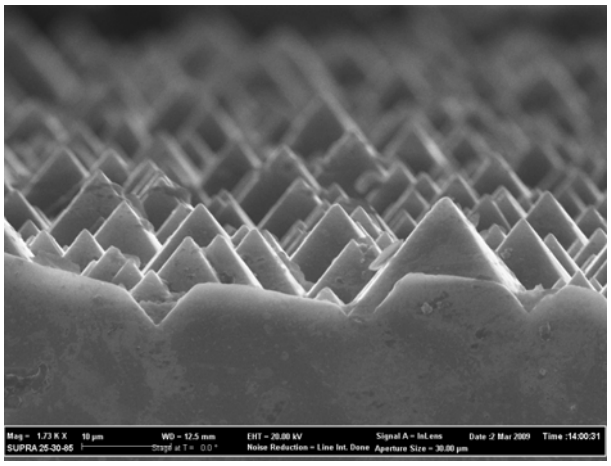


Рис. 2 . РЭМ-изображение микрорельефа поверхности кремниевой пластины используемой в процессе получения пористого кремния

Размеры полиэдров травления, «пирамид», образующих микрорельеф варьируются в пределах от 2 до 12 мкм.

Подложки монокристаллического кремния с уже сформированным микрорельефом подвергались электрохимическому травлению в водноспиртовом растворе плавиковой кислоты. В таблице 1 представлены параметры электрохимической обработки.

Таблица 1

Образец	Состав электролита	Плотность анодного тока, мА/см^2
№1	HF – 1 моль/л; C_2H_5OH – 4 моль/л	4,29
№2	HF – 1,2 моль/л; C_2H_5OH – 5,3 моль/л	
№3	HF – 1,4 моль/л; C_2H_5OH – 6,4 моль/л	
№4	HF – 1,9 моль/л; C_2H_5OH – 8,5 моль/л	
№5	HF – 2,4 моль/л; C_2H_5OH – 10,7 моль/л	

Анализ результатов

С помощью спектрофотометра Shimadzu IRPrestige-21 были получены спектры отражения в видимом диапазоне для исходной текстурированной пластины монокристаллического кремния и образцов №1-5 (рисунки 3-8).

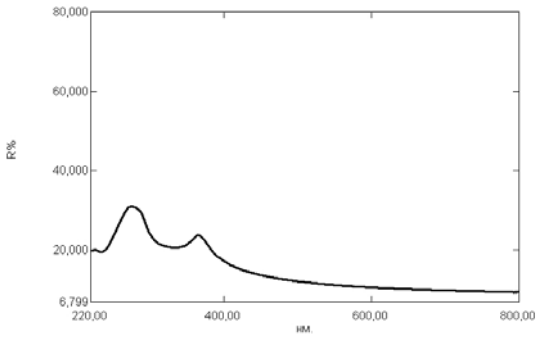


Рис. 3.– Спектр отражения исходной пластины в видимом диапазоне длин волн

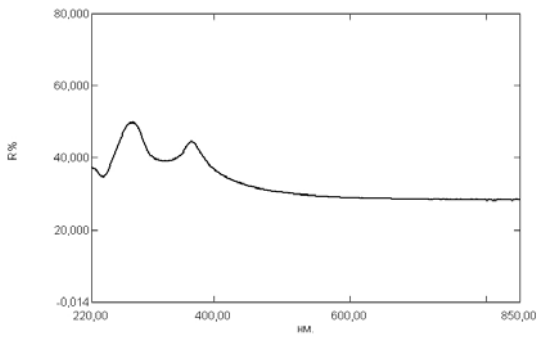


Рис. 4. Спектр отражения образца №1 в видимом диапазоне длин волн

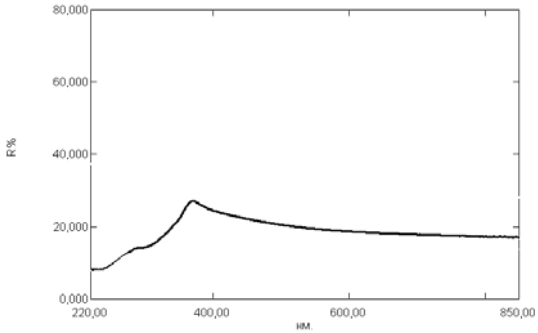


Рис.5. Спектр отражения образца №2 в видимом диапазоне длин волн

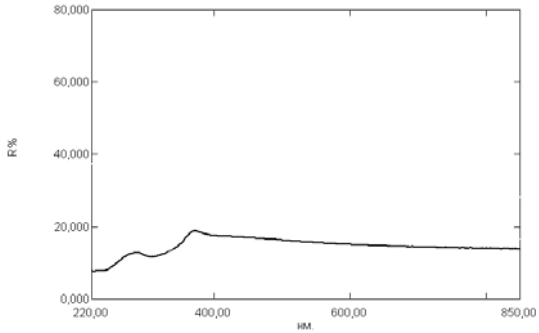


Рис. 6. Спектр отражения образца №3 в видимом диапазоне длин волн

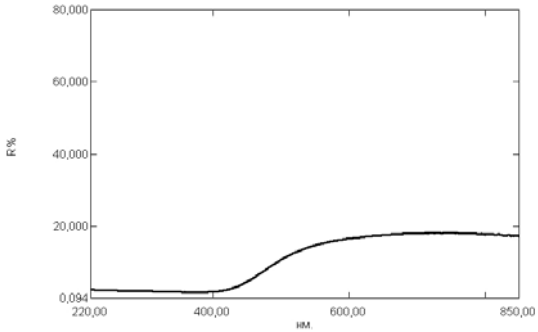


Рис.7. – Спектр отражения образца №4 в видимом диапазоне длин волн

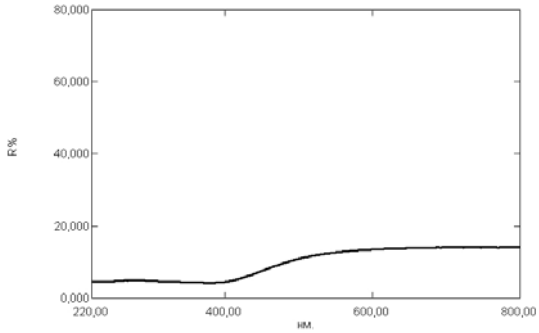


Рис.8. – Спектр отражения образца №5 в видимом диапазоне длин волн

Представленные выше графики иллюстрируют ярко выраженную зависимость оптических свойств получаемых наноструктурированных материалов от параметров процесса их формирования.

Так образцы, полученные при больших концентрациях HF в составе электролита, обладают значительно большей чувствительностью в фиолетовой и ультрафиолетовой области спектрального диапазона, чем исходный материал с микрорельефом, традиционно используемый в производстве солнечных батарей, или чем образцы, полученные при меньших концентрациях HF.

Анализ РЭМ-изображений поверхности и сколов образцов, полученных при одной и той же величине анодного тока показывает, что изменение концентрации плавиковой кислоты (HF) в составе электролита даже на доли процента приводит к существенным изменениям структуры и свойств наноструктурированного материала.

Так при низком содержании HF получаемая структура полностью повторяет изначальный микрорельеф пластины в виде полиэдров травления, поры же преимущественно располагаются на стыках пирамид (рисунок 9, рисунок 10).

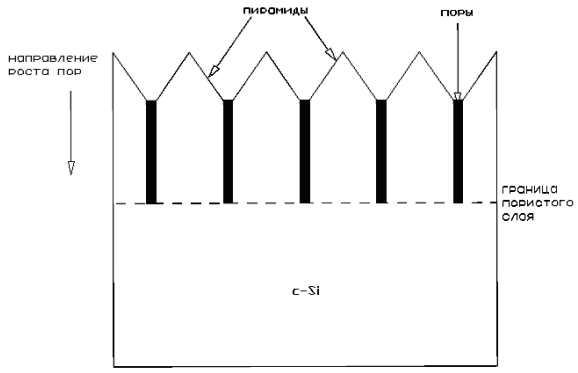


Рис. 9. Схематичное изображение структуры пористого слоя полученного на текстурированной поверхности кремниевой пластины при низких концентрациях HF

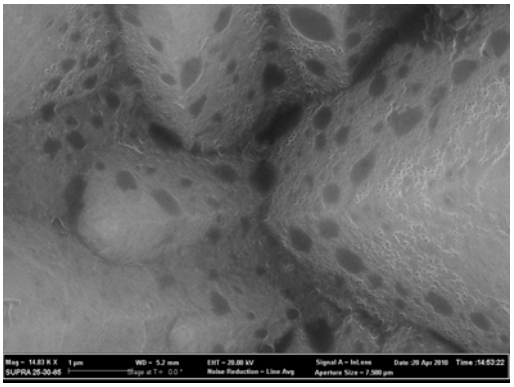


рис. 10. РЭМ-изображение поверхности образца №2 вблизи стыков «пирамид»

Подобные структуры можно рассматривать в виде матрицы микронных кремниевых «столбиков» отделенных друг от друга участками нанопористого кремния, имеющего заметно большее сопротивление. Наблюдается значительная анизотропия токопроводимости [6]. Поэтому перспективным направлением использования подобных структур является создание фоточувствительных матриц позволяющих регистрировать с высоким разрешением пространственное распределение интенсивности излучения.

Чем выше концентрация HF в составе электролита, тем большая доля поверхности «пирамид» подвергается протравливающему эффекту, который представляет собой совокупность двух процессов: порообразования непосредственно на гранях и сглаживания рельефа (рисунок 5).

В результате образованный пористый слой по сравнению со слоями полученными при меньших концентрациях HF обладает большей пористостью, более равномерным распределением пор и более развитой площадью поверхности [5,7], что важно для применения подобных структур в качестве светопоглощающих элементов солнечных батарей [9,10].

Литература

1. Головань, Л.А. Оптические свойства нанокомпозитов на основе пористых сисем/Л.А. Головань, В.Ю. Тимошенко, П.К. Кашкаров//УФН.

– 2007. – №6. – С. 619-638.

2. Lehmann, V. Formation Mechanism and Properties of Electrochemically Etched Trenches in n-Type Silicon/V. Lehmann, H.Foll//J.Electrocherm. Soc.

– 1990. – №2. – P. 653-659.

3. Ottow, S. Processing of Three-Dimensional Microstructures Using

Macroporous n-Type Silicon/S. Ottow, V. Lehmann, H. Föll//J.Electrocherm. Soc. – 1996. – №1. – P. 385-390.

4. Астрова, Е.В. Технология создания рисунка в макропористом кремнии и получение полос двумерных фотонных кристаллов с вертикальными стенками/Е.В. Астрова, Т.Н. Боровинская, В.А. Толмачев, Т.С. Перова//Физика и техника полупроводников. – 2004. – №9. – С. 1125-1128.

5. Vyatkin, A. Random and ordered macropore formation in p-type sisicon/A. Vyatkin, V. Starkov, V. Tzeitlin//J.Electrocherm. Soc. – 2002. – №1.

– P. 70-76.

6. Мартышев, М.Н. Исследование ориентационной зависимости электропроводности в слоях анизотропного пористого кремния/П.А. Форш, Л.В. Шапошников, В.Ю. Тимошенко, П.К. Кашкаров//Материалы электронной техники. – 2008. – №4. – С. 35-38

7. Ерохов, В.Ю. Кремний мультипористой текстуры для фотоэлектрических преобразователей солнечной энергии/А.А. Дружинин//Технология и конструирование в электронной аппаратуре. – 2009. – №3. – С. 21-23

8. Горячев Д.Н., Электролитический способ приготовления пористого кремния с использованием внутреннего источника тока/Д.Н. Горячев, Л.В. Беляков, О.М. Сресели//Физика и техника полупроводников. –2003. –№ 4. –С. 494-498.

9. Латухина, Н.В. «Фотоэлектрические свойства структур с микро- и нано-пористым кремнием»/Т.С.Дереглазова, С.В. Ивков, А.В.Волков, В.А. Деева//Известия СНЦ РАН. –2009. – Т.11. – №3.– С. 66-70.

10. Н.В. Латухина, Н.А.Нечаева, В.А.Храмков, А.В.Волков, А.Н.Агафонов Структуры с макропористым кремнием для фотопреобразователей на кремниевой подложке // Харьковская нанотехнологическая Ассамблея. Т.2. Сборник докладов. 18 Международного симпозиума «Тонкие пленки в оптике и наноэлектронике» г.Харьков: ННЦ «ХФТИ». ИПП «Контраст». – 2006. – С. 207-211.

Субволновая фокусировка с помощью планарной гиперболической секансной линзы

А.А. Ковалев, Я.Р. Триандафилов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара
Институт систем обработки изображений РАН, Самара

В последнее время возрос интерес к градиентным и фотонно-кристаллическим линзам, которые способны обеспечить субволновую фокусировку лазерного света [1-5] и применяются для изображения отдельных молекул [6] и для ультракомпактного сопряжения планарных волноводов разной ширины [7]. В качестве планарной градиентной линзы используется линза, показатель преломления которой зависит от поперечной координаты как гиперболический секанс. Гиперболическая секансная (ГС) линза имеет свою долгую историю. Еще в 1930 году П.С. Эпштейн [8] рассмотрел задачу расчета мод для градиентного волновода со сложным показателем преломления, обобщающим ГС-профиль. В 1951 году А.Л. Микаэлян [9] нашел, что ГС-профиль показателя преломления является оптимальным для фокусировки света. Поэтому ГС-линза Микаэляна является частным случаем градиентного волновода Эпштейна. Далее задача распространения света в ГС-волноводе и ГС-линзе решалась в геометрооптическом [10,11], квазиоптическом [12] и волновом [13-14] приближениях. В [15,16] описаны экспериментальные результаты по фокусировке света с помощью ГС-линзы. В [17] ГС-линза используется для сверхразрешения совместно с рефракционной и дифракционной линзами. В известной монографии М. Адамса [18] дан обзор работ по волноводам Эпштейна и ГС-волноводам.

В данной работе ТЕ-моды ГС-линзы выражены через ортогональные многочлены Якоби. Показано, что любое световое поле, распространяющееся в ГС-волноводе будет периодически повторяться с периодом Тальбота, который не зависит от длины волны света. В общем случае и на конкретном примере показано, что для любого четного светового поля в ГС-волновое на оптической оси через расстояние, равное четверти периода Тальбота, интенсивность будет иметь экстремумы. Для ГС-линзы, являющейся куском ГС-волновода, показано, что основная мода Фурье-инвариантна. Это позволило оценить ширину фокусного пятна по полуспаду интенсивности. Планарная ГС-линза - это двумерная градиентная среда с показателем преломления вида $n(x) = n / \mathrm{ch}(x / a)$, где x – поперечная координата линзы (продольная координата – z), a – ширина линзы, при которой показатель преломления на оптической оси n уменьшается в 1.54 раза. Уравнение Гельмгольца для амплитуды проекции электрического вектора световой волны с ТЕ-поляризацией $E_y(x,z)$ имеет вид:

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \frac{k^2 n^2}{ch^2(x/a)}\right] E_y(x,z) = 0, \qquad (1)$$

где k - волновое число света. Модовое решение уравнения (1) имеет вид (ТЕ-мода):

$$E_y(x,z) = \exp(i\beta z / a)(1 - y^2)^{\beta/2} P_m^{(\beta,\beta)}(y), \qquad (2)$$

где $2\beta = \sqrt{1 + 4(akn)^2} - (2m+1)$, m – целое неотрицательное число, β – постоянная распространения моды, $y = \mathrm{th}(x/a)$, $P_m^{(\beta,\beta)}(y)$ - многочлены Якоби [19]. Моды (2) ортогональны и имеют конечную энергию при $\beta > 0$. Две первые моды из (2) имеют вид: $\Psi_0(x) = \mathrm{ch}^{-\beta_0}(x/a)$, $\Psi_1(x) = \beta_0 \,\mathrm{sh}[x/a]\,\mathrm{ch}^{-\beta_0}[x/a]$, где

$2\beta_0 = \sqrt{1+4(akn)^2} - 1$ или $\beta_0 = \beta + m$. Любое решение уравнения (1) можно разложить по базису (2):

$$E_y(x,z) = \exp(i\beta_0 z/a)\sum_{m=0}^{\infty} C_m \exp(-imz/a)\Psi_m(x), \qquad (3)$$

где $\Psi_m(x) = (1-y^2)^{\beta/2} P_m^{(\beta,\beta)}(y)$, C_m – коэффициенты. Из (3) следует, что при $z_p = 2\pi ap$, p =1,2,3…, а для четных и нечетных функций $E_y(x,z)$ при $z_p = \pi ap$, модуль комплексной амплитуды (3) будет периодически повторяться. Поэтому расстояние $L=2\pi a$ называется периодом Тальбота. Заметим, что период Тальбота не зависит от длины волны. Это значит, что в ГС-линзе нет хроматической аберрации: все длины волн имеют одинаковое фокусное расстояние. Моделирование это подтвердило. Покажем, что на расстояниях $z_p = \pi ap/2$ в распределении интенсивности на оптической оси (x = 0) для четных световых полей (3) имеют место экстремумы. Действительно, выражение для интенсивности светового поля (3) имеет вид:

$$I(x=0,z) = \sum_{p=0}^{\infty} C_{2p}^2 \Psi_{2p}^2(0) \pm 2\sum_{p<q} \left| C_{2p} C_{2q} \Psi_{2p}(0)\Psi_{2q}(0) \right| \cos\left[2(p-q)z/a\right]. \qquad (4)$$

Величины C_{2p} и $\Psi_m(0)$ действительные. Это следует из действительности модовых функций (2) при z = 0 и из предположения, что рассматриваемое световое поле $E_y(x,z)$ – четное и действительное при z = 0. Четность светового поля требуется для того, чтобы на оптической оси не было тождественного нуля интенсивности. Производная от интенсивности имеет вид:

$$\frac{dI(x=0,z)}{dz} = \pm\frac{4}{a}\sum_{p<q} \left| C_{2p} C_{2q} \Psi_{2p}(0)\Psi_{2q}(0) \right| (p-q)\sin\left[2(p-q)z/a\right]. \qquad (5)$$

Из (5) следует, что производная равна нулю при $z_p = \pi ap/2$. То есть в этих точках на оптической оси будут иметь место максимумы и минимумы интенсивности. И так как в точках $z_p = \pi ap$ интенсивность будет повторяться, то в промежуточных точках $z_p = \pi a/2 + \pi ap$ световое поле будет либо коллимироваться (минимум интенсивности), либо фокусироваться (максимум интенсивности). Это можно показать также на конкретных примерах. Например, с помощью справочного ряда из [20] можно получить:

$$E_y(x,z) = \frac{\exp(i\beta_0 z/a)}{ch^{\beta_0}\left[x/a\right]} \sum_{m=0}^{\infty} \frac{\exp(-imz/a) P_m^{(\beta_0-m,\beta_0-m)}(y)}{(1-y^2)^{m/2}} = \frac{\exp(i\beta_0 z/a)}{ch^{\beta_0}\left[x/a\right]} \left[1 - sh\left[x/a\right] e^{-iz/a} + \frac{e^{-i2z/a}}{4} \right]^{\beta_0}. \qquad (6)$$

Световое поле (6) может распространяться в ГС-линзе, так как функция (6) удовлетворяет уравнению (1). На оптической оси будем иметь интенсивность:

$$I(x=0,z) = \left[\frac{17}{16} + \frac{\cos(2z/a)}{2} \right]^{\beta_0}. \qquad (7)$$

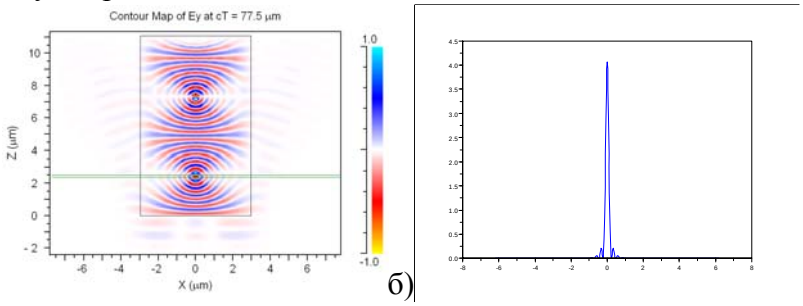
Из (7) следует, что при z = 0 на оси будет максимум $I_{\max} = (5/4)^{2\beta_0}$, а при $z_p = \pi a/2$ минимум $I_{\min} = (3/4)^{2\beta_0}$. Из наличия экстремумов осевой интенсивности в точках $z_p = \pi ap/2$ можно предположить , что амплитуды световых полей в ГС-линзе на плоскостях, разделенных расстоянием $z = \pi a/2$, связаны между собой преобразованием Фурье. Это также подтверждается тем, что с помощью лучевой оптики показано, что все лучи в ГС-линзе, параллельные оптической оси (плоская волна, падающая вдоль оптической оси) собираются на расстоянии $z = \pi a/2$ в одну точку фокуса на оптической оси [7]. Используя это предположение, оценим размер фокусного пятна в ГС-линзе. Пусть на входе ГС-линзы при z = 0 имеется основная мода $\Psi_0(x) = ch^{-1}(x/b)$, ширина которой b отличается от ширины ГС-линзы a. Известен интеграл [21]:

$$\int_{-\infty}^{\infty} \frac{\exp(idx\xi)}{\mathrm{ch}(x/b)}dx = \frac{\pi b}{\mathrm{ch}(\pi db\xi/2)}, \qquad (8)$$

где ξ – поперечная координата в месте фокуса. Интеграл (8) показывает, что основная мода ГС-линзы при $\beta_0 = 1$ Фурье-инвариантна. Далее найдем параметр d в (8). Если в ГС-линзе распространяется основная мода с шириной $b = a$, то она при распространении не меняет свою ширину. Тогда из (8) получим $d = 2/(\pi a^2)$. Фокусное расстояние в ГС –линзе равно ¼ от периода Тальбота $f = \pi a/2$, тогда получим: $d = 1/(af)$. Еще одно условие на ширину моды a следует из того, что $\beta_0 = 1$: $a = \sqrt{2}/(kn)$. Тогда получим окончательное выражение для параметра d: $d = kn/(f\sqrt{2})$. С помощью (8) найдем ширину фокусного пятна по полуспаду интенсивности, исходя из условия $|E_0(\xi, z = f)|^2 = 0.5(kb/2)^2$. Тогда ширина фокуса по полуспаду интенсивности на расстоянии $z = f = \pi a/2$ от начала $z = 0$ равна:

$$FWHM = \frac{\lambda a}{\sqrt{2}\pi nb}\ln(3+2\sqrt{2}) = 0.396\frac{\lambda a}{nb}. \qquad (9)$$

При $n = 3.47$ (кремний) получим $FWHM = 0.115\lambda a/b$. Заметим, что скалярный дифракционный предел в 2D среде равен $FWHM = 0.44\lambda/n$. Поэтому при $b = a$ значение ширины фокуса будет немного меньше, чем дифракционный предел. Предельная ширина фокуса в 2D среде $FWHM = 0.44\lambda/n$ получена при максимальной числовой апертуре $NA = n$. Хотя у ГС-линзы числовая апертура меньше. Числовую апертуру ГС-линзы можно найти из уравнения луча в градиентной среде $n(x)\cos\theta(x) = const$, где θ – угол наклона касательной к лучу к оптической оси z. Если начальный луч падает на линзу параллельно оси z на расстоянии от нее $x = R$, где R – радиус ГС-линзы, который находится из условия $n(R) = 1$ и равен $R = arcch(n)$. Тогда получим, что $n(R) = 1 = n(x)\cos\theta(x) = n(0)\cos\theta(0) = n\cos\theta$, где, как и раньше, n – показатель преломления на оптической оси ($x = 0$) и θ - угол между касательной к лучу и оптической осью в точке пересечения лучом этой оси. Из предыдущей цепочки равенств следует, что $\cos\theta = 1/n$ и поэтому числовая апертура ГС-линзы равна: $NA = n\sin\theta = \sqrt{n^2 - 1}$. Она зависит только от показателя преломления на оптической оси. Например, при $n = 3.47$ имеем $NA = 3.3$. Поэтому дифракционный предел для 2D ГС-линзы будет равен $FWHM = 0.44\lambda/NA = 0.44\lambda/\sqrt{n^2 - 1}$.



а) б)

Рис.1 Мгновенная картина реальной части амплитуды $E_y(x,z)$ в ГС-линзе (а), свет падает снизу вверх, и усредненное распределение интенсивности $|E_y(x,z)|^2$ в фокусе (на горизонтальной линии на рис.1а) (б)

На рис.1 показаны результаты моделирования с помощью FDTD-метода фокусировки ТЕ-волны основной моды с шириной $b \neq a$ с помощью ГС-линзы: мгновенная картина реальной части амплитуды $E_y(x,z)$ (а) и усредненное распределение интенсивности $|E_y(x,z)|^2$ в фокусе линзы (б). Параметры моделирования: длина волны

λ = 1.55 мкм, параметр линзы $a = \lambda$, $b = 4\lambda$, длина линзы вдоль оси z от начала до первого фокуса $L_1 = \pi a/2$ = 2435 нм, ширина линзы $W = 2a\ln\left(n_0 + \sqrt{n_0^2 - 1}\right)$ = 5.94 мкм, показатель преломления на оси n = 3.47, дискретность сетки отсчетов – 64 отсчета на длину волны. Ширина фокусного пятна (рис.1б) по полуспаду интенсивности равна FWHM = 0.132λ. Такая же ширина фокусного пятна будет в непосредственной близости от выходной поверхности линзы, если линзу отрезать по горизонтальной линии на рис.1а.

Дифракционный предел в среде для ГС-линзы $FWHM = 10.44\lambda / \sqrt{n^2 - 1}$ для кремния (n = 3.47) равен FWHM= 0.133λ.

ГС-линза была изготовлена в виде планарной фотонно-кристаллической (ФК) линзы, диаметр отверстий которой увеличивался от оптической оси к периферии. Линза была применена для сопряжения двух планарных волноводов разной ширины (рис.2).

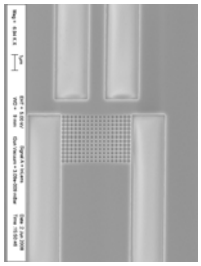
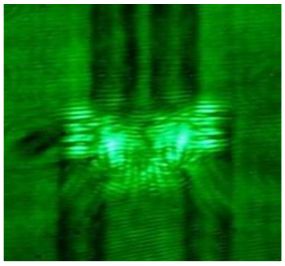
а)  б) 

Рис.2 Изображение сверху в СЭМ фотонно-кристаллической ГС-линзы в пленке кремния (а) вместе с двумя планарными волноводами и вид сверху прохождения света с длиной волны 532 нм через эту линзу (свет распространяется снизу вверх)

На рис.2а показан вид сверху в электронном микроскопе ГС-линзы, выполненной в виде фотонного кристалла (диаметры отверстий увеличиваются от центра к периферии от 180 нм до 220 нм) в пленке кремния толщиной 200 нм вместе с двумя планарными волноводами разной ширины (4.5 мкм и 1 мкм). На рис.2б показано сверху прохождение света с длиной волны 532 нм через эту линзу. На рис.2б видно, что внутри линзы образуется стоячая волна (интерференция волн, падающих и отраженных от задней грани ГС-линзы), у которой интерференционные полосы изогнуты в виде дуг окружностей, похожих на полосы на рис.1а (только период этих полос в 3 раза меньше, чем на рис.1а, так как длина волна в 3 раза меньше). Из рис.2б также видно, что интерференционные полосы в местах, где нет линзы, имеют вид отрезков прямых. Поэтому свет из широкого волновода (снизу на рис.2) фокусируется линзой в узкий волновод (сверху на рис.2).

Таким образом, в работе показано, что любое монохроматическое световое поле с ТЕ-поляризацией, распространяющееся в планарной ГС-линзе представимо в виде линейной комбинации ортогональных световых мод, амплитуда которых выражается через многочлены Якоби. Показано, что это световое поле будет периодически повторяться с периодом Тальбота и фокусироваться с периодом, равным половине периода Тальбота. Причем величина этого периода не зависит от длины волны. Световое поле, амплитуда которого пропорциональна основной моде ГС-линзы, но ширина которой не совпадает с шириной основной моды, будет Фурье-инвариантно. В фокусе ГС-линзы такое поле будет сохранять вид основной моды, отличаясь только масштабом. С помощью ГС-линзы можно получить фокусное пятно, с шириной равной дифракционному пределу в среде (для кремния этот предел равен 0.133 от длины волны). При этом интересно, что числовая апертура ГС-линзы зависит только от показателя преломления на оптической оси, и не зависит от ее геометрических параметров. ГС-линза была реализована в виде планарной фотонно-кристаллической линзы в пленке кремния для длины волны 1.55 мкм. Сходящейся волновой фронт в этой линзе был визуализирован с помощью света с длиной волны 532 нм.

Литература

1. X. Wang, Z.F. Ren, K. Kempe. Unrestricted superlensing in a triangular twodimensional photonic crystal. Opt. Express 2004. – Vol. 12. – p. 2919- 2924.

2. Q. Wu, J.M. Gibbons, W. Park. Graded negative index lens by photonic crystal. Opt. Express 2008. – Vol. 16., №21. – p.16941-16949.

