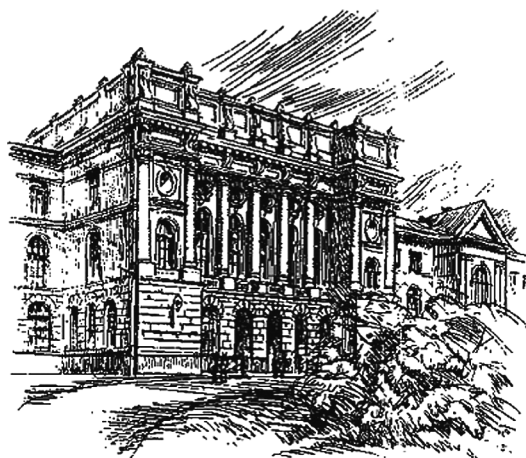




НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ

САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО
ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Информатика. Телекоммуникации.
Управление

6(162) 2012

НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

РЕДАКЦИОННЫЙ СОВЕТ ЖУРНАЛА

Алферов Ж.И., академик РАН; *Васильев Ю.С.*, академик РАН (председатель);
Костюк В.В., академик РАН; *Лопота В.А.*, чл.-кор. РАН;
Окрепилов В.В., академик РАН; *Рудской А.И.*, чл.-кор. РАН;
Патон Б.Е., академик НАН Украины и РАН; *Федоров М.П.*, академик РАН;
Фортов В.Е., академик РАН.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Васильев Ю.С., академик РАН (главный редактор); *Арсеньев Д.Г.*, д-р техн. наук, профессор;
Бабкин А.В., д-р экон. наук, профессор (зам. гл. редактора);
Боронин В.Н., д-р техн. наук, профессор; *Глухов В.В.*, д-р экон. наук, профессор;
Дегтярева Р.В., д-р ист. наук, профессор; *Иванов А.В.*, д-р техн. наук, профессор;
Иванов В.К., д-р физ.-мат. наук, профессор; *Козловский В.В.*, д-р физ.-мат. наук, профессор;
Рудской А.И., чл.-кор. РАН (зам. гл. редактора); *Юсупов Р.М.*, чл.-кор. РАН.

ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ

РЕДАКЦИОННЫЙ СОВЕТ

Юсупов Р.М., чл.-кор. РАН – председатель;
Абрамов С.М., чл.-кор. РАН;
Воеводин В.В., чл.-кор. РАН;
Заборовский В.С., д-р техн. наук, профессор;
Козлов В.Н., д-р техн. наук, профессор;
Фотиади А.Э., д-р физ.-мат. наук, профессор;
Черноруцкий И.Г., д-р техн. наук, профессор.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Юсупов Р.М., чл.-кор. РАН – председатель;
Арсеньев Д.Г., д-р техн. наук, профессор – зам. председателя;
Бабкин А.В., д-р экон. наук, профессор – зам. председателя;
Антонов В.И., д-р техн. наук, профессор;
Голландцев Ю.А., д-р техн. наук, профессор;
Карпов Ю.Г., д-р техн. наук, профессор;
Коротков А.С., д-р техн. наук, профессор;
Макаров С.Б., д-р техн. наук, профессор;
Устинов С.М., д-р техн. наук, профессор;
Цикин И.А., д-р техн. наук, профессор;
Шкодырев В.П., д-р техн. наук, профессор;
Клавдиев В.Е., канд. техн. наук, доцент.

Журнал с 1995 года издается под научно-методическим руководством Российской академии наук. С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» ISSN 1994-2354.

Журнал с 2002 года входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

Сведения о публикациях представлены в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich`s Periodical Directory».

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Подписной индекс **47517** в каталоге «Газеты. Журналы» Агентства «Роспечать».

Журнал включен в базу данных «Российский индекс научного цитирования» (РИНЦ), размещенную на платформе Национальной электронной библиотеки на сайте <http://www.elibrary.ru>

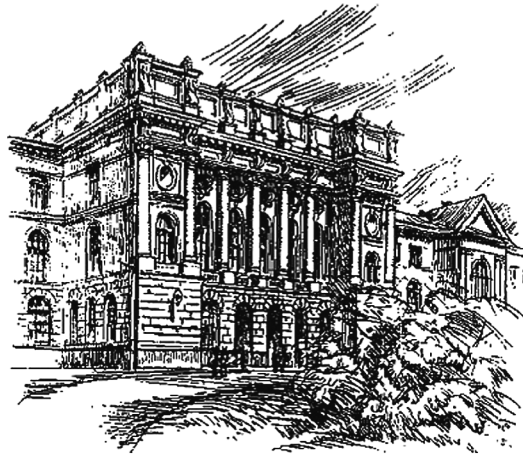
При перепечатке материалов ссылка на журнал обязательна.

Точка зрения редакции может не совпадать с мнением авторов статей.

Адрес редакции и издательства: Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.
Тел. редакции (812) 552-62-16.

© Санкт-Петербургский государственный политехнический университет, 2012

THE MINISTRY OF EDUCATION AND SCIENCE OF THE RUSSIAN FEDERATION



ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL

Computer Science.
Telecommunications and Control Systems

6(162) 2012

Polytechnical University Publishing House
Saint Petersburg
2012

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL

EDITORIAL COUNCIL

Y.S. Vasiliev – full member of the Russian Academy of Sciences, President of the St. Petersburg State Polytechnical University, editor-in-chief; *Zh.I. Alferov* – full member of the Russian Academy of Sciences; *V.V. Kostiuk* – full member of the Russian Academy of Sciences; *V.V. Lopota* – corresponding member of the Russian Academy of Sciences; *V.V. Okrepilov* – full member of the Russian Academy of Sciences; *B.E. Paton* – full member of the Russian Academy of Sciences and the National Academy of Sciences of Ukraine; *A.I. Rudskoy* – corresponding member of the Russian Academy of Sciences; *M.P. Fedorov* – full member of the Russian Academy of Sciences; *V.E. Fortov* – full member of the Russian Academy of Sciences.

EDITORIAL BOARD

Y.S. Vasiliev – full member of the Russian Academy of Sciences, President of the St. Petersburg State Polytechnical University, editor-in-chief; *D.G. Arseniev* – Dr.Sc.(tech.), prof.; *A. V. Babkin* – Dr.Sc. (econ.), prof., deputy editor-in-chief; *V.N. Boronin* – Dr.Sc.(tech.), prof.; *V.V. Glukhov* – Dr.Sc. (econ.), prof.; *R. V. Degtyareva* – Dr.Sc. (history), prof.; *A. V. Ivanov* – Dr.Sc.(tech.); *V.K. Ivanov* – Dr.Sc.(phys.-math.), prof.; *V.V. Kozlovsky* – Dr.Sc.(phys.-math.), prof.; *A.I. Rudskoy* – corresponding member of the Russian Academy of Sciences, deputy editor-in-chief; *R.M. Yusupov* – corresponding member of the Russian Academy of Sciences.

COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS

EDITORIAL COUNCIL

R.M. Yusupov – corresponding member of the Russian Academy of Sciences, head of the editorial council; *S.M. Abramov* – corresponding member of the Russian Academy of Sciences; *V.V. Voevodin* – corresponding member of the Russian Academy of Sciences; *V.S. Zaborovsky* – Dr.Sc.(tech.), prof.; *V.N. Kozlov* – Dr.Sc.(tech.), prof.; *A.E. Fotiadi* – Dr.Sc.(phys.-math.), prof.; *I.G. Chernorutsky* – Dr.Sc.(tech.), prof.

EDITORIAL BOARD

R.M. Yusupov – corresponding member of the Russian Academy of Sciences, head of the editorial board;
D.G. Arseniev – Dr.Sc.(tech.), prof., deputy head of the editorial board;
A. V. Babkin – Dr.Sc. (econ.), prof., deputy head of the editorial board;
V.I. Antonov – Dr.Sc.(tech.), prof.;
Y.A. Gollandtsev – Dr.Sc.(tech.), prof.;
Y.G. Karpov – Dr.Sc.(tech.), prof.;
A.S. Korotkov – Dr.Sc.(tech.), prof.;
S.B. Makarov – Dr.Sc.(tech.), prof.;
S.M. Ustinov – Dr.Sc.(tech.), prof.;
I.A. Tsikin – Dr.Sc.(tech.), prof.;
V.P. Shkodyrev – Dr.Sc.(tech.), prof.;
V.Ye. Klavdiev – Candidate of Technical Sciences, associate prof.

The journal is published under scientific and methodical guidance of the Russian Academy of Sciences since 1995. The journal is published since 2008 as part of the periodical edition «Nauchno-tekhnicheskie vedomosti SPbGPU» (ISSN 1994-2354).

The journal is included in the List of Leading Peer-Reviewed Scientific Journals and other editions to publish major findings of PhD theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

The publications are presented in the VINITI RAS Abstract Journal and Ulrich's Periodical Directory International Database.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ПИ № ФЦ77-51457 issued Oct. 19, 2012.

Subscription index **47517** in the «Journals and Magazines» catalogue, Rospechat agency.

The journal is on the Russian Science Citation Index (RSCI) database

© Scientific Electronic Library (<http://elibrary.ru/>).

No part of this publication may be reproduced without clear reference to the source.

The views of the authors can contradict the views of the Editorial Board.

The address: 195251 Polytekhnicheskaya Str. 29, St. Petersburg, Russia.

Содержание

Инфокоммуникационные технологии

Болдырев Ю.Я. Роль суперкомпьютерных технологий в инженерном образовании.....	9
Сметанин А.Г., Тормасов А.Г. Математическая модель эффективных алгоритмов поиска и размещения данных в децентрализованной распределенной файловой системе.....	16

Телекоммуникационные системы и компьютерные сети

Заборовский В.С., Кондратьев А.С., Мулюха В.А., Силиненко А.В., Ильяшенко А.С., Филиппов М.С. Удаленное управление робототехническими объектами в космических экспериментах серии «Контур».....	23
Гаазе М.В., Лычагин Н.И., Малаховский А.А. Интегрированные телекоммуникационные службы для ведомственных сетей связи	32

Проблемы передачи и обработки информации

Алексеев В.Г. Система цифровых «водяных» знаков для аудиосигналов на основе использования явления реверберации.....	43
Воронин В.В. Автоматизированное детектирование дефектов при реконструкции архивных фотодокументов.....	50
Мухин В.Е., Корнага Я.И. Многоуровневая модель с дифференциальным уровнем доверия к субъектам для повышения защищенности распределенных баз данных.....	54
Заборовский В.С., Кондратьев А.С., Мулюха В.А., Силиненко А.В., Ильяшенко А.С. Управление группировками напланетных роботов с борта пилотируемой орбитальной станции в рамках проекта «METERON»	59
Кислицын А.Б., Рашич А.В. Передача и прием сигналов с DOQPSK и CDOQPSK с использованием алгоритма Витерби.....	66

Радиотехника, антенны, СВЧ-устройства

Голубев А.Г., Смирнов А.С. О расчете верхней границы чувствительности интерференционных волоконно-оптических гидрофонов приемной гидроакустической антенны.....	73
--	----

Системный анализ и управление

Чупров С.Г., Иванов А.А. Моделирование динамики двухдвигательного сервопривода с эффектом самоторможения	81
---	----

Вычислительные машины и программное обеспечение

Петросян Г.С. Анализ объектно-ориентированной парадигмы и поиск лучшей альтернативы	87
Пышкин Е.В. Проблемы автоматизации приемочного тестирования программного обеспечения при использовании подхода «разработка, управляемая описанием поведения»	90

Лабошин Л.Ю., Лукашин А.А., Семеновский В.Б. Поддержка вычислений на графических ускорителях в виртуальной среде XEN для обработки потоков данных робототехнических систем.....	97
--	----

Математическое моделирование: методы, алгоритмы, технологии

Валынец Е.В., Орженовский М.В., Иванков А.А. Стохастическая модель структуры пула заявок поставщиков на закрытом аукционе спотовых цен	101
Иванов А.А. Метод планирования формы гиперизбыточного манипулятора с ограниченным диапазоном изменения шарнирных координат	112
Кадырова Н.О., Павлова Л.В. Эффективная методика обработки многомерных данных большого объема	118
Петрова И.Ю., Хоменко Т.В. Математическая модель выбора и метод оценки вариантов физического принципа действия чувствительного элемента	124
Черноруцкий И.Г. Некоторые стандартные схемы параметрической оптимизации.....	128
Марценюк М.А., Поляков В.Б., Селетков И.П. Матричная реализация алгоритмов нечеткого вывода	133
Черноруцкий И.Г. Алгоритмические проблемы жесткой оптимизации.....	141
Шебалов А.А., Баженов А.Н. Исследование производительности методов вычисления оптического потока	152

Приборы, информационно-измерительные системы

Аникеева М.С., Софронов А.Н., Дремов С.С., Тер-Мартirosян А.Л. Цифровая система визуализации для фотодинамической терапии	159
--	-----

Управление в социальных и экономических системах

Латыпова Э.А., Бондарева И.О. Имитационное моделирование как средство управления логистическим предприятием на примере грузового порта	165
Аксенов К.А., Аксенова О.П., Ван Кай Планирование портфеля проектов в строительстве на основе мультиагентного имитационного моделирования.....	171
Рыбак В.А., Сулейман Х.М. Анализ и оптимизация эколого-экономических параметров инновационных проектов.....	175
Маслов В.И., Минаков В.Ф. Критерий эффективности в системе управления качеством продукции предприятия	179
Калмыкова С.В., Соколицин А.С. Оптимизация содержания и структуры учебного процесса в вузе с использованием метамоделей	185
Сведения об авторах	193
Аннотации.....	197

Contents

Infocommunication technologies

Boldyrev Yu.Ya. <i>The role of the supercomputer technology in the engineering education</i>	9
Smetanin A.G., Tormasov A.G. <i>Mathematical model and efficient algorithm to find and distribute data in the decentralized distributed file system</i>	16

Telecommunication systems and computer networks

Zaborovsky V.S., Kondratiev A.S., Muliukha V.A., Silinenko A.V., Ilyashenko A.S., Filippov M.S. <i>Remote robotic objects control within space experiments «Kontur» series</i>	23
Gaaze M.V., Lychagin N.I., Malakhovskiy A.A. <i>Integrated telecommunication services for departmental network communications</i>	32

Information transfer and processing

Alekseyev V.G. <i>Digital watermarks system for audio signals based on the reverberation effect</i>	43
Voronin V.V. <i>Automated defect detection in case of reconstruction of archive documentary photographs ...</i>	50
Mukhin V.E., Kornaga Ya.I. <i>Multi-level model with differential trust level to subjects for the distributed databases data security improving</i>	54
Zaborovsky V.S., Kondratiev A.S., Muliukha V.A., Silinenko A.V., Ilyashenko A.S. <i>Surface robotic groups control from manned space station within «METERON» project</i>	59
Kislicin A.B., Rashich A.V. <i>Viterbi demodulator for DOQPSK and CDOQPSK signals</i>	66

Radio engineering, aerials, SHF-devices

Golubev A.G., Smirnov A.S. <i>Calculation of upper limit of sensitivity for interference fiber-optic hydrophone array</i>	73
--	----

System analysis and control

Chuprov S.G., Ivanov A.A. <i>Modelling of dynamics of two-motor servo drive with self-locking effect</i>	81
---	----

Computer systems and software

Petrosyan G.S. <i>Analysis of the object-oriented paradigm and search for a better alternative</i>	87
Pyshkin E.V. <i>Problems of software acceptance testing automation in behavior driven development</i>	90
Laboshin L.Yu., Lukashin A.A., Semenovskiy V.B. <i>Method to provide GPU support for virtual machines in XEN hypervisor for data stream processing in robotic systems</i>	97

Mathematical modelling: methods, algorithms, technologies

Valynets E.V., Orzhenovskiy M.V., Ivankov A.A. <i>Stochastic model of supplier bids pool at the energy spot price auction</i>	101
--	-----