3. I.V. Minin, O.V. Minin, Y.R. Triandafilov, V.V. Kotlyar. Subwavelength diffractive photonic crystal lens. Prog. Electromag. Res. B 2008. – Vol. 7. – p.257-264.

4. I.V. Minin, O.V. Minin, Y.R. Triandafilov, V.V. Kotlyar. A photonic crystal Mikaelian lens. Opt. Mem. Neur. Netw. 2008. – Vol. 17. – p.1-7.

5. S.Kameda, A. Mizutani, H. Kikuta. Numerical study on a micro prizm and micro lenses with metal-dielectric multilayered structures. J. Opt. Soc. Am. A, v.27, no.4, p.749-756 (2010).

6. J.J. Schwarz, S. Stavrakis, S.R. Quake. Colloidal lenses allow high-temperature single-molecule imaging and improve fluorophore photostability. Nature Nanotechnology, v.5, p.127-132 (2010).

7. M.I. Kotlyar, Y.R. Triandafilov, A.A. Kovalev, V.A. Soifer, M.V. Kotlyar, L. O'Faolain. Photonic crystal lens for coupling two waveguides. Appl. Opt. 2009. – Vol. 48. N 19. – p.3722-3730.

8. P.S. Epstein// Proc. Nat. Acad. Sci. USA 1930. – Vol. 16. – p. 627-237.

9. A.L. Mikaelian. Application of stratified medium for waves focusing. Doklady Akademii Nauk SSSR 1951. – Vol. 81. – P. 569–571.

10. G. Rawson, D.R. Herriott, J. MacKenne. Analysis of refractive index distributions in cylindrical, gradient-index glass rods used as image relays. Appl. Opt. 1970. – Vol. 3., N 3 – p.753-759.

11. G. Rawson, D.R. Herriott, J. MacKenne. Gradient-index tapered hyperbolic second planar waveguide for focusing, collimation and beam-size control. J. Opt. Sos. Am. A. 1997. – Vol. 14., N 8 – p.1754-1759.

12. J. Linares, C. Gomez-Reino. Optical propagator in a graded-index medium with a hyperbolic secant refractive-index profile. Appl. Opt. 1994. – Vol. 33., N 16 – p. 3427-3431.

13. W. Streifer, C.N. Kurz Scalar analysis of radially inhomogeneous guiding media. J.Opt. Soc. Am. 1967. – Vol. 57., N 6 – p. 779-786.

14. C.A. van Duin, J. Boersma, F.W. Sluijter. TM-modes in a planar optical waveguide with a graded index of the symmetric Epstein type. Wave Motion 1986. – Vol. 8., N 2 – p. 175-190.

15. D.W. Hewak, J.W.Y. Lit Solution deposited optical waveguide lens. Appl. Opt. 1989. – Vol. 28., N 19 – p. 4190-4198.

16. J.M. Rivas-Moscoso, D. Nieto, C. Gomez-Reino, C.R. Fernandez-Pousa. Focusing of light by zone plates in selfoc gradient-index lenses. Opt. Lett. 2003. – Vol. 28., N 22 – p. 2180-2182.

17. J.M. Rivaz-Moscoso, C.R. Fernandez-Pousa, C. Gomez-Reino. Hybrid refractive-diffractive-gradient-index superresolving focusing device. Appl. Opt. 2008. – Vol. 47., N 22 – p.E68-E75.

18. M. Adams. *An Introduction to Optical Waveguides*, J. Wiley and Sons, Chichester, New York, Brisbane, Toronto, 1981.

19. *Handbook of Mathematical Functions* / ed. M. Abramowitz, I.A. Stegun.- National Bureau of Standards, Washington, DC, 1964.- 1044p.

20. А.П. Прудников, Ю.А. Брычков, О.И. Маричев «Интегралы и ряды. Специальные функции», М., Наука, 1983.

21. А.П. Прудников, Ю.А. Брычков, О.И. Маричев «Интегралы и ряды», М., Наука, 1981.

Формирование светового поля для линейного перемещения микрообъекта

А.А. Морозов
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В микромеханике есть задача по перемещению и сортировке микрочастиц с помощью ДОЭ, формирующего изображение полосы с градиентом фазы вдоль нее. При сортировке частиц таким ДОЭ более тяжелые частицы будут проходить через полосу, а более легкие останутся зафиксированными. Существуют различные варианты формирования такого светового поля. Например, ДОЭ, формирующие заданное распределение интенсивности в одной плоскости, они известны довольно давно [1-3]. Есть работы, в которых фокусаторы формировали заданную интенсивность, и при этом световое поле обладало и некоторой заранее известной фазой [4-5]. В [4] рассматривался метод расчета фокусаторов, которые позволяют формировать заданное амплитудно-фазовое распределение практически без ограничений на форму этого распределения, но при этом использовалось амплитудное кодирование, за счет которого была достаточно большая потеря энергии. В [5] световой пучок имел вихревую фазу с заданным угловым орбитальным моментом, т.е. фактически фокусатор формировал вихревое световое поле с заданной интенсивностью, в данном методе, как отмечают сами авторы, могут быть использованы только многоуровневые ДОЭ, которые гораздо сложней в изготовлении. Также есть различные работы напрямую связанные с сортировкой микрочастиц [6-7]. В [7] изображение полоски для сортировки частиц формируется вообще без использования ДОЭ за счет преломления в оптической системе микрообъектива, но при этом сильно ограничены длина полоски и область ее формирования, а также происходят сильные потери энергии. В настоящей работе рассматривается метод расчета ДОЭ формирующего световое поле для линейного перемещения микрообъекта. В данной работе вместо расчета специального фокусатора, требуемый элемент рассчитывается на основе линейно закодированного оптического «вихря» и формирует полоску с направленным вдоль нее градиентом фазы. Также рассматриваются эксперименты по микроманипуляции в световых полях сформированных такими ДОЭ.

Расчет элемента

Расчет элемента производится на основе линейно кодированного оптического «вихря». Кодирование фазовой функции оптического «вихря» для приведения ее к бинарному виду осуществляется по формуле:

$$\varphi_b(x,y) = \arg\left(\varphi^*(x,y)\cdot e^{-i\alpha x} + \varphi(x,y)\cdot e^{i\alpha x}\right), \qquad (1)$$

где α- частота несущей, φ - исходная фаза, φ_b – новая бинарная фаза.

При линейном кодировании оптического «вихря», для получения бинарного элемента, на получаемом изображении вместо одного кольца формируется сразу два, причем их орбитальные угловые моменты направлены в противоположные стороны. Если менять такой параметр как частота линейного кодирования будет меняться расстояние между кольцами (рис. 1), при определенной частоте можно получить изображение, на котором кольца будут соприкасаться, а в месте соприкосновения будет формироваться полоска. При этом то, что орбитальные угловые моменты колец направлены в противоположные стороны обеспечивает сильный градиент фазы направленный вдоль полоски, как показано на рис. 2.

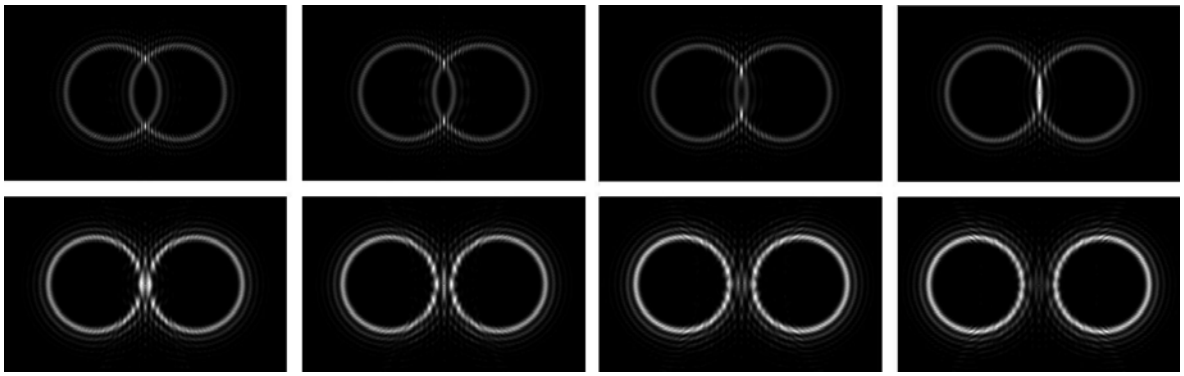


Рис. 1. Изменения дифракционной картины для ДОЭ формирующего оптический вихрь с n=60 при изменении частоты кодирования от 18 500 м$^{-1}$ до 36 000 м$^{-1}$ с шагом 2 500 м$^{-1}$ (справа налево и сверху вниз частота увеличивается)

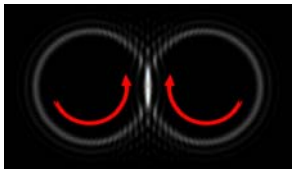


Рис. 2 . Направления закрутки колец для ДОЭ формирующего вихрь с n=60

Также следует отметить, что для формирования необходимой полоски подходят оптические «вихри» только с четным порядком сингулярности, т.к. при нечетном порядке в месте соприкосновения колец вместо одной прямой полоски формируется две дуги (рис. 3).

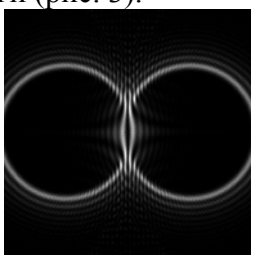
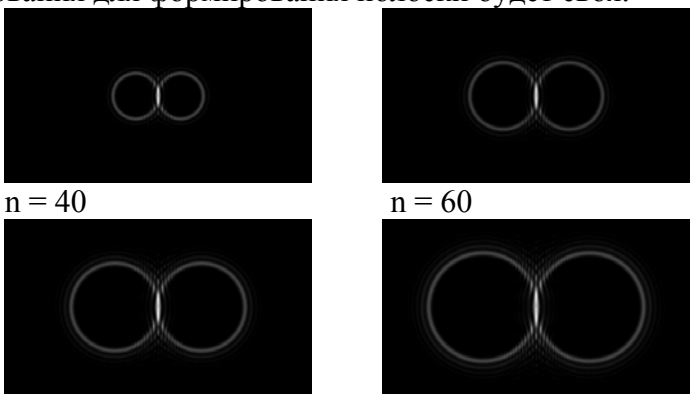


Рис. 3. Картина, возникающая при нечетном порядке сингулярности (n=101)

При увеличении порядка сингулярности увеличивается диаметр колец, а также и длина полоски (рис. 4), но при этом для каждого порядка оптимальная частота кодирования для формирования полоски будет своя.



n = 40 n = 60

n = 80 n = 100

Рис. 4. Дифракционные картины, формирующиеся при разных порядках оптических «вихрей»

Расчет производился с помощью моделирующей программы, которая позволяла рассчитывать ДОЭ с заданными параметрами кодирования и порядка сингулярности, а

также моделировала дифракционную картину формируемую таким элементом. Для более удобного расчета в данную программу была добавлена возможность рассчитывать совместно несколько ДОЭ с различной частотой кодирования с заданным шагом, получать от них изображения колец и измерять длину и эффективность полоски. Под эффективностью в данном случае понималось отношение энергии попадающей в полоску к энергии, идущей на формирование всей дифракционной картины. По результатам расчетов программы было выбрано пять ДОЭ формирующие оптические «вихри» с различным порядком сингулярности. Эти ДОЭ имеют оптимальную частоту кодирования, при которой образуется наиболее длинная и эффективная полоска.

Таблица 1. Зависимость эффективности и длины полоски от сингулярности при оптимальной частоте кодирования

сингулярность	частота кодирования, $м^{-1}$	Эффективность, %	длина полоски, pixel
20	9060	18,244	17
40	17760	16,004	24
60	26260	11,386	29
80	34320	8,4575	38
100	42360	6,5517	46

Из таблицы 1 можно отметить, что при возрастании сингулярности помимо увеличения длины полоски происходит уменьшение ее эффективности, поэтому в зависимости от условий задачи нужно будет выбирать что важнее: длина полоски или ее эффективность.

Экспериментальная часть

Все пять ДОЭ рассчитанных выше были изготовлены на стекле и с помощью них были получены изображения дифракционной картины. Для этого на столе была собрана соответствующая оптическая схема, представленная на рис. 5.

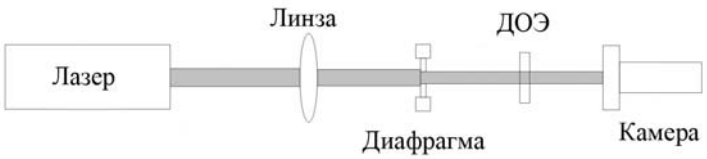
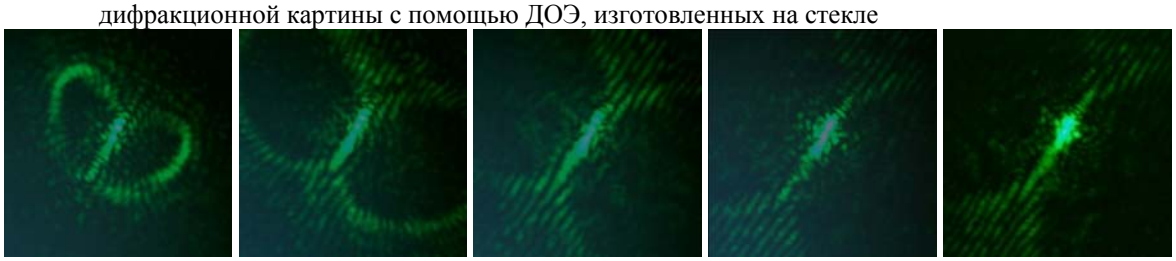


Рис. 5. Общий вид оптической схемы для получения дифракционной картины с помощью ДОЭ, изготовленных на стекле



а б в г д

Рис. 6. Изображения, полученные с помощью ДОЭ, изготовленных на стекле и формирующих оптические «вихри» с n=20(а), n=40(б), n=60(в), n=80(г), n=100(д)

Как видно из рис. 6, данные ДОЭ тоже дают не самое качественное изображение, и в середине полоски также наблюдается усиление интенсивности. Это может быть связано с небольшой погрешностью глубины элементов полученной при

изготовлении. Тем не менее, с этими элементами был произведен эксперимент по манипуляции с помощью них микрочастицами.

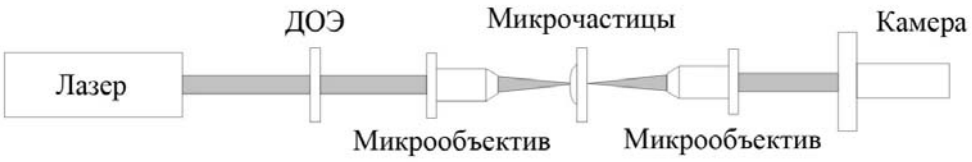


Рис. 7. Общий вид оптической схемы для осуществления микроманипуляции

Для этого на оптическом столе была собрана специальная схема (рис. 7). ДОЭ располагался на удалении от лазера, так чтобы выходящий луч успел расшириться и перекрыть весь элемент. После ДОЭ лазерный луч фокусировался микрообъективом до размеров подходящих для манипуляции частицами размером 5мкм. В области фокуса этого микрообъектива находилась стеклянная подложка с нанесенной на нее взвесью микрочастиц, а следом за ней находился еще один микрообъектив строящий изображение рабочей области. Далее лазерный луч перенаправлялся зеркалом в камеру, перед которой стоял набор светофильтров с минимальным коэффициентом пропускания на длине волны используемого лазера, так чтобы при этом было видно изображение частиц и полоски сопоставимое по яркости с фоновой подсветкой.

Манипуляция микрочастицами осуществлялась с помощью ДОЭ формирующего оптический «вихрь» 20-го порядка. Удалось осуществить захват и передвижение частицы вдоль полоски. Момент захвата и перемещения представлен на изображениях ниже.

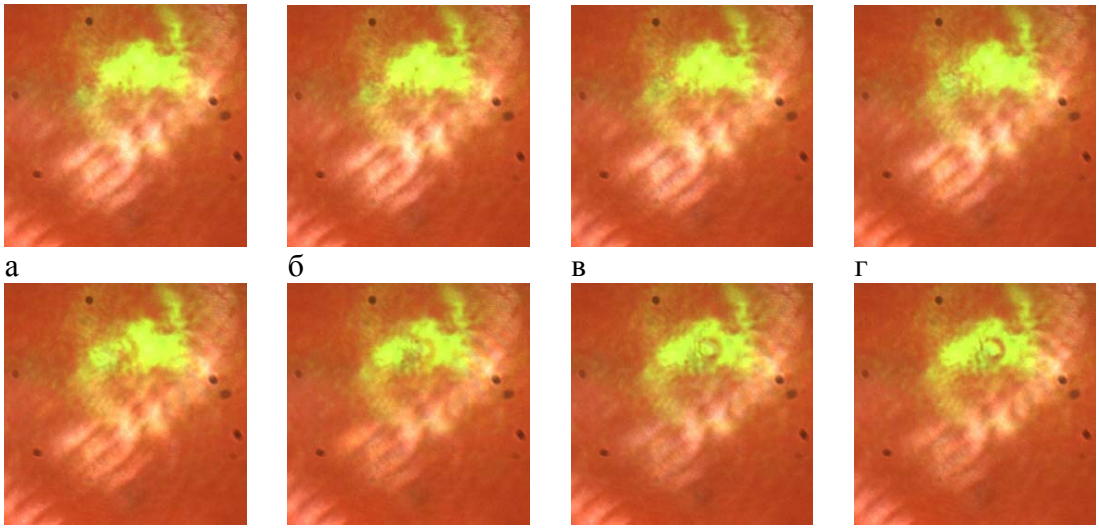


Рис. 8. Моменты захвата и перемещения частицы с промежутком между кадрами 0.25с

На рис. 8 видно как частица захватывается слева в область повышенной интенсивности и движется направо вдоль полоски примерно до ее середины. Средняя скорость движения частицы составила 7,6мк/с, что является достаточно хорошим результатом, учитывая интенсивность используемого пучка.

Заключение

В ходе работы на основе линейно кодированных оптических «вихрей» были рассчитаны ДОЭ формирующие полосу с градиентом фазы направленным вдоль нее. Для этого была усовершенствована моделирующая программа, чтобы с помощью нее было проще подобрать оптимальную частоту кодирования для получения изображения полоски на основе оптического «вихря» с заданным порядком сингулярности.

Рассчитанные ДОЭ были изготовлены на стекле и с помощью них были осуществлены захват и перемещение микрочастиц.

В данной работе для получения ДОЭ, формирующих изображение полоски, использовались линейно кодированные оптические «вихри», что дало преимущество в эффективности в несколько раз по сравнению с элементами, рассчитанными с помощью других алгоритмов дающих эффективность порядка 1-3%.

Литература

1. Голуб, М.А. Фокусировка излучения в заданную область пространства с помощью синтезированных на ЭВМ голограмм / М.А. Голуб [и др.] // Письма в ЖТФ, 1981. - Т. 7, № 10, - С. 618-623.
2. Гончарский, А.В. Решение обратной задачи фокусировки лазерного излучения в произвольную кривую / А.В. Гончарский [и др.] // ДАН СССР, 1983. - Т. 273, № 3, - С. 605-608.
3. Воронцов, М.А. К расчету фокусаторов лазерного излучения в дифракционном приближении / М.А. Воронцов, А.Н. Матвеев, В.П. Сивоконь // Компьютерная оптика, 1987. - №. 1, - С. 74-78.
4. Скиданов, Р.В. Оптическая микроманипуляция с использованием бинарных фокусаторов /Р.В. Скиданов, А.П. Порфирьев // Компьютерная оптика, том 34, №. 1
5. Абрамочкин, Е.Г. Формирование вихревых световых полей с заданной формой интенсивности для задач лазерной манипуляции микрообъектами / Е.Г. Абрамочкин // материалы XXVI школы по когерентной оптике и голографии «Голография: фундаментальные исследования, инновационные проекты и нанотехнологии», Иркутск: Папирус. 2008. - С. 203-207.
6. Cheng-Shan Guo Optical sorting using an array of optical vortices with fractional topological charge / Cheng-Shan Guo, Ya-Nan Yu, Zhengping Hong // Optics Communications. 283 (2010). - p. 1889–1893.
7. Marchington, R.F. Optical deflection and sorting of microparticles in a near-field optical geometry / R. F. Marchington , M. Mazilu , S. Kuriakose , V. Garces-Chavez , P. J. Reece , T. F. Krauss , M. Gu , and K. Dholakia // Optics Express. 2008. - Vol.16. №6. - p. 3712-3726.

Применение подвижных фокусаторов излучения для интенсификации диффузионных процессов лазерным воздействием

С.П. Мурзин, Е.В. Шокова, Н.В. Трегуб
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

В работах [1-2 и др.] на основании результатов проведенных теоретических и экспериментальных исследований установлено, что воздействие лазерного излучения с высокой частотой следования импульсов предоставляет возможность интенсификации процессов массопереноса в твердой фазе металлических материалов. При этом не происходит значительного укрупнения зерен, существенно снижается интенсивность окисления и насыщения поверхностного слоя материала газами, а также других негативных процессов, сопровождающих длительный высокотемпературный нагрев обрабатываемого материала. Применение данного метода лазерной обработки представляется перспективным для осуществления дегазации металлических материалов конструкционного назначения. Известно, что формирование лазерного воздействия с помощью дифракционных элементов компьютерной оптики [3, 4 и др.] – фокусаторов излучения позволяет регулировать тепловые процессы в обрабатываемых материалах с целью улучшения их физико-механических свойств. В данной работе представлены результаты исследований процессов дегазации алюминиевого деформируемого сплава лазерным нагревом с применением фокусаторов излучения.

Проведение теоретических и экспериментальных исследований перераспределения плотности мощности лазерного излучения в фокальной плоскости подвижного фокусатора

Проведено исследование перераспределения плотности мощности лазерного излучения в фокальной плоскости дифракционного оптического элемента с формой рабочей поверхности на основе гиперболического параболоида при его повороте вокруг центральной нормали к рабочей поверхности. При проведении расчетов использовались следующие исходные данные: длина волны излучения $\lambda = 10{,}6\cdot10^{-6}$ м, максимальный радиус фокусируемого пучка $R_{\max} = 17{,}5\cdot10^{-3}$ м, угол между оптической осью излучения и нормалью к плоскости фокусатора $\theta = 45^{\circ}$, плотность мощности излучения $Q = 1000$ Вт. В рамках геометрической оптики задача фокусировки лазерного излучения эквивалентна поиску функций отображения области оптического элемента в область плоскости регистрации: область падения излучения на оптический элемент разбивается на малые площади, в которых определяется интенсивность излучения, после вычисления координат точки падения луча на плоскость регистрации определяется распределение интенсивности в фокальной плоскости отраженного пучка. При дискретизации поверхности оптического элемента и расчетной области размер элементарной ячейки составлял $(0{,}25\times0{,}25)\cdot10^{-3}$ м. Проведены также экспериментальные исследования перераспределения плотности мощности лазерного излучения с использованием технологического CO_2-слэб лазера Rofin DC010. При обработке результатов экспериментальных исследований использовался метод параллельного анализа, т. е. проводилась визуализация распределения плотности мощности лазерного излучения одновременно по всему сечению пучка с последующей оцифровкой изображения. Результаты экспериментальных исследований удовлетворительно коррелируют с расчетными данными. Установлено, что при повороте фокусатора излучения вокруг центральной нормали к его рабочей

поверхности на угол от 0 до 90º происходит изменение угла поворота отрезка в фокальной плоскости, уменьшение его длины и фокусного расстояния. Так, для углов поворота в φ = 30º; 60º; 90º фокусное расстояние составило: f = 0,69 м; 0,48 м; 0,42 м соответственно. При этом максимальная плотность мощности в фокальном пятне также изменяется, превышая $q \geq 5 \cdot 10^7$ Вт/м^2.

Разработка метода дегазации путем интенсификации массопереноса в твердой фазе кристаллических материалов лазерным воздействием с высокой частотой следования импульсов

Разработан метод дегазации путем интенсификации массопереноса в твердой фазе кристаллических материалов лазерным воздействием с высокой частотой следования импульсов. Проведены экспериментальные исследования по определению влияния импульсно-периодического лазерного воздействия на процесс дегазации алюминиевого деформируемого сплава. Энергетическое воздействие осуществлялось на предварительно очищенные участки поверхности выбранных форм и размеров; для повышения эффективности поглощения лазерного излучения поверхность материала подвергали предварительной обработке полированием, наносили поглощающие покрытия. Температурно-скоростные режимы обработки регулировали следующими параметрами: энергия в импульсе, длительность импульса, частота следования импульсов, форма импульса, форма лазерного пятна и плотность мощности. Определенная при испытаниях на растяжение образцов средняя концентрация газов в алюминиевом сплаве АМг6 после обработки лазерным воздействием с высокой частотой следования импульсов составила C_0 = 0,19 см^3/100 г, что на 30 % меньше, чем в образцах в исходном состоянии. В результате проведенных металлографических исследований установлено, что после лазерной обработки структура материала существенно не изменяется.

Заключение

Разработан метод дегазации путем интенсификации массопереноса в твердой фазе кристаллических материалов лазерным воздействием с высокой частотой следования импульсов. При подводе энергии лазерного излучения обеспечивается локальность физических процессов по глубине и площади при сохранении исходных свойств материала в остальном объеме, а применение специальных оптических систем позволяет избирательно проводить обработку областей необходимой геометрии. Формирование лазерного излучения осуществляется дифракционным оптическим элементом с формой рабочей поверхности на основе гиперболического параболоида. Температурно-скоростные режимы обработки регулируются следующими параметрами: энергия в импульсе, длительность импульса, частота следования импульсов, форма импульса, форма лазерного пятна и плотность мощности, причем изменение перераспределения плотности мощности достигается при повороте фокусатора излучения вокруг центральной нормали к его рабочей поверхности. Условием для интенсификации массопереноса в твердой фазе металлических материалов является нестационарная локальная деформация, вызываемая воздействием лазерного излучения воздействием с высокой частотой следования импульсов. Определенная при испытаниях на растяжение образцов средняя концентрация газов в алюминиевом сплаве АМг6 после обработки лазерным воздействием с высокой частотой следования импульсов составила C_0 = 0,19 см^3/100 г, что на 30 % меньше, чем в исходных образцах. В результате проведенных металлографических исследований установлено, что после лазерной обработки размер и распределение структурных составляющих материала существенно не изменяются.

Благодарности

Работа выполнена при поддержке Министерства образования и науки Российской Федерации в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009 – 2013 годы (государственный контракт П819).

Литература

1. Мурзин, С.П. Лазерное наноструктурирование металлических материалов с применением подвижных фокусаторов излучения / С.П. Мурзин, В.И. Трегуб, А.В. Меженин, Е.Л. Осетров // Компьютерная оптика, 2008. - Т. 32, № 4. - С. 353-356. – ISSN 0134-2452.

2. Мурзин, С.П. Создание нанопористых металлических материалов с применением лазерного воздействия / С.П. Мурзин, Е.Л. Осетров, Н.В. Трегуб, А.М.Никифоров // Известия СамНЦ РАН, 2009. - Т. 11, № 5. - С. 102-105. – ISSN 1990-5378.

3. Методы компьютерной оптики / под редакцией В.А. Сойфера. – М.: Физматлит, 2000. – 688 с.

4. Golub, M.A. Infra-red radiation focusators / M.A. Golub, I.N. Sisakian, V.A. Soifer // Optics and lasers in engineering, 1991. – Vol. 15. N 5. – P. 297-309.

Сравнение методов расчета распространения рентгеновского излучения через ДОЭ в параксиальной области

А.Г. Налимов, С.Н. Хонина
Институт систем обработки изображений РАН, Самара
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

В статье приведены результаты различных вариантов моделирования распространения света после прохождения через зонную пластинку из меди и кремния для рентгеновского излучения. Моделирование проводилось с помощью расчета одномерного преобразования Френеля, одномерного непараксиального распространения (через угловой спектр), двумерного интеграла Релея-Зоммерфельда, RSoft BeamProp из пакета FullWave, производитель RSoft Design Group, США. Показано, что первые три метода дают очень близкие результаты. Результат, полученный с помощью программы BeamProp, отличается приблизительно на 10%.

Введение

Использование мягкого и жесткого рентгеновского излучения в оптических системах позволяет получать высокое разрешение в задачах микроскопии и высокоразрешающей оптики [1]. Короткая длина волны рентгеновского излучения по сравнению с видимым светом накладывает более жесткие условия на расчет и изготовление оптики – линз Френеля, зонных пластинок, дифракционных решеток [2-4].

Зонные пластинки являются одним из самых простых элементов для фокусирования когерентного рентгеновского излучения [5-8]. Теория их расчета известна давно [9], однако интерес представляет так же и моделирование распространения света через рассчитанную зонную пластинку. Основной проблемой для расчета дифракции рентгеновского излучения на зонной пластинке является выбор необходимого метода расчета, а так же время моделирования. В статьях [1-9] не были рассмотрены различные методы расчета дифракции излучения.