Ivanov A.A. <i>Method of a form's planning of the hyper redundant manipulator's with a limited variation range of joint coordinates.....</i>	112
Kadyrova N.O., Pavlova L.V. <i>Methods for efficient processing and mining of big data.....</i>	118
Petrova I.Yu., Khomenko T.V. <i>Mathematical model of the selection and method of evaluation of options for the physical principle of the action of the sensing element.....</i>	124
Chernorytsky I.G. <i>Standard schemes of parametric optimization.....</i>	128
Martsenyuk M.A., Polyakov V.B., Seletkov I.P. <i>Matrix implementation of fuzzy inference alghorithms</i>	133
Chernorytsky I.G. <i>Algorithmic problems of stiff optimization.....</i>	141
Shebalov A.A., Bazhenov A.N. <i>Performance analysis of optical flow techniques</i>	152

Devices, information-measuring systems

Anikeeva M.S., Sofronov A.N., Dremov S.S., Ter-Martirosyan A.L. <i>Digital imaging system for photo-dynamic therapy</i>	159
--	-----

Management in social and economic systems

Latypova E.A., Bondareva I.O. <i>Simulation as a means of logistics enterprises by the example of the cargo port.....</i>	165
Aksyonov K.A., Aksyonova O.P., Wang Kai <i>House-building projects portfolio scheduling based on multi agent simulation.....</i>	171
Rybak V.A., Suleyman H.M. <i>Analysis and optimization of environmental and economic parameters of innovative projects</i>	175
Maslov V.I., Minakov V.F. <i>An efficiency criterion in the enterprise production quality control system.....</i>	179
Kalmykova S.V., Sokoliscin A.S. <i>Optimizing the content and structure of the educational process at the university with using metamodels</i>	185
<i>About the authors.....</i>	193
<i>Abstracts.....</i>	197

РОЛЬ СУПЕРКОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В ИНЖЕНЕРНОМ ОБРАЗОВАНИИ

На протяжении XX века произошла смена технологических укладов: от уклада, характеризующегося массовым использованием паровых двигателей в начале века, до технологического уклада массового применения компьютерных технологий в его конце. В нашем кратком исследовании мы не задаемся целью точного определения существа всех сменивших друг друга технологических укладов и тем более определения их временных границ. Попытаемся лишь оценить ту роль, которую сыграло и продолжает играть внедрение математического моделирования на базе компьютеров и их наиболее мощного сегмента – суперкомпьютеров на основы инженерной деятельности и на инженерное образование.

Начнем с того, что интриги не нужно ждать до конца, мы сразу же ответим на поставленный вопрос: появление суперкомпьютерных технологий сыграло и пока играет слабую роль в инженерном образовании, более того, сегодня не просматриваются полномасштабные пути исправления такого положения дел. Причем это проблема не только отечественной высшей школы, но и высшей школы практически всех передовых стран мира. Почему так происходит? На этот важный вопрос мы и попытаемся найти здесь ответ.

Начнем с тех важнейших обстоятельств, которые видятся ключевыми в понимании существа поставленного вопроса. Этими обстоятельствами являются следующие. Во-первых, нам необходимо определиться с пониманием того, а что же есть суперкомпьютерные технологии? Но это только небольшая часть проблемы. Во-вторых, необходимо установить связь между суперкомпьютерными технологиями и математическим моделированием как некоторым «надтехнологическим инструментарием». И это, второе обстоятельство,

является, без сомнения, важнейшим фактором, в определенном смысле препятствующим широкому внедрению рассматриваемых технологий как в инженерный анализ и проектирование, так и в собственно инженерное образование. Почему так происходит, мы также попытаемся дать ответ в нашей статье.

Суперкомпьютерные технологии

Итак, что же мы понимаем под суперкомпьютерными технологиями? Детальный ответ на этот вопрос сформулирован в [1, 2], поэтому здесь приведем наше видение ответа кратко. Под суперкомпьютерными технологиями мы понимаем совокупность технологий, включающую в себя три блока:

I – технологии разработки и создания суперкомпьютеров¹;

II – технологии разработки и развития системного и прикладного программного обеспечения для суперкомпьютеров;

III – технологии для проблемно-ориентированного применения суперкомпьютеров.

Содержание блоков I и II относительно понятно. Технологии блока III требуют некоторого пояснения. По сути дела, его содержание – это широчайшая сумма технологий, основанных на вычислительных (математических) методах или, говоря точно, на математическом моделировании (см. ниже). Каждая из таких отдельных технологий ориентирована на ту или иную предметную отрасль научного или инженерного знания, медицины, культуры и т. д. Что их объединяет?

¹ Здесь уместно отметить, что большой вклад в развитие суперЭВМ в СССР и в России внес выпускник Политехнического института академик А.В. Каляев, 90-летие которого недавно отмечалось [3].

Объединяющим началом служит высокая вычислительная ресурсоемкость, необходимая для решения задач из той или иной предметной отрасли. При этом, вполне очевидно, что естественнонаучные вычисления (физика, механика, биология и т. д.), как и инженерные расчеты, по содержанию отличаются от кредитно-финансовых (банковских), а первые и вторые разнятся от управления в режиме реального времени с помощью суперкомпьютеров большими объемами сообщений в среде сотовой связи и т. д.

Поскольку здесь нас интересуют инженерные аспекты проблемы и в первую очередь инженерно-образовательные, то сразу же отметим, что подавляющее большинство направлений инженерной деятельности, в которых идет разработка машин, систем, технологий и т. д., прямо связано с естественнонаучными отраслями знания. Последние образуют те области, которые принято считать фундаментальными основами инженерного знания. Сюда в первую очередь следует отнести механику сплошных сред, с такими областями, как аэрогидродинамика, механика твердого и деформируемого тела и т. д.

Подчеркнем, что как естественнонаучные, так и инженерные отрасли объединяет необходимость проведения больших объемов вычислений (а в инженерных отраслях, как правило, и в сжатые сроки), хотя, конечно, в каждой из них имеется своя специфика. Например, в естественнонаучных и инженерных вычислениях особо выделяются методологии решения задач о движении жидкостей и газов, с которыми связано подавляющее большинство задач живой и неживой природы, – от течения крови по сосудам, до движения межгалактических газовых облаков, а в инженерных задачах большинство процессов в машинах и механизмах. Отметим, что исходя из этого, часто конструируют суперкомпьютеры, ориентированные на преимущественное решение задач динамики жидкости и газа. Так, один из наиболее мощных суперкомпьютеров первого десятилетия XXI века – Blue Gene фирмы IBM учитывает особенности решения подобного рода задач [4].

Суперкомпьютерные технологии часто отождествляют с технологиями высокопроизводительных вычислений (High Performance Computing – HPC).

Теперь нам предстоит ответить на вопрос о связи суперкомпьютерных технологий с матема-

тическим моделированием, но прежде всего мы ответим на вопрос о том, а что же мы понимаем под математическим моделированием.

Математическое моделирование

Подавляющее большинство студентов, да и не только студентов, не могут дать внятного ответа на вопрос о том, что служит основой, фундаментом применения компьютеров. Лучший ответ – «вычислительная математика», не является вполне точным и полным, хотя вычислительная математика служат инструментом создания программного обеспечения для компьютеров в очень большой мере. Наиболее полным и точным является ответ – «математическое моделирование».

Современная концепция математического моделирования разработана, главным образом, в трудах академика А.А. Самарского [5]. Приведем основные позиции (блоки) этой концепции, к которым мы в той или иной мере будем обращаться далее.

I. Составление математической модели явления, процесса, задачи и т. д. Это, вообще говоря, предметная область естественных и других наук, где составляются количественные соотношения для описываемых явлений и процессов. И это труднейшая область научной деятельности, представляющая основу, например, естественных наук. Итак, выделяя самое основное, этот блок составляет существо некоторой «замены» реального физического процесса или явления его математической моделью, т. е. описание всех его свойств-характеристик с помощью математических соотношений¹.

II. Анализ математической корректности построенной математической модели, описывающей нашу задачу. Весь спектр рассматриваемых здесь проблем носит сугубо математический характер и в подавляющем большинстве обходится исследователями и инженерами. У опытных вычислителей, как правило, имеется понимание

¹ Например, исследование аэродинамики полноразмерного самолета в аэродинамической трубе при скоростях более 200 м/с практически нереально, но можно, разработав математическую модель его обтекания, получать характеристики обтекания с помощью компьютера. При этом вычислительный эксперимент позволяет нам исследовать поведение самолета при таких скоростях, которые относятся к гиперзвуковым (более 2-3 км/с) и какие принципиально невозможно получить в натурном эксперименте в аэродинамической трубе.



того, что должно получиться. Однако такой путь опасен, и студенту на примерах следует показать, к чему может привести недопонимание важности проблем данного блока.

III. Переход от непрерывной математической модели к модели дискретной. Отметим, что в ряде задач этот блок может отсутствовать, т. е. он, как правило, характерен для таких моделей – явлений, которые описываются дифференциальными и интегральными соотношениями. При переходе от непрерывной модели к дискретной мы, по сути дела, переформулируем нашу задачу и получаем некоторую новую, в которой присутствует один или несколько параметров, характеризующих новую дискретную задачу, и которых не было в исходной задаче. Примером этого является задача вычисления интеграла, когда мы вычисляем его с помощью формул трапеций или Симпсона, в которых имеется параметр N , связанный с разбиением промежутка, как раз и характеризующий степень точности такого приближения.

IV. Анализ математической корректности вновь полученной дискретной задачи (на основе исходной). По сути дела этот блок есть предмет вычислительной математики. Здесь рассматриваются те же проблемы, что и в блоке II, к которым добавляется еще одна весьма важная задача. Решение этой задачи должно ответить на такой вопрос: будет ли построенная в блоке III дискретная задача, например, величина конечной суммы в упомянутом методе трапеций или Симпсона при числе разбиений – параметре $N \rightarrow \infty$, стремиться к числу, равному значению исходного интеграла.

V. Написание алгоритма для дискретной задачи, т. е. последовательности вычислительных шагов и его перенос на компьютер, – программирование. Этот блок носит совершенно особый характер и его роль со времени внедрения вычислительных машин в научную и инженерную практику непрерывно возрастала.

VI. Отладка программы, т. е. ее тестирование, получение результатов и их анализ. Назначение этого блока вполне очевидно и поэтому оставим его без комментариев.

Изложенная схема, конечно, является очень общей и ее применение в полном объеме отдельным исследователем при решении какой-либо задачи, в общем, не характерно и является, как правило, функцией научной группы или коллек-

тива исследователей. Но с позиций тех вопросов, которые мы затрагиваем, описание концепции математического моделирования весьма и весьма важно.

Прежде чем переходить к рассмотрению связи математического моделирования и суперкомпьютерных технологий, попытаемся установить связь математического моделирования с основами инженерного знания.

Математическое моделирование и инженерное знание

Все современное естественнонаучное и инженерное знание представляет собой совокупность многотысячелетнего человеческого опыта. Причем в своем подавляющем большинстве все бесчисленные составляющие этого опыта есть продукт физического эксперимента. Именно многовековой экспериментальный опыт служил основой инженерному сообществу с незапамятных времен, примерно до конца XVIII века, в создании всех машин и механизмов, служивших человеку.

В начале XIX века массовое развитие железнодорожного транспорта и парового флота привело к необходимости создания теоретических основ того, что мы сегодня понимаем под инженерным анализом и проектированием. Полномасштабная потребность и значимость инженерного анализа окончательно сформировались во второй половине – конце XIX века, а с началом XX века, когда очень быстрыми темпами стали появляться принципиально новые механизмы и машины, начали формироваться и новые требования к самому существованию инженерного анализа. В связи с последним важно отметить, что именно глубокое понимание этих проблем и стремление вывести Россию на передовые мировые научно-технические и экономические позиции лежали в основе самой идеи создания Санкт-Петербургского политехнического института [6].

В течение первой половины XX века многовековая практика физического эксперимента подошла к практическому исчерпанию своих возможностей при создании новых типов машин и систем, поскольку при решении все большего числа инженерных, да и многих естественнонаучных проблем, не позволяла провести сам физический эксперимент либо из-за его сложности, либо

по причине принципиальной его невозможности¹. Последние обстоятельства привели инженерное и научное сообщество к необходимости использования все более и более мощных методов математики или, говоря современным языком, методов математического моделирования. Таким образом, в течение XX века произошло революционное изменение инструментария исследования природы и всего того, что создается человеком, т. е. многовековая практика физического эксперимента стала активно дополняться и заменяться экспериментом математическим. Этому способствовали два обстоятельства: во-первых, новейшие достижения в фундаментальной и вычислительной математике как инструменте количественного описания явлений и процессов и, во-вторых, возникновение и мощное развитие вычислительной техники.

Активное внедрение математического моделирования в инженерную практику способствовало переходу на качественно новый уровень в инженерных расчетах и проектировании всего множества машин и систем во всех отраслях промышленности. Особенно сильно это новое качество стало проявлять себя в последние десятилетия в связи с широким применением суперкомпьютеров. Но прежде чем раскрывать существо этого качества, необходимо понять, как и в каких направлениях компьютерные технологии повлияли на развитие промышленности. При этом отметим, что сфера применения компьютеров в промышленности практически сразу с их внедрением в нее стала быстро расширяться и, конечно, не испытывалась инженерными расчетами.

Компьютерные технологии и развитие промышленности во второй половине XX века

Эта тема крайне широка и принципиально сводится к рассмотрению научно-технического существа CALS-технологий или технологий непрерывной поддержки поставок и жизненного цикла продукции (Continuous Acquisition and Life cycle Support), (также часто говорят о Product

Life Management – PLM-технологиях). Сущность CALS- или PLM-технологий состоит в попытке описать и систематизировать в рамках компьютерных технологий все производственные и иные процессы, в которых рассматривается жизнь любого изделия от идеи его создания, проектирования, изготовления, жизни за пределами предприятия (т. е. его эксплуатации и обслуживания), вплоть до ее окончания и, наконец, утилизации. Естественно, что при такой широкой постановке проблемы CALS-технологии содержат в себе массу экономических, управленческих и иных проблем, которые, конечно, также составляют важные аспекты инженерного образования. Но здесь мы сосредоточимся на том, что составляет основу – ядро инженерной деятельности, – на инженерном анализе и проектировании. Этот блок CALS-технологий включает в себя цепочку CAD/CAE/CAM-технологий (CAD/CAE/CAM – Computer Aided Design/Engineering/Manufacturing). То есть проектирование, инженерный анализ и передача его итогов в производство. Таким образом, именно эта «триада» содержит в себе самое наукоемкое, естественнонаучное и инженерное ядро CALS-технологий, в котором исторически традиционный, последовательный подход к разработке изделий заменен принципиально новым, интегрированным подходом, который иногда называют «параллельным проектированием».

Итак, мы подошли к самому главному. Именно цепочка CAD/CAE/CAM-технологий, из которых наиболее трудными как в освоении, так и в реализации, безусловно, являются CAE-технологии, технологии компьютерного инжиниринга, лежит в основе современного инженерного анализа и проектирования. И реализации их на основе суперкомпьютеров в промышленности и «встраиванию» таких технологий в процесс подготовки инженерных кадров по сути дела и посвящена наша статья.

Суперкомпьютерные технологии и инженерный анализ

Развитие суперкомпьютерных технологий с позиций чисто вычислительных мощностей (ресурсов) привело в наши дни научное и инженерное сообщество к возможности ставить и решать такие классы естественнонаучных и инженерных задач, которые были немыслимы для инженеров

¹ Сказанное не следует понимать как полное отрицание физического эксперимента. Он, конечно, нужен и востребован как важный инструментарий в научных исследованиях и во многих натурных экспериментах. Более того, там, где мы плохо знаем природу, такой эксперимент незаменим.



и исследователей даже 15–20 лет тому назад. При этом в части инженерных задач главным инструментом здесь являются названные только что САЕ-технологии, в их суперкомпьютерных реализациях, которые в производственном цикле прямо взаимодействуют с CAD- и САМ-технологиями.

Что же нового привнесли суперкомпьютерные реализации САЕ-технологии компьютерного инжиниринга в промышленность? Вычислительная мощность, главная черта таких технологий, привела к принципиально новому качеству их применения в инженерной практике. И существо этого нового качества состоит в том, что появилась возможность решать междисциплинарные задачи, которые по своей постановке близки к задачам реального физического мира, в нашем случае – мира машин, механизмов и систем, создаваемых в сфере промышленного производства.

Чтобы пояснить роль и новое качество такого подхода в инженерной практике, приведем несколько характерных примеров.

Первым примером характерной задачи, демонстрирующей эффективность применения больших вычислительных ресурсов, является расчет течения в проточной части турбины Саяно-Шушенской ГЭС [7], которая решалась в лаборатории прикладной математики и механики СПбГПУ. Чтобы уяснить масштабы времени, необходимого для проведения расчетов, укажем на то, что для характерного варианта расчет проводился в течение трех месяцев (!) на шести узлах вычислителя, т. е. с использованием 24-процессоров (каждый узел вычислителя оснащен двумя двухъядерными процессорами AMD Opteron 275). При этом область течения разбивалась на 24 части с числом расчетных узлов в 5 624 563. Отметим, что расчеты проводились в 2006–2007 гг. В процессе решения задачи не удалось получить результаты в рамках ее нестационарной постановки, т. к. для этого требовалось как минимум около 38 млн расчетных узлов. Причиной невозможности получения решения на таком количестве узлов было отсутствие на тот момент времени у рабочей группы адекватных поставленной задаче вычислительных ресурсов. Рассматривая эти результаты, следует иметь в виду, что решение нестационарных задач является крайне важным практически. Вспомним аварию на Саяно-Шушенской ГЭС в 2009 г., одной из причин которой явились именно нестационарные процессы в проточной части турбины.

В качестве второго примера приведем расчеты процессов горения, выполненные в рамках магистерской работы на кафедре прикладной математики физико-механического факультета [8]. Рассматривалось численное моделирование пламени, которое в известном смысле является эталонным. Таким пламенем выбрано пламя из «архива»Livermoreской национальной лаборатории США (Sandia National Laboratories, Livermore, California, USA [9]). Число Рейнольдса¹ Re , характеризующее процесс течения газов, сопровождающих горение, в этой задаче равно 22 400, т. е. оно много больше его критического значения, равного примерно 1000, при котором течение становится турбулентным, т. е. рассматриваемое течение заведомо турбулентно. Чтобы представить себе масштабы необходимых для решения этой задачи объемов вычислений, приведем такие данные. Характерный расчет занял пять недель на 15 узлах кластера (т. е. на 60 процессорных ядрах) для одной из моделей горения, при этом в процессе этого расчета было сделано около 450 000 итераций.

Приведем еще один пример, носящий ярко выраженный характер прорывных разработок, которые могут быть реализованы только с использованием суперкомпьютеров. Этот пример подтверждает и недальновидность высказываемых иногда мнений о том, что для удовлетворения потребностей отечественной промышленности и науки суперкомпьютеров сверхвысоких «петафлопсных» мощностей ($1 \text{ PFlops} = 10^{15}$ операций в секунду с плавающей точкой) не требуется.

Итак, существо примера связано с рядом оценок по проектированию и многокритериальной оптимизации авиационного двигателя на нестационарных режимах их работы. Решение такой задачи потребует полной загрузки наиболее мощного на сегодняшний день отечественного суперкомпьютера «Ломоносов» на протяжении семи лет (!). Заметим, что производительность «Ломоносова» составляет в пиковом режиме свыше 1 PFlops . Поэтому для решения такого класса задач в приемлемые для промышленного производства сроки требуются вычислительные системы с производительностью не менее, чем в 1000 раз выше, т. е. системы производительностью порядка 10^{18} операций в секунду с плавающей точкой («экзафлопсные» системы).

¹ $Re = VL/\nu$, где V и L – характерные скорость и размер течения; ν – кинематическая вязкость среды.

Приведенных примеров достаточно для того, чтобы уяснить тот факт, что влияние суперкомпьютерных технологий на сферу промышленного производства становится решающим и уже является реальным фактором в борьбе за конкурентоспособность на мировом рынке, и именно на основе суперкомпьютерных технологий осуществляются все прорывные разработки в ведущих странах мира.

Теперь мы подошли к главному вопросу, предмету статьи: как и в каких формах суперкомпьютерные технологии должны влиять на образовательный процесс подготовки инженерных кадров.

Суперкомпьютерные технологии и их роль в инженерном образовании

Центральной проблемой, обуславливающей трудности и вопросы, встречающиеся на пути широкого внедрения суперкомпьютерных технологий в инженерное образование, является их универсальный и всеобъемлющий характер. Сущность этой универсальности заключается в том, что суперкомпьютерные технологии позволяют нам в максимальном объеме реализовать технологии математического моделирования. А это, в свою очередь, дает возможность ставить и решать во всей полноте междисциплинарные задачи, которые, как указывалось, дают возможность весьма близко подойти к описанию реального физического мира. При этом реализуется вся универсальная цепочка от постановки задачи, выбора эффективных вычислительных схем, обеспечивается требуемая быстрота получения результата и, что особенно важно, инженер и исследователь получают возможность полномасштабной визуализации результатов вычислений.

Но такое высокое, и все более нарастающее в своих возможностях качество описания явлений природы и мира техники требует и все более высокого уровня освоения инженерным сообществом самих фундаментальных основ инженерного знания! При этом сами эти основы становятся малоразличимыми со знанием естественнонаучным. Именно здесь, как нам видится, и «кроется» проблема слабого внедрения суперкомпьютерных технологий в инженерное образование. Высшая техническая школа, в своем подавляющем большинстве, не готова и во многих случаях принципиально не способна к радикальной и

быстрой перестройке учебного процесса на такие технологии.

Итак, как полномасштабное внедрение САЕ-технологий в их связке с САД- и САМ-технологиями, так и внедрение суперкомпьютерных технологий, революционно изменяющих весь процесс инженерной деятельности, требует радикальной перестройки образовательного процесса подготовки инженерных кадров. И сущность этой перестройки состоит в том, что инженерная деятельность в рамках названных технологий все более и более приобретает характер деятельности исследовательской, опирающейся на фундаментальное знание. При этом, как нам представляется, в недалеком будущем произойдет смена самой парадигмы инженерной деятельности. Что же имеется в виду? Если мы рассмотрим все последовательные этапы деятельности инженеров – разработчиков новой техники и систем (конструкторов, расчетчиков, проектировщиков и т. д.), т. е. тех категорий, которые заняты созданием новых машин, механизмов и систем, то эта деятельность в ближайшем будущем «обречена» на все более и более творческий характер, при которой сама проектно-конструкторская, рутинная, расчетно-технологическая работа будет возложена на расчетно-технологические вычислительные среды. Роль инженера-разработчика будет определяться такой творческой деятельностью, как разработка концепции нового изделия, включая эффективность и быстроту его производства, а также определение его эксплуатационно-технических характеристик.

Отметим, что под расчетно-технологическими вычислительными средами нами понимаются аппаратно-программные системы, реализующие САД/САЕ/САМ-технологии. Сегодня многие из составляющей таких аппаратно-программных систем преподаются в высшей школе. По крайней мере практически массовым и в некоторой мере устоявшимся стало преподавание САД-систем различного уровня, на основе которых формируется некоторый образ будущего изделия. Также в ряде вузов имеется и определенный опыт преподавания САМ-систем. Значительно хуже обстоит дело с преподаванием САЕ-технологий компьютерного инжиниринга, причем и с их изучением как универсального вычислительного инструментария, и с их использованием для проведения на качественно новом уровне практических, лабораторных и курсовых работ по очень большо-



му числу курсов. Чтобы не быть голословными, укажем только на те базовые фундаментальные дисциплины, где может быть в полной мере использован потенциал CAE-технологий. Причем отметим, что здесь к числу технологий компьютерного инжиниринга естественно отнести комплексы уровня ANSYS Mechanical, ANSYS Fluent и т. д., которые по уровню их массового использования мы относим к программным комплексам мирового уровня.

Итак, это такие базовые фундаментальные дисциплины:

механика жидкости газа;
механика твердого и деформируемого тела;
основы тепло- и массообмена;
вычислительная гидроаэромеханика;
вычислительная механика
и т. д.

Очевидно, что этот список может быть продолжен с большей предметной ориентацией на те или иные группы инженерных дисциплин. Заметим, что практически все дисциплины из приведенного списка при рассмотрении серьезных постановок задач требуют большого вычислительного ресурса, а значит, и использования суперкомпьютеров.

Но использование CAE-технологий, даже

в их полномасштабной связке с CAD- и CAM-технологиями, есть только один из сегментов того, что в недалеком будущем станет основой нового содержания инженерного образования. Как нам видится с позиций сегодняшнего дня, само содержание инженерного образования должно включать в себя преподавание реализованной в суперкомпьютерном варианте группы дисциплин, в основе которых лежит триада CAD/CAE/CAM-технологий. Задача ведущих технических вузов состоит в том, чтобы, во-первых, предложить концепцию инженерного образования на базе суперкомпьютерных технологий и, во-вторых, определить те группы дисциплин, которые будут составлять некоторое ядро – основу такого образования.

Санкт-Петербургский политехнический университет подготовил проект создания Суперкомпьютерного центра на базе вуза, поддержанный Минэкономразвития, Минобрнауки и Минкомсвязи России, который планируется реализовать в 2013–2014 гг. Важнейшим аспектом работы Суперкомпьютерного центра станет переход всей системы учебной работы на его основе, что позволит вузу готовить кадры мирового уровня с использованием наиболее передовых технологий.

СПИСОК ЛИТЕРАТУРЫ

1. **Болдырев, Ю.Я.** Суперкомпьютерные технологии как современное воплощение междисциплинарного подхода в научно-образовательной деятельности [Текст] / Ю.Я. Болдырев // Научно-технические ведомости СПбГПУ. Сер. Информатика. Телекоммуникации. Управление. –2010. –№ 4. –С. 99–106.
2. **Болдырев, Ю.Я.** Суперкомпьютерные технологии и их приложения: Учеб. пособие [Текст] / Ю.Я. Болдырев, Е.П. Петухов. –СПб.: Изд-во СПбГПУ, 2010. –92 с.
3. Академик Анатолий Васильевич Каляев. Ученый, воин, учитель. К 90-летию со дня рождения. –Ростов-на-Дону, 2012. –328 с.
4. **Sugavanam, Suga.** CFD on the Blue Gene IBM Systems & Technology Group [Электронный ресурс] / Suga Sugavanam, Claude Oudet, Pascal Vezolle // Пресс-релиз компании IBM, март 2006.
5. **Самарский, А.А.** Математическое моделирование: Идеи. Методы. Примеры [Текст] / А.А. Самарский, А.П. Михайлов. –2-е изд., испр. –М.: Физматлит, 2001. –320 с.
6. **Смелов, В.А.** Санкт-Петербургский политехнический. Дореволюционный [Текст] / В.А. Смелов. –СПб.: ООО «Береста», 2012. –620 с.
7. **Болдырев, Ю.Я.** Разработка методических вопросов расчета потерь энергии в элементах проточной части гидротурбин с применением вычислительной гидродинамики [Текст] / Ю.Я. Болдырев, Ю.К. Шиндер, С.В. Лупуляк [и др.] // Отчет о НИР с ОАО «Силовые машины ЛМЗ». –СПб.: Изд-во СПбГПУ, 2006. –70 с.
8. **Фролов, А.С.** Численное моделирование естественно-конвективного пламени методом крупных вихрей: Дис. ... магистра [Текст] / А.С. Фролов, А.Ю. Снегирев. –СПб.: Изд-во СПбГПУ, 2010.
9. Сайт Ливерморской лаборатории (Livermore, California, USA) [Электронный ресурс] / Режим доступа: <http://www.sandia.gov>

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ЭФФЕКТИВНЫХ АЛГОРИТМОВ ПОИСКА И РАЗМЕЩЕНИЯ ДАННЫХ В ДЕЦЕНТРАЛИЗОВАННОЙ РАСПРЕДЕЛЕННОЙ ФАЙЛОВОЙ СИСТЕМЕ

В последние годы быстрое развитие получили технологии организации распределенной обработки информации и высокопроизводительных вычислений. Одним из классов таких технологий являются облачные технологии хранения данных – это модель обеспечения повсеместного и удобного сетевого доступа по требованию к общему пулу конфигурируемых вычислительных ресурсов (например, к сетям передачи данных, серверам, устройствам хранения данных), которые могут быть оперативно предоставлены и освобождены с минимальными эксплуатационными затратами и/или обращениями к провайдеру.

Ключевые особенности, которыми должны обладать такие технологии:

- децентрализация – форма системы коллективных узлов без координации;
- масштабируемость – система будет одинаково эффективно функционировать при тысячах или миллионах узлов;
- отказоустойчивость – система будет одинаково надежна (в некотором смысле) с узлами, постоянно включающимися, выключающимися и выдающими ошибки.
- мобильность доступа – система должна быть доступна с тем же качеством из любой географической точки мира;
- расширяемость – новые вычислительные узлы с легкостью добавляются в систему;
- безопасность – авторизация доступа, контроль доступа, контроль целостности, защита от несанкционированных вторжений извне;
- изоляция – данные пользователя должны быть изолированы от всех других данных.

К облачным технологиям хранения данных относят различные системы, использующие разнообразные алгоритмы поиска и размещения данных:

облачные хранилища (Amazon S3, EMC Atmos, OceanStore);

распределенные файловые системы (GFS, Lustre, Ceph, HDFS);

децентрализованные распределенные файловые системы (FreeNet, Tahoe-LAFS);

p2p системы обмена файлами (BitTorrent, Gnutella).

В данной статье рассматриваются p2p децентрализованные файловые системы, в которых алгоритм поиска и размещения данных реализован с помощью DHT (Distributed Hash Table – «распределенная хеш-таблица»). В дальнейшем под термином *децентрализованная распределенная файловая система* будет подразумеваться децентрализованная распределенная файловая система в классе p2p DHT.

DHT – это класс децентрализованных распределенных систем, которые обеспечивают поисковый сервис, похожий по принципу работы на хеш-таблицу, и имеют структуру ассоциативного массива: (ключ, значение), хранящиеся в DHT, а каждый участвующий узел может рационально искать значение, ассоциированное с данным именем. Ответственность за поддержку связи между именем и значением распределяется между узлами. Таким образом, изменение набора участников является причиной минимального количества разрывов. Это позволяет легко масштабировать DHT и постоянно отслеживать добавление/удаление узлов и ошибки в их работе. Примеры DHT систем: Chord [1], Pastry[2], CAN [3], Tapestry [4].

Однако у DHT систем есть недостатки, связанные с тем, что данные, которые размещаются в системе, жестко привязаны к узлам системы.

Chord [1] присваивает идентификаторы узлам и блокам данных из одного одномерного пространства идентификаторов. Узел, который отвечает за хранение блока данных, является узлом, идентификатор которого наиболее близок к идентификатору блока данных.

Pastry [2] присваивает каждому узлу случайный идентификатор, который определяет позицию узла во множестве идентификаторов с круговой топологией. Для поиска блока данных по его идентификатору поисковый запрос отсылается узлу с идентификатором, ближайшим к идентификатору блока данных.

CAN [3] использует n -мерное пространство декартовых координат для осуществления DHT



абстракции. Координатное пространство разбивается на гиперпрямоугольники, называемые *зонами*. Каждый узел в системе отвечает за зону и определяет границы своей зоны. Блок данных отображается в точку в данном пространстве координат и хранится в узле, зона которого содержит координаты точки.