В данной статье приводятся результаты моделирования дифракционной решетки, находящейся на одномерной зонной пластинке, различными способами. Как показало моделирование, результаты расчета эффективности фокусирования излучения программным пакетом BeamPROP близки к результатам моделирования распространения излучения через зонную пластинку с помощью остальных методов.

1. Постановка задачи

Длина волны при моделировании была принята λ=0,229 нм, в связи с возможностью использовать рентгеновскую трубку с анодом из хрома. На рис. 1 представлена схема рассматриваемой задачи.

Линза представляет собой цилиндрическую дифракционную решетку с переменным периодом, фокусирующую свет в первом порядке (зонная пластинка). Микрорельеф выполнен следующим образом. В кремниевой пластине протравлена цилиндрическая дифракционная решетка, затем канавки в решетки заполняются медью.

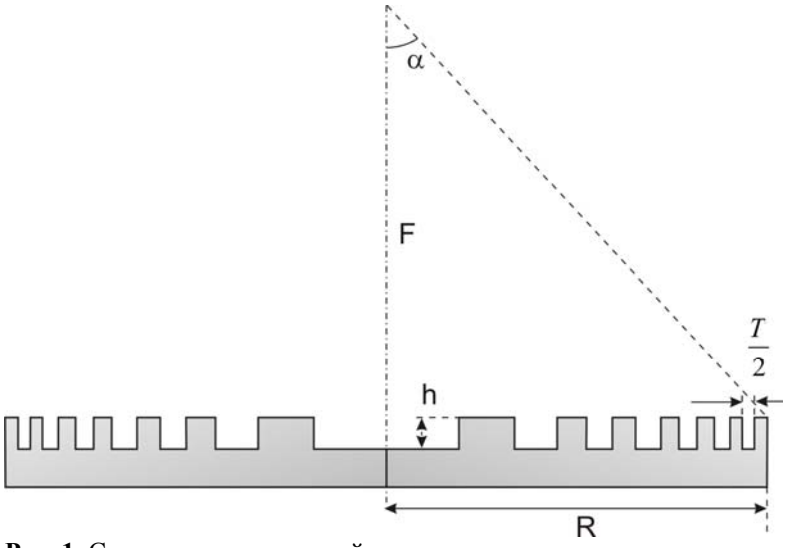


Рис. 1. Схема рассматриваемой задачи.

Показатель преломления меди для выбранной длины волны составляет $n = 1 - 1{,}56 \cdot 10^{-5} + 2{,}5 \cdot 10^{-6} i$, показатель преломления в кремнии составляет $n = 1 - 1{,}7 \cdot 10^{-5} + 8{,}5 \cdot 10^{-7} i$. Таким образом, оптимальная высота решетки равна 2,9 мкм. При радиусе R=50 мкм фокусное расстояние составляет F=0,45 м. Медь относительно прозрачна для данной длины волны, затухание света в два раза происходит на расстоянии 10 мкм, поэтому тонкий слой меди в дифракционной решетке можно использовать в качестве оптического элемента, работающего на пропускание.

Моделирование проводилось с помощью расчета одномерного преобразования Френеля, одномерного непараксиального распространения (через угловой спектр), двумерного интеграла Релея-Зоммерфельда, RSoft BeamProp из пакета FullWave, производитель RSoft Design Group, США. Расчет одномерного преобразования Френеля производился по формуле:

$$E(u,z) = \frac{k}{iz}\exp(ikz)\int_{-R}^{R}\tau(x)\exp\left(\frac{ik(x-u)^2}{2z}\right)\mathrm{d}x, \quad (1)$$

где $\tau(x)$ - комплексная функция пропускания оптического элемента, микрорельеф которого показан на рис. 1 (при плоском ограниченном апертурой плоском освещающем пучке), z – расстояние от поверхности оптического элемента вдоль оптической оси. Расчет с помощью одномерного непараксиального распространения производился по формуле:

$$E(u,z) = -\frac{i}{\lambda}\int_{-R}^{R}\tau(x)\left\{\int_{-1}^{1}\exp\left(ikz\sqrt{1-\xi^2}\right)\times\right. \quad (2)$$
$$\left.\times\exp\left(ik\left[\xi(u-x)\right]\right)\mathrm{d}\xi\right\}dx$$

Расчет с помощью двумерного интеграла Рэлея-Зоммерфельда первого типа производился по формуле:

$$E(u,v,z) = -\frac{z}{2\pi}\iint_{\Sigma}\tau(x,y)\frac{e^{ik\ell}}{\ell^2}\left(ik-\frac{1}{\ell}\right)dxdy, \quad (3)$$

где $\tau(x,y)$ - цилиндрическая комплексная функция пропускания оптического элемента, полученная из одномерной функции $\tau(x)$, $\ell = \sqrt{(u-x)^2+(v-y)^2+z^2}$, $\Sigma: 2R \times 2R$ - область входной апертуры.

2. Моделирование

2.1. Расчет с помощью одномерного преобразования Френеля.

На рис. 2 представлены результаты моделирования распространения света после прохождения через зонную пластинку с помощью одномерного преобразования Френеля.

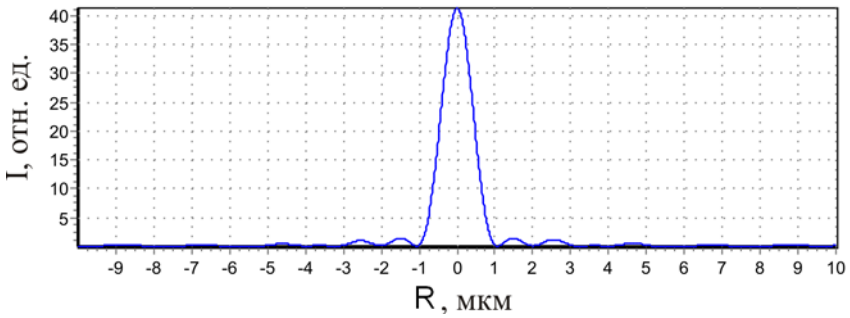


Рис. 2. Моделирование с помощью одномерного преобразования Френеля: интенсивность в фокальной плоскости (Z=450 мм).

Ширина интенсивности пучка в фокусе (рис. 2) по полуспаду: FWHM=0,936 мкм; ширина между первыми нулями 2,14 мкм, эффективность 39,5%. Под расстоянием между первыми нулями L_0 понимается расстояние между ближайшими минимумами интенсивности с обеих сторон от основного пика интенсивности в фокусе.

2.2. Моделирование с помощью 2D интеграла Релея-Зоммерфельда

На рис. 3 представлены результаты моделирования прохождения света через зонную пластинку с помощью интеграла Рэлея-Зоммерфельда.

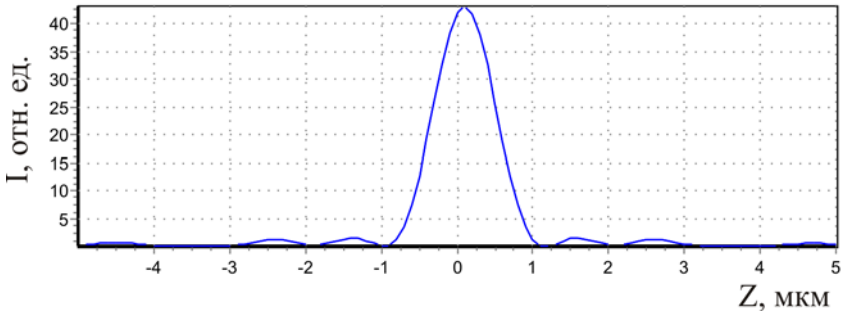


Рис. 3. Моделирование с помощью двумерного интеграла Релея-Зоммерфельда (распределение интенсивности) в фокальной плоскости (450 мм).

Как видно из рис. 3, Ширина интенсивности пучка в фокусе по полуспаду: FWHM=0,927 мкм; ширина между первыми нулями 2,2 мкм, эффективность 40,86%.

2.3 Моделирование с помощью программного пакета BeamPROP

При помощи программного пакета BeamProp производился расчет распространения света через линзу и далее до фокусной плоскости. Для уменьшения времени вычисления сетка бралась с шагом по поперечной к направлению распространения оси X: 20λ, вдоль оптической оси Y: 20000λ. Такие шаги сеток были выбраны как оптимальные. Уменьшение шагов дискретизации сетки вдвое изменяет результат вычислений менее чем на 5%. Ширина области расчета была равна апертуре зонных пластинок. На рис. 4 представлено распределение интенсивности рентгеновского излучения в фокальной плоскости после прохождения через зонную пластинку.

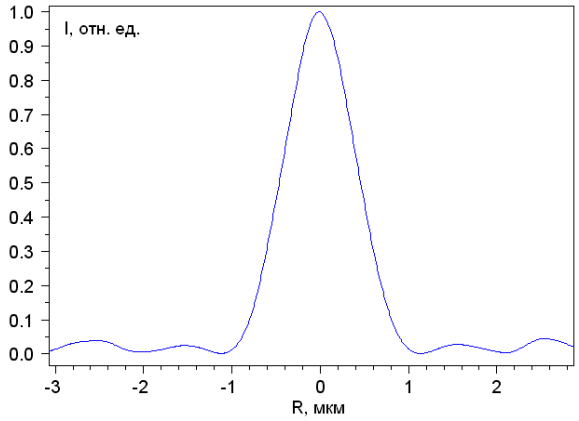


Рис. 4. Распределение интенсивности рентгеновского излучения в фокальной плоскости.

В фокальной плоскости содержится 66% энергии, излученной источником и прошедшей через зонную пластинку.

Как видно из рис. 4, ширина пика интенсивности по полуспаду составляет 0,85 мкм, энергетическая эффективность составляет 54,7%. Энергетическая эффективность в пересчете к полной прошедшей энергии через зонную пластинку составляет 28,9%.

Таблица 1. Сводные результаты моделирования.

	FWHM	Полная ширина между нулями	Эффективность
1D преобразование Френеля	0,936 мкм	2,14 мкм	39,5%
1D непараксиальное распространение (через угловой спектр)	0,927 мкм	2,12 мкм	39,8%
2D интеграл Рэлея-Зоммерфельда	0,927 мкм	2,2 мкм	40,86%
RSoft BeamProp из пакета FullWave	0,85 мкм	2,26 мкм	54,7%

Как видно из Таблицы 1, первые три метода дают очень близкие результаты, что вполне ожидаемо для дальней зоны, в которой проводятся исследования. Разумеется, использование выражения (3), как и выражения (2) в данной области избыточно по точности и затратно по ресурсам, но непараксиальные модели были применены для корректного сравнения полученных результатов.

Упомянутые выше методы применялись в приближении тонкого оптического элемента, не учитывающего поглощение и отражение на элементе. В пакете FullWave учитывается поглощение, поэтому результаты отличаются примерно на 10%.

3. Заключение

Первые три метода дают очень близкие результаты (различие на несколько процентов) – в среднем ширина центрального пика по полуспаду от максимума интенсивности составила 0,93 мкм, а полная ширина между нулевыми значениями – около 2 мкм, средняя эффективность – 40%. Непараксиальные методы, избыточные по точности затратные по ресурсам были применены для корректного сравнения полученных результатов. Данные методы использовались в приближении тонкого оптического элемента, не учитывающего поглощение и отражение на элементе.

В пакете FullWave учитывается поглощение, поэтому результаты отличаются примерно на 10% от полученных вышеупомянутыми методами: ширина центрального пика по полуспаду от максимума интенсивности составила 0,85 мкм, полная ширина между нулевыми значениями – 2,26 мкм, эффективность – 55%.

Учитывая результаты расчета можно сделать вывод, что для приближенного моделирования распространения рентгеновского излучения через зонную пластинку достаточно одномерного преобразования Френеля. Если необходимо более точное решение, то лучше использовать метод BPM – он учитывает поглощение в материале, что для рентгеновского диапазона актуально.

Благодарности

Работа выполнена при поддержке российско-американской программы "Фундаментальные исследования и высшее образование" (грант CRDF PG08-014-1), гранта Президента РФ поддержки ведущих научных школ (НШ-3086.2008.9) и гранта РФФИ 08-07-99007.

Литература

1. Attwood, D. Soft x-rays and extreme ultraviolet radiation: principles and applications / D. Attwood // Cambridge University Press, 1999.

2. Vaschenko, G. Sub-38 nm resolution tabletop microscopy with 13 nm wavelength laser light / G. Vaschenko, C. Brewer, F. Brizuela, Y. Wang [and oth.] // Opt. Lett., 2006. – Vol. 31. – PP. 1214–1216.

3. Vaschenko. G. Nano-imaging with a compact extreme ultraviolet laser / G. Vaschenko, F. Brizuela, C. Brewer, M. Grisham [and oth.] // Opt. Lett., 2005. – Vol. 30. – PP. 2095–2097.

4. M. Kuyumchyan, A. Isoyan, A. Shulakov, E. Aristov, V. Kondratenkov, M. Snigirev, A. Snigireva, and I. Yabashi, “High efficiency and low absorption Fresnel compound zone plates for hard X-ray focusing,” http://arxiv.org/pdf/cs.OH/0503005 (Cornell University, 2005).

5. M. Kuyumchyan, A. Isoyan, A. Shulakov, E. Aristov, V. Kondratenkov, M. Snigirev, A. Snigireva, and I. Yabashi, “High efficiency and low absorption Fresnel compound zone plates for hard X-ray focusing,” http://arxiv.org/pdf/cs.OH/0503005 (Cornell University, 2005).

6. Aoki, S. The 8th International Conference on X-ray Microscopy (Institute of Pure and Applied Physics, Tokyo, 2006); and references therein / S. Aoki, Y. Kagoshima, and Y. Suzuki [and oth.].

7. Chao, W.L. Soft X-ray microscopy at a spatial resolution better than 15 nm / W. L. Chao, B. D. Harteneck, J. A. Liddle, E. H. Anderson [and oth.] // Nature, 2005. – Vol. 435. – N. 7046. – PP. 1210–1213.

8. Jefimovs, K. Zone-doubling technique to produce ultrahigh-resolution x-ray optics / K. Jefimovs, J. Vila-Comamala, T. Pilvi, J. Raabe, M. Ritala [and oth.] // Phys. Rev. Lett., 2007. – Vol. 99. – N. 26. – PP. 264801.

9. Barret, H.H. Fresnel zone plate imaging of gamma rays; theory / H.H. Barret, F.A. Horrigan // Appl. Opt., 1973. – Vol. 12. – N. 11. – PP. 2686-2702.

Исследование влияния аберраций на высокоапертурные фокусирующие системы

Е. А. Пелевина[1] 1, С.Н. Хонина[1,2]
[1]*Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*
[2]*Институт систем обработки изображений РАН, Самара*

Введение

В последнее время появилось много работ, посвященных уменьшению поперечного размера фокального пятна высокоапертурной фокусирующей системы [1-6]. Наименьший поперечный размер фокального пятна был получен при радиальной поляризации, когда вклад продольной компоненты в суммарную интенсивность является максимальным.

Существуют различные способы еще уменьшить вклад поперечных компонент в общую интенсивность на оптической оси. Применяют изменение функции пропускания линзы – блокирование центральной части [3, 4] и более энергетически эффективное дополнение фазовыми оптическими элементами [5, 6].

Также большое количество работ посвящены изучению влияния различных аберраций на высокоапертурные оптические системы. В работах [7, 8] было проведено исследование влияния сферической аберрации и дефокусировки. Так было выяснено, что пик интенсивности является более узким, чем предсказано классической теорией.

Поэтому интересно рассмотреть возможность аналогичного выделения отдельной компоненты электрического поля с целью достижения наиболее компактной осевой локализации интенсивности на основе дополнительного введения пропускающих функций, аналогичных волновым аберрациям.

Базис Цернике и его свойства

Существует полное множество ортогональных функций с угловыми гармониками в круге радиуса r_0, которые используют для описания волновых аберраций. Это круговые полиномы Цернике:

$$\Psi_{nm}(r,\varphi)=\sqrt{\frac{n+1}{\pi r_0^2}}R_n^m(r)\exp(im\varphi), \tag{1}$$

где $$R_n^m(r)=\sum_{p=0}^{(n-m)/2}(-1)^p(n-p)!\left[p!\left(\frac{n+m}{2}-p\right)!\left(\frac{n-m}{2}-p\right)!\right]^{-1}\left(\frac{r}{r_0}\right)^{n-2p}. \tag{2}$$

В формуле (2) приведены радиальные полиномы Цернике $R_n^m(r)$. Их свойства:

– $R_n^{-m}(r)=R_n^m(r)$;

– $R_n^{\pm1}(r_0)=1$;

– $R_{2k+1}^{2l}(r)=0, R_{2k+1}^{2l+1}(r)=0, |m|\le n$;

– $R_0^0(r)=1$.

Непараксиальная векторная модель в приближении Дебая

Для высокоапертурной фокусирующей оптической системы векторное электрическое поле в однородной диэлектрической среде вблизи фокуса можно описать в приближении Дебая:

$$E(\rho,\varphi,z)=-\frac{if}{\lambda}\int_0^{\alpha}\int_0^{2\pi}B(\theta,\phi)T(\theta)P(\theta,\phi)\exp\left[ik(\rho\sin\theta\cos(\phi-\varphi)+z\cos\theta)\right]\sin\theta\,d\theta\,d\phi, \tag{3}$$

где (ρ, φ, z) – цилиндрические координаты фокальной области; (θ, ϕ) – сферические угловые координаты выходного зрачка; $B(\theta, \phi)$ – функция пропускания; $T(\theta)$ - функция аподизации зрачка; $P(\theta, \phi)$ - матрица поляризации.

Матрица поляризации фокусирующей системы $P(\theta, \phi)$ из (3) имеет следующий вид:

$$P(\theta,\phi) = \begin{bmatrix} 1+\cos^2\phi(\cos\theta-1) & \sin\phi\cos\phi(\cos\theta-1) & \sin\theta\cos\phi \\ \sin\phi\cos\phi(\cos\theta-1) & 1+\cos^2\phi(\cos\theta-1) & \sin\theta\sin\phi \\ -\sin\theta\cos\phi & -\sin\theta\cos\phi & \cos\theta \end{bmatrix} \begin{bmatrix} a(\theta,\phi) \\ b(\theta,\phi) \\ c(\theta,\phi) \end{bmatrix}, \quad (4)$$

где $a(\theta,\phi), b(\theta,\phi), c(\theta,\phi)$ - функции поляризации для *x*-, *y*- и *z*- компоненты падающего пучка.

Линейная x-поляризация

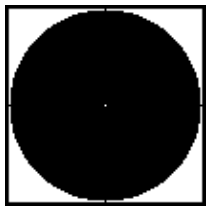
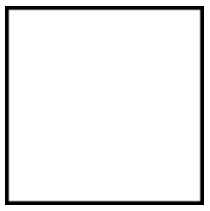
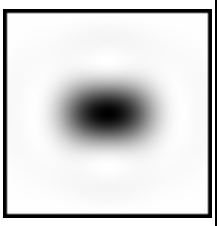
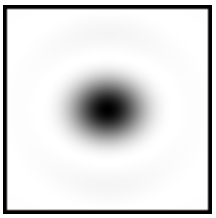
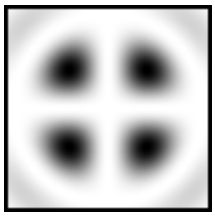
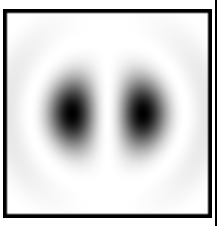
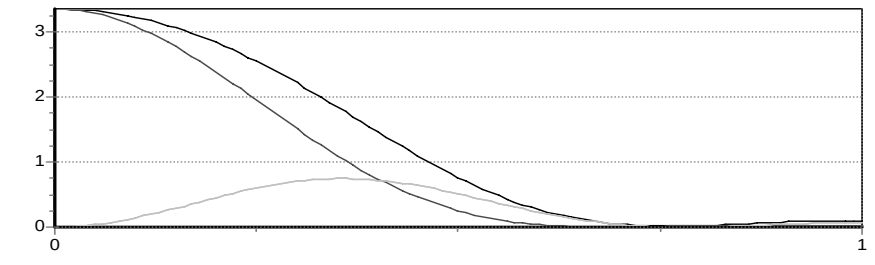
Поскольку большинство современных лазеров имеют линейную поляризацию, то остается актуальным вопрос уменьшения размера фокального пятна при линейной поляризации.

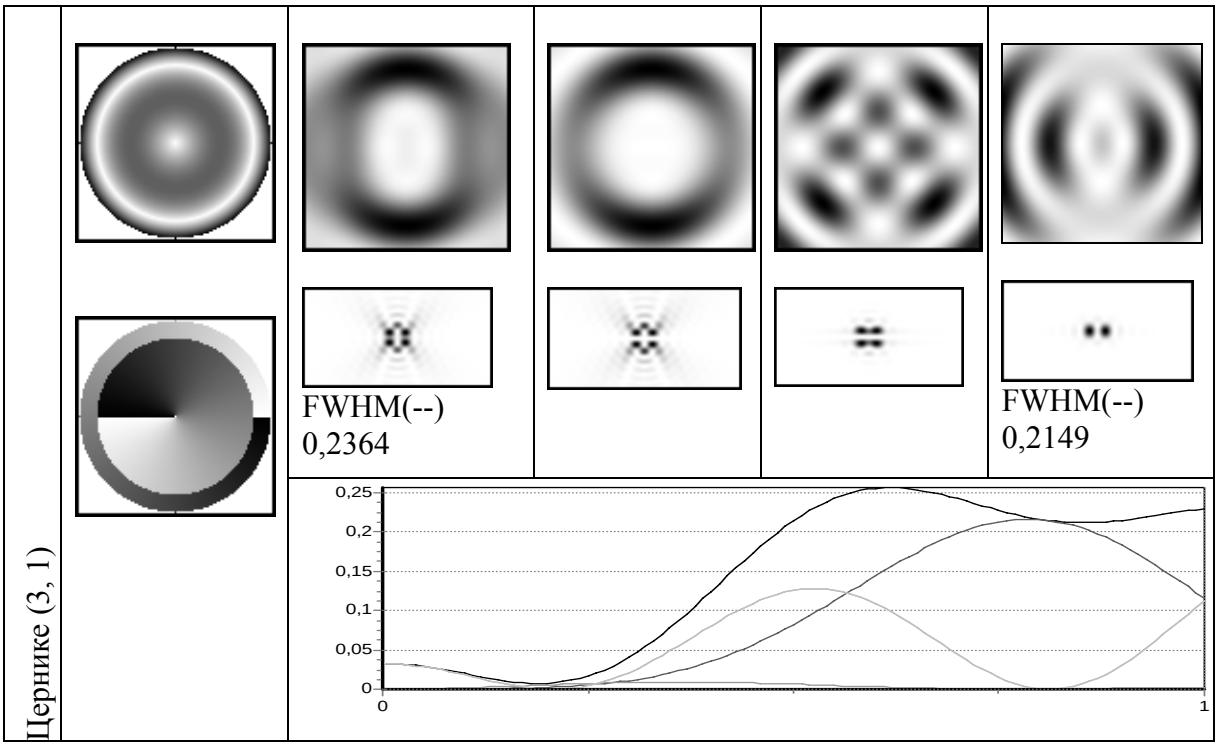
При линейной *x*-поляризации коэффициенты входной поляризации имеют вид для формулы (4): $a(\phi) = 1,\ b(\phi) = 0,\ c(\phi) = 0$. А матрица $P(\theta, \phi)$ сводится к виду:

$$P(\theta,\phi) = \begin{bmatrix} 1+\cos^2\phi(\cos\theta-1) \\ \sin\phi\cos\phi(\cos\theta-1) \\ -\cos\phi\sin\theta \end{bmatrix}.$$

Рассмотрим самые интересные из полученных результатов. На всех графиках черным цветом изображена полная интенсивность поля, темно-серым – *x*-составляющая, серым – *y*-составляющая, светло-серым – *z*-составляющая.

Таблица 1

	Амплитуда, фаза	$\lvert E\rvert^2$	E_x	E_y	E_z
Цернике (0, 0)	 	  FWHM(--) 0,7438	  FWHM(--) 0,5609	 -	 -
					

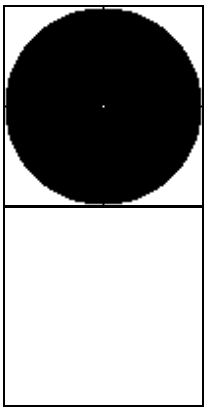
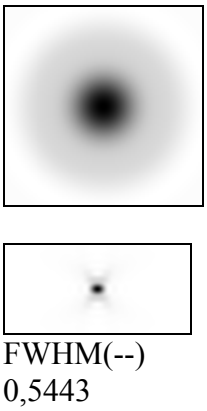
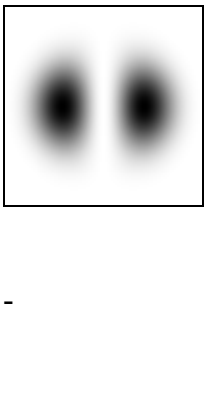
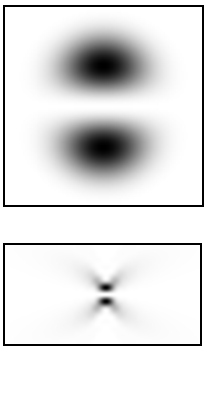
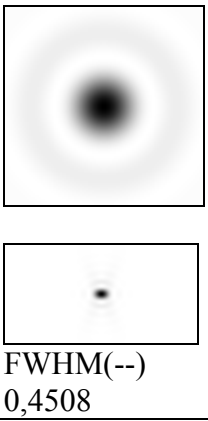


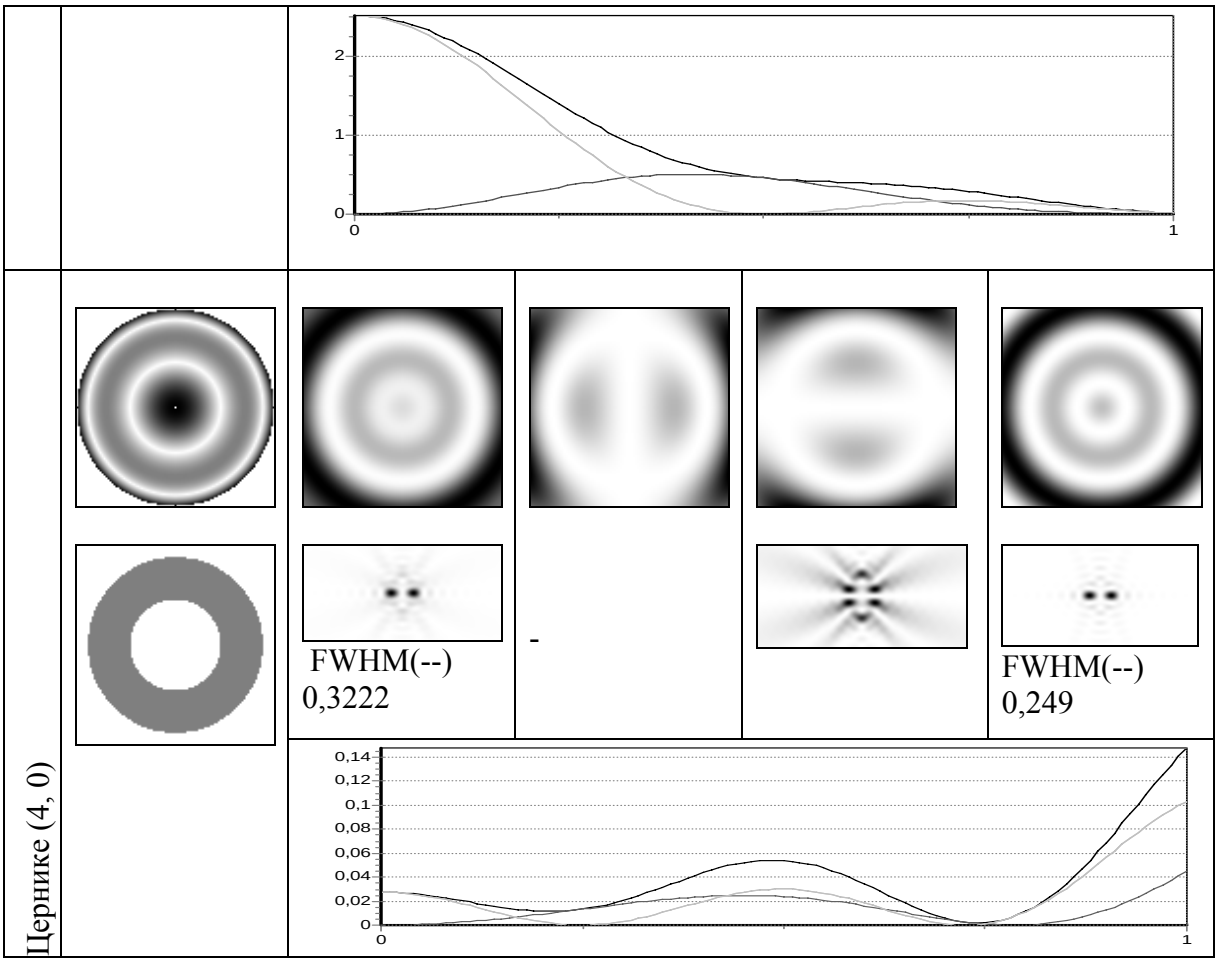
Радиальная поляризация

При радиальной поляризации излучения, падающего на высокоапертурную фокусирующую систему, коэффициенты входной поляризации записываются как $a(\phi) = \cos\phi, b(\phi) = \sin\phi, c(\phi) = 0$, и тогда матрица $P(\theta, \phi)$ из (4) для радиальной поляризации примет вид:

$$P(\theta,\phi) = \begin{bmatrix} \cos\phi\cos\theta \\ \sin\phi\cos\theta \\ -\sin\theta \end{bmatrix}.$$

Таблица 2

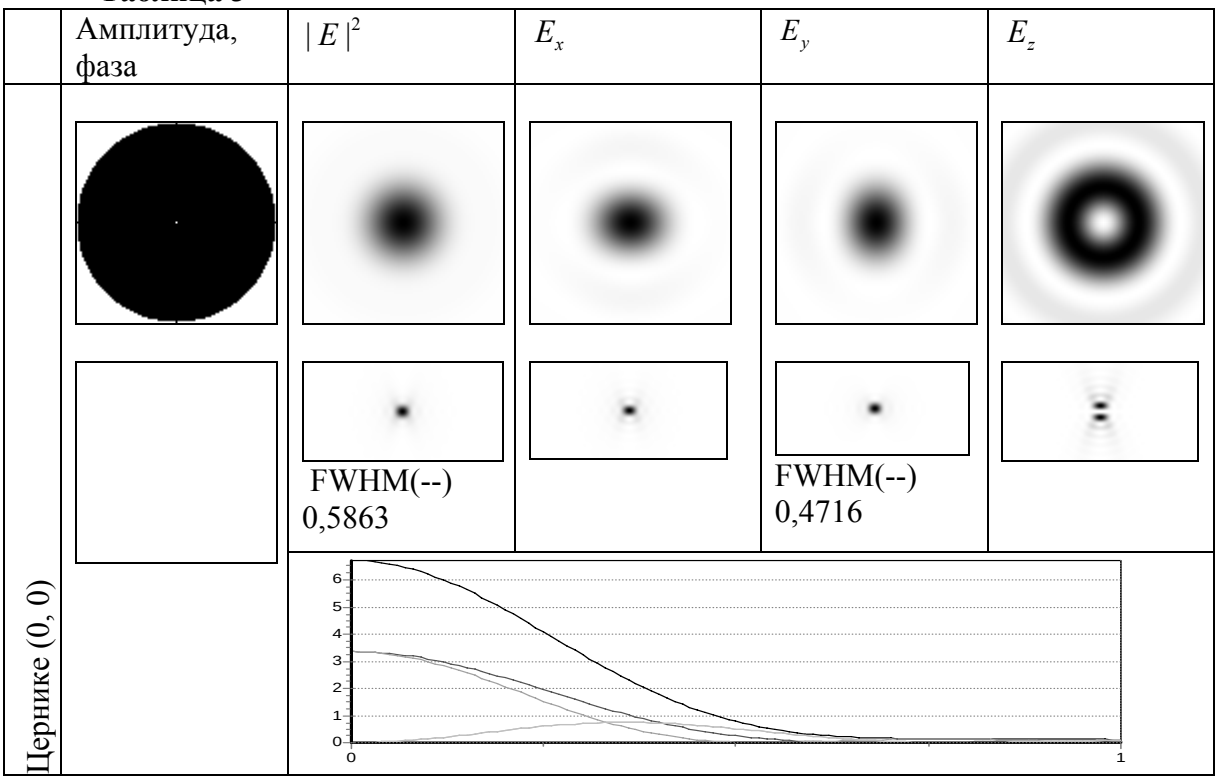
	Амплитуда, фаза	$\lvert E\rvert^2$	E_x	E_y	E_z
Цернике (0, 0)		 FWHM(--) 0,5443	 -		 FWHM(--) 0,4508

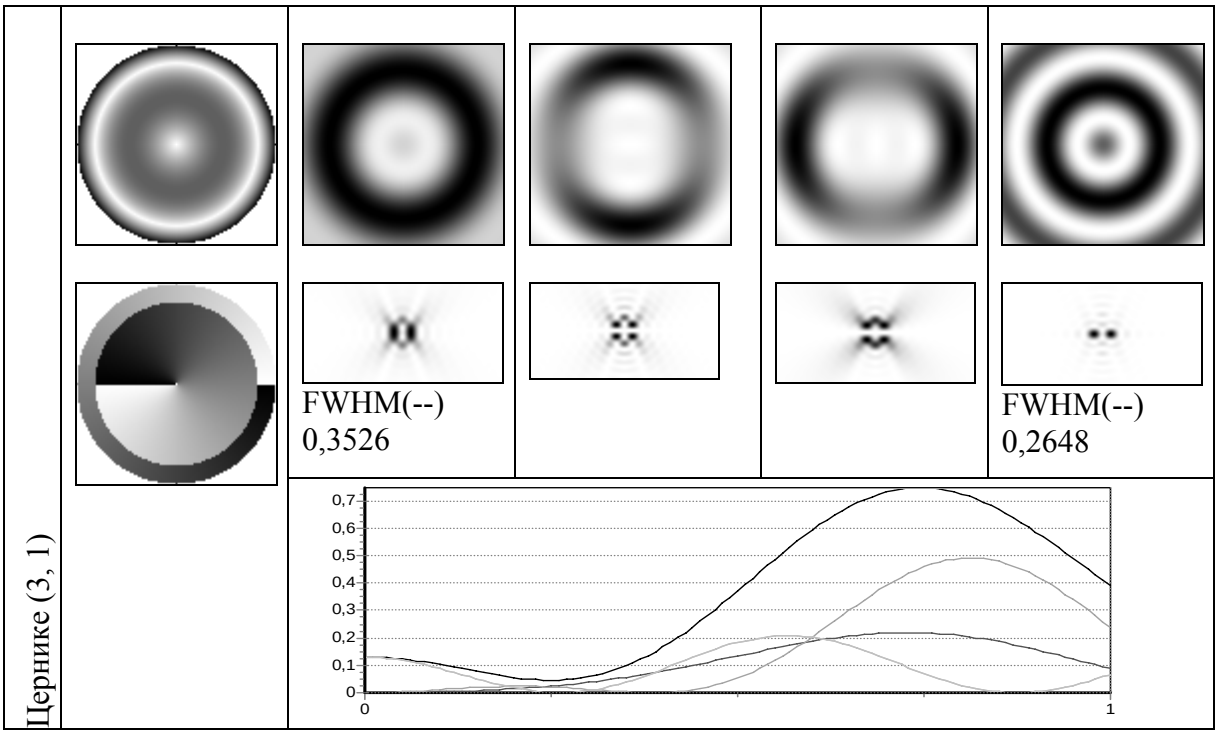


Круговая поляризация

Круговая поляризация также является распространённым и простым в реализации типом поляризации.

Таблица 3





Заключение

В работе рассмотрено введение пропускающих функций, соответствующих волновым аберрациям в высокоапертурных фокусирующих системах с различной поляризацией в модели Дебая. В результате подтверждена возможность уменьшения размера фокального пятна. Недостатком такого подхода является перераспределение энергии в боковые лепестки. Рассмотренные возможности могут быть полезны в области оптического манипулирования микрочастицами.

Литература

1. Хонина С.Н., Волотовский С.Г. Исследование применения аксиконов в высокоапертурной фокусирующей системе // Компьютерная оптика. 2010. Т. 34. № 1. С. 35-51.

2. Хонина С.Н., Волотовский С.Г. Управление вкладом компонент векторного электрического поля // Компьютерная оптика. 2010. Т. 34, № 1. С. 58-68.

3. Quabis, S. Focusing light to a tighter spot // Opt. Commun. 2000. V. 179. P. 1-7.

4. Dorn, R. Sharper focus for a radially polarized light beam // Phys. Rev. Lett. 2003. V. 91. P. 233901.

5. Sheppard, Colin J.R. Annular pupils, radial polarization, and superresolution // Appl. Opt. 2004. V. 43. 22. P. 4322-4327.

6. Хонина, С.Н., Волотовский, С.Г. Полноапертурное векторное формирование продольного поля с помощью линзакона: I. Острая фокусировка в приближении Дебая (представлено в «Компьютерную оптику»).

7. Visser T.D., Wiersma S.H. Spherical aberration and the electromagnetic field in high-aperture systems // J. Opt. Soc. Am. A. 1991. V. 8. No. 9. P. 1404-1410.

8. Visser T.D., Wiersma S.H. Diffraction of converging electromagnetic waves // J. Opt. Soc. Am. A. 1992. V. 9. No. 11. P. 2034-2047.

Травление кремния во внеэлектродной плазме высоковольтного газового разряда

В.В. Подлипнов, В.А. Колпаков
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Плазменное травление полупроводниковых и диэлектрических материалов широко применяется в технологических процессах дифракционной оптики и нанофотоники при формировании различного рода микро- и наноструктур [1].

В работе [2] показана перспективность применения внеэлектродной плазмы высоковольтного газового разряда (ВГР) для формирования оптического микрорельефа поскольку она свободна от недостатков, характерных для традиционных методов генерации плазмы, таких как эффекты загрузки и микрозагрузки, невозможность травления разных материалов в одной рабочей камере, контроль большого количества регулируемых и нерегулируемых рабочих параметров.

В данной работе предлагается использовать внеэлектродную плазму ВГР для травления кремния с целью формирования на его поверхности микроструктур. Схема реактора представлена на рисунке 1 [3, 4].

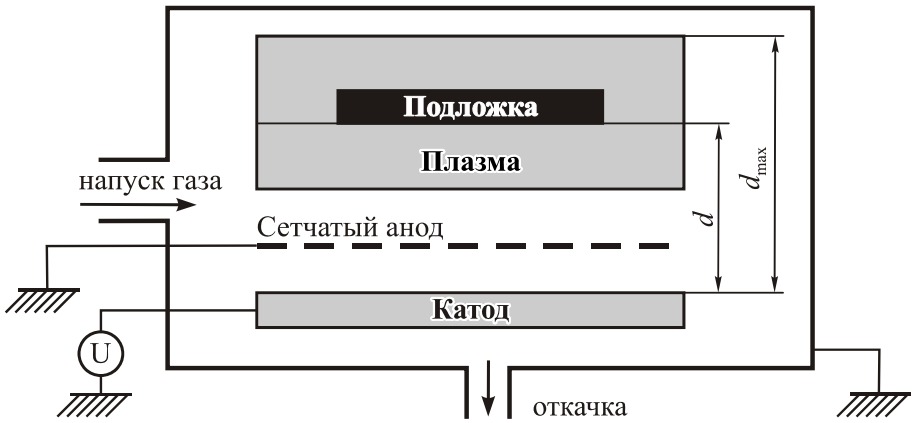


Рис. 1. Схема реактора

Для проведения эксперимента по травлению были подготовлены соответствующие образцы, представляющие собой кремниевые пластины со сформированным на их поверхности с помощью стандартной технологии фотолитографии [5] маскирующим слоем.

Параметры образцов до и после обработки в плазме высоковольтного газового разряда измерялись растровым электронным микроскопом (РЭМ) Carl Zeiss Supra 25.

Вид резистивной маски с периодом зон 3 мкм, сформированной на поверхности кремниевой пластины, показан на рисунке 2. Угол наклона поверхности образца 60-80 градусов позволяет видеть в отверстиях дно микроструктуры и измерить толщину слоя резиста, которая составляет 0,9-1,1 мкм.

Травление кремния осуществлялось во внеэлектродной плазме ВГР в смеси газов CF_4 и O_2 при различных значениях основных параметров газоразрядного прибора: напряжения на электродах (U) и тока разряда (I), которые изменялись в пределах U = 700÷2000 В и I = 50÷190 мА соответственно. Содержание газов CF_4 и O_2 в плазме поддерживалось постоянным [4], время травления для всех режимов составляло 10 мин.

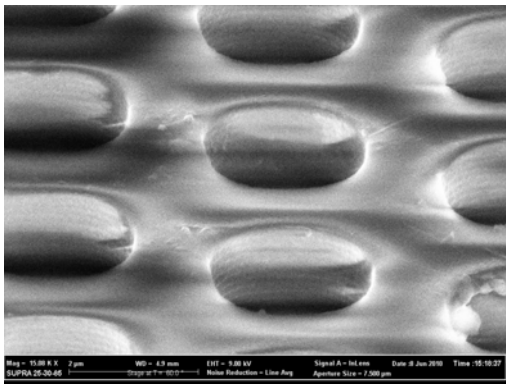
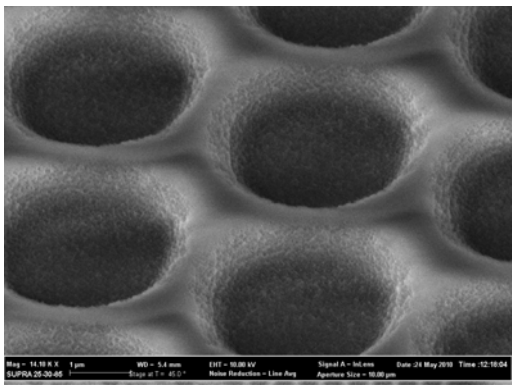
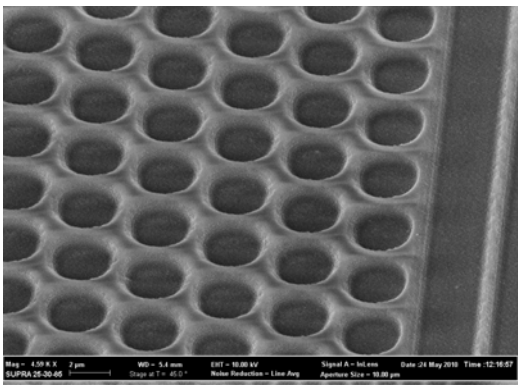


Рис 2. Изображение резистивной маски на поверхности кремния до обработки в плазме

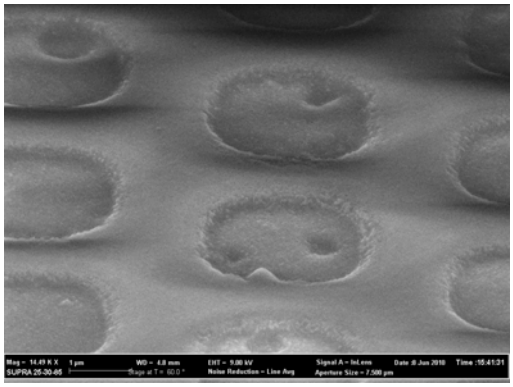
На рисунке 3 показаны микрорельефы поверхности подложек после травления образцов в различных режимах и удаления резистивной маски.



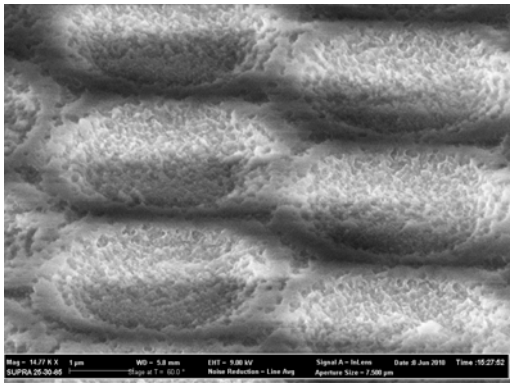
a



b



c



d

Рис 3. Изображения поверхности кремния после обработки в плазме ВГР при различных значениях тока разряда и напряжения:

a – I = 80 мА, U = 2000 В; b – то же, что и a) (I = 80 мА, U = 2000 В); c – I = 190 мА, U = 700 В; d – I=190мА, U= 1100 В

Анализ изображений микрорельефов на поверхности кремния, полученных с помощью РЭМ Carl Zeiss Supra 25, показывает, что при значении напряжения 2 кВ (см. рис. 3, a) и тока разряда 80 мА наблюдается лучшее качество травления кремния по сравнению с другими режимами. Геометрия полученного профиля в наибольшей степени повторяет геометрию исходного маскирующего слоя. Тем не менее, более точное сравнение геометрических параметров резистивной маски и сформированного

микрорельефа (рис. 3, a) позволяет утверждать, что незначительное растравливание материала все же существует, которое можно объяснить деградацией краев маски под действием температурного нагрева и энергетических ионов фтора (100 – 500 эВ), взаимодействующих с её углами. В большей степени влияние данного эффекта на ухудшение качества формируемого микрорельефа наблюдается при токе разряда I = 190 мА, несмотря на то, что значение напряжения на электродах было понижено до 1100 В (см. рис. 3, d). Дальнейшее снижение напряжения на электродах до 700 В (см. рис. 3, c) уменьшает плотность прямолинейных участков силовых линий электрического поля, достигающих поверхности обработки, на которых возникает и существует высоковольтный газовый разряд [2]. Следовательно, снижается и концентрация отрицательных ионов фтора в области существования газового разряда, принимающих непосредственное участие в образовании химически активных частиц на поверхности и травлении материала. При этом, геометрия профиля идентична геометрии исходной резистивной маски, что свидетельствует об отсутствии влияния температурного нагрева и физического распыления на деградацию и разрушение маски, а величина тока разряда не оказывает значительного влияния на скорость травления (см. рис. 3, c). Изображение микрорельефа, представленного на рис. 3, b подтверждает постоянство геометрических параметров канавок, сформированных на поверхности кремния от периода к периоду, что свидетельствует о высокой равномерности травления материала, объясняемой равномерным распределением заряженных частиц по сечению потока внеэлектродной плазмы.

Таким образом, основываясь на полученных результатах можно сделать следующие выводы:

- безусловно, на сегодняшний день, оптимальным режимом для травления кремния во внеэлектродной плазме является: I = 80 мА, U = 2000 В;

- для дальнейшего повышения качества микрорельефа, получаемого на поверхности кремния, необходимо более подробное исследование влияния качественного и количественного состава внеэлектродной плазмы рабочих газов, традиционно применяемых для данной цели в микро- и наноэлектронике, и температуры обрабатываемого материала на геометрические параметры формируемых микроструктур;

- при переходе к формированию наноструктур во внеэлектродной плазме необходимо значительно повысить роль химических реакций в процессе удаления травимого материала, предельно минимизировав влияние физического распыления, что возможно осуществить за счет повышения плотности внеэлектродной плазмы и снижения энергии частиц, взаимодействующих с поверхностью.

В целом, результаты, полученные в данной работе, хорошо согласуются аналогичными результатами по травлению диоксида кремния, карбида кремния и алмазоподобных пленок, опубликованными в монографии [2]. Значения скоростей травления кремния во внеэлектродной плазме, полученные экспериментально, находятся в разумном соответствии с аналогичными данными, опубликованными в работах [7-9]. Представленные в настоящей работе результаты свидетельствуют о перспективности применения внеэлектродной плазмы для формирования на поверхности оптических материалов различного рода дифракционных микро- и наноразмерных структур.

Литература

1. Методы компьютерной оптики / под ред. В.А. Сойфера. – М.: Физматлит, 2003. – 688 с.

2. Казанский Н.Л., Колпаков В.А. Формирование оптического микрорельефа во внеэлектродной плазме высоковольтного газового разряда. Монография. – М.: Радио и связь, 2009. – 220 с. – ISBN 5-89776-011-X.

3. Колпаков В.А. Моделирование процесса травления диоксида кремния в плазме газового разряда высоковольтного типа // Микроэлектроника. – 2002. – Т. 31. – № 6. – С. 431-440.

4. Казанский Н.Л., Колпаков А.И., Колпаков В.А. Исследование особенностей процесса анизотропного травления диоксида кремния в плазме газового разряда высоковольтного типа // Микроэлектроника. – 2004. – Т. 33. – № 3. – С. 218-233.

5. Моро У. Микролитография. Ч.1; пер. с англ. – М.: Мир, 1990. – 605 с.

6. Пат. 2328707 Российская Федерация, МПК7 H 01 B 7/18. Способ измерения температуры поверхности образца, облучаемого газоразрядной плазмой / Сойфер В.А., Казанский Н.Л., Колпаков В.А., Колпаков А.И., Паранин В.Д.; заявитель и патентообладатель ИСОИ РАН. – № 2005118364; заявл. 14.06.05; опубл. 10.07.08, Бюл. № 8. – 5 с.

7. Ивановский Г.Ф., Петров В.И. Ионно-плазменная обработка материалов. – М.: Радио и связь, 1986. – 232 с.

8. Орликовский А.А. Плазменные процессы в микро- и наноэлектронике. Плазмохимические реакторы нового поколения и их применение в технологии микроэлектроники // Микроэлектроника. – 1999. – Т.28. – № 6. – С. 415-426.

9. Ковалевский А.А., Малышев В.С., Цыбульский В.В., Сорокин В.М. Исследование процесса изотропного плазмохимического травления пленок диоксида кремния // Микроэлектроника. – 2002. – Т. 31. – № 5. – С. 344-349.

ДОЭ для автоматической оптической сборки микросистем

А.П. Порфирьев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Широко известно о возможности воздействия света на микрочастицы [1]. При этом на частицу можно оказывать влияние как путем формирования амплитудной составляющей светового поля, так и путем формирования его фазовой составляющей. Поэтому с помощью фокусаторов, формирующих заданное амплитудно-фазовое распределение, можно оказывать двойное влияние на движение микрочастиц. С их помощью можно сформировать и переместить устойчивую структуру из микрочастиц. В последнее время появляется все больше работ, посвященных использованию динамических модуляторов света [2-4], которые могут, в том числе, формировать набор световых пучков, образующих в совокупность область заданной формы. Есть и работы, в которых фокусаторы формировали заданную интенсивность, и при этом световое поле обладало и некоторой заранее известной фазой [5]. В [5] световой пучок имел вихревую фазу с заданным угловым орбитальным моментом, т.е. фактически фокусатор формировал вихревое световое поле с заданной интенсивностью. Вихревой характер формируемого поля означает, что на распределение интенсивности накладывается ряд ограничений, которые не позволяют формировать световое поле с произвольным распределением интенсивности. В настоящей работе рассматривается метод расчета фокусаторов, которые позволяют формировать заданное амплитудно-фазовое распределение практически без ограничений на форму этого распределения. Также рассматриваются эксперименты по микроманипуляции в световых полях, сформированных такими фокусаторами.

Задача формирования светового поля с заданным распределением интенсивности и фазы

Оптически более плотная, чем окружающая среда, микрочастица перемещается в световом поле в направлении градиента интенсивности. Поэтому, если фокусатор формирует световое поле в виде некоторой кривой, микрочастицы будут втягиваться в это световое поле и удерживаться им. Если световое поле будет иметь также градиент фазы направленный вдоль сформированной кривой, то микрочастица после втягивания в этот световой пучок продолжит движение вдоль линии. Направление этого движения - в сторону возрастания фазы в пучке.

Поэтому задача расчета фокусатора формулируется следующим образом. Необходимо рассчитать фокусатор, формирующий произвольную кривую с фазой, которая линейно изменяется вдоль этой кривой. При этом необходимо сформировать бинарный фазовый элемент, т.к. технология изготовления многоуровневых элементов в настоящее время не отработана.

Метод расчета фокусаторов

Для расчета амплитудно-фазовых распределений в данной работе использовалась процедура, которая содержит следующие шаги:

из заданных распределений амплитуды $A_0(\xi,\eta)$ и фазы $\varphi_0(\xi,\eta)$ формируется комплексная амплитуда $W(\xi,\eta) = A_0(\xi,\eta)\cdot\exp(i\varphi_0(\xi,\eta))$;

полученная комплексная амплитуда подвергается обратному преобразованию Фурье, чтобы получить комплексное распределение в плоскости ДОЭ;

осуществляется кодирование амплитуды фазовой функцией;

для приведения фазы ДОЭ к бинарному виду осуществляется кодирование фазы по формуле:

$$\varphi_b(x,y)=\arg\left(\varphi^*(x,y)\cdot e^{-i\alpha x}+\varphi(x,y)\cdot e^{i\alpha x}\right) \quad (1)$$

где α - частота несущей, $\varphi(x,y)$ - исходная многоуровневая фаза, $\varphi_b(x,y)$ – новая бинарная фаза.

В качестве процедуры кодирования амплитуды был выбран метод локального фазового скачка [6].Суть этого метода в следующем: каждый отсчет амплитуды заменяется на участок ДОЭ шириной Δx со ступенчатым скачком фазы $\Delta\varphi_n$ (рис. 1). В данной работе при расчете число пикселей, приходящихся на отрезок дискретизации Δx, бралось равным 16. Величина скачка фазы - π.

Ширина скачка δx_n вычисляется по формуле:

$$\delta x_n=\frac{\Delta x}{2}\left(1-A_n\right), \quad (2)$$

где $0\le A_n\le 1$ - нормированная амплитуда, Δx - период дифракционной решетки. При этом вводится уровень амплитуды $0\le L_c\le 1$, выше которого операция кодирования не применяется. Изменение значения данного параметра позволяет варьировать значения ошибки формирования распределения и дифракционной эффективности в широких пределах.

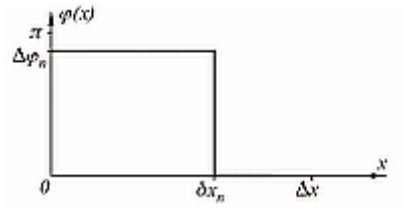
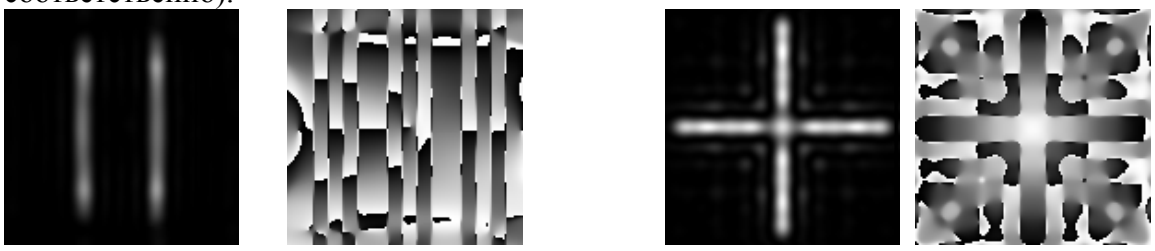


Рис.1. Кодирование амплитуды.

Расчет фокусаторов и их численное моделирование.

На рисунке 2 представлены рассчитанные распределения для случаев фокусатора в крест с линейным увеличением фазы вдоль линий и фокусатора в две линии, также с линейным увеличение фазы вдоль линий (фаза меняется на 2π и 4π соответственно).



а б в г

Рис. 2. Интенсивность (а) и фаза (б), сформированные фокусатором в две линии; интенсивность (в) и фаза (г), сформированные фокусатором в крест.

При моделировании дифракционная эффективность вычислялась по формуле:

$$\varepsilon=\frac{\iint\limits_{Q}A^2(\xi,\eta)d\xi d\eta}{\iint\limits_{D}A^2(\xi,\eta)d\xi d\eta}, \quad (3)$$

где Q – область, в которой заключается полезное изображение, D – выходная область, $A(\xi,\eta)$ - амплитудное распределение.

Для случая фокусатора в две линии при расчете удалось добиться наилучшего значения эффективности равного 37%, для случая фокусатора в крест 98%. Заметим, что в данном случае под полезным изображением понимались суммарно два порядка, которые появляются в результате кодирования фазы ДОЭ.

Натурный эксперимент по оптическому захвату микрочастиц с использованием ДОЭ, изготовленного на стекле

Для манипулирования использовался He-Ne лазер с длиной волны излучения 532 нм, соответствующей зеленому излучению, и мощностью излучения 500 мВт. Однако опыт показывает, что из-за отражения на оптических поверхностях и из-за поглощения среды до ДОЭ доходит около трети исходной мощности. В качестве микрочастиц были выбраны полистироловые шарики диаметром 5 мкм. На рисунке 3 представлена оптическая схема, которая была использована в эксперименте.



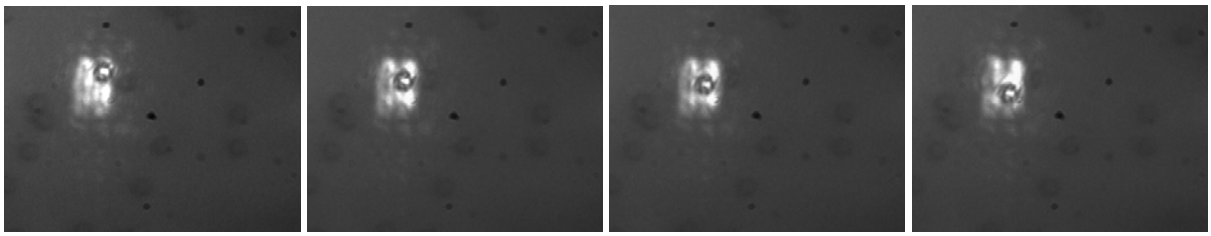
Рис. 3. Оптическая схема, используемая для экспериментов.

L – твердотельный He-Ne лазер, M_1, M_2, M_3 – поворотные зеркала, L_1-фокусирующий микрообъектив, L_2- микрообъектив для построения изображения рабочей области, DOE – фокусатор, V – стеклянная подложка, I – осветитель, L_3– микрообъектив, TV – видеокамера модели VSTT-252, PC – персональный компьютер.