Tapestry [4] осуществляет поиск блока данных по его идентификатору, сравнивая определенные биты идентификатора с определенными битами идентификатора узла, поисковый запрос отсылается тому узлу, битовое совпадение с которым наилучшее.

В случае добавления новых узлов к DHT требуется перераспределение размещенных данных по узлам системы. Кроме того, миграция данных между двумя узлами осложнена по той же причине.

Лишенная этих недостатков распределенная система рассмотрена в [6]. Предлагаемая система имеет три основных свойства. Во-первых, данные единицы образуют сеть типа «малого мира», гарантирующую, что любой блок данных доступен из любого другого блока данных за небольшое количество шагов. Это помогает быстро найти любой блок данных, начиная с любого другого блока данных. Во-вторых, все блоки данных упорядочены и задана метрика (расстояние) между двумя блоками данных. Это позволяет построить систему таким образом, что простой жадный алгоритм поиска может использоваться для поиска любого блока данных. В-третьих, блоки данных не привязаны к конкретным узлам, а топология системы основана на графе блоков данных, а не на графе узлов, как, например, в случае DHT (Chord) [1].

В работе [6] не рассмотрены вопросы обеспечения отказоустойчивости системы в том случае, если узлы системы склонны выключаться на время, что вполне реально при практическом применении.

В данной статье изучается способ обеспечения отказоустойчивости системы за счет построения двухуровневой сети малого мира. Первый уровень состоит из групп узлов, и образуют сеть, аналогичную [6], нижележащий второй уровень состоит из блоков данных, которые также образуют сеть типа [6]. В работе проводится математическое исследование и моделирование алгоритмов поиска и размещения данных в построенной распределенной файловой системе.

Моделирование поиска и размещения данных в распределенной файловой системе

Для моделирования компьютеров, соединенных по сети Internet, с хранимыми на них данными используется граф, в вершинах которого находятся отдельные компьютеры системы, а ребра графа представляют собой каналы коммуникации между компьютерами системы. Для создания более реалистичной модели считается, что компьютеры могут как появляться и присоединяться к распределенной системе, так и пропадать, причем пропадать компьютеры могут не только штатно (например, компьютер выключен для проведения технического обслуживания), но и в случае возникновения сбоев в компьютере или в каналах связи между ними. В данной статье рассматривается построение системы масштаба 10^5 географически удаленных компьютеров. Для эффективной коммуникации такого числа компьютеров весьма важна эффективная топология, обеспечивающая быструю связь между любыми двумя компьютерами системы. Одним из вариантов такой топологии системы может быть модель «малого мира» [5].

«Малый мир» – это тип связного графа, в котором каждая вершина достижима из любой другой за малое количество шагов порядка $\log(N)$, где N – число вершин графа, при этом любая вершина не обязана соседствовать с другой напрямую. Существует множество работ, описывающих свойства «малого мира», мы рассматривали работу [6], в которой описаны эффективные алгоритмы поиска данных в «малом мире». За основу взята эта модель, т. к. она обладает следующими важными свойствами:

коротким путем от одной вершины графа к любой другой, пропорциональным $\log(N)$, где N – число вершин графа [6];

надежностью сети, образуемой данным графом. То есть если, например, удалить какую-либо вершину из такого графа, в большинстве случаев граф останется связанным, а длина пути из одной вершины до любой другой значительно не изменится.

Модель «малого мира» предполагает, что вершины графа не будут удаляться или добавляться слишком часто. Тем не менее компьютеры (участники распределенной системы) могут появляться и удаляться значительно чаще, чем предполагается в модели [6]. Для того чтобы сгладить данное явление, компьютеры объединяются в небольшие группы, а сами группы объединяются в граф «малого мира». Группы компьютеров хранят информацию о данных распределенной

файловой системы с помощью так называемой (n, k) -схемы, описанной в работе [7]. (n, k) -схема – это схема избыточного кодирования, позволяющая разбить исходную информацию на n частей и через продолжительное время восстановить из любых k частей исходную информацию.

Таким образом, можно разбить любую информацию на n частей, затем эти части разложить по компьютерам группы для последующего хранения. Возникающие вопросы репликации данных и синхронизации между компьютерами группы в данном методе хранения выходят за рамки этой статьи. Для упрощения будем считать, что компьютеры группы хранят идентичные копии информации.

Рассмотрим n компьютеров, объединенных в одну группу. Считая что компьютер работоспособен (доступен) с вероятностью p , можно оценить работоспособность группы, как вероятность того, что хотя бы один из компьютеров группы включен:

$$P_g = 1 - (1 - p)^n.$$

Например, пусть $p = 0,5$, уже при $n \geq 5$ вероятность $P_g \geq 0,96875$ и близка к единице, что и позволяет объединить группы в «малый мир» и воспользоваться его свойствами.

Нами предложена математическая модель, описывающая алгоритмы поиска и размещения данных в описанной распределенной системе. Модель описывает систему, состоящую из двух-уровневой сети «малого мира»: первый уровень состоит из групп узлов, образующих сеть, аналогичную [6]; ниже лежащий второй уровень состоит из блоков данных, также образующих сеть типа [6]. Маршрутизация между двумя блоками данных, не лежащих в одной группе, строится с помощью маршрутизации на уровне групп узлов.

Модель системы состоит из следующих частей:

1) модели одного узла – близкие по географической метрике узлы системы объединяются в группы узлов;

2) модели группы узлов – группы узлов образуют т. н. «small world» структуру [6], когда каждая группа связана лишь с небольшим количеством других групп;

3) модели (n, k) -схемы [7] преобразования данных для обеспечения сохранности данных;

4) модели кэширования данных – миграции и дубликации данных внутри распределенной системы для более эффективной работы с точки зрения расхода ресурсов.

Математическая модель поиска и размещения информации

В нашей модели данные хранятся в виде пар: ключ и соответствующее ключу значение. Зададим множество уникальных ключей K . Пары ключ-значение в распределенной системе хранятся на серверах системы, а сами серверы системы объединяются в группы. Для простоты будем считать, что потери на коммуникацию между серверами внутри группы незначительны в сравнении с потерями на коммуникацию между группами. Тогда будем считать, что ключи лежат внутри групп серверов, пропуская уровень серверов в нашей модели.

Определение 1. Множество $GUID$ – множество уникальных целых чисел из числового отрезка $[2^{127}, 2^{128} - 1]$.

Алгоритм генерации уникальных 128-битных идентификаторов известен и описан в работе [8] (версия 4). Можно оценить для этого алгоритма вероятность того, что два сгенерированных идентификатора совпадут, с учетом «парадокса дней рождений» [9], приближенной формулой:

$$p(n) \approx 1 - \exp\left(-\frac{n^2}{2x}\right),$$

где n – число сгенерированных идентификаторов; $x = 2^{122}$, где 122 – число бит в алгоритме, генерируемых случайным образом. Например, при $n = 2^{36} \approx 6,9 \cdot 10^{10}$, $p \approx 4 \cdot 10^{-16}$. Хотя уникальность каждого отдельного идентификатора не гарантируется, будем считать, что генерируемые идентификаторы уникальны. Для упрощения считается, что существует функция $gen_guid()$, которая возвращает уникальный, ранее не использовавшийся идентификатор из множества $GUID$.

Определение 2. Ключом k системы называется кортеж $(id, data, gid)$, где $id \in GUID$, $data \in B$, B – множество слов из алфавита $\{0, 1\}$, $gid \in GUID$.

Будем в дальнейшем обозначать для ключа k , что $k.id$ – идентификатор ключа; $k.data$ – данные ключа; $k.gid$ – идентификатор группы, к которой относится данный ключ.

В качестве данных ключа $k.data$ могут выступать любые бинарные данные. Например, содержимое каждого файла распределенной файловой системы может храниться в поле $k.data$ ключа, ассоциированного с файлом.

Определение 3. Группа ключей g системы называется *кортеж* $(id \in GUID, keys \subset K)$.

Будем в дальнейшем обозначать для группы g , что $g.gid$ – это идентификатор группы.

Зададим два множества – множество уникальных ключей K и множество групп G . Будем



считать, что множество групп G состоит из групп ключей, так что никакие две группы не имеют общих ключей. Далее введем неориентированный граф ключей системы – $G_k(V_k, E_k), V_k \subset K$ и неориентированный граф групп системы – $G_g(V_g, E_g), V_g \subset G$.

Введем функцию M :

$M : GUID \times GUID \rightarrow Z^+, Z^+$ – множество целых положительных чисел

$$M(x, y) = |x - y|,$$

где $x - y$ – арифметическое вычитание целых чисел; $|x - y|$ – взятие модуля целого числа.

На множестве групп определим функцию M_g – метрика близости групп по их идентификаторам:

$$M_g : V_g \times V_g \rightarrow Z^+ \\ M_g(g_x, g_y) = |g_x.id - g_y.id|,$$

где $g_x, g_y \in V_g$.

На множестве ключей определим функцию M_k – метрика близости ключей по их идентификаторам:

$$M_k : V_k \times V_k \rightarrow Z^+ \\ M_k(k_x, k_y) = |k_x.id - k_y.id|,$$

где $k_x, k_y \in V_k$.

Лемма 1 Функции $M_g(g_x, g_y), M_k(k_x, k_y)$ удовлетворяют аксиомам метрики:

1. $M_g(g_x, g_y) = 0 \Leftrightarrow g_x = g_y$
и $M_k(k_x, k_y) = 0 \Leftrightarrow k_x = k_y$.
2. $M_g(g_x, g_y) = M_g(g_y, g_x)$
и $M_k(k_x, k_y) = M_k(k_y, k_x)$.
3. $M_g(g_x, g_z) \leq M_g(g_x, g_y) + M_g(g_y, g_z)$
и $M_k(k_x, k_z) \leq M_k(k_x, k_y) + M_k(k_y, k_z)$.

Доказательство. Доказательство следует из свойств арифметического вычитания и модуля целых чисел.

Для ключей и групп из графов $G_k(V_k, E_k), G_g(V_g, E_g)$ определим функции neighbours, successor, predecessor, nearest следующим образом:

$$\text{neighbours}(x) = \Gamma^+(x),$$

где $\Gamma^+(x)$ – множество смежности вершины x .

$$\text{successor}(x, id) = \\ = z \in \Gamma^+(x) : M(z.id, id) = \min(M(y.id, id)), \\ \forall y \in \Gamma^+(x) : y.id > id.$$

$$\text{predecessor}(x, id) = \\ = z \in \Gamma^+(x) : M(z.id, id) = \min(M(y.id, id)), \\ \forall y \in \Gamma^+(x) : y.id < id. \\ \text{nearest}(x, id) = \\ = z \in \Gamma^+(x) : M(z.id, id) = \min(M(y.id, id)), \\ \forall y \in \Gamma^+(x).$$

Приведем описание алгоритмов добавления и поиска ключа/группы в систему в виде псевдокода.

Алгоритм поиска группы по ее идентификатору. Поиск группы реализован как последовательное приближение к группе с идентификатором, равным искомому. Для этого на каждом шаге алгоритма из соседей текущей группы выбирается сосед с идентификатором ближе по метрике к искомому идентификатору и осуществляется переход в эту соседнюю группу. Алгоритм итеративно повторяется до достижения цели.

group_nearest_search(curr_g, id):

#curr_g – любая группа, уже находящаяся в графе групп

#id – идентификатор группы, по которому требуется найти группу

1. d = M(curr_g.id, id)

2. if d = 0 return curr_g

3. nearest_g = nearest(curr_g, id)

4. if M(nearest_g.id, id) < d : curr_g = nearest_g; goto 1;

5. return nearest_g

Алгоритм добавления новой группы. Для добавления новой группы в граф достаточно найти группу с идентификатором, ближайшим к идентификатору новой группы, затем у этой ближайшей группы найти соседа который является predecessor или successor для новой группы, далее связать эти две группы с новой:

group_append(curr_g, new_g):

#curr_g – любая группа, уже находящаяся в графе групп

#new_g – новая группа, которую необходимо добавить в граф групп

1. new_g.id = gen_guid()

2. nearest_g = group_nearest_search(curr_g, new_g.id)

3. if nearest_g.id = new_g.id return ALREADY_EXISTS

4. if nearest_g.id < new_g.id: neighbour = successor(nearest_g, new_g.id)

5. else: neighbour = predecessor(nearest_g, new_g.id)

6. mutually connect new_g and neighbour

7. mutually connect new_g and nearest_g
8. return SUCCESS

Алгоритм поиска ключа по его идентификатору. По аналогии с группами строится граф ключей. Ключи связаны между собой близкими по метрике идентификаторами, и для поиска достаточно переходить от одного ключа к другому с идентификатором, более близким к искомому, пока не будет найден искомый или ближайший. Отличием является то, что следующий в пути поиска ключ может находиться в другой группе, и эту группу надо найти в графе групп по ее идентификатору:

key_nearest_search(curr_g, id):

#curr_g – любая группа, уже находящаяся в графе групп

#id – идентификатор ключа, по которому требуется найти ключ

1. curr_k = curr_g.keys[0]
2. d = M(curr_k.id, id)
3. if d = 0 return curr_k
4. nearest_k = nearest(curr_k, id)
5. if M(nearest_k.id, id) >= d: return nearest_k
6. if nearest_k not in curr_g:
7. curr_g = group_search(curr_g, nearest_k.gid)
8. return key_nearest_search(curr_g, id)
9. else: curr_k = nearest_k; goto 2

Алгоритм добавления нового ключа. Для добавления нового ключа в систему достаточно найти ключ с идентификатором, ближайшим к идентификатору нового ключа, затем у этого ближайшего ключа найти соседа, который является predecessor или successor для нового ключа, и связать эти два ключа с новым:

key_append(curr_g, new_k):

#curr_g – любая группа, уже находящаяся в графе групп

#new_k – новый ключ, который необходимо добавить в граф ключей

1. new_k.id = gen_guid()
2. nearest_k = key_nearest_search(curr_g, new_k.id)
3. if nearest_k.id = new_k.id return ALREADY_EXISTS
4. if nearest_k.id < new_k.id: neighbour = successor(nearest_k, new_k.id)
5. else: neighbour = predecessor(nearest_k, new_k.id)
6. mutually connect new_k and neighbour
7. mutually connect new_k and nearest_k
8. return SUCCESS

Далее докажем, что приведенные выше алгоритмы корректно и эффективно работают.

Определение 4. Предыдущим ключом p по

отношению к ключу k называется такой ключ p , что $p.id < k.id : p \in \Gamma^+(k)$.

Последующим ключом s по отношению к ключу k называется такой ключ s , что $s.id > k.id : s \in \Gamma^+(k)$.

Предыдущей группой p по отношению к группе g называется такая группа p , что $p.id < g.id : p \in \Gamma^+(g)$.

Последующей группой s по отношению к группе g называется такая группа s , что $s.id > g.id : s \in \Gamma^+(g)$.

Определение 5. Множество $KeyIds = \bigcup_{k \in V_k} k.id$.

Множество $GroupIds = \bigcup_{g \in V_g} g.id$.

Определение 6. Максимальным по идентификатору ключом из множества $V_k \subset G_k(V_k, E_k)$ называется ключ $k_{\max}$, такой, что $k_{\max}.id = \max_{id \in KeyIds} id$.

Минимальным по идентификатору ключом из множества $V_k \subset G_k(V_k, E_k)$ называется ключ $k_{\min}$, такой, что $k_{\min}.id = \min_{id \in KeyIds} id$.

Максимальной по идентификатору группой из множества $V_g \subset G_g(V_g, E_g)$ называется группа $g_{\max}$, такая, что $g_{\max}.id = \max_{id \in GroupIds} id$.