На рисунках 4-5 представлен захват в область линий одной и четырех микрочастиц соответственно. В таблице 1 приведены измеренные значения скоростей движения частиц. Из их сравнения видно, что по линии с большим изменением фазы частицы движутся быстрее.

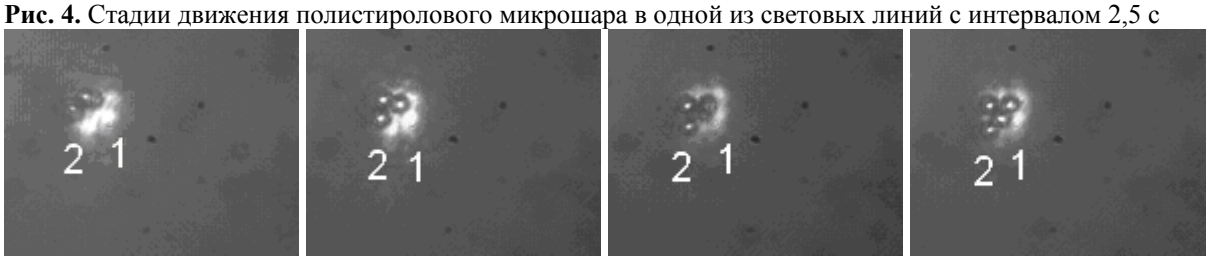
Также были проведены аналогичные эксперименты и с фокусатором в крест. На рисунке 6 можно видеть стадии движения одной микрочастицы в световом пучке в виде креста с фазой, которая линейно уменьшается от центра к концам креста. Как видно из приведенных изображений при попадании в окрестность любого из лучей креста частица захватывается в пучок и двигается к центру креста, где и фиксируется пучком. Полученные значения скоростей микрочастиц представлены в таблице 2.

При кодировании фазы, как отмечалось ранее, возникало два порядка изображений, при этом фаза одного из них являлась инвертированной фазой другого. Поэтому был также проведен эксперимент с распределением в крест, фаза которого уменьшалась при движении к центру (рисунок 7). Видно, что частицы при попадании в световой пучок перемещаются от центра на края лучей.



а б в г

Рис. 4. Стадии движения полистиролового микрошара в одной из световых линий с интервалом 2,5 с

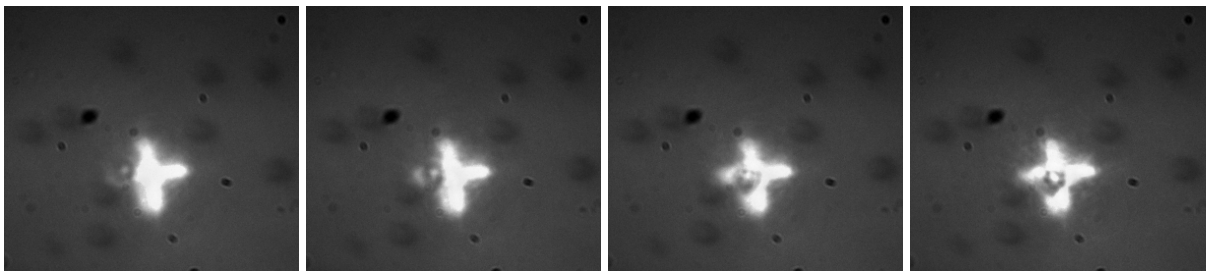


а б в г

Рис. 5. Стадии движения четырех полистироловых микрошаров в световых линиях с интервалом 1,5 с. Задержка времени между кадрами б, в 11,5 с

Таблица 1. Результаты экспериментального исследования

Рассматриваемый случай	Пройденное расстояние, мкм	Время наблюдения, с	Средняя скорость, мкм/с	Относительная погрешность, %
Рисунок 4	6,40	7,75	0,82	7,2
Движение «нижней» частицы по линии «1» рисунок 5	3,80	3,75	1,01	12,2
Движение «верхней» частицы по линии «2» рисунок 5	1,80	1,25	1,44	26,8
Движение «нижней» частицы по линии «2» рисунок 5	2,20	1,25	1,76	22,8

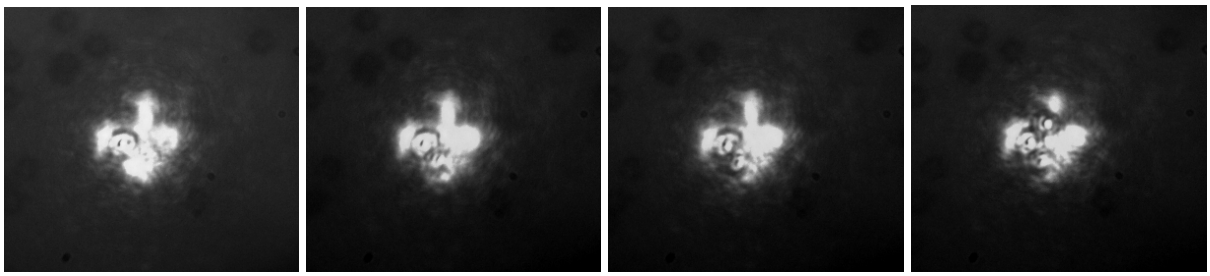


а б в г

Рис.6. Стадии движения полистиролового микрошара в световом пучке в виде креста с линейным уменьшением фазы от центра к лучам с интервалом 0,5 с

Таблица 2. Результаты экспериментального исследования

Рассматриваемый случай	Пройденное расстояние, мкм	Время наблюдения, с	Средняя скорость, мкм/с	Относительная погрешность, %
Рисунок 6	4,8	1,25	3,84	16,5



а б в г

Рис.7. Стадии движения трех полистироловых микрошаров в световом пучке в виде креста с линейным увеличением фазы от центра к лучам с интервалом 3,0 с.

Заключение

В работе описан метод расчета ДОЭ для формирования световых пучков для произвольно заданного перемещения микрообъектов. Рассмотрены эксперименты по такому перемещению. Рассмотренные фокусаторы могут найти применение в микромеханике и микробиологии. Фокусатор в две линии рассмотренный в работе, может быть применен для сортировки микрочастиц по размерам, когда частицы с большим размером, попав в область световой линии, отклоняются от общего потока и отводятся в сторону. Фокусатор в крест с линейным увеличением фазы к центру может быть использован для точного позиционирования микрочастиц. Фокусатор в крест с линейным уменьшением фазы к центру может разводить частицы по краям, очищая центральную область от частиц. Комбинируя эти распределения, можно сформировать более сложную микросистему, которая будет автоматически разводить частицы и фиксировать их в нужных точках.

Литература

1. Лебедев, П. Н. Максвелло-бартолиевские силы давления лучистой энергии [Текст]/П.Н. Лебедев//Журнал Русского Физико-Химического Общества. - 1900. - Т. 32 - Вып. 8. - С. 211.

2. Davis J.A., Slovick B.A., Tuvey C. S., and Cottrell D.M. High diffraction efficiency from one- and two-dimensional Nyquist frequency binary-phase gratings Appl. Optics, 2008 V. 47. N. 15, P. 2829 - 2834.

3. Ohtake Y., Ando T., Fukuchi N., Matsumoto N., Ito H. , and Hara T. Universal generation of higher-order multiringed Laguerre–Gaussian beams by using a spatial light modulator // Opt. Lett., 2007 V. 32, N. 11, P. 1411-1413.

4. Gibson G., Carberry D.M., Whyte G., Leach J., Courtial J., Jackson J.C., Robert D., Miles M. and Padgett M. Holographic assembly workstation for optical manipulation // J. Opt. A: Pure Appl. Opt., 2008 V. 10, P. 1-6.

5. Е.Г. Абрамочкин, К.Н. Афанасьев, А.В. Коробцов, С.П. Котова, Н.Н. Лосевский, В.Г. Волостников, А.М. Майорова, Е.В. Разуева. Формирование вихревых световых полей с заданной формой интенсивности для задач лазерной манипуляции микрообъектами // материалы XXVI школы по когерентной оптике и голографии «Голография: фундаментальные исследования, инновационные проекты и нанотехнологии», Иркутск: Папирус, 2008, с. 203-207.

6. Котляр В.В., Хонина С.Н., Мелехин А.С., Сойфер В.А. // Кодирование дифракционных оптичских элементов методом локального фазового скачка // Компьютерная оптика - 1999, Вып.19, №9, С.54-64.

7. Скиданов Р.В. Расчет силы взаимодействия светового пучка с микрочастицами произвольной формы // Компьютерная оптика - 2005, Вып.28, С.18-22.

Численное моделирование движения объектов в задачах оптического захвата

М.А. Рыков
Институт систем обработки изображений РАН, г. Самара

Введение

Оптические ловушки (также *оптический пинцет*) — бесконтактный инструмент, использующий один или несколько лазерных пучков для генерации сил, порядка сотни пН, достаточно мощных для манипулирования объектами мезоскопического масштаба([1]). Эта технология представляет особый интерес в ряде областей: атомная физика, биология, микрохирургия, нанотехнология и микромеханика, а также при изучении свойств коллоидов, микро-потоков и световых пучков([2]).

Биологические микрообъекты, наиболее интересные в качестве препарата для оптического захвата имеют характерные размеры 5-7мкм. Неровности этих микрообъектов – скажем, вогнутые поверхности эритроцитов – имеют характерные размеры порядка 1-0,5мкм, что сравнимо с длиной волны лазера. Следовательно, каждый микрообъект можно аппроксимировать подходящим эллипсоидом, поскольку неровности порядка длины волны лазера не будут иметь сколько-нибудь заметного значения для геометрического приближения, которое используется нами для вычисления сил светового давления([3]). Органеллы, входящие в состав клетки – например, ядро клетки дрожжей – имеют характерные размеры ещё меньшие, чем клетка, и соответствующих размеров неровности на своей поверхности, поэтому также могут быть аппроксимированы эллипсоидом.

Физическое решение задачи

Промоделируем движение модели эллипсоида в водном растворе (например, в изотоническом растворе NaCl), т.е., в присутствии сил вязкого трения. Это движение при определённых допущениях можно разделить на два независимых друг от друга процесса: вращение вокруг центра масс эллипсоида и поступательное движение эллипсоида. Допущения, которые необходимо при этом сделать, следующие: органеллы неподвижны внутри клетки; органеллы не оказывают заметного влияния на расположение центра масс клетки: он совпадает с геометрическим центром эллипсоида; при вычислениях силы сопротивления поступательному движению не учитывается ориентация клетки. Уравнение поступательного движения клетки будет выглядеть следующим образом:

$$m\ddot{\vec{r}} = F - k(\dot{\vec{r}} - \vec{v}_{\text{т}}), \qquad (1)$$

где под F подразумевается результирующая всех сил, действующих на тело, кроме сил сопротивления, под k — коэффициент сопротивления, который обеспечивает вязкость жидкости, а под $\vec{v}_{\text{т}}$ — скорость потока жидкости относительно лазерного пучка. Скорость потока окружающей жидкости и сила со стороны лазера предполагаются на каждом шаге постоянными, поскольку продолжительность этого шага много меньше времени, за которые обе величины успеют серьёзно измениться.

Решив задачу Коши для этого уравнения (предполагая, что $\vec{v}_0$ — начальная скорость), найдём положение центра масс объекта и его скорость через время τ от начала движения(т.е., в нашем случае — в конце единичного шага):

$$\vec{v} = \vec{v}_0 \cdot e^{-\frac{k}{m}\tau} - \frac{\sum_n F_i + k\vec{v}_т}{k} \cdot (1 - e^{\frac{-k}{m}\tau});$$ (2)

$$\vec{r} = \vec{v}_0 \frac{m}{k} \cdot (1 - e^{\frac{-k}{m}\tau}) - \frac{\sum_n F_i + k\vec{v}_т}{k} \cdot (\tau - \frac{m}{k}(1 - e^{\frac{-k}{m}\tau}))$$ (3)

Найдём коэффициент k сопротивления. Число Рейнольдса для нашей задачи весьма мало, поэтому при равномерном движении для объекта будет верен закон Стокса([4]):

$k \approx 3\pi d\eta,$

где d — наибольший диаметр объекта, η — динамическая вязкость жидкости, в которой частица двигается.

Произведём некоторые численные оценки. Как уже упоминалось, характерный размер объекта ~ 5 мкм. В этом случае $k \sim 10^{-8}\,\text{Н}\frac{c}{\text{м}}$. Масса объекта порядка нескольких десятков пикограммов ($(10^{-6}\,\text{м})^3 \cdot 1000\frac{\text{г}}{\text{м}^3} = 10^{-15}\,\text{г}$). Характерное время релаксации $\frac{m}{k}$ при этом оказывается пренебрежимо малым, следовательно, $e^{\frac{-k}{m}\tau} \ll 1$, а уравнения (2) и (3) упрощаются следующим образом:

$$\vec{v} = \frac{\sum_n F_i + k\vec{v}_т}{k};$$ (4)

$$\vec{r} = \frac{\sum_n F_i + k\vec{v}_т}{k} \cdot \tau.$$ (5)

Уравнение вращения клетки очень похоже на уравнение движения. Для момента сопротивления будем иметь следующее выражение([5]):

$$M_c = \frac{1}{2} C_\omega \rho_f (\frac{\delta}{2})^5 |\Omega_R| \Omega_R,$$ (6)

где M_c — момент диссипативных сил, C_ω — коэффициент сопротивления, ρ_f — плотность жидкости, δ — характерный размер объекта, Ω_R — скорость объекта относительно окружающей жидкости. Для малых чисел Рейнольдса $C_\omega = \frac{64\pi}{Re_\omega}$. Так как

$Re_\omega = \frac{\Omega(\frac{\delta}{2})^2 \rho_f}{\eta}$, то

$$M_c = 32\pi\eta(\frac{\delta}{2})^3 \Omega_R.$$ (7)

Обозначим в дальнейшем $32\pi\eta(\frac{\delta}{2})^3$ как κ. Пренебрежём вращательными потоками вблизи захваченного объекта.

Таким образом, обозначив за J тензор инерции объекта, получаем аналогичное уравнение вращения:

$$J\dot{\vec{\Omega}} = \vec{M} - \kappa\vec{\Omega}$$ (8)

В общем случае вектор $\vec{\Omega}$ не сонаправлен ни одной из главных осей объекта-эллипсоида, поэтому необходимо использовать полный тензор инерции при вычислении:

$$J_0 = \begin{pmatrix} J_x & 0 & 0 \\ 0 & J_y & 0 \\ 0 & 0 & J_z \end{pmatrix} \qquad (9)$$

Индекс 0 указывает в (9) на то, что такой вид тензор имеет только в той системе отсчёта, в которой оси симметрии эллипсоида параллельны координатным осям. Для получения тензора в текущей системе отсчёта необходимо выполнить соответствующее преобразование.

Помножим обе части (8) на J^{-1} и перейдём в систему отсчёта 0, где матрица J диагональная (обозначим $*$ вектора в этой системе отсчёта). Получим, решая дифференциальное уравнение:

$$\vec{\Omega}^* = \vec{\Omega}_0^* e^{-J_0^{-1}\kappa\tau} + J_0^{-1} M \int_0^\tau e^{-J_0^{-1}\kappa t} dt \qquad (10)$$

В силу диагональности J_0 (10) распадается на три отдельных независимых уравнения для каждой составляющей, которые легко интегрируются([6]). В результате получим:

$$\vec{\Omega}^* = \begin{pmatrix} \Omega_{0x}^* e^{-J_0^{-1}{}_x\kappa\tau} + \dfrac{M_x}{\kappa}\left(1 - e^{-J_0^{-1}{}_x\kappa\tau}\right) \\ \Omega_{0y}^* e^{-J_0^{-1}{}_y\kappa\tau} + \dfrac{M_y}{\kappa}\left(1 - e^{--J_0^{-1}{}_y\kappa\tau}\right) \\ \Omega_{0z}^* e^{--J_0^{-1}{}_z\kappa\tau} + \dfrac{M_z}{\kappa}\left(1 - e^{--J_0^{-1}{}_z\kappa\tau}\right) \end{pmatrix} \qquad (11)$$

Однако, для каждой из трёх составляющих в (21) верно всё, сказанное относительно (2), поэтому, окончательно:

$$\vec{\Omega} = \frac{\vec{M}}{\kappa}$$

Численное моделирование движения связанной системы микроэллипсоидов в вязкой среде

Благодаря тому, что само движение эллипсоида можно разделить на поступательное и вращательное, информацию о положении каждого отдельного эллипсоида также можно хранить в виде координаты его геометрического центра и матрицы, являющейся композицией всех поворотов, которые этот эллипсоид претерпел.

Рассмотрим вначале описание поступательного движения. Понятно, что движение объекта при моделировании будет дискретным с некоторым шагом по времени τ.Из уравнения () легко получить элементарный сдвиг центра эллипсоида относительно предыдущего положения:

$\Delta\vec{x} = \vec{v}\,\tau$.

После чего объект сдвигается на расстояние $\Delta\vec{x}$:

$\vec{P}_{i+1} = \vec{P}_i + \Delta\vec{x}$,

где $\vec{P}_i$ – радиус-вектор центра эллипсоида на *i*-том шаге.

Для моделирования вращательного движения подход должен быть несколько другим. Угловая скорость $\vec{\Omega}$ в общем случае не является сонаправленной ни с одним

из базисных векторов. Вычисление в этом случае требует нескольких операций (см. рисунок 1).

Так процесс выглядит для одиночного эллипсоида. На каждом из шагов A, B, C мы получаем матрицу перехода к новому базису. Поворот объекта, следовательно, будет осуществляться так:

$$T_{i+1} = A \times B \times C \times T_i ,$$

где T_i – такая матрица поворота, которая совмещает оси эллипсоида с координатными осями на *i*-том шаге.

В задаче для связанных эллипсоидов (к примеру, один эллипсоид содержит несколько внутренних) необходимо учитывать ещё два фактора:

- эллипсоиды передвигаются не только сами по себе, но и как часть целого, вместе с внешним объектом;
- связанные эллипсоиды не могут нарушать границ друг друга.



Рисунок 2 . Порядок вычисления движения одиночного эллипсоида

Соответственно, для вложенного эллипсоида алгоритм будет другим(см Рис. 2.).

В этой диаграмме все шаги тривиальны, кроме «Обработки столкновения» и «Вращения вместе с родительским объектом». Способ обработки столкновения зависит от конкретной задачи. При передаче вращения от родительского объекта нужно учитывать два момента:

- поворот вектора, связывающего центры родительского и дочернего объектов (перемещение дочернего объекта из-за вращения родительского);
- собственно поворот дочернего объекта.



Рисунок 3 . Порядок вычисления движения дочернего объекта

Практическая реализация

При практической реализации изложенного подхода были получены численные результаты, качественно согласующиеся с экспериментом. Конкретно, был проведён

практический эксперимент по захвату клетки дрожжей *Saccharomyces cerevisiae* и промоделирован захват схожего по форме и размерам эллипсоида. Относительная ошибка определения скорости передвижения объекта в лазерном пучке составила в среднем 30%.

При моделировании захвата системы, изображённой на рисунке 3, гауссовым пучком наблюдалось следующее:

1) система смещалась как целое в центр интенсивности пучка;
2) на органеллы действовали силы светового давления, противоположные по направлению действующим на клетку в целом.

Следовательно, захват в оптическом пучке должен приводить к появлению напряжений в клетке. Силы оптического давления, действующие на внутренние объекты, невелики(на два порядка меньше действующих на клетку в целом), поэтому не оказывают заметного влияния на движение клетки в целом, однако в случае с захватом биологических объектов *in vitro* могут негативно воздействовать на жизнедеятельность последних.

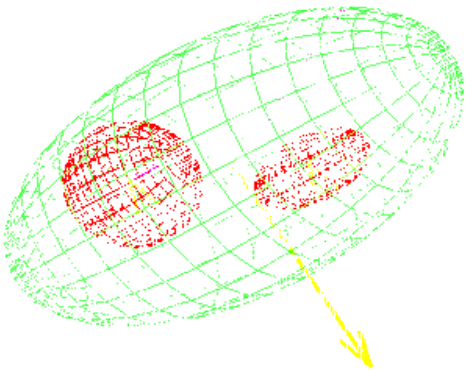


Рисунок 4 . Система из трёх эллипсоидов, использованная для моделирования

Заключение

Проведено численное моделирование движения системы связанных эллипсоидов в лазерном пучке. Получены оценки для величин, действующих на эллипсоиды сил и скоростей эллипсоидов.

Литература

1. A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu, "Observation of a single-beam gradient force optical trap for dielectric particles," Opt. Lett. 11, 288-290 (1986).
2. Zhang H, Liu K. K. Optical tweezers for single cells. J R Soc Interface 5:671–690 (2008)
3. Скиданов Р.В. Расчет силы взаимодействия светового пучка с микрочастицами произвольной формы // Компьютерная оптика. 2005. Вып. 28. С. 18-22.
4. Ландау Л.Д., Лифшиц Е. М. Теоретическая физика: Учебное пособие. В 10т. Т. VI. Гидродинамика. М.: Наука, 1986. 736с.
5. Наумов. В.А. Динамика дисперсной частицы вязкой среде // Математическое моделирование, 2006. №18. С.27–36.
6. Бреннер Г., Хаппель. Дж. Гидродинамика при малых числах Рейнольдса. М.: Мир, 1976. 631с

Улучшение разрешения на основе использования пространственных фильтров

Д.А.Савельев[1], С.Н. Хонина[1,2]
[1]*Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара*
[2] *Институт систем обработки изображений РАН, Самара*

Введение

В последние годы в связи с развитием нанотехнологии, клеточной медицины возрос интерес к проблемам улучшения разрешения в оптике. В настоящее время достигнута реализация сверхразрешения порядка нескольких нанометров в изображениях биообъектов при использовании ближнепольных и атомных микроскопов, но обработка изображения занимает длительный срок и все более востребованы оптические микроскопы, позволяющие визуализировать структуру в целом. В данной работе проводятся исследования различных оптических установок и типов пространственных фильтров для улучшения разрешения.

1 Фильтры Торалдо

В работе мы используем несколько вариантов установок. Первая из них – для исследования пространственных фильтров Торалдо:

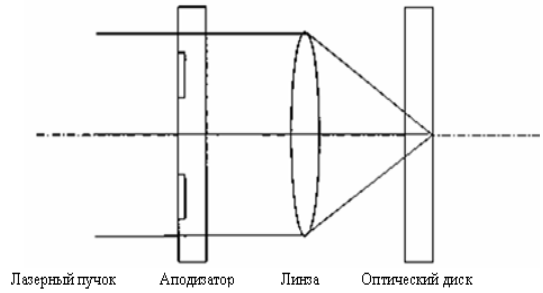


Рис. 1. Системная установка

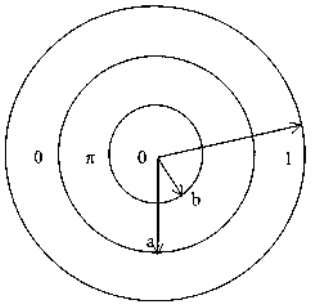


Рис. 2. Структура фильтра Торалдо

Чисто-фазовые пространственные фильтры используются, чтобы оптимизировать осевое распределение интенсивности и увеличить глубину фокуса для оптического захвата. С использованием данного вида фильтра ожидаемая глубина фокуса может быть в 2-4.88 раза больше чем из оригинальной системы, и размер пятна рассеивания будет меньшим [1]. Для нас актуальным является то, что использование данного типа фильтров позволяет добиться улучшения разрешения путем уменьшения пятна рассеивания. Использовалась изображенная на рис. 1 системная установка и фильтр с изображенной структурой – данная структура фильтра с чередованием 0-π впервые была предложена Торалдо [2]. Нормализованное амплитудное распределение в плоскости сужения пучка имеет следующий вид:

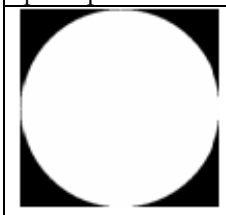
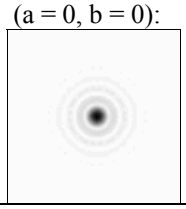
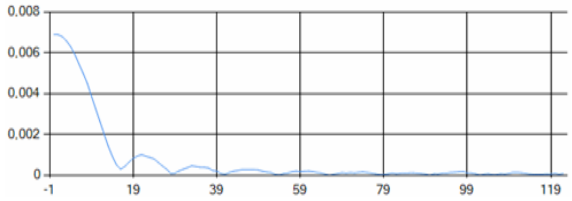
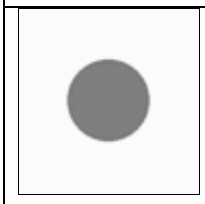
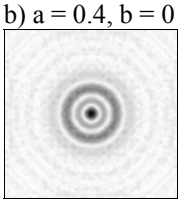
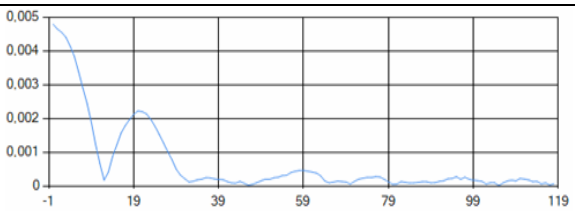
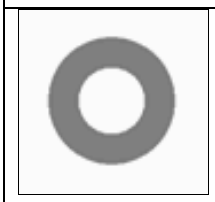
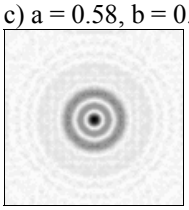
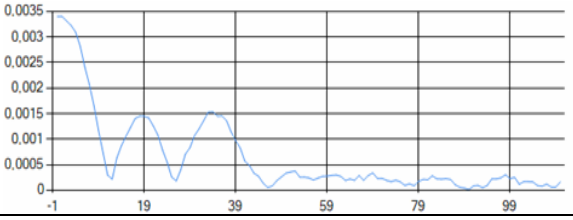
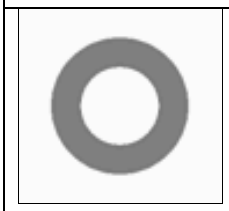
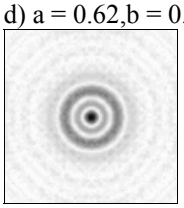
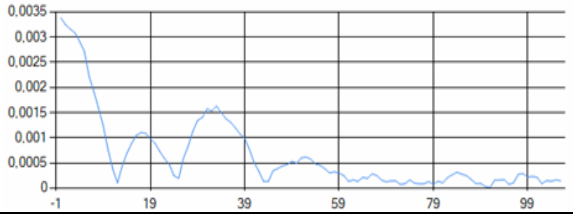
$$G(\rho,u)=2\sum_{j=1}^{N}\exp(i\phi_j)\int_{j-1}^{r_j} rJ_0(\rho r)\times\exp[-(1/2)iur^2]dr \qquad (1)$$

Когда число скачков равно 2 (рис. 2) осевое распределение интенсивности зависит от двух коэффициентов – числа Штреля и внутренний радиуса фильтра b.

$S=[G_{a\neq b}(0,0)/G_{a=b=0}(0,0)]^2=[1-2(a^2-b^2)]^2$ - коэффициент Штреля. (2)

Производились эксперименты в соответствии с [1]. Использовалась круглая апертура, с приведенными радиусами равными 10. На вход подавался вписанный круг, вид которого вместе с преобразованием Фурье от него и изображениями пространственных фильтров показаны в таблице 1. Фильтр является чисто фазовым элементом. Здесь и в дальнейшем все результаты – инвертированы.

Таблица 1. Исследования пространственного фильтра Торалдо

На вход/Вид фильтра	После линзы	Сечение амплитуды
	а) без фильтра (a = 0, b = 0): 	
	b) a = 0.4, b = 0 : 	
	c) a = 0.58, b = 0.3 : 	
	d) a = 0.62,b = 0.35 : 	

Исследования проводились исследования при различных параметрах a и b. Визуально, лучшие распределения в случаях c и d. Чтобы убедиться в этом рассмотрим их сечения, что подтверждает наше предположение. Случай d является наилучшим – более четкий провал по сравнению с c.

2 Вихревая и низкочастотная/высокочастотная фильтрация

Вторая установка – это классическая четырехфокусная система (показана на рис.3), где A_1 – входная плоскость, A_3 – выходная плоскость, L – сферическая линза, A_2 – плоскость фильтра, f – фокусное расстояние линзы.