Минимальной по идентификатору группой из множества $V_g \subset G_g(V_g, E_g)$ называется группа $g_{\min}$, такая, что $g_{\min}.id = \min_{id \in GroupIds} id$.

Лемма 2

1. $\forall k : k \in V_k \setminus \{k_{\min}, k_{\max}\} \subset G_k(V_k, E_k)$ существуют последующий и предыдущий ключи по отношению к ключу k .

2. У ключа $k_{\min} \in V_k \subset G_k(V_k, E_k)$ существует последующий ключ по отношению к ключу $k_{\min}$.

3. У ключа $k_{\max} \in V_k \subset G_k(V_k, E_k)$ существует предыдущий ключ по отношению к ключу $k_{\max}$.

Доказательство. Следует из построения графа $G_k(V_k, E_k)$, непосредственно из процедуры key_append. Для вставки нового ключа процедура key_append проходит по вершинам графа в поисках места, где новый ключ станет соседом двух вершин или одной вершины. В случае двух вершин эти вершины являются предыдущим и последующим ключами по отношению к новому ключу. Случай нахождения одной вершины-соседа – это случай, когда вставляемый ключ будет являться $k_{\min}$ либо $k_{\max}$.

Лемма 3

1. $\forall g : g \in V_g \setminus \{g_{\min}, g_{\max}\} \subset G_g(V_g, E_g)$ существуют последующая и предыдущая группы по отношению к группе g .

2. У группы $g_{\min} \in V_g \subset G_g(V_g, E_g)$ существует последующая группа по отношению к группе $g_{\min}$.



3. У группы $g_{\max} \in V_g \subset G_g(V_g, E_g)$ существует предыдущая группа по отношению к группе $g_{\max}$.

Доказательство. Следует из построения графа $G_g(V_g, E_g)$, непосредственно из процедуры group_append. Для вставки новой группы процедура group_append проходит по вершинам графа в поисках места, где новая группа станет соседом двух вершин или одной вершины. В случае двух вершин эти вершины являются предыдущим и последующим группами по отношению к новой группе. Случай нахождения одной вершины-соседа – это случай, когда вставляемая группа будет являться $g_{\min}$ либо $g_{\max}$.

Определение 7. $D(k_x, k_y)$ – функция расстояния между ключами системы определена на множестве ключей и равна числу межгрупповых переходов в процедуре key_nearest_search, то есть числу вызовов group_search в key_nearest_search.

Таким образом,

$D(k_x, k_y) = 0$, если $k_x = k_y$.

$D(k_x, k_y) = 0$, если $k_x.gid = k_y.gid$, то есть принадлежат одной группе.

Теорема 1 $\forall k_x \neq k_y : k_x, k_y \in G_k(V_k, E_k)$ верно, что

1) существует путь от ключа k_x до ключа k_y ;

2) путь от ключа k_x до ключа k_y конечен;

3) если идентификаторы ключей и групп, которые добавлялись последовательно в систему, равномерно распределены на множестве GUID, то длина пути – $D(k_x, k_y) \sim O(\log(|V_g|))$.

Доказательство. Докажем п. 1. Для любого ключа $k \neq k_y$ существует ключ $x \in \Gamma^+(k)$, который ближе по метрике M к k_y . Действительно, в качестве такого ключа x для ключа k мы можем выбрать его предыдущий ключ или последующий ключ на основе леммы 2. Таким образом, можно построить путь от любого ключа $k \neq k_y$ к ключу k_y , в том числе путь от k_x до k_y . Если $k_x.id < k_y.id$, то мы можем построить путь $k_x, k_i, k_{i+1}, \dots, k_N, k_y$ такой, что $k_x.id < k_i.id < k_{i+1}.id < \dots < k_N.id < k_y.id$.

Иначе $k_x.id > k_y.id$, и мы можем построить путь $k_x, k_i, k_{i+1}, \dots, k_N, k_y$ такой, что $k_x.id > k_i.id > k_{i+1}.id > \dots > k_N.id > k_y.id$.

Докажем п. 2. Из доказательства п. 1, построенный путь будет конечен, т. к. последовательность идентификаторов ключей, проходимых в построенном пути, конечна. Действительно, на отрезке $[k_x.id, k_y.id]$ найдется лишь конечное число идентификаторов ключей, т. к. сами идентификаторы уникальны и являются целыми положительными числами. В построенном пути ни

один из ключей не встречается дважды.

Докажем п. 3. Вычисление длины пути от k_x до k_y упрощается, если рассматривать межгрупповой путь от группы, к которой принадлежит k_x , до группы, к которой принадлежит k_y . К такому упрощению можно перейти из определения функции $D(k_x, k_y)$: из того факта, что путь между ключами одной группы равен нулю. Следовательно, длина пути $D(k_x, k_y)$ равна длине межгруппового пути от группы, к которой принадлежит k_x , до группы, к которой принадлежит k_y . Граф групп $G_g(V_g, E_g)$ аналогичен структуре, описанной в [6], причем выполняются те же самые предположения относительно распределения входных данных. В работе [6] доказывается, что длина пути от одной вершины структуры до другой $\sim O(\log(N))$, где N – число элементов структуры. Для аналогичного графа $G_g(V_g, E_g)$ справедлива такая же оценка длины пути. Следовательно, $D(k_x, k_y) \sim O(\log(|V_g|))$.

Результаты имитационного моделирования

Нами разработан комплекс программ, реализующий имитационную модель системы, проведены эксперименты по измерению длины пути поисковых запросов в системе.

Цель имитационного моделирования – подтверждение полученных математических оценок длины пути поискового запроса. Для этого провели измерения длины пути поискового запроса в зависимости от числа групп.

Алгоритм проведения испытания следующий.

1. Пусть число групп 2, число ключей 1024.

2. Создается заданное количество групп, которые объединяются в сеть «малого мира». Идентификаторы групп генерируются с помощью функции gen_guid(). Считаем, что качество используемого генератора случайных чисел приемлемо для эксперимента.

Средняя длина пути запроса search, число серверов = 1024, число ключей=1024

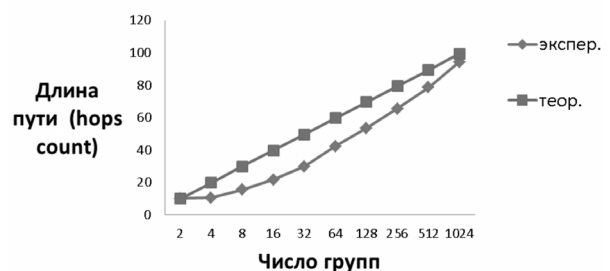


Рис. 1. Зависимость длины пути запроса search от числа групп в системе

3. В каждой группе размещается заранее заданное количество ключей (для каждой группы одинаковое число ключей), которые объединяются в сеть «малого мира». Число ключей, размещаемых в каждой группе, равно (число ключей)/(число групп). Идентификаторы генерируются с помощью функции `gen_guid()`.

4. Далее случайным образом выбираются два различных ключа в системе. С первого ключа запускается поиск второго ключа по его идентификатору. Записывается длина пути запроса.

5. Повторяем с шага 3 сто раз для различных пар ключей.

6. Усредняем по числу запросов длину пути запроса.

7. Если число групп не превышает число ключей, то удваиваем число групп и переходим к шагу 2. Если число групп превышает число ключей, то измерения закончены.

На графике (рис. 1) приведены результаты измерения длины поискового запроса в зависимости от числа групп в системе. Видно, что длина поискового запроса (длина пути от одного ключа к другому) в среднем логарифмически зависит от числа групп в системе.

Далее мы добавили в имитационную модель уровень серверов, хранящих ключи с данными, которые объединялись в группы серверов. На графике (рис. 2) приведены результаты измерения длины поискового запроса в зависимости от числа серверов в системе при фиксированном числе групп. Видно, что длина поискового запроса (т. е. путь от одного ключа к другому) в среднем логарифмически зависит от числа серверов в системе.

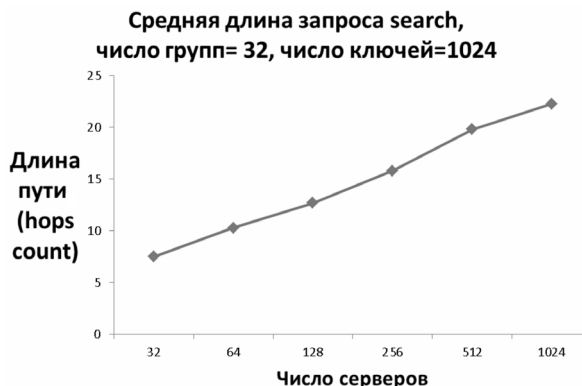


Рис. 2. Зависимость длины пути запроса search от числа серверов в системе

Предложена модель распределенной системы, построенная как двухуровневая сеть «малого мира», первый уровень которой состоит из групп узлов, образующих сеть, аналогичную [6], нижележащий второй уровень состоит из блоков данных, также образующих сеть типа [6]. Изучены алгоритмы поиска и добавления данных, а также поиска и добавления узлов системы.

На основе предложенной модели и алгоритмов построена математическая модель системы. Рассматривая данную математическую модель системы, доказана теорема о работоспособности и эффективности данной системы: получена математическая оценка длины поиска в данной системе.

Разработан комплекс программ, реализующий имитационную модель системы, проведены эксперименты по измерению длины поискового запроса в системе. Установлено, что экспериментальные результаты соответствуют полученной оценке длины поискового запроса.

СПИСОК ЛИТЕРАТУРЫ

1. **Stoica, I.** Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications [Text] / I. Stoica, R. Morris, D. Karger [et al.] // Proc. ACM SIGCOMM, Aug. 2001.
2. **Owstron, A.** Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems [Электронный ресурс] / A. Owstron, P. Druschel // In Proc. of the 18th IFIP/ACM International Conf. on Distributed Systems Platforms (Middleware 2001), Nov. 2001.
3. **Ratnasamy, S.** A scalable contentaddressable network [Text] / S. Ratnasamy, P. Francis, M. Handley [et al.] // In Proc. ACM SIGCOMM. –San Diego, CA, Aug. 2001. –P. 161–172.
4. **Hildrum, K.** Distributed Object Location in a Dynamic Network [Электронный ресурс] / K. Hildrum, J.D. Kubatowicz, S. Rao [et al.] // In Proc. 14th ACM Symp. on Parallel Algorithms and Architectures, Aug. 2002.
5. **Watts, Duncan J.** Collective dynamics of 'small-

- world' networks [Электронный ресурс] / Duncan J. Watts, H. Steven Strogatz. –June 1998. –Nature 393 (6684). –P. 440–442.
6. **Krylov, V.** Metrized Small World Properties Data Structure [Электронный ресурс] / V. Krylov, A. Logvinov, A. Ponomarenko, D. Ponomarev // Proc. Software Engineering and Data Engineering, SEDE 2008.
7. **Тормасов, А.Г.** Обеспечение отказоустойчивости в распределенных средах [Текст] / А.Г. Тормасов, М.А. Хасин, Ю.И. Пахомов // Программирование. – 2001. –№ 5. –С. 26–34.
8. RFC 4122, A Universally Unique IDentifier (UUID) URN Namespace [Электронный ресурс] / Режим доступа: <http://www.ietf.org/rfc/rfc4122.txt>
9. **Klamkin, M.** Extensions of the Birthday Surprise [Text] / M. Klamkin, D. Newman // J. of Combinatorial Theory 3. –1967. –P. 279–282.



Телекоммуникационные системы и компьютерные сети

УДК 004.5+004.77

*В.С. Заборовский, А.С. Кондратьев, В.А. Мулюха,
А.В. Силиненко, А.С. Ильяшенко, М.С. Филиппов*

УДАЛЕННОЕ УПРАВЛЕНИЕ РОБОТОТЕХНИЧЕСКИМИ ОБЪЕКТАМИ В КОСМИЧЕСКИХ ЭКСПЕРИМЕНТАХ СЕРИИ «КОНТУР»

Роботы традиционно используются в неблагоприятных для человека условиях, а также при реализации механических и умственных монотонных операций, требующих больших временных затрат. По мере развития вычислительных и телекоммуникационных технологий роботам и их группировкам становятся доступны принципиально новые возможности выполнения многоцелевых операций, связанных, например, с применением в экстремальных условиях космических полетов и реализацией исследовательских миссий при освоении планет Солнечной системы [1, 2]. При интеграции таких робототехнических систем в глобальную информационно-телематическую среду существенно повышается эффективность их применения, что приводит к сокращению затрат на проведение дорогостоящих космических экспедиций.

Необходимость функционирования в сложных условиях недетерминированной среды на большом удалении от центра принятия решений и при сравнительно небольшой длительности сеанса связи выдвигает особые требования к системе управления космическими робототехническими объектами (РО). Такая система должна быть надежной и устойчивой к воздействиям неблагоприятных факторов, основными из которых являются задержки при передаче команд управления и данных телеметрии. Наличие задержек в каналах связи затрудняет возможность управления роботом в реальном масштабе времени и снижает точность выполнения операций. При этом в зависимости от характеристик канала управления задержки пакетов данных имеют различные статистические свойства, включая степенные рас-

пределения значений дисперсий, изменяющиеся в широком диапазоне и обладающие фрактальным характером, типичным для сети Интернет [3, 4]. Изучение таких свойств привело к появлению нового класса информационных структур, получивших название *телематических сетей* [5, 6], развитие которых связано с использованием стека протоколов TCP/IP.

Управление РО с использованием телематических сетей – перспективное направление научных исследований в области робототехники, которое активно развивается на протяжении последних лет [7]. Возможными приложениями такого типа управления в робототехнике являются дистанционное обучение, удаленное управление автоматизированными производствами, телемедицина, телеуправление РО в экстремальных средах.

Один из первых экспериментов в этой области осуществлен специалистами Центрального научно-исследовательского института робототехники и технической кибернетики (ЦНИИ РТК, Санкт-Петербург) в 1997 г. во время работы Международного аэрокосмического салона в Турине (Италия). Был проведен сеанс удаленного управления в реальном времени через сеть Интернет находившимся в ЦНИИ РТК космическим роботом-манипулятором многоразовой космической системы «Буря» [8].

Важной частью исследования систем удаленного управления РО стали космические эксперименты (КЭ) по проекту «ROKVISS», проводимые Институтом робототехники и мехатроники Немецкого аэрокосмического центра (DLR-RMC), г. Оберпфaffenхофен (Германия). Специали-

стами этой организации реализованы различные варианты дистанционного управления роботом «ROBOTIC», установленным на внешней поверхности Российского сегмента (РС) Международной космической станции (МКС), с Земли. Для этой цели использовался специальный выделенный радиоканал, работающий в диапазоне S-band и обеспечивающий дуплексную несимметричную передачу данных телеметрии и информации мультимедиа в режиме реального времени.

Настоящая статья посвящена космическим экспериментам серии «Контур» по удаленному управлению робототехническими объектами.

1. Цель, задачи и описание космического эксперимента «Контур»

Цель космического эксперимента (КЭ) «Контур», проведенного в 2009–2011 гг. ЦНИИ РТК совместно с DLR-RMC и Ракетно-космической корпорацией (РКК) «Энергия» имени С.П. Королёва, – апробация возможностей использования телекоммуникационных подходов при управлении сетевым РО на борту РС МКС с использованием задающего человеко-машинного интерфейса с силовой обратной связью [9]. Технологической базой КЭ «Контур» являются решения, использующие глобальную сеть Интернет для имитации статистических характеристик задержки и пропускной способности спутниковых каналов связи, используемых для передачи команд управления и дан-

ных телеметрии, формируемых в процессе функционирования РО, расположенного на внешней поверхности РС МКС.