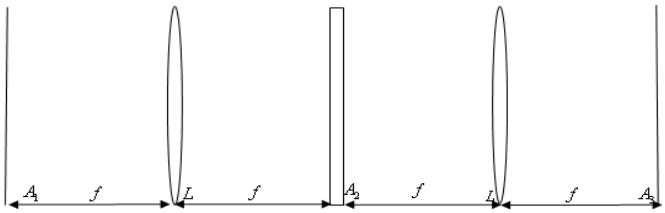


Рис. 3. Схема оптической установки (4f система)

Для моделирования линзы мы используем преобразование Фурье:

$$G(u,v)=\frac{k}{f}\int_{-\infty}^{\infty}\int g(x,y)\exp\left[-\frac{ik}{f}(xu+yv)\right]\mathrm{d}x\,\mathrm{d}y \qquad (3)$$

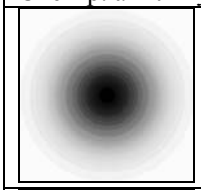
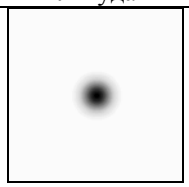
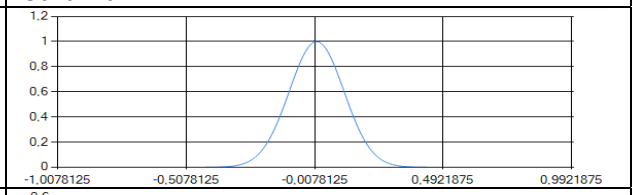
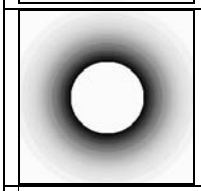
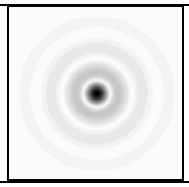
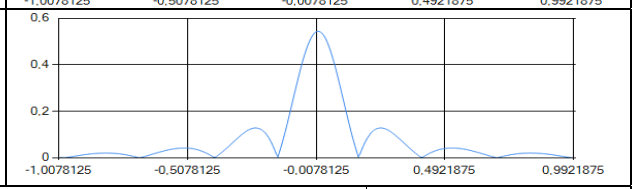
В качестве источника мы брали одинарные и двойные гауссовы пучки. В качестве фильтра мы использовали фильтр низких частот, он же теневой фильтр и различные его дополнения: производилось усечение низких частот, т.е. высокочастотная фильтрация. Т.о., по сути, производилось вырезание «кольца» в спектре. Также было применялось совмещение теневого и вихревого фильтров.

Нами были проведены исследования, показывающие зависимость радиуса вырезания «кольца» в спектре, на различение двух близко расположенных источников, причем нас интересовали сечения амплитуд, фазы после обратного преобразования Фурье за отсутствием информативности не приводятся. Эксперименты проводились над изменением внутреннего радиуса кольца; над изменением внешнего радиуса, при неизменном внутреннем. Исследовали изменение зависимости ZR (радиус от центра гаусса, т.е. максимальной интенсивности, до спада интенсивности до 0.) и FWHM от радиуса вырезания.

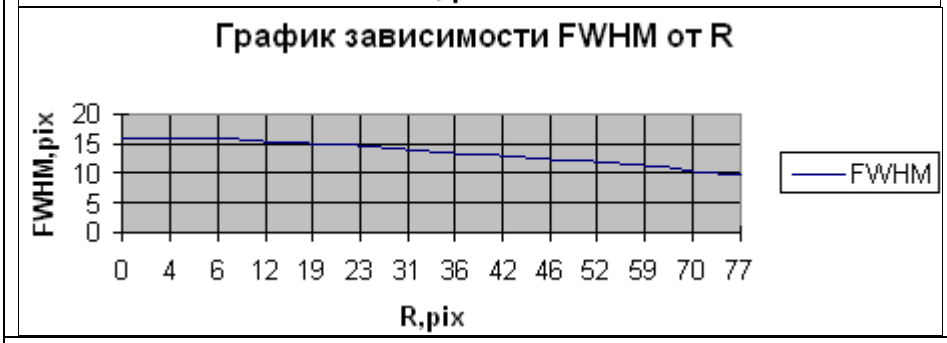
Учитывая то обстоятельство, что нам требуется получать в спектре достаточно хорошо прописанный пучок, было произведено дополнение первоначальной картинки размером 256x256 до 2048x2048. Внешний вид фильтров (спектр обрезан до 256x256), амплитуды пучков и их сечения после прохождения системы демонстрируются в таблице 2.

Таблица 2. Вихревая и низкочастотная/высокочастотная фильтрация

1) Входное изображение: $F = e^{-\frac{\left((x-x_1)^2+(y-y_1)^2\right)}{\sigma^2}}$, $\sigma = 0.15$, фаза нулевая.

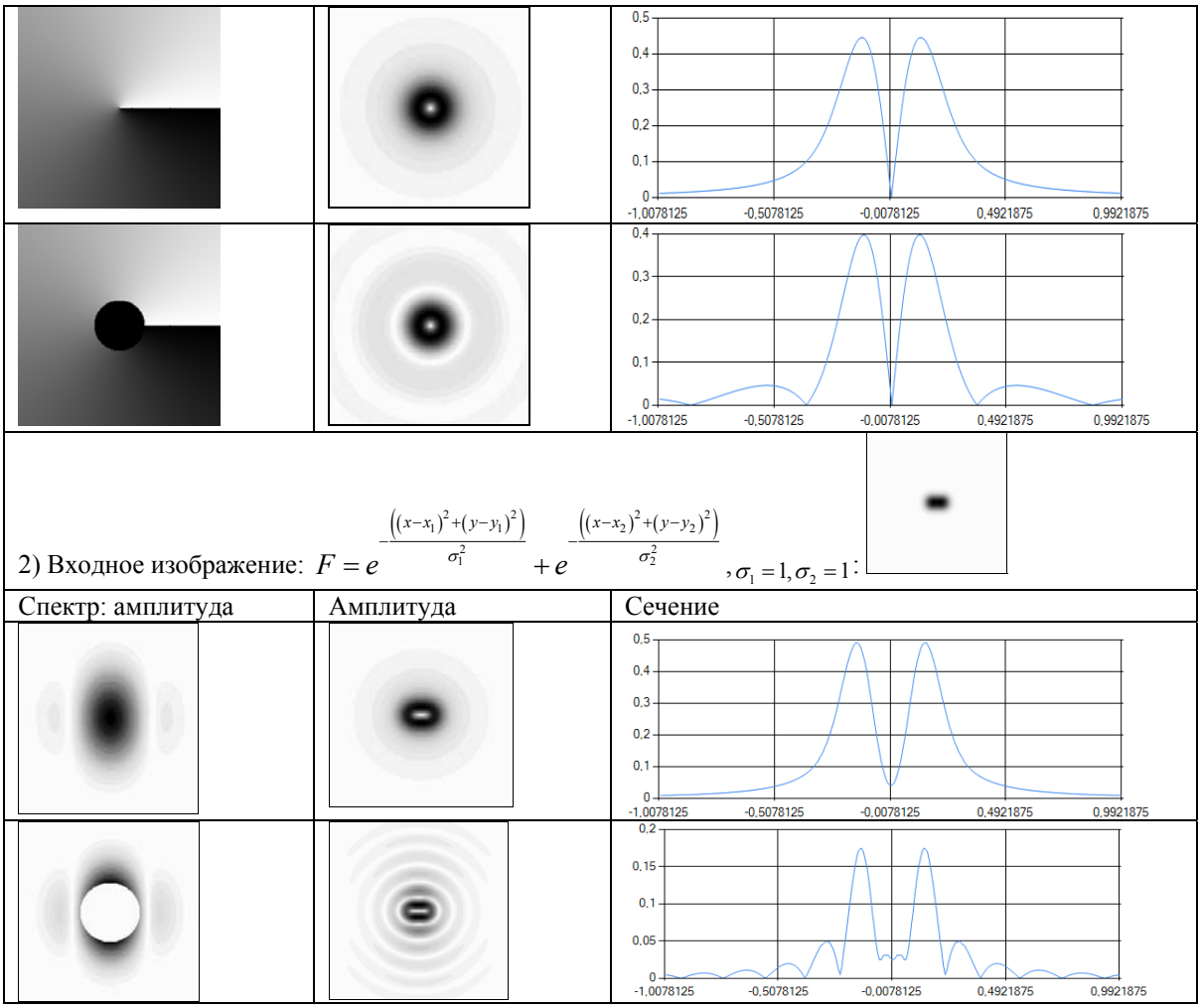
Спектр: амплитуда	Амплитуда	Сечение
		
		





Теневой вихревой фильтр (приведена фаза в спектре, у амплитуды вырезание аналогично)

Спектр: фаза	Амплитуда	Сечение



Вообще говоря, вырезание «кольца» в спектральной области, которое мы проводили, соответствует в объектной области аксикону и часто в оптических системах используют аксикон.

3 Аксикон

Аксикон – это оптический элемент, делящий волновой фронт. Он описывается формулой: $e^{-i\alpha r}$. Из всех возможных видов аксиконов наиболее прост в изготовлении и подходит для многих задач – конус. Для конуса есть интересное свойство – с увеличением расстояния от конуса увеличивается различимость[3].

В данной части работы мы использовали следующую схему оптической системы (рис. 4) – аксикон и идущие после него две линзы с различным фокусным расстоянием [4]. Эксперименты проводились с плоской освещающей волной, т.е. получали функцию Бесселя нулевого порядка [5].

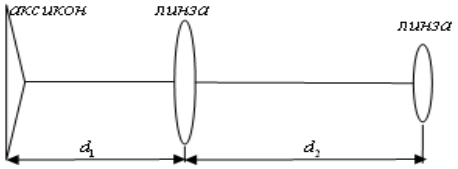
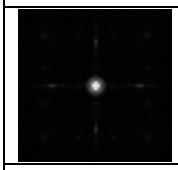
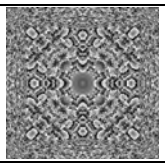
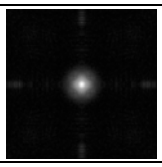
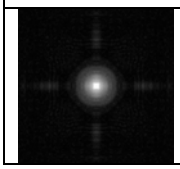
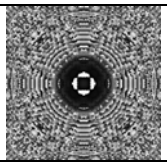
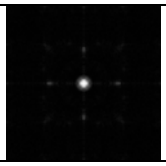
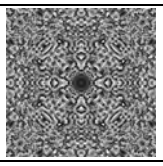


Рис. 4. Схема для экспериментов с аксиконом

Теперь приведем результаты моделирования схемы на рис.4. На входе – круг с радиусом R = 5 мм, расстояние до аксикона равно 300 мм, расстояние после аксикона - 300 мм. Далее полученные картинки подавались на вход системе из 2 линз, с разными фокусными расстояниями - с такими, чтобы образовывать требуемую систему. В дальнейшем ограничимся тем, что будем просто указывать фокусные расстояния

соответствующих линз (Для первой - f_1, для второй - f_2). Фокус второй линзы должен быть меньше чем первой [4]. Будем демонстрировать изображения, полученные после второй линзы. Результат показан в таблице 3.

Таблица 3 – Исследования аксиконов

Амплитуда	Фаза	Амплитуда	Фаза
Фокусное расстояние: $f_1 = 400; f_2 = 100$		$f_1 = 400; f_2 = 200$	
			
Фокусное расстояние: $f_1 = 400; f_2 = 250$		$f_1 = 500; f_2 = 100$	
			

Заключение

В работе были рассмотрены различные способы улучшения разрешения. Учитывая следующее определение различимости - две точки различимы, если центральный максимум одной точки лежит не ближе первого минимума другого – при помощи компьютерного моделирования было достигнуто улучшение разрешения в рассмотренных случаях.

Литература

1. Wang H., Gan F., High focal depth with a pure-phase apodizer, Applied optics 40(31), 5658 – 5662 (2001)
2. Cox I. J., Sheppard C. J. R., Wilson T., Reappraisal of arrays of concentric annuli as superresolving filters, Optical Society of America, Vol. 72(9),1287 – 1291 (1982)
3. Mcleod J.H., The Axicon: A New Type of Optical Element, Journal of the optical society of America, Vol 44(8), 592 – 597 (1954)
4. Lei M., Yao B., Multifunctional darkfield microscopy using an axicon, Journal of Biomedical Optics, 13(4), 044024-1 – 044024-4 (2008)
5. Хонина С.Н., Волотовский С.Г., Исследование применения аксиконов в высокоапертурной фокусирующей системе // Компьютерная оптика. – 2010. –Т. 34, № 1. – С. 35-51.

Моделирование R-FDTD методом субволновой фокусировки света с радиальной поляризацией зонной пластинкой

С.С. Стафеев
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Зонные пластинки используются в оптике уже более ста лет, однако ранее в работах преимущественно рассматривалась фокусировка в дальней зоне, и до недавнего времени оставался неисследованным вопрос по фокусировке зонной пластинкой в ближней зоне. В 2008 году в [1] для субволновой фокусировки была использована зонная пластинка с нанесенным на нее металлическим покрытием (плазмонная зонная пластинка). В [2] было показано, что фазовая зонная пластинка лучше подходит для субволновой фокусировки в ближней зоне, чем плазмонная зонная пластинка. В обеих работах пластинки освещались линейно поляризованными пучками, однако радиально поляризованный свет лучше подходит для субволновой фокусировки, чем линейно поляризованный, потому что не разрушает радиальную симметрию фокального пятна. В работе [3] для субволновой фокусировки была использована зонная пластинка, освещаемая радиально поляризованным светом, однако использовался метод 3D FDTD – медленный и ресурсоемкий по сравнению с методом, используемым в нашей статье.

В представленной работе с помощью R-FDTD (radial finite-difference time domain) метода моделировалось прохождение радиально-поляризованной моды R-TEM_{01} (с длиной волны λ=0,532мкм) через зонную пластинку с фокусным расстоянием f=10мкм и радиусом R=23мкм. Было показано, что диаметр фокального пятна, формируемого такой пластинкой, достигает FWHM=0,71λ, это значение хорошо согласуется с теоретическим значением, полученным с помощью формул Ричардса-Вольфа (FWHM=0,73λ).

1. Моделирование с помощью R-FDTD метода

Метод FDTD представляет собой численное решение уравнений Максвелла и был впервые предложен K. S. Yee в работе [4]. Позднее для этого метода была решена проблема ввода излучения в вычисляемую область [5] и проблема граничных условий [6]. Модификация метода для цилиндрической системы координат была предложена в работе [7], а в работе [8] данная модификация была применена для расчета распространения радиально-поляризованного света.

При вычислении распределения электромагнитного поля по методу FDTD расчетная область покрывается сеткой, в каждом узле которой рассчитывается отдельная компонента поля [4]. Для радиально-поляризованного света в цилиндрической системе координат из шести компонент электромагнитного поля отличны от нуля только три компоненты, а уравнения Максвелла приобретают вид [8]:

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = -\mu\mu_0 \frac{\partial H_\phi}{\partial t}, \quad (1)$$

$$-\frac{\partial H_\phi}{\partial z} = \varepsilon\varepsilon_0 \frac{\partial E_r}{\partial t} + \sigma E_r, \quad (2)$$

$$\frac{1}{r}\frac{\partial (rH_\phi)}{\partial r} = \varepsilon\varepsilon_0 \frac{\partial E_z}{\partial t} + \sigma E_z, \quad (3)$$

где E_r и E_z – радиальная и продольная составляющие напряженности электрического поля, H_φ – азимутальная составляющая напряженности магнитного

поля, ε и μ – диэлектрическая и магнитная проницаемости среды, ε_0 и μ_0 – диэлектрическая и магнитная постоянные, σ – электрическая проводимость.

В данной работе моделировалось прохождение радиально поляризованной моды R-TEM$_{01}$ через зонную пластинку с фокусным расстоянием f=10мкм и радиусом R=23мкм. Радиальное сечение такой зонной пластинки, а также ее расположение в вычисляемой области показаны на рисунке 1, числовая апертура такой зонной пластинки равна NA=sin(arctg(R/f))=0,9171.

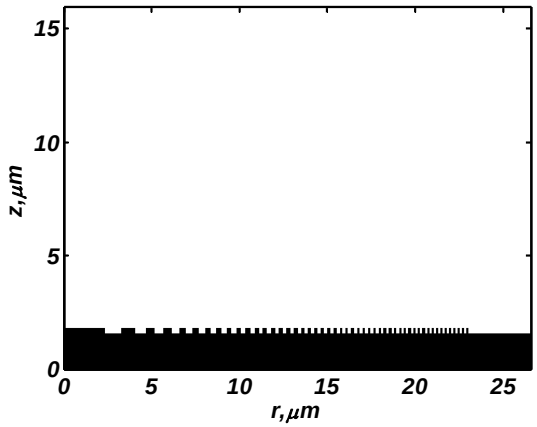


Рис. 1. радиальное сечение зонной пластинки и ее расположение в вычисляемой области при моделировании методом R-FDTD

Радиально-поляризованная мода R-TEM$_{01}$ имеет вид:

$$E_r = Ar\exp\left\{-\frac{r^2}{\omega^2}\right\}, \tag{4}$$

где ω – радиус гауссова пучка, A – константа. На рисунке 2 показано радиальное сечение такой моды, при ω=10,64мкм.

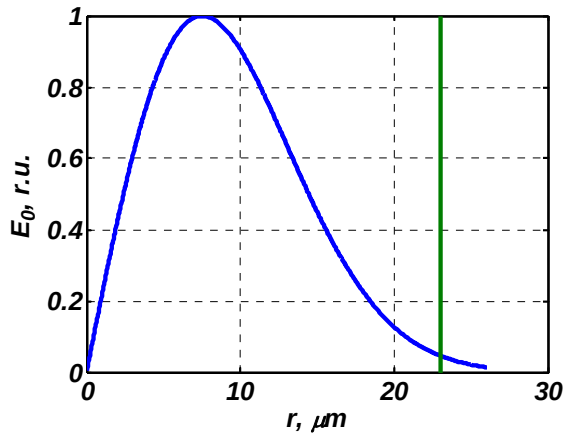


Рис.2. распределение радиальной составляющей напряженности электрической компоненты моды R-TEM$_{01}$ (зеленой прямой показан радиус зонной пластинки)

Размеры вычисляемой области при расчете по методу R-FDTD составляли 30λ*50λ (λ – длина волны), на краях помещались поглощающие слои Беренгера толщиной λ [6], длина волны была равна λ=0.532мкм, шаг координатной сетки по осям r и z был одинаков и составлял 20 отсчетов на длину волны.

На рисунках 3-4 показаны результаты моделирования: на рисунке 3а – распределение интенсивности в вычисляемой по методу R-FDTD области, на рисунке 3б – показано распределение интенсивности вдоль продольной оси z, а на рисунке 4 распределение интенсивности и отдельных компонент в фокальной плоскости. Все значения интенсивности нормировались на значение интенсивности на оси в фокальном пятне.

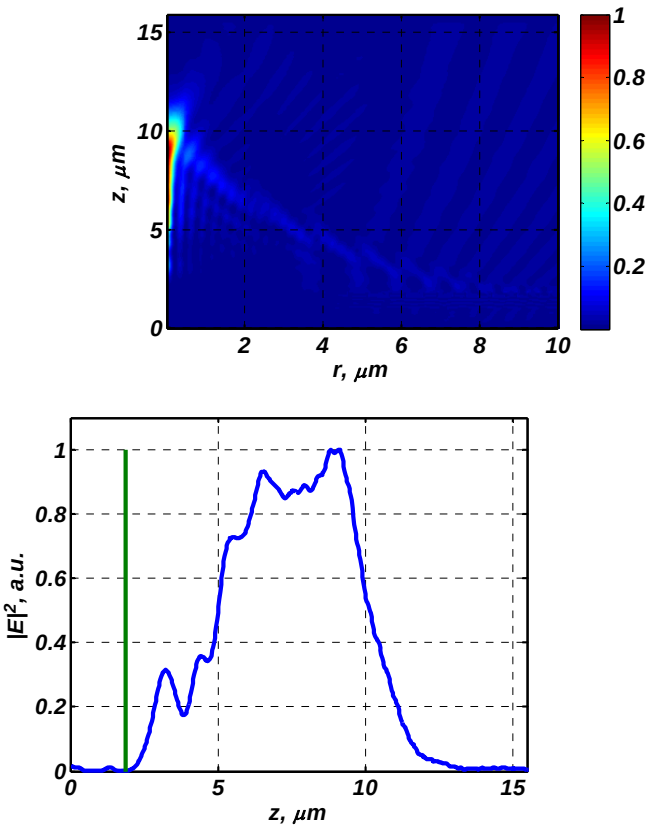


Рис. 3. распределение интенсивности (а) в области и (б) вдоль оси *z*. Зеленая прямая показывает границу зонной пластинки

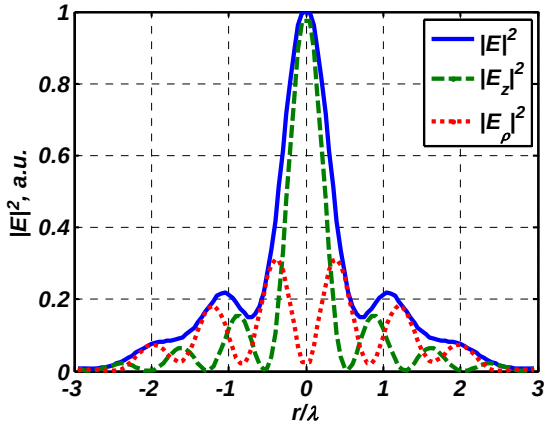


Рис. 4. распределение фокальной плоскости радиальной (красная пунктирная кривая), продольной (зеленая штриховая кривая) составляющих интенсивности и полной интенсивности (синяя сплошная кривая).

Поперечная ширина фокального пятна по полуспаду интенсивности составила FWHM=0,71λ (FWHM = full width at half maximum). Как видно из рисунка 3 фокус вытянут вдоль оси *z* ($FWHM_z$=9,69λ)

2. Сравнение с результатами на основе формул Ричардса-Вольфа

Чтобы проверить правильность моделирования методом R-FDTD была проведена другая серия моделирований на основе формул Ричардса-Вольфа [9]. Модификация формул Ричардса-Вольфа для фокусировки радиально поляризованных пучков была предложена в работе [10], позднее было показано, что с помощью радиально-поляризованных пучков можно преодолеть дифракционный предел [11]. А в работе [12] было показано, что дифракционный предел может быть преодолен и с помощью зонной пластинки.

Уравнения Ричардса-Вольфа для радиально поляризованного света имеют вид:

$$E_r(r,z) = A\int_0^{\theta_{\max}} T(\theta)B(\theta)\sin(2\theta)l_0(\theta)J_1(kr\sin\theta)\exp\{ikz\cos\theta\}d\theta \quad (5)$$

$$E_z(r,z) = 2iA\int_0^{\theta_{\max}} T(\theta)B(\theta)\sin^2(\theta)l_0(\theta)J_0(kr\sin\theta)\exp\{ikz\cos\theta\}d\theta \quad (6)$$

где $l_0(\theta)$ – начальное распределение электрической составляющей поля в координатах выходного зрачка, $J_n(x)$ – функция Бесселя порядка n, A – константа, $T(\theta)$ – функция аподизации зрачка (для зонной пластинки $T(\theta)=\cos^{-3/2}\theta$), $B(\theta)$ – функция пропускания.

На рисунках 5-6 показаны результаты распределения интенсивности в фокусе, полученные на основе уравнений (5)-(6).

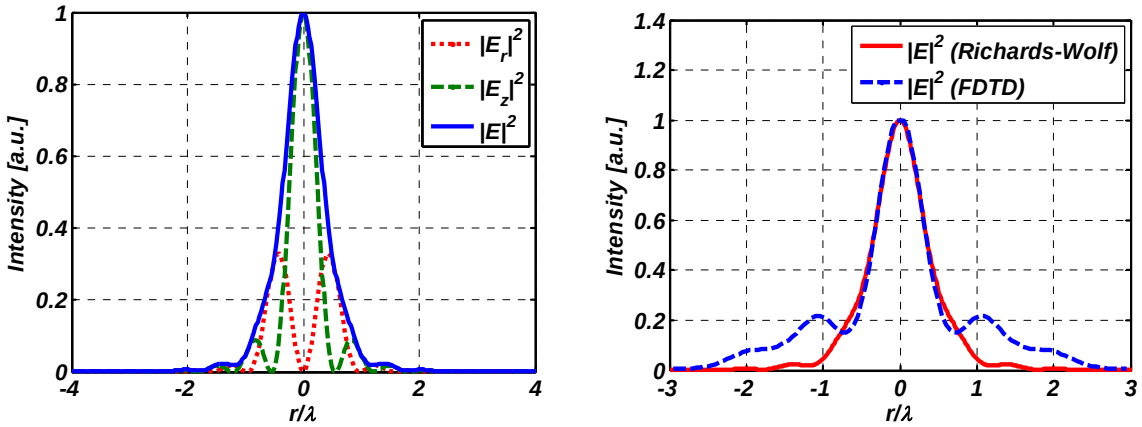


Рис. 5. (а) распределение в фокальной плоскости радиальной (красная пунктирная кривая), продольной (зеленая штриховая кривая) составляющих интенсивности и полной интенсивности (синяя сплошная кривая) (б) сравнение распределения полной интенсивности, полученной разными методами: методом R-FDTD (синяя штриховая кривая) и с помощью формул Ричардса-Вольфа (красная сплошная кривая)

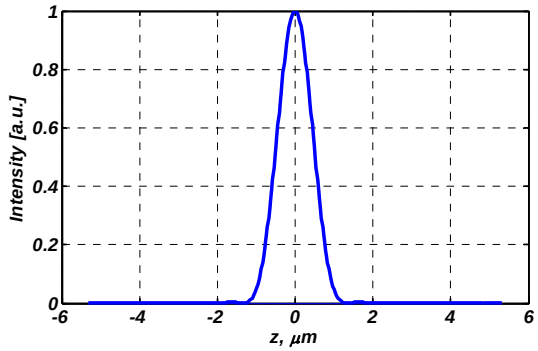


Рис. 6 . распределение интенсивности в фокусе вдоль оси z

Теоретическое значение диаметра фокального пятна по полуспаду интенсивности (рисунок 5а) составило FWHM=0,73λ, что согласуется с предыдущим результатом (FWHM=0,71λ). Однако продольная ширина фокального пятна (рисунок 6) оказалась значительно меньше, чем полученная ранее методом R-FDTD, и составила FWHM=1,95λ.

Заключение

В данной работе с помощью R-FDTD метода моделировалось прохождение радиально-поляризованной моды R-TEM$_{01}$ через зонную пластинку с фокусным расстоянием f=10мкм и радиусом R=23мкм. Было показано, что диаметр фокального пятна, формируемого такой пластинкой, достигает FWHM=0,71λ, это значение хорошо согласуется с теоретическим значением, полученным с помощью формул Ричардса-Вольфа (FWHM=0,73λ). Однако моделирование с помощью метода R-FDTD показывает большие значение интенсивности боковых пиков (рисунок 5б). Также моделирование с

помощью R-FDTD метода показало, что фокус имеет значительную продольную ширину $FWHM_z$=9,69λ (рисунок 3), в то время как теоретическое значение продольной ширины фокального пятна (на основе формул Ричардса-Вольфа) $FWHM_z$=1,95λ (рисунок 6).

Литература

1. Y. Fu, W. Zhou, L. E. N. Lim "Near-field behavior of zone-plate-like plasmonic nanostructures", JOSA A, v25, p238-249 (2008)
2. R. G. Mote, S. F. Yu, B. K. Ng, W. Zhou, S. P. Lau "Near-field focusing properties of zone plates in visible regime – New insights", Opt. Express, v16, p9554-9564 (2008)
3. R. G. Mote, S. F. Yu, W. Zhou, X. F. Li "Subwavelength focusing behavior of high numerical-aperture phase Fresnel zone plates under various polarization states", Appl. Phys. Lett, v95, p191113 (2009)
4. K. S. Yee "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", IEEE Trans. Antennas Propag, AP-14, p302–307 (1966)
5. A. Taflove "Computational Electromagnetics: The Finite-Difference Time Domain Method", (Artech House, Boston, Mass., 1995).
6. J. P. Berenger "A perfectly matched layer for the absorption of electromagnetic waves", J. Comput. Phys., v114, p185–200 (1994).
7. D. W. Prather, S. Shi "Formulation and application of the finite-difference time-domain method for the analysis of axially symmetric diffractive optical elements", v16, JOSA A, p.1131-1142 (1999)
8. V.V. Kotlyar, S.S. Stafeev "Sharply focusing a radially polarized laser beam using a gradient Mikaelian's microlens", Optics Communications, v282, p459–464 (2009)
9. B. Richards, E. Wolf "Electromagnetic diffraction in optical systems II. Structure of the image field in an aplanatic system", Proc. Roy. Soc. A, v253, p358-379 (1959).
10. K. S. Youngworth, T. G. Brown "Focusing of high numerical aperture cylindrical vector beams", Optics Express, v7, p77-87 (2000)
11. R. Dorn, S. Quabis, G. Leuchs "Sharper Focus for a Radially Polarized Light Beam", Phys. Rev. Lett., v91, p233901 (2003).
12. N. Davidson, N. Bokor "High-numerical-aperture focusing of radially polarized doughnut beams with a parabolic mirror and a flat diffractive lens" Optics Letters, v29, p1318-1320 (2004)

Моделирование волноводных режимов в многослойных структурах

Т.С. Стрилец, В.В. Котляр
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

Введение

Многослойные планарные волноводы являются важными элементами современных интегральных оптических устройств, таких как полупроводниковые лазеры, модуляторы, светофильтры, поляризаторы света. Разрабатывая такие устройства необходимо знать точные свойства распространения мод в волноводных структурах. Возможные типы мод: моды малых потерь, моды с потерями на вытекание[1, 2] и моды с потерями на поглощение и рассеивание.

Для нахождения мод многослойных планарных структур существуют несколько численных методов[3-8], которые разделяют на строгие и нестрогие методы. Строгие численные методы[3-6] непосредственно ищут корни в комплексной плоскости, погрешность такого метода ограничена только численной точностью, они затратные по времени. Нестрогие методы не дают точных решений, они учитывают резонансы при поперечном освещении, которые соответствуют вытекающим модам. Такие методы в сравнении со строгими методами требуют меньше времени вычисления. Рассматриваемый в данной работе метод TTR относится к нестрогим методам. Он предполагает добавление слоя с высоким показателем преломления к покрытию, не требует выбора направляющего слоя. Снятие этого ограничения позволяет применять метод TTR к произвольной многослойной волноводной структуре, и этот метод обеспечивает прямой и простой подход к определению постоянных распространения вытекающих мод и профилей мод волновода.

1. Матрица переноса

Матрица переноса применяется при систематическом подходе к анализу многослойных структур. На рис. 1 представлена геометрия многослойной структуры, где n_c и n_s - показатели преломления слоя покрытия и подложки, n_i - показатель преломления i-го слоя волновода.

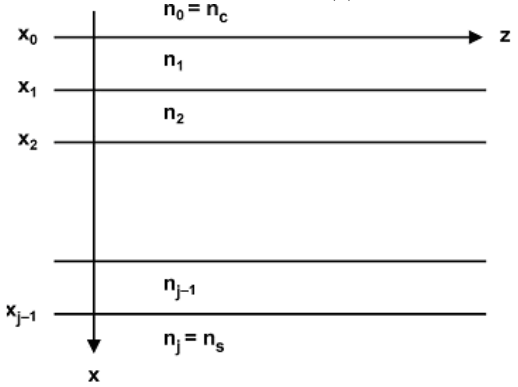


Рис. 1. Геометрия многослойной структуры.

Для поперечно-электрической (TE) волны, распространяющейся в направлении $+z$ на i-ом слое, $\left(x_{i-1} \le x \le x_i\right)$ напряженность электрического поля

$$E_i = yE_{yi}(x)\exp\left[j\left(\omega t - \gamma z\right)\right], \quad (1)$$

где y - единичный вектор в направлении $+y$ и $\gamma = \beta - j\alpha$ - комплексная постоянная распространения, где β и α - фазовый коэффициент и коэффициент затухания соответственно. Эффективный показатель преломления N определяется как

$N \equiv \beta / k_0$, где $k_0 = 2\pi / \lambda_0$ и λ_0 - длина волны. Напряженность электрического поля i-ого слоя, суммирующего волны, распространяющиеся от $-x$ до $+x$, может быть выражена следующим образом

$$E_{yi}(x) = A_i \exp\left[-\kappa_{x,i}\left(x - x_{i-1}\right)\right] + B_i \exp\left[+\kappa_{x,i}\left(x - x_{i-1}\right)\right], \quad (2)$$

где A_i и B_i являются комплексными амплитудами поля, распространяющихся в отрицательном и положительном направлениях оси x в i-ом слое, соответственно, $\kappa_{x,i} = \left(\gamma^2 - k_0^2 n_i^2\right)^{1/2}$ - поперечный волновой вектор i-ого слоя, и x_i определяет границу между i-ым и $i + 1$-ым слоем.

Через сравнение тангенциальных компонент электрического и магнитного полей на границе раздела, комплексные амплитуды полей в слое покрытия и подложки могут быть связаны с помощью матричного уравнения

$$\begin{pmatrix} A_s \\ B_s \end{pmatrix} = M_{j-1} M_{j-2} \ldots M_1 M_0 \begin{pmatrix} A_c \\ B_c \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} A_c \\ B_c \end{pmatrix}, \quad (3)$$

где матрица переноса для i-ого слоя M_i выглядит следующим образом

$$M_i = \frac{1}{2} \begin{pmatrix} \left[1 + f_i \frac{\kappa_{x,i}}{\kappa_{x,i+1}}\right] \exp\left(-\kappa_{x,i} d_i\right) & \left[1 - f_i \frac{\kappa_{x,i}}{\kappa_{x,i+1}}\right] \exp\left(\kappa_{x,i} d_i\right) \\ \left[1 - f_i \frac{\kappa_{x,i}}{\kappa_{x,i+1}}\right] \exp\left(-\kappa_{x,i} d_i\right) & \left[1 + f_i \frac{\kappa_{x,i}}{\kappa_{x,i+1}}\right] \exp\left(\kappa_{x,i} d_i\right) \end{pmatrix}, \quad i = 0, \ldots, j-1, \quad (4)$$

где для ТЕ волны $f_i = 1$, а для ТМ волны $f_i = n_{i+1}^2 / n_i^2$, d_i - толщина i-ого слоя. Для ТМ волн вместо $E_{y,i}$ в формуле (2) будет $H_{y,i}$.

Формула (3) через коэффициент пропускания и коэффициент отражения перепишется в виде

$$\begin{pmatrix} t_s \\ 0 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1 \\ r_c \end{pmatrix}, \quad (5)$$

$$\begin{pmatrix} r_s \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 0 \\ t_c \end{pmatrix}, \quad (6)$$

для случаев падающей плоской волны для покрытия и для подложки, соответственно [9,10].

В случае падения света на слой покрытия компонента поля в направлении $-x$ в подложке равна нулю $\left(B_s = 0\right)$. В случае падения света на подложку компонента поля в направлении $+x$ в слое покрытия равна нулю $\left(A_c = 0\right)$. В результате получим

$$r_c = -\frac{m_{21}}{m_{22}}, \qquad r_s = \frac{m_{12}}{m_{22}}, \quad (7)$$

$$t_c = \frac{1}{m_{22}}, \qquad t_s = \frac{m_{11} m_{22} - m_{12} m_{21}}{m_{22}}.$$

Доля энергии отражения

$$R_c = \left|r_c\right|^2, \qquad R_s = \left|r_s\right|^2. \quad (8)$$

Доля энергии пропускания

$$T_c = T_s = \left|t_c\right|^2 \frac{n_c \cos\left(\theta_c\right)}{n_s \cos\left(\theta_s\right)} = \left|t_s\right|^2 \frac{n_s \cos\left(\theta_s\right)}{n_c \cos\left(\theta_c\right)} = \left|t_c t_s\right|, \quad (9)$$

где θ_c - угол падения света на слой покрытия и θ_s - угол выхода света из подложки, и, наоборот, θ_s - угол падения света на подложку и θ_c - угол выхода света из слоя покрытия. Равенство (9) вытекает из обратимости хода лучей. Из (7) видно, что при m_{22} = 0 коэффициенты отражения и пропускания становятся бесконечными. Поэтому величина γ, при которой $m_{22} = 0$ (или минимальная), будет равна константе распространения каналируемой моды (если γ вещественная) или вытекающей моды (если γ мнимая) [11].

2. Метод поперечного пропускания/отражения (TTR)

Метод TTR определяет все моды только в том случае, если показатели n_c, и n_s больше, чем эффективный показатель N интересующих мод. Это условие достигается за счет добавления слоя (слоёв) с показателем выше, чем эффективный показатель любой волноводной моды, к покрытию многослойной структуры. В результате структура будет возмущенной версией начальной структуры с немного смещенной действительной частью постоянной распространения и чуть большим коэффициентом затухания по сравнению с постоянной распространения начальной моды. Следует отметить, что α и β, определяемые методом TTR, рассматриваются для возмущенной структуры, которая содержит добавленный слой с высоким показателем.

Для получения решений вытекающих мод волноводной структуры с добавленным слоем с высоким показателем методом TTR необходимо отыскать минимум m_{22} как функции от двух аргументов α и β численным методом. В данной работе для отыскания минимума был использован метод спуска по координатам [12].

3. Результаты моделирования мод многослойной планарной структуры

Для нахождения вытекающих мод методом TTR был произведен расчет для структуры ARROW (anti-resonant reflecting optical waveguide). Структура ARROW представляет собой многослойный планарный оптический волновод, свойства которого определяются антирезонансным отражением от многослойной оболочки [13]. Структуры ARROW используют тонкие и с высокой отражательной способностью слои между направляемым слоем и подложкой для уменьшения потерь излучения мод.

В данном конкретном случае была рассмотрена структура, представленная на рис. 2(a), первоначально покрытие - воздушное, поэтому слой с высоким показателем должен быть добавлен выше воздушного слоя для возбуждения вытекающих мод с эффективным показателем больше 1. Толщина воздушного зазора, d_a, выбирается $0{,}5\lambda$ для достижения соответствующей интенсивности. Рис. 2(b) показывает структуру с добавленным слоем с высоким показателем. Длина волны в свободном пространстве для этой структуры составляет 0.6328 *мкм*.

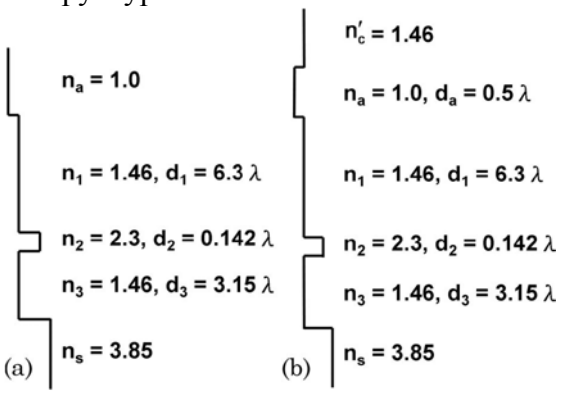


Рис. 2. (a) Структура ARROW. (b) Структура ARROW с добавленным слоем с высоким показателем.

Рассчитанные с точностью $\varepsilon = 10^{-13}$ постоянные распространения вытекающих мод для структуры, представленной на рис. 2(b) приведены в табл. 1. В табл. 2 представлены результаты моделирования TTR методом структуры ARROW, описанной в работе [1]. Полученные константы распространения для вытекающей моды исследуемой структуры совпадают с аналогичными константами в работе [1] с точностью 10^{-2}.

Таблица 1. Расчет нормированной постоянной распространения $\gamma_n / k_0 = (\beta_n - j\alpha_n)/k_0 = N_n - jHWHM_n$ вытекающей моды TE_n для структуры ARROW, представленной на Рис. 3 (b).

Мода TE_n	γ_n / k_0
TE_1	1.457837895957358 - j1.269362923598716 e-013
TE_2	1.45188 - j4.96e-005
TE_3	1.45123 - j1.75e-004
TE_4	1.440422432800073 - j5.2e-006
TE_5	1.4269 - j1.995e-004
TE_6	1.424889613177957 - j7.036e-004

Таблица 2. Результаты моделирования TTR методом структуры ARROW, приведенные в работе [1].

Мода TE_n	γ_n / k_0
TE_1	~1.45794
TE_2	1.45192 - j4.97e-005
TE_3	1.45122 - j1.78e-004
TE_4	1.44137 - j5.47e-006
TE_5	1.42740 - j2.02e-004
TE_6	1.42462 - j7.15e-004

Также для каждой из рассчитанных мод $TE_1 - TE_4$ из табл. 1 был найден профиль, который содержит волны, распространяющиеся от волновода в слой с высоким показателем и в подложку. Профили указанных мод представлены на рис. 3-6.

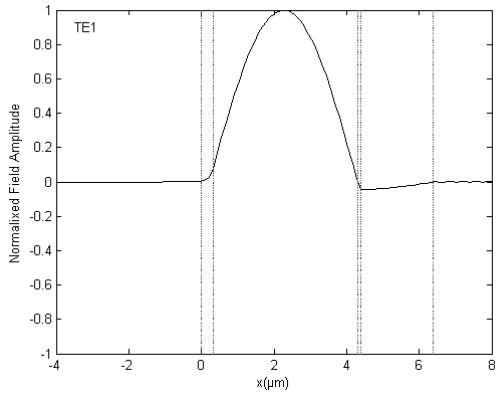


Рис. 3. Профиль моды TE_1 из табл. 1 структуры ARROW, представленной на Рис. 2 (b).

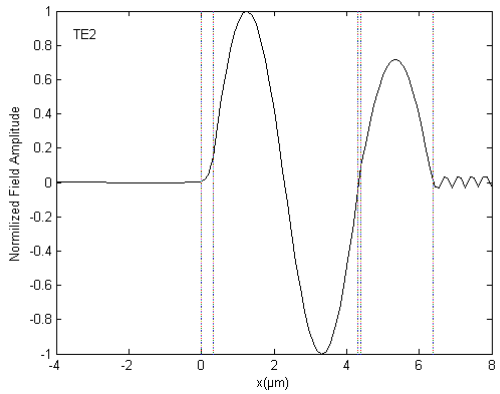


Рис. 4. Профиль моды TE_2 из табл. 1 структуры ARROW, представленной на Рис. 2 (b).

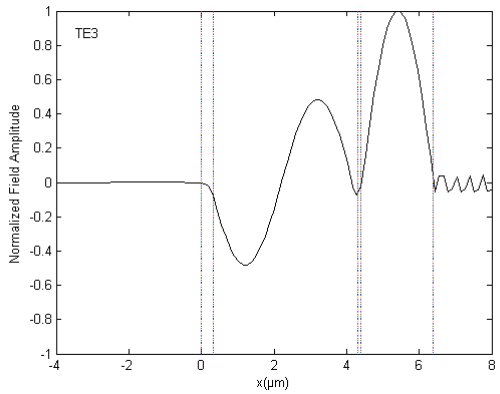


Рис. 5. Профиль моды TE_3 из табл. 1 структуры ARROW, представленной на Рис. 2 (b).

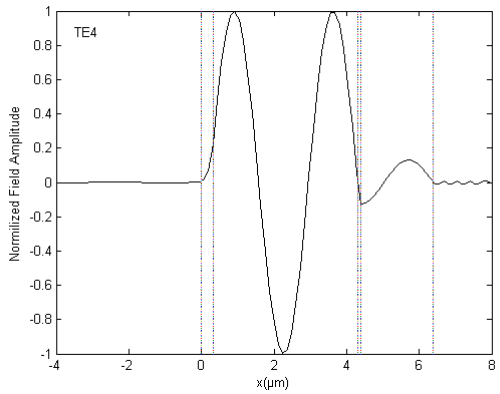


Рис. 6. Профиль моды TE_4 из табл. 1 структуры ARROW, представленной на Рис. 2 (b).

Заключение

Предложен численный метод расчета констант распространения волноводных и вытекающих мод многослойного планарного волновода, где минимум функции двух переменных отыскивается методом спуска по координатам. Проведено моделирование 6-ти слойной антирезонансной отражающей структуры ARROW и рассчитаны комплексные константы распространения для 6-ти первых вытекающих мод ; эти константы отличаются от полученных в [1] в третьем знаке после запятой (отклонение 10^{-2}) в реальной части. Также для 4 первых вытекающих мод планарного 6-ти слойного волновода ARROW построены их нормированные профили. Из этих 4 мод только фундаментальная мода (основная) распространяется без потерь и является каналируемой модой.

Литература

1. C.Lin, Gaylord T.K. Attenuation and mode profile determination of leaky/lossy modes in multilayer planar waveguides by a coupling simulation method // J. Appl.Opt., 19, 3603–3613 (2009).
2. Marcuvitz N. On field representations in terms of leaky modes or eigenmodes // IRE Trans. Antennas Propag. 4, 192–194 (1956)
3. Tamir T., Kou F. Y. Variety of leaky waves and their excitation along multilayered structures // J. Quantum Electron. 22, 544–551 (1986).
4. Anemogiannis E., Glytsis E.N. Multilayer waveguide: efficient numerical analysis of general structures // J. Lightwave Technol. 10, 1344–1351 (1992).
5. Smith R.E., Houde-Walter S.N. Leaky guiding in nontransparent waveguides // J. Opt. Soc. Am. A 12, 715– 724 (1995).
6. Anemogiannis E., Glytsis E. N., Gaylord T. K. Efficient solution of eigenvalue equations of optical waveguiding structures // J. Lightwave Technol. 12, 2080–2084 (1994).
7. Koshiba M., Kumagami H., Suzuki M. Finite-element solution of planar arbitrarily anisotropic diffused optical waveguide // J. Lightwave Technol. 3, 773–778 (1985).
8. Ghatak A. K., Thyagarajan K., Shenoy M. R. Numerical analysis of planar optical waveguides using matrix approach // J. Lightwave Technol. 5, 660–667 (1987).
9. A. W. Snyder, A. W., Love J. D., Optical Waveguide Theory (Chapman and Hall, 1983).
10. Ulrich R., Theory of prism-film coupler by plane-wave analysis // J. Opt. Soc. Am. 60, 1337–1350 (1970).
11. Azzam R. M. A., Bashara N. M., Ellipsometry and Polarized Light (Elsevier, 1977).
12. Калиткин, Н.Н. Численные методы: учебное пособие для вузов. - М.: Наука, 1978. - 512 с.
13. Kersten R. T. The prism-film coupler as a precision instrument Part I. Accuracy and capabilities of prism coupler as instruments // Opt. Acta 22, 503–513 (1975).

Исследование методов контроля твердости слоев при формировании наноструктур

А.Д. Щербатенко
Самарский государственный аэрокосмический университет имени академика С.П. Королева (национальный исследовательский университет), Самара

При формировании микрорельефов дифракционных оптических элементов плазмохимическим травлением, важную роль играет плазмостойкость, которая в свою очередь характеризуется нанотвердостью материала.

В данной работе исследовались методы контроля и измерения нанотвердости фоторезиста и электронного резиста. В ходе экспериментов были определены твердости наносимых покрытий в зависимости от скорости вращения центрифуги и сделаны выводы о методике нанесения светочувствительных материалов на подложки при создании микрорельефов.

При выполнении работы использовалась измерительная установка NanoTest, для эксплуатации которой были изучены принципы и режимы работы, методики измерений механических свойств поверхностей.

При провидении экспериментов на платформе NanoTest, предварительно необходимо выполнить определенный ряд калибровок. Данные калибровки необходимы для точности измерений при проведении экспериментов.

Процесс индентирования – внедрение одного тела (в идеале, абсолютно твердого) в другое, является относительно простым, оперативным и неразрушающим способом испытания материалов. Метод позволяет проводить измерения в весьма малых объемах, что делает его особенно привлекательным для определения механических свойств отдельных фаз, тонких пленок, слоев деталей, подвергшихся поверхностному воздействию, и т.п.

В ходе работы были проведены контрольные измерения твердости кварцевого стекла (в качестве исследуемого образца было выбрано кварцевое стекло, т.к. при калибровке NanoTest кварц используется как контрольный образец с хорошо известными механическими свойствами) и получены значения близкие к действительным, погрешность измерений составила 1.2%. Было определенно, что, для более точных и близких к табличным значениям твёрдости кварцевого стекла результатам, необходимо максимально оградить эксперимент от внешних помех и, проводить индентирование при больших нагрузках или на больших глубинах. Также была поставлена задача по определению необходимой толщины наносимых покрытий для точного определения их механических характеристик.

Опираясь на полученные в ходе экспериментов значения твердости кварцевого стекла, можно сделать вывод что измерения, проводимые на установке NanoTest дают достаточную точность при индентировании. Результаты измерений показали, что выбранный материал кварцевой подложки удовлетворяет требованиям предъявляемых к материалам используемых в качестве основы для оценки нанотвердости исследуемых материалов.

В качестве исследуемых тонких пленок были взяты электронные резисты и фоторезисты, применяемые в микро- и нанотехнологии в качестве маскирующих слоев, наносимых методом центрифугирования на кварцевые подложки. Исследуемые в ходе работы образцы представляли собой подложки с нанесенными на них тонкими пленками фоторезиста и электронного резиста. Нанесение резистов на подложку, в качестве которой использовалось кварцевое стекло с нанесенным на нее слоем хрома (для более хорошей адгезии), осуществлялось методом центрифугирования. В методе

центрифугирования предварительно растворенный в растворителе полимер наносился на подложку и, за счет действия центробежных сил в центрифуге, равномерно ложился на поверхность, образуя ровную тонкую пленку.

Было изготовлено 6 образцов. Три образца электронного резиста 495PMMA 11%A при скорости вращения центрифуги 3000, 5000 и 7000об/мин. И три фоторезиста ФПТ1-40 при таких же скоростях вращения центрифуги. После нанесения пленки образцы помещались в термостатическую печь, в которой поддерживалась постоянная температура 180ºС. В печи происходило испарение растворителя в течение 40мин.

В ходе экспериментов были определены твердости наносимых покрытий в зависимости от скорости центрифуги. Была определена зависимость твердости у электронного резиста в моменты до задубливания и после.

При исследовании электронного резиста до стадии задубливания был проведен ряд внедрений при различных нагрузках и различных глубинах. Внедрение осуществлялось до углубления в 1000nm со скоростью нагружения индентора 0.1mN/s.

Для простоты будем обозначать резисты, нанесенные при скорости вращения центрифуги 3000, 5000 и 7000об/мин, электронными резистами №1, №2 и №3 соответственно.

При индентировании на 1000nm электронных резистов получились следующие значения нанотвердости: для электронного резиста №1 - 0.701±0.054 GPa, для №2 - 0.760±0.086 GPa и для №3 - 0.689±0.029 GPa. Следовательно, можно сделать вывод, что твердость в электронном резисте 495PMMA без задубливания с увеличением числа оборотов практически не меняется, т.к. пленка резиста не до конца структурировалась и стабилизировалась.

В качестве индентора использовался трехгранный алмаз Берковича.

При индентировании задубленных электронных резистов была получена следующая толщина нанесенных пленок: для резиста №1 - 1800nm, для №2 - 1650nm и для №3 - 1050nm. Следовательно, можно сделать вывод, что с увеличением скорости вращения центрифуги при нанесении полимера толщина наносимых пленок уменьшается.

Также при индентировании электронных резистов на глубину 1000nm получились следующие значения для твердости, только уже после стадии задубливания: электронный резист №1 - 0.906±0.053 GPa, №2 - 1.617±0.285 GPa. Для резиста №3, толщина которого составляет 1050нм, при индентировании на 1000нм значении твердости имеет не точное значение и меняется в пределах от 1 до 3 GPa. Это связанно с тем, что толщина этого резиста слишком мала для измерения, т.к. при внедрении на 1000nm уже начинается деформация подложки, что влияет на результат измерения.

В качестве исследуемого образца фоторезиста был выбран ФПТ1-40. Было нанесено три фоторезиста на подложку при скорости вращения центрифуги 3000, 5000 и 7000об/мин. Задубливание данных резистов не было произведено, т.к. на данный момент отсутствует технология задубливания данного резиста.

Для удобства аналогично будем обозначать фоторезисты нанесенные центрифугированием со скоростью 3000, 5000 и 7000об/мин – фоторезист №1, №2 и №3 соответственно.

Индентирование всех трех резистов проводилось при различных нагрузках и глубинах внедрения трехгранным алмазом Берковича.

При индентировании резиста ФПТ1-40 нанесенного на подложки методом центрифугирования со скоростью 3000, 5000 и 7000об/мин при внедрении индентора на 1000nm получились следующие значения нанотвердости: для 1-ого фоторезсита - 0.306±0.005 GPa, для 2-ого - 0.389±0.004 GPa и для 3-его - 0.408±0.008 GPa.

Исходя из полученных в ходе экспериментов значений твердости фоторезистов можно сделать вывод, что твердость наносимой пленки резиста, аналогично электронному резисту, возрастает с увеличением скорости вращения центрифуги при нанесении резиста.

В ходе экспериментов были определены твердости наносимых покрытий в зависимости от скорости центрифуги. Была определена зависимость твердости у электронного резиста в моменты до задубливания и после.

Разработан метод контроля твердости слоев светочувствительных материалов, позволяющий получать необходимую твердость за счет изменения скорости вращения центрифуги при их нанесении.

Полученные при исследовании результаты в дальнейшем будут использоваться при создании микрорельефов методом плазмохимического травления.

СОДЕРЖАНИЕ

Секция 1. АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ

Подсекция 1. «Информационные технологии в научных исследованиях»

SCAN: Software package of spectral-correlation analysis
Sergey A. Prokhorov, Ilona M. Kulikovskikh……3

Пропаганда открытых кодов
Е.В. Авдеев……9

Информационная технология сравнительной оценки доказательств
К.А. Арутюнян, А.А. Ильичев……10

К задаче оптимального выбора элементов псевдослучайной последовательности
М.С. Воронов……14

Перспективы применения программного комплекса аппроксимативного корреляционно-спектрального анализа для обработки результатов испытаний камер сгорания ГТД
Я.В. Газетова……15

Сравнительный анализ методов вычисления спектральной плотности мощности сигналов
В.В. Графкин……20

Извлечение нечетких правил из числовых данных и структуризация базы знаний для задачи классификации
А.Н. Даниленко, А.Н. Коварцев……24

Фреймовая база знаний информационной компьютерной среды
В.П. Дерябкин, А.И. Белоусов……27

Математическая модель и решение обобщенной задачи о покрытии
Б.А. Есипов……31

Автоматизированная система выполнения математических расчетов в сети Интернет с элементами социальной сети
А.В. Камынин……33

Модель для создания системы поддержки принятия решений в области планирования и управления инновационными программами
В.Б. Киреев, Т.В. Кондранин, А.В. Николаев, Д.И. Озол, Д.Б. Пайсон, А.В. Киселёв, Ю.И. Полищук, С.С. Семёнов……37

О некоторых методах прогнозирования нестационарных временных рядов
А.А. Курбакова……43

Применение сети Вольтерри для решения задачи идентификации случайных последовательностей
И.В. Лёзина……47

Система научной визуализации химических процессов
А.В. Смирнов, Л.С. Зеленко, В.В. Варфоломеева, А.В. Терентьев……52

Разработка пакетов прикладных программ и использующих их приложений на основе онтологического подхода
С.В. Смирнов……56

Графо-аналитический метод моделирования сигма-дельта преобразователей информации замкнутой структуры
Б.В. Чувыкин, А.В. Селезнев, Р.Г. Тер-Аракелян……60

Организация информационной поддержки процесса принятия решений в условиях неопределенности и риска
А.А. Шкромадо……..64

Подсекция 2. «Автоматизированные информационные системы»

Автоматизированная информационная система обслуживания посетителей и статистического анализа библиотечных процессов
В.В. Васильев……..69

Автоматизированная система мониторинга состояния здоровья и уровня физического развития студентов ВУЗов
В.В. Горожанов, В.М. Богданов, А.В.Соловов……71

Программное обеспечение процесса сертификации систем менеджмента
А.В. Графкин, В.В. Графкин……..72

Автоматизированная система оценки эффективности венозного оттока
Д.А. Загуменнов, Л.С. Зеленко, П.Ф. Кравцов ……74

Автоматизированная система управления ремонтами оборудования
И.В. Занин……..79

Применение информационных технологий при выборе схемы расположения опор в конструктивно-силовой системе ГТД
В.А. Зрелов, А.С. Миронов, М.Е. Проданов……82

Создание многофункциональных средств измерения физических величин для промышленных систем управления
С.В. Ионов, В.В. Козлов, М.Г. Мясникова, А.А. Трофимов……85

Компьютерное моделирование цифроуправляемых генераторов опорных сигналов
А.М. Косолапов, В.П. Рузанов……88

Компьютерное моделирование датчиков тока мобильных объектов
А.М. Косолапов, Д.Н. Франтасов……92

Анализ погрешности измерения геометрических размеров лопаток газотурбинных двигателей
И.А. Лёзин……95

Автоматизированная система обработки акустических сигналов при диагностике сварных соединений
А.Ю. Модин……..100

Функциональная модель береговой системы наблюдения
А.И. Моисеев……105

Анализ алгоритмов восстановления частотной синхронизации с использованием преамбулы
И.М. Нестеров……..108

Автоматизированная система обработки результатов исследований по влиянию предпосевного воздействия на жизнеспособность растений
М.В. Пантюхина, Л.С. Зеленко……111

Разработка систем измерений и анализа сигналов на базе свободного программного обеспечения
А.И. Попов, А.В. Родионов……..115

Адаптивный квадратно-корневой ковариационный алгоритм фильтрации для навигационных комплексов
И.В. Семушин, Ю.В. Цыганова……118

Исследование и разработка алгоритмов распознавания номерных знаков для системы контроля доступом
А.В. Симкин……..123

Подсистема расчета молекулярной площадки адсорбата
М.С. Стребков, Л.С. Зеленко, В.В. Варфоломеева, А.В. Терентьев……127

Применение преобразование Гильберта-Хуанга в обработке медицинских изображений
А.Ю. Тычков……132

Роль систем Service Desk в IT–менеджменте предприятия
Т.О. Шевина, Л.С. Зеленко……138

Методы оперативного рационального перестрахования особо серьёзного риска на базе эволюционных алгоритмов как новое научное направление. Риски авиационные и космические
М.В. Шептунов……142

Формат *.dxf как основа технологического проектирования
В.Н. Шерстобитова……147

Методы и средства обеспечения безопасности полетов по упреждающей технологии
Р.К. Мирзаев……151

Автоматизированная система обработки полетной информации: проблемы и перспективы
Р.К. Мирзаев……156