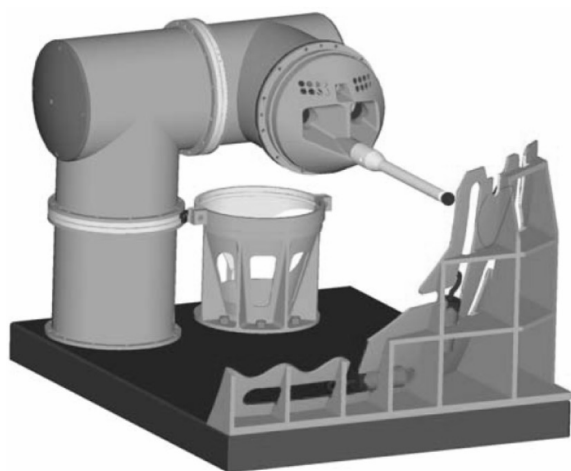
Управляемым объектом служил двух-шарнирный робот манипулятор «ROBOTIC» (рис. 1), разработанный DLR-RMC и установленный на поверхность РС МКС в 2004 г. в рамках КЭ «ROKVISS» [10–12]. Копия этого робота (робот-дублер, квалификационная модель) установлена на территории DLR-RMC. Технические характеристики робота «ROBOTIC» приведены в табл. 1.

Другой важной задачей КЭ являлось исследование принципов, методов и алгоритмов supervisory управления, при котором человек-оператор ставит РО локальные, простые задачи, которые он исполняет в автоматическом режиме, а оператор осуществляет визуальный или иной контроль выполнения операции.

Помимо задач, связанных с компенсацией нестационарных задержек в телематическом канале управления РО, решались задачи разработки протокола взаимодействия программно-математического обеспечения оператора с роботом, определения пропускной способности Интернет-канала, требуемой для команд управления и данных обратной связи, а также обеспечения информационной защищенности взаимодействующих сторон и безопасности передаваемых данных.

На рис. 2 представлена структура взаимодействия различных элементов системы в рамках КЭ «Контур».

а)



б)

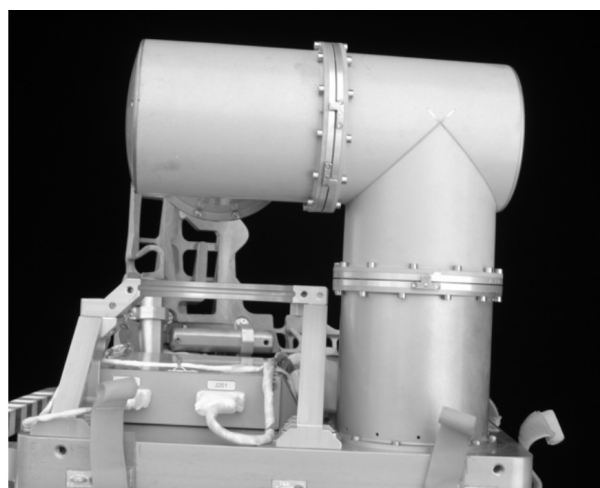


Рис. 1. Робот «ROBOTIC»:
а – компьютерная модель; б – «ROBOTIC» на внешней поверхности РС МКС

Таблица 1

Технические характеристики робота «ROBOTIC»

Наименование	Значение
Количество степеней свободы	2
Напряжение питания, В	28,5
Выходной момент для каждого шарнира, Нм, не более	40
Скорость перемещения, м/с, не более	0,3
Ускорение, м/с ² , не более	3
Управляющий интерфейс	SERCOS
Масса, кг, не более	7
Мощность, Вт, не более	34

Для реализации КЭ «Контур» и отработки технологии супервизорного управления удаленными РО с использованием открытых каналов сетей общего доступа применялось специализированное программно-математическое обеспечение (ПМО), состоящее из двух основных компонент: наземной и бортовой. Наземная компонента осуществляла задание операций роботу, а бортовая – планирование его траектории и обеспечение обратной связи при отработке координат робота в шарнирах (углов в шарнирах). В КЭ «Контур»

наземная компонента ПМО размещалась на компьютере в ЦНИИ РТК, бортовая – на борту РС МКС, в аппаратных компонентах манипулятора «ROBOTIC».

Для обеспечения безопасности и надежности технологический цикл проведения КЭ «Контур» включал в себя три этапа: математическое моделирование, макетирование на наземном роботе-дублере, расположенном в DLR-RMC, и выполнение реальных операций в космосе при помощи робота «ROBOTIC».

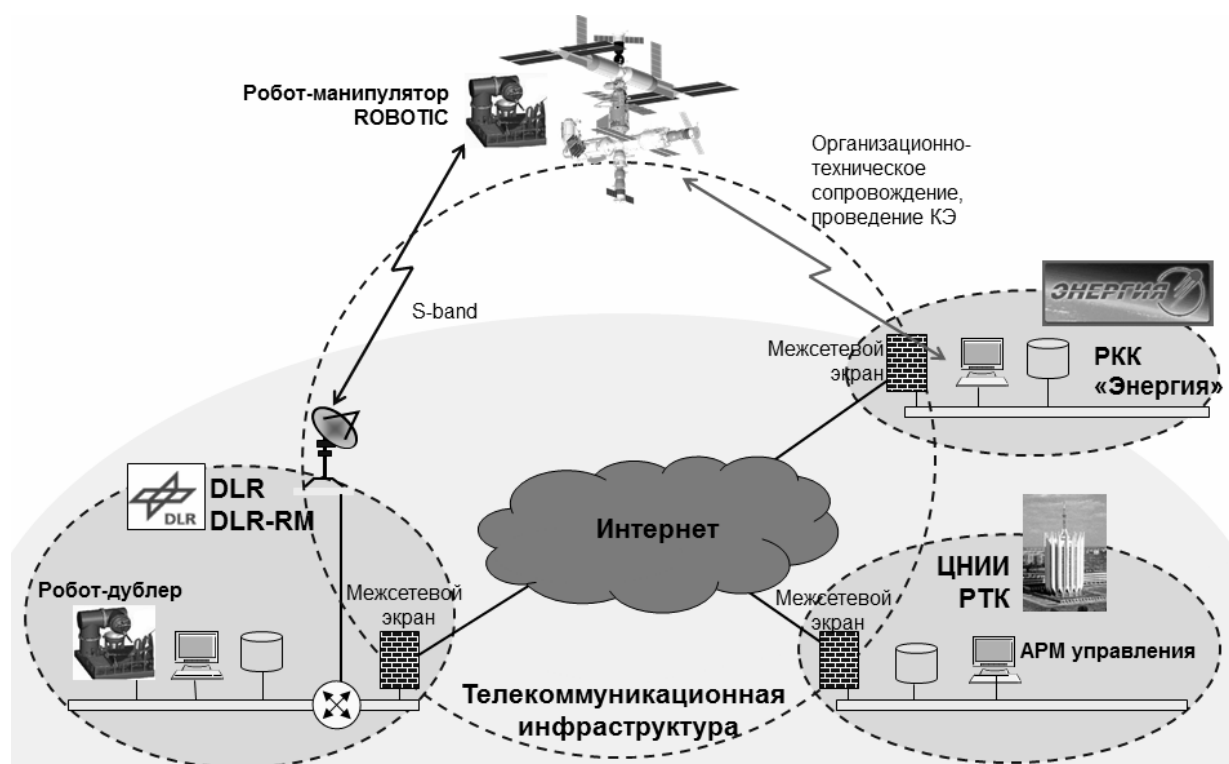


Рис. 2. Космический эксперимент «Контур» – структурные элементы системы и связи

Комплекс математических моделей состоял из кинематической и динамической модели, разработанных в DLR-RMC, и гибридной системы моделирования, разработанной в ЦНИИ РТК. Кинематическая модель DLR-RMC разработана в системе программирования «Pro Engineer» и визуально имитирует робот и его рабочую зону. Основными элементами модели являются плоское квадратное основание РО и «контур» в виде изрезанной каналами цилиндрической поверхности, внутри которого перемещается рабочий орган робота в виде штыря. Эта модель использовалась при предварительной отработке геометрии и кинематики операций в режиме моделирования, позволяя оценить отсутствие нежелательных столкновений и силовых воздействий. Динамическая модель DLR-RM разработана в системе Matlab-Simulink и учитывает динамику робота, в т. ч. регуляторы приводов.

Система математического моделирования манипулятора «ROBOTIC», созданная в ЦНИИ РТК, имеет гибридную структуру: динамическая модель манипулятора создана в системе программирования Matlab-Simulink, а программа управления манипулятором – на Delphi. Для организации совместного функционирования программ управления и моделирования организован циклический обмен данными между этими программами, включающий начальную инициализацию.

Полученные в результате моделирования дан-

ные [13] позволили оценить качество регуляторов при операциях свободного движения в пространстве и при силомоментном взаимодействии с объектом рабочей зоны «контур» при выполнении контактных и силомоментных операций.

2. Исследования каналов связи в КЭ «Контур»

За время реализации КЭ в 2009 г. выполнено 18 успешных сеансов связи с бортом МКС для проведения экспериментов и обслуживания программного и аппаратного обеспечения; из них 6 сеансов – для проведения экспериментов по управлению роботом из ЦНИИ РТК. В 2010 г. организовано 26 сеансов связи с бортом МКС, из которых признаны состоявшимися 23 сеанса и успешными – 21 сеанс. Из ЦНИИ РТК эксперименты по управлению роботом проводились шесть раз.

Для проведения КЭ «Контур» разработан и реализован протокол передачи команд управления, данных телеметрии и видеоданных, поступающих от расположенных на роботе видеокамер. Данный сеансовый протокол регламентирует следующий порядок обмена данными (рис. 3).

1. Перед началом обмена данными между компьютером управления и наземным компьютером формируются два виртуальных соединения по протоколу TCP: первое используется для передачи команд управления для РО, второе – для передачи оператору данных телеметрии.

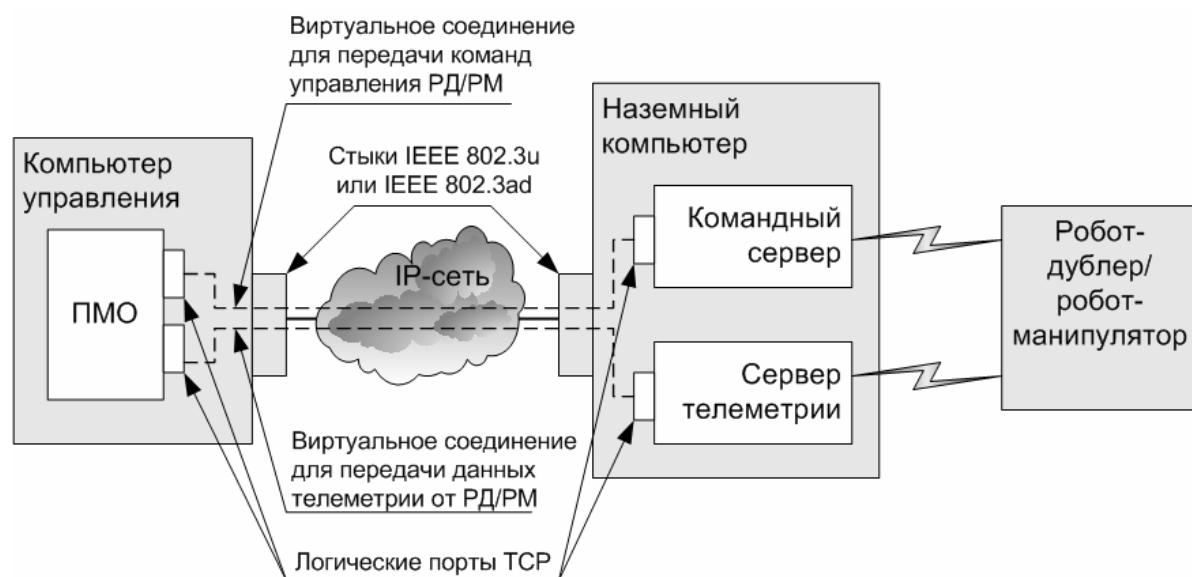


Рис. 3. Схема обмена данными между компьютером управления (КУ) и наземным компьютером (НК) в DLR-RM



2. При организации виртуальных соединений программно-математическое обеспечение, функционирующее на стороне DLR-RMC, выступает в качестве сервера, т. е. ожидает запросов на установление соединения со стороны программного обеспечения оператора на заранее оговоренные логические порты протокола TCP.

3. При организации виртуальных соединений программно-математическое обеспечение оператора (ПМО) выступает в качестве клиентов, т. е. инициирует организацию виртуального соединения с сервером. Данное требование продиктовано соображениями безопасности для предотвращения перехвата управления третьей стороной в процессе управления РО.

4. Корректный разрыв соединения происходит по инициативе клиентского программного обеспечения оператора.

В ходе эксперимента исследованы статистические характеристики телематических каналов связи, получены оценки пропускной способности и времени прохождения пакетов через виртуальные соединения, организованные между ПМО и РО в различных режимах управления. Графики вариаций значений пропускной способности виртуальных соединений и значений задержек для сеанса, проведенного 9 октября 2009 г., приведены на рис. 4. В табл. 2 приведены полученные данные о задержках в канале управления во время проведения сеансов.

Таблица 2

Задержки в канале управления

Дата	Среднее значение, мс	Среднеквадратичное отклонение, мс	Максимальное значение, мс	Минимальное значение, мс
08.07.2009	72,923	66,882	1218	46
10.07.2009	246,749	22,811	593	203
16.07.2009	75,203	51,011	718	46
23.07.2009	146,808	155,160	1969	46
29.07.2009	273,597	428,048	3875	46
07.10.2009	256,539	136,011	1812	46
09.10.2009	64,088	13,478	207	41
09.10.2009	257,962	92,841	938	46
02.12.2009	66,180	36,881	453	46
03.12.2009	109,363	138,485	2266	46
03.12.2009	280,379	80,309	938	62
04.12.2009	219,811	71,746	625	46
04.02.2010	112,33	47,063	890	46
08.02.2010	91,18	30,902	767	48
09.02.2010	74,50	17,013	508	48
24.03.2010	108,72	45,902	1093	46
24.03.2010	90,67	30,892	633	51
10.05.2010	210,48	110,110	1460	64
14.05.2010	150,89	89,903	865	46
19.08.2010	133,18	67,678	791	46
20.08.2010	95,01	40,568	508	46
06.09.2010	89,19	29,390	659	48
09.09.2010	130,90	75,093	1004	51
13.09.2010	125,15	53,095	1978	46
13.09.2010	99,07	20,90	703	46
14.09.2010	90,10	32,749	615	48
14.09.2010	167,67	56,049	1609	48