Подсекция 3. «Информационные технологии в образовании»

Интернет система мониторинга учебного процесса по информационным специальностям
И.Ф. Батаев, В.В. Козлов……161

Автоматизированная информационная система управления учебным процессом на факультетах СГАУ
М.С. Глянько, Д.С. Оплачко, Е.И. Чигарина……163

Опыт обучения информационным технологиям в условиях зарубежной сетевой академии
А.В. Горшков, У.М. Тулемисов……166

Автоматизированная система выявления индивидуального профиля студентов «деятельность-мотивация»
Е.А. Гресько……170

Разработка подсистема инвентаризации автоматизированного рабочего места инженера кафедры
Д.Н. Докшин, Л.С. Зеленко……173

Автоматизированная система проведения онлайн-тестирования по иностранному языку «С.О.Т.А»
М.В. Ефремов, Л.С. Зеленко ……175

Концептуальные основы построения дистанционных обучающих систем
Л.С. Зеленко, Т.И. Михеева……179

Интернет-портал «Авторский Комикс»
А.А. Козлов, Л.С. Зеленко……184

Модель кластеризации студенческой группы на основе непрерывного мониторинга успеваемости
В.В. Козлов……186

Интерактивное обучение информатике и основам программирования с использованием когнитивной компьютерной графики
Р.В. Малышев, Л.С. Зеленко……189

Программный комплекс формирования региональных образовательных целевых программ
М.А. Мартынов……191

Разработка виртуального тренажёра для дистанционного обучения основам информатики и программирования
А.В. Невинский, М.А. Шершнев, Л.С. Зеленко……192

Рейтинговая оценка текущей успеваемости студентов СГАУ
В.В. Никишкин, Д.С. Оплачко, Е.И. Чигарина……196

Информационная технология формирования исследовательских компетенций студентов НИТУ
С.А. Пиявский……199

Управление учебной и внеучебной деятельностью студента в ВУЗе на основе информационно-коммуникационных технологий
С.А. Пиявский, З.Ф. Камальдинова……204

Об автоматизированной информационной системе управления компетентностным развитием студенческого коллектива
А.А. Тимофеев, Д.С. Шишков……209

Дистанционная интеллектуальная система, реализующая интерактивное обучение в виртуальном пространстве
А.В. Топунов, Д.А. Загуменнов, Л.С. Зеленко……210

Реализация дистанционного образования на базе системы дистанционного обучения кафедры программных систем СГАУ
Р.Д. Фахрутдинов, Л.С. Зеленко……212

Подсистема электронного документооборота в составе автоматизированного рабочего места инженера кафедры
О.В. Федюнев, Л.С. Зеленко……215

Роль системы дистанционного обучения в проекте «Школа информатики СГАУ»
А.А. Флигин, Л.С. Зеленко……218

Интерактивный тренажёр «Конструктор алгоритмов»
И.В. Юдин, Л.С. Зеленко ……221

Подсекция 4. «Автоматизированные системы реального времени»

Многоуровневая модель разработки ПО АСРВ средствами операционной системы реального времени
А.В. Баландин……222

К проблеме построения оптимизирующих преобразований спецификации управляющих алгоритмов реального времени
А.Ю. Богатов……226

Измерительная система для прочностных испытаний элементов авиационных конструкций
Л.Э. Вилоп……229

Формализованное описание производственного процесса предприятия
А.П. Зеленин……233

Применение метода экспресс-анализа для архивирования трендов АСУТП

А.Ю. Зенов, М.П. Строганов……..*236*

Разработка автоматизированной системы прогнозирования термодеформационного состояния технологического оборудования
К.В. Марусич……..*239*

Модель оценивания достоверности значений параметров АСУТП
А.А. Сидоров, В.Е. Захарченко……*242*

Логический вывод в системе верификации требований к управляющим алгоритмам реального времени
А.А. Тюгашев, А.В. Шулындин……*247*

Подсекция 5. «Сетевые технологии»

Перспективы внедрения протокола IPv6 и исследование параметров сетей нового поколения
С.С. Болдырев, П.В. Веселовский……*248*

Методы построения оптимальной структуры HRBF нейронной сети для решения задачи прогнозирования
В.Г. Литвинов, О.П. Солдатова……*252*

Информационная система структуризации знаний из Интернета
А.С. Лубышев, А.Х. Галеев……*255*

Нейроимитатор нейронных сетей на основе нечеткого многослойного персептрона
В.В. Новиков, Е.А. Расщепкин, О.П. Солдатова……*257*

Математическое моделирование динамического последовательного заполнения сетей с кластерной топологией
И.И. Курочкин, Я.Р. Гринберг……*262*

Повышение безопасности передачи данных в сетях с временной синхронизацией через NTP протокол
М.А. Макаров, Е.А. Симановский……*266*

Модель сверточной нейронной сети для извлечения паттернов признаков на основе рецептивных полей
Е. Пудикова, О.П. Солдатова……*268*

Пути повышения качества передачи видео в беспроводных сетях
Е.С. Сагатов……*273*

Объектно-сегментированный анализ телекоммуникационных данных на основе апостериорных вероятностей
А.А. Салмин……*278*

Самоорганизующиеся нейронные сети в обработке пространственных данных
О.Н. Сапрыкин……*281*

Исследование качества видеопотоков в беспроводных сетях 3G
А.А. Семенов……*286*

Исследование влияния параметров алгоритмов обучения нейронной сети на точность решения задачи семантического анализа текста
О.П. Солдатова, Э.В. Кольбова……*291*

Экспериментальная проверка модели измерения доступной пропускной способности
Т.Г. Султанов……*296*

Исследование производительности компьютерных сетей на основе анализа их трафика
В.Н. Тарасов, Н.Ф. Бахарева, А.Л. Коннов……*301*

Подсекция 6. «Информационные технологии в управлении движением»

Фреймовая модель представления знаний в системе планирования и оптимизации маршрутов грузоперевозок
А.И. Белоусов……306

Методы зональной оптимизации координированного управления
С.В. Михеев……310

Синтез интегрированных систем мониторинга и управления транспортной инфраструктурой
Т.И. Михеева, О.М. Батищева……315

ГИС в составе информационной системы паспортизации объектов транспортной инфраструктуры
Т.И. Михеева, Н.В. Гаманова ……320

Система автоматизированного управления интерактивным контентом
Т.И. Михеева, О.К Головнин……325

Методы оценки уровня безопасности транспортной инфраструктуры
Т.И. Михеева, А.В. Золотовицкий……326

Алгоритмы координации управления транспортными потоками на связанных перекрестках
Т.И. Михеева, С.В. Михеев……331

Информационная система дислокации светофорных объектов с использованием цифровых моделей местности
Т.И. Михеева, О.А. Япрынцева……336

Бортовая система учета пассажиров общественного транспорта
Б.В. Скворцов, Е.С. Перцович……340

Подсекция 7. «CALS/ИПИ-технологии»

Использование 3D-моделей переменной плотности в задачах оптимального проектирования авиационных конструкций
А.В. Болдырев……344

Программное обеспечение онтологии концептуального проектирования самолета
Н.М. Боргест, А.В. Глушков, А.В. Петровский, Д.В. Шустова……348

Тезаурус как основа онтологии концептуального проектирования самолета
Н.М. Боргест, Д.В. Шустова……353

Математическое моделирование весовой и аэродинамической эффективности телескопических крыльев
А.А. Вырыпаев……357

Информационная модель технологических данных предприятия ракетно-космической техники
А.А. Греднева……358

Методы и средства интеграции информационных систем и оценки эффективности инновационных проектов предприятий наукоемкого машиностроения
В.Ф. Денисов……362

Проблемы обеспечения согласованного взаимодействия в интегрированной информационной среде предприятия
А.В. Иващенко……369

Мультиагентные технологии управления распределением производственных ресурсов
М.В. Андреев, И.О. Бабанин, А.В. Иващенко, П.О. Скобелев……373

Разработка РКТ с использованием прототипов в системе нисходящего проектирования
Л.А Комарова……377

Особенности электронного макетирования космических аппаратов дистанционного зондирования
Е.В. Космодемьянский……378

Методы и средства программной инженерии при проектировании интегрированной информационной системы предприятия аэрокосмической отрасли
М.Е. Кременецкая, П.В. Сафин……379

Формирование облика самолёта на основе высокоточного математического моделирования
А.С. Кузнецов……383

Применение генетических алгоритмов для проектирования подкрепленных панелей из композиционных материалов
А.В. Черняев……384

Подсекция 8. «Информационные технологии в задачах проектирования ракетоносителей и космических аппаратов»

Усовершенствование системы управления угловой скоростью космического аппарата
Д.А. Азарных……385

Моделирование динамики сложной системы «экипаж – космический аппарат с постоянной закруткой»
В.А. Акулов……387

Расчет продолжительности оптимального по времени разворота космического аппарата дистанционного зондирования Земли с использованием инерционных исполнительных органов-маховиков, при смене режимов ориентации
Р.С. Алиев……390

Повышение эффективности авиационных работ по распределению веществ в сельском хозяйстве на основе оптимизации траектории движения воздушного судна
Р.Б. Алтынбаев……393

Управление колебаниями спутника на стационарной орбите
Л.К. Бабаджанянц, Ю.Ю. Пупышева……399

Программный комплекс расчета параметров противометеорной защиты
А.В. Бабенко, Н.Т. Каргин……404

Обработка измерений спутниковых навигационных систем
С.Ю. Гиндин……405

Последовательно – оптимальные программы навигационных определений КАДЗЗ
А.П. Долгинцев, И.В. Елисеев……407

Разработка программного комплекса для моделирования движения многофункционального малого космического аппарата
А.А. Масловская, И.С. Ткаченко, Л.С. Зеленко……412

Мультиагентная подсистема планирования грузопотока Международной космической станции
А.В. Иващенко, А.В. Леонтьев, А.Р. Диязитдинова, П.О. Скобелев, М.В. Сычева, И.И. Хамиц……417

Мультиагентная подсистема построения программы полета Международной космической станции
А.В. Иващенко, К.Ю. Матвеев, А.Л. Новиков, Е.В. Симонова, П.О. Скобелев, М.В. Сычева, И.И. Хамиц……422

Модели и алгоритмы для оценки среднесуточного времени радиовидимости спутника-ретранслятора с космического аппарата наблюдения с учетом затенения антенн
В.И. Куренков, Н.Р. Стратилатов, О.Г. Федоренко, Л.Б. Шилов……426

Программное обеспечение проектно-баллистической оптимизации межпланетных КА с ЭРДУ, использующих гравитационные маневры
Д.В. Курочкин, О.Л. Старинова……431

Разработка проблемно-ориентированного программного обеспечения для задач проектирования космических аппаратов наблюдения
А.С. Кучеров, В.И. Куренков, В.Б. Белкин, А.А. Якищик……433

Система моделирования вращательного движения микроспутника с магнитной системой ориентации в геомагнитном поле
А.С. Лебедев……437

Об одной системе автоматического управления ориентацией солнечных батарей микроспутника на источник излучения
В.В. Любимов……441

Определение оптимальных мест установки демпферов для снижения вибраций оборудования космических аппаратов на основе конечно элементного моделирования
К.В. Пересыпкин……444

Исследование деформации боковых блоков ракет-носителей при падении в воду с использованием компьютерных конечно-элементных технологий
М.А. Петровичев, В.К. Дуплихин, Н.Т. Каргин, А.М. Петровичев……449

Программный комплекс для решения задач оптимизации перелетов между эллиптическими орбитами с использованием двигателей большой и малой тяги
К.В. Петрухина……452

Автоматизация расчета проектных характеристик энергодвигательного электроракетного модуля низкоорбитального космического аппарата
В.В. Салмин, В.В. Волоцуев……454

Программный комплекс для выбора оптимальных проектно-баллистических параметров многоразового межорбитального транспортного аппарата с электрореактивной двигательной установкой
В.В. Салмин, А.С. Четвериков……456

Мультиагентная система для исследования методов взаимодействия космических аппаратов дистанционного зондирования Земли
П.О. Скобелев, А.В. Соллогуб, А.В. Иващенко, Е.В. Симонова, М.Е. Степанов, А.В. Царев……460

Компьютерная технология имитации и анимации углового движения информационного спутника с упругими панелями солнечных батарей
С.Е. Сомов……465

Компьютерные технологии для анализа устойчивости движения и колебаний упругого спутника при широтно-импульсном управлении ориентацией с физическим запаздыванием
С.Е. Сомов……471

Программный комплекс моделирования и оптимизации движения космического аппарата с малой тягой в системе Земля-Луна

О.Л. Старинова, И.Л. Матерова……476

Использование информационных технологий твёрдотельного геометрического проектирования в позиционных задачах проектирования летательных аппаратов
А.И. Шулепов……477

Секция 2. ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ ВЫЧИСЛЕНИЯ И GRID-ТЕХНОЛОГИИ

Грид-среда Самарского государственного аэрокосмического университета имени академика С.П. Королева (национального исследовательского университета). Состояние и перспективы
А.В. Баскаков, Е.А. Симановский……481

Автоматизация параллельного программирования с использованием системы Graphplus templet
С.В. Востокин……486

Графический редактор системы визуального моделирования параллельных вычислений (PGRAPH)
П.В. Аболмасов, А.Н. Коварцев……490

Идентификация с последовательной схемой отбора согласованных оценок
С.А. Бибиков, А.В. Гаврилов, В.А. Фурсов……494

Некоторые аспекты разработки гибридных параллельных эвристических алгоритмов оптимизации
О.И. Булычов……499

Исследование аэродинамических характеристик снаряда при активном старте
Ю.В. Ганзий, А.Ш. Нагуманов, Н.В. Митюков……502

Автоматический синтез параллельных программ на основе их графических моделей в системе PGRAPH
В.В. Жидченко, А.Н. Коварцев……505

Численное решение уравнений газодинамики неравновесных сред на GPU с использованием технологии CUDA
В.Г. Макарян……510

Алгоритм структурного преобразования графа управления параллельными вычислениями для систем с распределенной памятью MPI
Д.А. Попова-Коварцева, А.Н. Коварцев……515

Параллельная реализация алгоритма Камеды-Вейнера
А.В. Цыганов……521

Интеграция контроля доступа для распределенной системы управления физическим экспериментом
И.С. Ястребов……524

Построение быстрых алгоритмов преобразований в разностно упорядоченных системах дискретных функций Уолша
М.Ю. Васильева……529

Применение схем вычислительных процессов в системе GraphPlus templet на примере схемы Pipeline
С.В. Востокин, А.Р. Хайрутдинов……533

Обработка изображений в распределенной массивно многопоточной CUDA-среде
А.В. Никоноров, П.Ю. Якимов……538

Программный комплекс имитационного моделирования алгоритмов балансировки
И.Д. Зекцер……543

Континуальный и дискретный подходы при имитационном моделировании соударений высокоскоростных частиц с герметичными конструкциями космического аппарата
А.А. Потапкин, Л.С. Зеленко……549

Реализация FDTD-метода по технологии CUDA
С.А. Малышева……554

Секция 3. ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И МАТЕМАТИЧЕСКАЯ ФИЗИКА

Расчет параметров ракетных двигателей с учетом вероятностных отклонений начальных данных
Д.М. Борисов, А.М. Камай……560

Численное моделирование противотока в разделительной газовой центрифуге
И.В. Ериклинцев, С.А. Козлов, О.В. Трошкин……562

Ударные волны и автоволновые структуры в газовой среде с релаксационным процессом и источником энергии
Р.Н. Галимов, В.Г. Макарян, Н.Е. Молевич……565

Численное моделирование структуры вихревого потока в присутствии разряда
И.П. Завершинский, В.Г. Макарян, Н.Е. Молевич, Д.П. Порфирьев……569

Расчет стеклопластиковой панели с учетом ползучести при одностороннем высокотемпературном нагреве
А.А. Абдрахманова, В.П. Павлов……574

Численное моделирование пространственно-временной динамики широкоапертурного лазера с отстройкой частоты генерации
Д.А. Анчиков, А.А. Кренц, Н.Е. Молевич……579

Об адекватности некоторых моделей самоподобного трафика
А.В. Благов, А.Ю. Привалов……584

Приближенное решение гиперсингулярных интегральных уравнений
И.В.Бойков, А.И. Бойкова……588

Применение современных моделей турбулентности к расчету аэродинамических характеристик профиля при умеренных числах Рейнольдса
И.Н. Варюхина……593

Конечно-элементный подход к расчету ресурсной прочности огневой стенки камеры двигателя 14Д23
С.А. Владимиров, С.И. Трефилов……596

Определение условий неустойчивости типа жесткой моды и анализ уравнения для параметров порядка в модели морфогенеза Гирера Майнхардта во внешнем случайном поле
К.А. Головкин……597

Исследование погрешности численного моделирования процесса прохождения электромагнитного излучения в волноводе с распределенным коэффициентом преломления
А.А. Дегтярев, Е.С. Козлова……601

Разработка и исследование алгоритма поиска множества решений нелинейного дифференциального уравнения эллиптического типа
А.А. Дегтярев, Г.М. Комендантян……606

Дисперсия стохастического процесса распределения ресурса
А.В. Докучаев……610

Вычисление параметра регуляризации методом перекрёстной значимости на основе разложения Холесского
В.В. Долишний, А.И. Жданов……615

Исследование механизмов ценообразования опционов в моделях обобщенного случайного блуждания в непрерывном времени с негауссовскими распределениями скачков логарифмов цен
А.Н. Евдокимов, А.П. Зубарев……619

Исследование фотодиссоциирующих областей в межзведной среде
Д.И. Завершинский, Н.Е. Молевич……623

Метод L-кривой для итерационных алгоритмов решения плохообусловленных задач
А.А. Иванов……628

Теоретический анализ динамики систем реакция-диффузия вблизи точки бифуркации в поле внешних флуктуаций
С.Е. Курушина……633

Исследование процессов формо- и фазообразования в неравновесных малых металлических частицах
Н.И. Лиманова, Е.А. Талалова, М.М. Ворона, Ю.Р. Хайруллова, П.А. Шишкин……637

Моделирование напряженно-деформированного состояния многослойных гетерогенных структур
А.Н. Литвинов, М.А. Литвинов……642

Программный комплекс для исследования пространственно временной динамики систем реакция-диффузия в поле мультипликативных флуктуаций параметров
В.В. Максимов……647

Статистическая модель и программа оценки вероятностных характеристик вращательного движения капсулы на тросе
Ю.М. Заболотнов, О.Н. Наумов……651

Моделирование логопериодических колебаний с помощью иерархических структур
А.С. Пивоварова……656

Использование HASH функций при моделировании стохастических процессов
Д.А. Рыбаков……661

Применение метода вычислительного эксперимента для исследования влияния параметров правой части нелинейного уравнения теплопроводности на его решение
А.А. Дегтярев, М.В. Силакова……664

Имитационное моделирование фондовых рынков
А.А. Стеряков……668

Исследование устойчивости закрученных течений со скачками скорости и плотности потока
С.С. Сугак, И.П. Завершинский……673

Термические напряжения в ледяном массиве при замерзании воды в цилиндрической трубе
А.Ф. Федечев, Н.Л. Федечева……676

CFD-моделирование охлаждения лопатки с вихревой матрицей
Л.С. Шаблий……680

Исследование динамики фондового рынка с применением радиально базисных нейронных сетей

П.В. Левченко……684

Модель движения материальной точки в поле ядерных сил для построения алгоритма глобальной многоэкстремальной оптимизации
А.Н. Коварцев, Д.С. Оплачко……689

Метод повышения устойчивости среднеквадратичного оценивания коэффициентов разностного уравнения колебаний систем с турбулентным трением
А.А. Егорова, В.Е. Зотеев……691

Принцип сведения в задаче об устойчивости при постоянно действующих возмущениях
Д.М. Щепакин……696

Исследование устойчивости динамической системы с задержкой в модели рынка одного товара с учетом спекулятивной активности
И.П. Завершинский, М.Р. Майлян……700

Распространение звука в неоднородной тепловыделяющей среде
Н.Е. Молевич, Н.В. Трошкин……705

Критические явления в модели кубического автокаталатора
Н.М. Шемонаева, Е.А. Щепакина……709

Об одном методе выделения информативных фрагментов сигнала
П.С. Сидоров……713

Бесконтактное измерение концентрации газа по анализу спектров излучения синглетного кислорода
А.С. Инсапов……716

Бесконтактное измерение температуры газа по анализу спектров излучения кислорода в диапазоне 600-800 нм
М.С. Малышев……719

Фильтрация шумов, вызванных нелинейными колебаниями модели в аэродинамическом эксперименте при помощи искусственных нейронных сетей
Д.В. Ливерко……722

Секция 4. ГЕОИНФОРМАТИКА И ОБРАБОТКА ИЗОБРАЖЕНИЙ

Спектрально-энергетический метод идентификации линейной модели наблюдения при дистанционном зондировании Земли
В.В. Сергеев, А.Ю. Денисова……726

Разработка геоинформационной системы агропромышленного комплекса Самарской области для учета и контроля земель сельхозназначения
Н.С. Воробьева, Е.И. Тимбай……731

Детектирование бликов в системах видеонаблюдения с выделением последовательности опорных кадров
Е.Ю. Минаев, А.В. Никоноров, В.А. Фурсов……733

Алгоритм определения периодов повышенной сейсмической опасности по «шаблонам малой динамики»
А.Н. Коварцев, О.А. Никонорова……738

Метод идентификации минных полей по аэрокосмическим снимкам
Н.С. Булдаков……743

Автоматическая географическая привязка спутниковых изображений с пиксельной точностью в задаче минимизации объема хранимой информации
Р.К. Газизов, И.Р. Фатхулисламов……748

Анализ пространственного взаимодействия оптических шкал с использованием преобразования Радона
В.М. Гречишников, А.А. Юдин, О.Ю. Борисов……753

Модифицированный усредняющий фильтр. Определение мощности шума на видеоизображении
Р.Р. Диязитдинов, В.А. Буцких……758

Байесовский подход для оценивания заднего плана и обнаружения посторонних объектов на видеоизображении
Р.Р. Диязитдинов, В.А. Буцких……764

Интеграция модулей системы зашиты распределённых сетей
А.А. Иванов, М.В. Кузнецов……769

Технология автоматического наполнения банка данных космических снимков в режиме постоянного приема
В.Н. Копенков, А.В. Чернов, Н.В. Чупшев……773

Анализ изображений на наличие локальных искусственных изменений со свойствами JPEG сжатия
А.В. Кузнецов……778

Организация распределённых вычислений
И.И. Меркульев, М.В. Кузнецов……783

Разработка метода снижения размерности пространства признаков на основе опорных узлов
Е.В. Мясников……785

Модификация алгоритма «Split – Merge – Pixelwise refinement» для сегментации спутниковых изображений
А.Ю. Петрова, В.А. Федосеев……790

Построение текстурных признаков на основе регрессионных моделей
А.И. Пластинин……795

К вопросу обеспечения безопасности распределённых сетей и технической поддержки
М.С. Воронов……800

Определение динамических характеристик капли жидкости, нанесенной на поверхность подложки, в задаче оценки степени ее чистоты
М.С. Глянько, П.Ю. Изотов……802

Разработка алгоритмов локализации инородного тела на рентгенографических изображениях с использованием искусственных нейронных сетей
Е.С. Залеткин, А.В. Куприянов, Н.Ю. Ильясова……806

Исследование свойств внутриглазного инородного тела на основе анализа рентгенографических изображений черепа
Н.Ю. Ильясова, А.В. Куприянов, С.Г. Засканов……810

Система комплексной информационной защиты распределённых сетей
М.В. Кузнецов, М.С. Воронов, А.А. Иванов, А.А. Солонщиков, Н.В. Студеникин, С.А. Бобров, М.В. Яшин, И.И. Меркульев……815

Требования к антивирусной защите в современных корпоративных сетях
А.А. Солонщиков, М.В.Кузнецов……818

Анализ современных технологий защиты от спама
Н.В. Студеникин, М.В. Кузнецов……821

Применение локального корректирующего фильтра для устранения «блочности» на изображениях, сжатых методом JPEG

Е.И. Тимбай……824

Алгоритм обработки изображений при фотограмметрических измерениях
П.П. Чураков……829

Анализ способов применения централизированного управления в системах комплексной информационной защиты распределённых сетей
М.В. Яшин……835

Системы защиты от утечек данных (DLP-системы)
С.А. Бобров, И.С. Кузнецова……837

Секция 5. КОМПЬЮТЕРНАЯ ОПТИКА И НАНОФОТОНИКА

Решение задач компьютерной оптики на графических вычислительных устройствах
Д.Л. Головашкин, Н.Л. Казанский……840

Расчет пластмассово-линзовых микрообъективов для массовых видеокамер наблюдения
Г.И. Грейсух, Е.Г. Ежов, И.А. Левин, С.А. Степанов……845

Коррекция хроматизма оптических систем, содержащих дифракционные элементы
Г.И. Грейсух, Е.Г. Ежов, С.В. Казин, С.А. Степанов……850

Возможности суперахроматизации пластмассово линзовых микрообъективов
Г.И. Грейсух, Е.Г. Ежов, И.А. Левин, С.А. Степанов……855

Методика формирования падающей волны при разностном решении уравнения Даламбера
М.А. Астраханцева……860

Резонансное усиление затухающих порядков дифракции в дифракционных спектральных отражающих фильтрах для наноразмерной фотолитографии
Е.А. Безус, Л.Л. Досколович, Н.Л. Казанский……865

Расчет вторичной оптики светодиодного прожектора для формирования заданного распределения освещенности
А. А. Белоусов, М.А. Моисеев……868

Изучение возможности применения хрома в качестве маскирующего слоя при глубоком травлении кремния
Н.А. Болобанов, А.В. Волков, Г.Ф. Костюк, С.Д. Полетаев……872

Магнитооптические свойства периодических дифракционных структур, намагниченных в плоскости. Эффект резонансного изменения фазы
Д.А. Быков, Л.Л. Досколович……874

Моделирование распространения лазерных пучков с фазовой сингулярностью в оптических волноводах
С.А. Дегтярев, С.Н. Хонина……877

Расчет преломляющих поверхностей для формирования диаграммы направленности в виде линии
А.Ю. Дмитриев, Л.Л. Досколович……882

Синтез и исследование трехмерных металлодиэлектрических фотонных кристаллов
П.Н. Дьяченко, С.В. Карпеева, Е.В. Фесик, Ю.В. Микляев, В.С. Павельева, Г.Д. Мальчиков….886

Астигматическое преобразование одномерных распределений, согласованных с функциями Эйри
О.В. Зотеева, С.Н. Хонина……890

Анализ методов экспресс-контроля чистоты поверхности
Н.А. Ивлиев, В.А. Колпаков, С.В. Кричевский……895

Исследование формирования осевого отрезка с помощью оптимизированного бинарного ДОЭ
Д.Г. Качалов, В.С. Павельев, С.Н. Хонина, Р.В. Скиданов, О.Ю. Моисеев……899

Наноструктуры на базе пористого кремния для фотоэлектрических преобразователей
В.А. Китаева……902

Субволновая фокусировка с помощью планарной гиперболической секансной линзы
А.А. Ковалев, Я.Р. Триандафилов……907

Формирование светового поля для линейного перемещения микрообъекта
А.А. Морозов……912

Применение подвижных фокусаторов излучения для интенсификации диффузионных процессов лазерным воздействием
С.П. Мурзин, Е.В. Шокова, Н.В. Трегуб……917

Сравнение методов расчета распространения рентгеновского излучения через ДОЭ в параксиальной области
А.Г. Налимов, С.Н. Хонина……920

Исследование влияния аберраций на высокоапертурные фокусирующие системы
Е.А. Пелевина, С.Н. Хонина……925

Травление кремния во внеэлектродной плазме высоковольтного газового разряда
В.В. Подлипнов, В.А. Колпаков……930

ДОЭ для автоматической оптической сборки микросистем
А.П. Порфирьев……934

Численное моделирование движения объектов в задачах оптического захвата
М.А. Рыков……939

Улучшение разрешения на основе использования пространственных фильтров
Д.А.Савельев, С.Н. Хонина……944

Моделирование R-FDTD методом субволновой фокусировки света с радиальной поляризацией зонной пластинкой
С.С. Стафеев……949

Моделирование волноводных режимов в многослойных структурах
Т.С. Стрилец, В.В. Котляр……954

Исследование методов контроля твердости слоев при формировании наноструктур
А.Д. Щербатенко……960

ПЕРСПЕКТИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ АВИАЦИИ И КОСМОСА
(ПИТ-2010)

ТРУДЫ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ С ЭЛЕМЕНТАМИ НАУЧНОЙ ШКОЛЫ ДЛЯ МОЛОДЕЖИ

Печатается в авторской редакции

Ответственный за выпуск Э.И. Коломиец
Компьютерная верстка Н.Л. Додонова, И.М. Куликовских

Подписано в печать 22.09.2010 г. Формат 60x84 1/ 8.
Электронное издание.
Печ. л. .
Тираж 500 экз. Заказ №

Самарский государственный аэрокосмический университет
443086 Самара, Московское шоссе, 34

Издательство Самарского государственного аэрокосмического университета
443086 Самара, Московское шоссе, 34