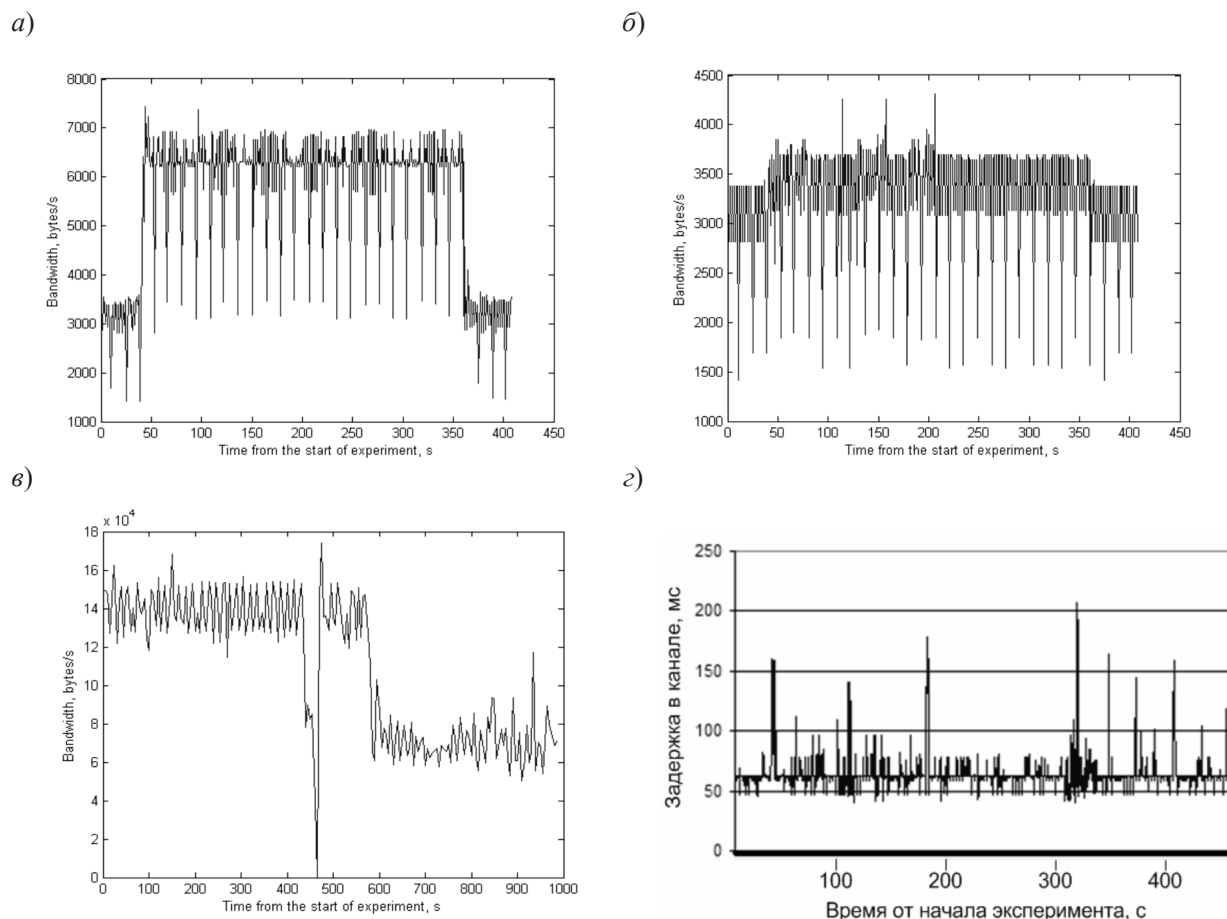


Рис. 4. Вариации значений полосы пропускания и задержки в канале управления от 9 октября 2009 г.:
а – канал передачи команд ЦНИИ РТК → DLR-RM, протокол TCP; б – данные телеметрии DLR-RM → ЦНИИ РТК, протокол TCP; в – видеоданные DLR-RM → ЦНИИ РТК, протокол UDP; з – зависимость значения задержек в канале управления от времени

Экспериментальные исследования, проведенные в рамках КЭ «Контур», показали, что при среднем значении времени задержки, не превышающем 120 мс, и значениях среднеквадратического отклонения, не превышающих 70 мс, возможно применение задающих манипуляторов с силовой обратной связью для удаленного телеуправления РО с использованием стандартных сетевых и транспортных протоколов стека TCP/IP. При увеличении среднего значения или среднеквадратического отклонения времени задержки передачи, для использования человеко-машинных интерфейсов с силовым осязанием, требуется применение специализированных механизмов приоритезации данных, протоколов и алгоритмов обработки сетевого трафика, обеспечивающих требуемый уровень качества передачи информации.

3. Демонтаж и исследование шарниров робота «ROBOTIC»

Существенный интерес представляли шарниры робота «ROBOTIC» как технического объекта, шесть лет функционирующего на внешней поверхности МКС. По этой причине в рамках завершающей стадии КЭ «Контур» были запланированы и проведены работы по демонтажу, возвращению на Землю и исследованию шарниров. Эти работы выполнялись совместно специалистами РКК «Энергия», DLR-RMC и ЦНИИ РТК.

2 ноября 2010 г. состоялся последний сеанс связи с бортом МКС в рамках КЭ «Контур», в ходе которого робот «ROBOTIC» перевели в «парковочное» положение, обеспечивающее минимальный риск повреждения для космонавтов при проведении демонтажа. В ходе внекорабель-



ной деятельности 15 ноября 2010 г. моноблок «ROBOTIC» демонтировали с внешней поверхности и поместили на временное хранение в гермоотсек станции, а 24 мая 2011 г. возвратили на Землю в спускаемом аппарате «Союз-ТМА-20».

С 25 сентября 2011 г. специалистами DLR-RMC и ЦНИИ РТК проводятся совместные работы по исследованию состояния возвращенных шарниров, включающие функциональные испытания и анализ основных составных частей шарниров. Цель функциональных тестов – исследование изменений параметров системы после спуска на Землю по сравнению с последними измерениями в открытом космосе. Анализ составных частей проводится для оценки надежности электронных и механических компонентов, а также пригодности использованных смазок.

Большинство тестов на момент написания статьи (01.04.2012) были завершены, окончательный отчет, содержащий результаты исследований шарниров, планируется подготовить в ближайшем будущем. Однако уже сейчас можно отметить значительный запас надежности шарниров в целом и отдельных его компонентов. Будучи установленными на испытательный стенд, возвращенные шарниры функционировали без каких-либо отклонений, как и находившаяся на Земле квалификационная модель (робот-дублер), а параметры практически всех электронных и механических компонентов находились в пределах номинальных значений. Таким образом, можно считать успешным завершение КЭ «Контур», результаты которого планируется использовать при разработке новых робототехнических изделий космического назначения и систем управления ими.

4. Развитие космического эксперимента «Контур»

В ходе выполнения КЭ «Контур» установлено, что важной составляющей эффективного телеуправления РО является высокая степень «погружения» оператора в среду функционирования робота [13]. При этом телеприсутствие оператора в такой среде обеспечивается путем вывода на рабочее место видеoinформации, поступающей с видеокамер робота, а также за счет моделирования осязательных способностей робота на основе задающих человеко-машинных интерфейсов с силомоментной обратной связью для воспроизведения значений моментов, полученных от

соответствующих датчиков. В связи с этим перспективным развитием КЭ «Контур» признано исследование влияния условий невесомости на возможность космонавтом-оператором удаленно управлять напланетным РО при помощи силомоментного задающего манипулятора.

Исследование и освоение околоземного космического пространства, планет Солнечной системы и дальнего космоса требуют разработки эффективных систем удаленного управления объектами различного назначения, учитывающих условия, в которых находится оператор в процессе управления. В связи с этим актуальной научно-технической проблемой нового КЭ «Контур-2» является создание методов и отработка технологий телеуправления напланетными ГО и их группировками с орбитальных космических аппаратов для решения задач исследования планет.

Возможность эффективного управления роботами, находящимися на поверхности планет, с борта пилотируемого орбитального космического аппарата определяется следующими основными факторами:

- надежностью телекоммуникационных каналов, используемых для передачи команд управления, данных телеметрии и видеoinформации в системе «оператор – объект управления», их способностью к реконфигурации и масштабируемости;
- степенью реалистичности, с которой для оператора в условиях невесомости воспроизводятся физические свойства среды функционирования робота;
- адекватностью реакций оператора в условиях невесомости на воздействия со стороны задающего манипулятора с силомоментным оцувствлением с учетом временных задержек и дискретности видеoinформации обратной связи, поступающей от объекта управления.

Одним из возможных сценариев освоения планет Солнечной системы является управление РО с борта орбитального пилотируемого космического аппарата на поверхности планеты путем формирования и последующего использования телекоммуникационной инфраструктуры на основе группировки мобильных и/или стационарных РО. При этом связь объекта управления на поверхности планеты с космическим аппаратом обеспечивается за счет использования каналов космической связи и наземной телематической сети (рис. 5). Такая инфраструктура передачи

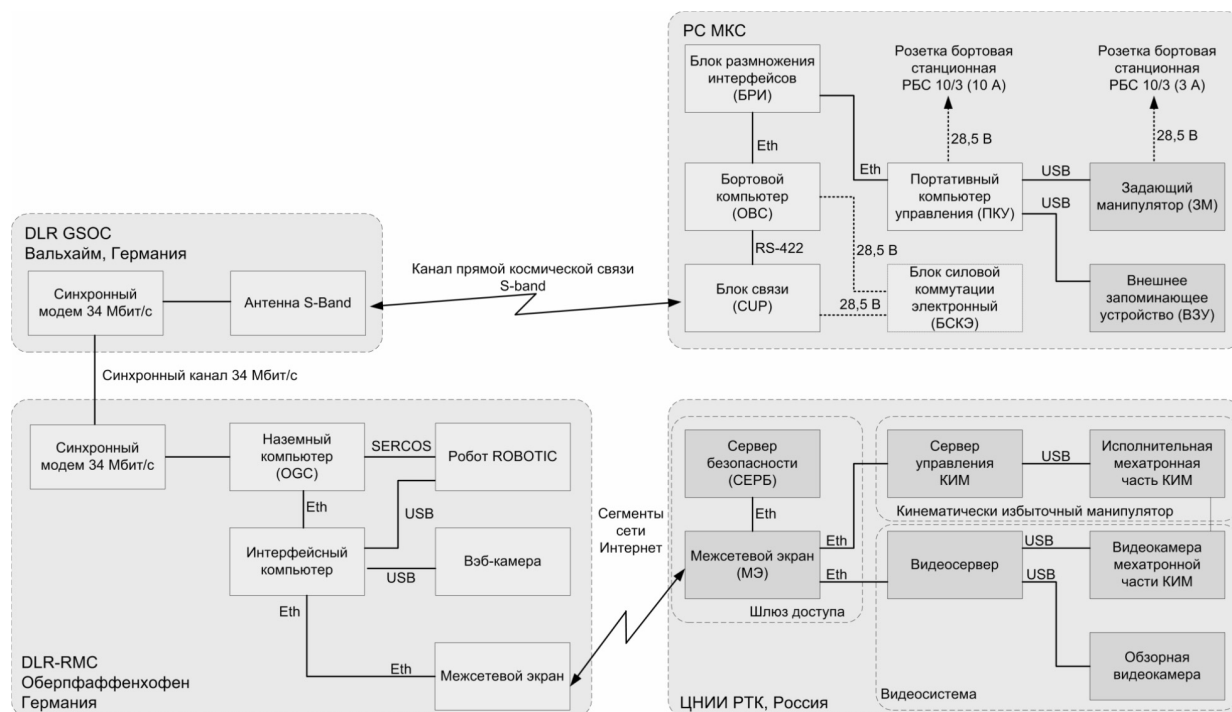


Рис. 5. Структурная схема системы телеуправления в КЭ «Контур-2»

данных характеризуется рядом параметров (пропускная способность, задержки и др.), значения которых могут существенно варьироваться во время сеанса управления. Таким образом, актуальной задачей КЭ «Контур-2» является исследование влияния нестационарных характеристик телекоммуникационных каналов на эффективность телеуправления напланетным РО с борта РС МКС, а также разработка методов адаптации используемых сетевых ресурсов к таким параметрам на основе управления пропускной способностью и приоритетами виртуальных соединений.

Успешное решение этих научно-технических задач позволит отработать технологии и создать новые средства, предназначенные для исследования планет Солнечной системы с помощью планетарных роботов, управляемых с борта пилотируемых орбитальных космических аппаратов.

КЭ «Контур-2» представляет собой моделирование одного из возможных сценариев исследования планет (например, Марса) при помощи находящихся на их поверхности роботов-манипуляторов, управляемых человеком-оператором с орбитального космического корабля. В рамках КЭ Земля является исследуемой планетой, МКС представляет собой орбитальный пилотируемый

корабль, космонавт на МКС является человеком-оператором, который управляет при помощи задающего манипулятора с силовой обратной связью, находящимися на поверхности исследуемой планеты РО или группировкой РО, связанных между собой телекоммуникационными каналами связи. На рис. 5 представлена структурная схема системы телеуправления в КЭ «Контур-2».

Объектом исследований КЭ «Контур-2» является система телеуправления, изображенная на рис. 5, которая позволяет осуществить управление с борта МКС РО, находящимися на поверхности Земли. В качестве объектов телеуправления выступают:

реконфигурируемый змеевидный робот на основе унифицированных двухступенных мехатронных модулей производства ЦНИИ РТК;

двухшарнирный робот-манипулятор со встроенными сило-моментными датчиками производства DLR-RMC (робот-дублер в КЭ «Контур»).

В рамках КЭ «Контур-2» предлагается решить следующие задачи:

- разработать систему устойчивого телеуправления роботами с обратной связью по силе/моменту в условиях ограничений коммуникационных каналов;
- разработать программно-математическое



обеспечение для телеуправления роботами с борта РС МКС;

- разработать задающий манипулятор с силовой моментной обратной связью для использования в составе рабочего места оператора, обеспечивающего телеуправление с борта МКС;
- обеспечить выполнение сеансов КЭ по удаленному управлению с борта РС МКС роботами, находящимися в ЦНИИ РТК и DLR-RM;
- провести анализ результатов КЭ по телеуправлению в режиме сило-моментного и позиционно-скоростного управления роботами на Земле.

В ходе эксперимента планируется регистрация информации двух типов: визуальной и телеметрической. Качество регистрируемой информации определяется ограничениями канала космической связи S-band: пропускная способность восходящего потока данных (с Земли на МКС) – 256 Кбит/с, пропускная способность нисходящего потока данных (с МКС на Землю) – 4 Мбит/с. Для передачи команд управления роботами необходимо использовать нисходящий поток данных, для передачи данных телеметрии и визуальной информации – восходящий поток данных. При этом для передачи данных телеметрии планируется использовать часть полосы пропускания восходящего потока шириной 128 Кбит/с, для передачи визуальной информации – часть полосы пропускания восходящего потока шириной 128 Кбит/с.

Проведенные эксперименты по удаленному управлению РО на РС МКС с использованием наземных сегментов глобальной сети открытого доступа и канала прямой космической связи S-band DLR-RM являются уникальными в мировом масштабе. Использование интерактивных телекомму-

никационных взаимодействий непосредственно с бортом МКС без использования ретрансляторов, расположенных на геостационарной орбите, позволило существенно расширить спектр выполняемых космических экспериментов, связанных с необходимостью передачи больших объемов обрабатываемых данных при малых задержках (до 60–100 мс). К числу таких экспериментов можно отнести наблюдение за наземными объектами в реальном масштабе времени, супервизорное управление робототехническими объектами как с Земли, так и с борта МКС, мониторинг состояния МКС и др.

Из результатов исследований следует, что супервизорное управление РО, используемое в КЭ «Контур», может успешно применяться для задач удаленного управления объектами в космосе с использованием стандартных телеметрических методов и технологий передачи и обработки данных для значений времени задержки, не превышающих 120 мс. При небольшом значении задержки могут успешно компенсироваться применением специализированных технологий управления и передачи данных.

Отработанные в ходе КЭ «Контур» технологии обработки и передачи данных и супервизорного управления могут успешно применяться в будущих космических российских робототехнических системах.

Результатом текущего космического эксперимента «Контур-2» должна стать отработка технологии удаленного управления с борта РС МКС робототехническими объектами на поверхности Земли с использованием телекоммуникационных каналов связи и средств цифровой обработки информации.

Работа проведена при финансовой поддержке Министерства образования и науки РФ.

СПИСОК ЛИТЕРАТУРЫ

1. **Юревич, Е.И.** Космическая робототехника: состояние и перспективы развития [Текст] / Е.И. Юревич // Rational Enterprise Management. –04.2011. –С. 24–26.
2. **Минаков, Е.П.** Концепция развития робототехнических систем в интересах пилотируемой космонавтики, исследования Луны и планет Солнечной системы. Экстремальная робототехника. Нано- микро- и макророботы (ЭР-2009) [Текст] / Е.П. Минаков, В.А. Лопота, Е.И. Юревич, А.С. Кондратьев // Матер. XX Междунар. науч.-технич. конф. –Таганрог: Изд-во ТТИ ЮФУ, 2009. –С. 103–106.
3. **Вишневецкий, В.М.** Теоретические основы про-

ектирования компьютерных сетей [Текст] / В.М. Вишневецкий. –М.: Техносфера, 2003. –512 с.

4. **Городецкий, А.Я.** Информатика: фрактальные процессы в компьютерных сетях [Текст] / А.Я. Городецкий, В.С. Заборовский. –СПб.: Изд-во СПбГТУ, 2000. –102 с.

5. **Заборовский, В.С.** Сети ЭВМ и телекоммуникации. Моделирование и анализ компьютерных сетей: телематический подход: Учеб. пособие [Текст] / В.С. Заборовский, В.А. Мулюха, Ю.Е. Подгурский. –СПб.: Изд-во Политехнического ун-та, 2010. –92 с.

6. **Zaborovsky, V.** Active Queuing Management for

Telematics Space Network Robotics System (Управление телематическими робототехническими системами космического назначения на основе приоритетной обработки сетевого трафика) [Текст] / V. Zaborovsky, O. Zayats, V. Mulukha // Тр. XXI Междунар. науч.-технич. конф. Экстремальная робототехника. –СПб.: Изд-во «Политехника-сервис», 2010. –С. 340–349.

7. **Белоусов, И.Р.** Алгоритмы управления роботоманипулятором через Интернет [Текст] / И.Р. Белоусов // Математическое моделирование. –2002. – Т. 14. –№ 8. –С. 10–15.

8. **Заборовский, В.С.** Международный АТМ-проект с участием России [Текст] / В.С. Заборовский, Ю.Е. Подгурский, В.Б. Семеновский // Глобальные сети и телекоммуникации. –1998. –№10. –С. 56–58.

9. **Ландцеттель, К.** Эксперимент «Контур» на Российском сегменте МКС [Электронный ресурс] / К. Ландцеттель, В.С. Заборовский, А.С. Кондратьев, А.В. Силиненко, М.Ю. Беляев, Е.В. Бабкин // Научные чтения памяти К.Э. Циолковского, г. Калуга. –16.09.2009. –Режим доступа: <http://readings.gmik.ru/lecture/2009-EKSPERIMENT-KONTUR-NA-ROSSIYSKOM-SEGMENTE-MKS>

10. **Schäfer, B.** ROKVISS – Space Robotics Dynamics and Control Performance Experiments [Электрон-

ный ресурс] / B. Schäfer, B. Rebele, K. Landzettel // At the ISS. Proc. ACA2004 IFAC Symp. on Automatic Control in Aerospace. –St. Petersburg, 13-18 June 2004. –P. 333–338.

11. **Hirzinger, G.** ROKVISS – Robotics component verification [Электронный ресурс] / G. Hirzinger, K. Landzettel, D. Reintsema [et al.] // On iss. Proceedings of the 8th International symp. on Artificial Intelligence, Robotics and Automation in Space, 5-8 Sept/ 2005, Munich, Germany. –European Space Agency (ESA) SP-603 Aug. 2005. Session 2b – Orbital Missions (03). –11 p.

12. **Albu-Schaffer, A.** ROKVISS - Robotics component verification on ISS current experimental results on parameter identification [Электронный ресурс] / A. Albu-Schaffer, W. Bertleff, B. Rebele [et al.] // IEEE International Conf. Robotics and Automation. –2006. –P. 3879–3885.

13. **Бабкин, Е.В.** Супервизорное управление космическим роботом на Международной космической станции (МКС) с использованием сети Интернет [Текст] / Е.В. Бабкин, М.Ю. Беляев, И.А. Васильев, В.С. Заборовский, В.П. Макарычев, А.В. Силиненко, К. Ландцеттель // Матер. Всерос. молодежной конф. Экстремальная робототехника. –2011. –С. 32–42.

УДК 621.39;004.7

М.В. Гаазе, Н.И. Лычагин, А.А. Малаховский

ИНТЕГРИРОВАННЫЕ ТЕЛЕКОММУНИКАЦИОННЫЕ СЛУЖБЫ ДЛЯ ВЕДОМСТВЕННЫХ СЕТЕЙ СВЯЗИ

Высокие темпы развития информационных технологий создают объективные предпосылки для дальнейшего совершенствования телекоммуникационных услуг в автоматизированных системах управления (АСУ), создаваемых в интересах различных ведомств. Характерным примером интегрированной услуги связи является видеоконференцсвязь (ВКС). Ситуационные центры верхнего уровня иерархии в АСУ многих ведомств к настоящему времени оснащены зарубежным оборудованием ВКС производства ведущих мировых фирм. Высокая стоимость зарубежного оборудования ВКС, отсутствие открытого программного кода, разнотипность установленных средств существенно затрудняют оснащение средних и нижних уровней АСУ требуемыми терминальными

и серверными средствами. Комплексное оснащение различных уровней ведомственных АСУ оборудованием, предоставляющим современные интегрированные телекоммуникационные услуги, целесообразно проводить на основе отечественных разработок.

В статье рассматривается методология создания и реализация отечественных комплексов предоставления современных интегрированных телекоммуникационных услуг с учетом особенностей функционирования ведомственных АСУ на примере реализации услуг ВКС.

Основополагающие принципы

Интегрированные телекоммуникационные службы подразумевают предоставление ряда те-



лекоммуникационных услуг, таких, как аудио- и видеосвязь, аудио- и видеоконференцсвязь, передача файлов, передача текстовых сообщений, индикация присутствия и др. При этом в максимальной степени должна обеспечиваться унификация аппаратного и программного обеспечения (ПО), а также основных системотехнических решений, включая:

- единые механизмы управления сеансами;
- единые технологии сетевого взаимодействия;
- единые подходы по управлению и обслуживанию комплексов связи.

В методологию проектирования интегрированных служб положены следующие принципы:

- 1) открытая архитектура и принципы построения в строгом соответствии с перспективными международными стандартами и рекомендациями;
- 2) единая технологическая основа для реализации всего спектра телекоммуникационных услуг;
- 3) иерархический модульный принцип создания переносимого кроссплатформенного ПО, независимого от аппаратной платформы и операционной среды;
- 4) распределенная архитектура для обеспечения надежности функционирования. Центральное серверное оборудование служит только для расширения функциональности предоставляемых услуг связи.

Методология рассматривается на одном виде связи – ВКС, которая является наиболее сложной как с точки зрения реализации алгоритмов и методов, используемых при создании ПО, так и в части реализации аппаратных решений.

Выбор протокольного стека и нормативная база

В настоящее время основными концепциями развития телекоммуникационных сетей являются NGN (Next-Generation Network) и IMS (IP Multimedia Sub-system), в рамках которых реализация телекоммуникационных услуг базируется на единой технологии управления сеансами – протоколе SIP (Session Initiation Protocol). Большинство создаваемых в последнее время систем телефонии и видеотелефонии являются SIP-ориентированными. В унаследованных системах аудио- и видеоконференцсвязи используется семейство протоколов H.323. Типовой режим работы таких конференций – централизованный. Передача аудио-, видео-

информации и данных управления производится непосредственно между каждым терминалом и сервером управления конференцией (MCU), что предъявляет повышенные требования к производительности и надежности MCU.

Децентрализованный режим, реализованный во многих системах ВКС на базе H.323 посредством механизмов многоадресной рассылки, не обеспечивает возможности построения полностью распределенной системы мультимедийной связи. Несмотря на то, что в режиме многоадресной рассылки не требуется центральный сервер MCU, в каждом сегменте сети остаются устройства, выполняющие централизованные функции, – это коммутаторы и маршрутизаторы, обеспечивающие регистрацию и функционирование групп многоадресной рассылки. Данный режим обеспечивает эффективное мультиплексирование аудио- и видеотрафика без использования централизованного сервера MCU, однако обладает следующими недостатками:

- отсутствует возможность независимого управления потоками аудио- и видеоинформации;

- отсутствует возможность организации обратной связи по каждому потоку, необходимой для обеспечения качества обслуживания;

- отсутствует возможность создания меток для трафика в направлении различных абонентов с целью обеспечения требований безопасности;

- ограничены возможности по управлению услугой вследствие того, что управление многоадресной рассылкой производится не на уровне управления услугами, а на уровне сетевых элементов (коммутаторы, маршрутизаторы);

- показатели надежности системы практически не отличаются от систем с централизованным сервером.

Достоинствами SIP являются четкое отделение процессов обработки сигнализации от процессов обработки медиаданных, а также разделение системы на функциональные элементы, такие, как фокус конференции, фабрика фокусов, служба уведомления, служба политики, медиасервер и т. д. [1, 2]. При этом обеспечивается расширение функциональных возможностей систем конференцсвязи на базе SIP по сравнению с H.323-системами.

Стек протоколов SIP обеспечивает также органичную и естественную поддержку возможности организации распределенных конференций с полной децентрализацией передачи и обработки

медиаданных. Данный режим недавно получил наименование «simulcast» (режим одновременной рассылки) и в настоящее время рассматривается для применения в системах интерактивного телевидения.

Многоточечные сеансы ВКС в режиме одновременной рассылки не могут реализовываться с использованием стандартных протоколов H.323. Для этого могут применяться SIP и SDP, которые в рамках базовых стандартов поддерживают организацию и управление сеансами многоточечной ВКС в данном режиме.

Использование режима одновременной рассылки позволяет избавиться от дорогого и сложного с технической точки зрения устройства (медиасервер) и повысить надежность системы в целом за счет устранения единой точки отказа и увеличения топологической надежности системы. Однако при этом повышаются требования по производительности терминального оборудования и пропускной способности каналов связи. Оба фактора в настоящее время перестают быть препятствием к динамичному развитию распределенных технологий в связи с постоянным увеличением возможностей базовых средств вычислительной техники и увеличением пропускной способности каналов. Протокол SIP обеспечивает также полнофункциональную поддержку каскадных конференций [2]. Причем каскадирование может осуществляться как в части обработки медиа данных, так и в части управления сессиями конференций. Возможны любые сочетания этих режимов в зависимости от характеристик сети связи и территориально-организационных аспектов создаваемой системы. Совместное использование технологий децентрализации и каскадирования обеспечивает динамическое управление медиапотоками в сессии конференцсвязи, позволяет в наибольшей степени оптимизировать передачу медиаданных по сетям с любой топологией и эффективно использовать ресурсы сети, неоднородной в части пропускных способностей каналов. При этом медиасервер может выполнять различные группы функций, начиная от микширования медиаданных, получаемых от различных источников, и заканчивая процедурами распределения медиапотоков в сети без транскодирования и промежуточной обработки.

Исходя из представленных выше преимуществ, в качестве протокола сигнализации для программной реализации услуги видеоконфе-

ренцсвязи выбран протокол SIP. Для организации и управления сессиями конференцсвязи использован ряд рекомендаций организации IETF, которые можно разбить на следующие области: архитектура конференцсвязи: RFC 4245, RFC 4353; управление вызовом третьей стороной: RFC 3725, RFC 4579; управление выступлениями: RFC 4376, RFC 4582, RFC 4583; система уведомлений: RFC 3265, RFC 4575.

Выбор протоколов для передачи по сети аудио- и видеоинформации достаточно тривиален, протоколы RTP (Real-time Transport Protocol) и RTCP (RTP Control Protocol) являются общепринятыми международными стандартами для решения этой задачи в IP-сетях с любыми видами сигнализации. Форматы и структуры сообщений протоколов RTP и RTCP, а также порядок их следования и обработки представлены в рекомендации RFC 3550 [3].

В рамках международных стандартов услуги передачи мгновенных сообщений и индикации статусов пользователей в сети (услуга «Присутствие») с использованием протокола SIP увязаны в единую систему SIMPLE (SIP for Instant Messaging and Presence Leveraging Extensions), обеспечивающую наибольшие функциональные возможности по сравнению с системами, базирующимися на других протоколах (H.323, XMPP и т. д.).

Кроме стандартной статусной информации PIDF (Presence Information Data Format) с использованием SIP может передаваться дополнительная информация, включающая расширения RPID (Rich Presence Extensions to the Presence Information Data Format), CIPID (Contact Information for the Presence Information Data Format), GEOPRIV. RPID позволяет индизировать текущее состояние и возможности оконечных устройств, а также обеспечивает привязку статуса к стандартным адресным книгам, календарям и плановикам. CIPID описывает расширения PIDF в части представления абонента, позволяющего воспроизводить графическую и аудиоинформацию, связанную с индизируемым пользователем. GEOPRIV позволяет описывать географическое местоположение с использованием языка GML (Geography Markup Language).

Доставка текстовых сообщений с использованием протокола SIP может выполняться в двух режимах: пейджинговом и сессионном. Первый режим описан в документе RFC 3428 [4]. Сессионный режим обеспечивает гарантированную до-

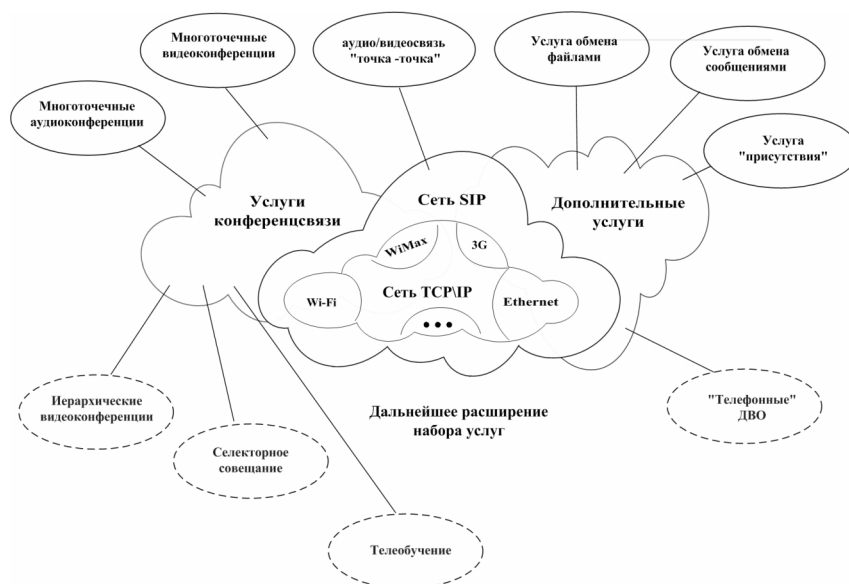


Рис. 1. Возможности платформы интегрированных телекоммуникационных услуг

ставку сообщений во времени, близком к реальному, и реализуется при помощи протокола MSRP, специфицированного в рекомендации RFC 4975 [5]. Для передачи файлов целесообразно использовать только сессионный режим, который реализуется с использованием протокола MSRP.

Таким образом, анализ нормативной базы показывает, что широкий спектр телекоммуникационных услуг, связанных с предоставлением возможности обмена мультимедийной информацией, включая аудио- и видеоданные, может быть реализован на базе единого сигнализационного протокола в соответствии с международными стандартами и рекомендациями (рис. 1).

Функциональная модель предметной области

Цель создания функциональной архитектуры системы в любой предметной области – декомпозиция ее общей функциональности на ряд функциональных модулей, каждый из которых реализует группу тесно взаимосвязанных функций. Предлагаемая функциональная архитектура позволяет минимизировать и четко специфицировать связи между отдельными функциональными модулями системы, реализующей предоставление телекоммуникационных услуг на базе протокола сигнализации SIP. Это обеспечивает возможность независимой проработки различных аспектов реализации каждого из функциональных модулей и в будущем возможность замены одного из них без существенного влияния на остальные.

Общую функциональность системы представляется целесообразным разделить на функциональные модули и функциональные плоскости. Такое разделение позволяет представлять процессы, проходящие в рамках системы, в виде взаимодействия функциональных модулей. При этом взаимодействие модулей группируется по функциональным областям на функциональных плоскостях.

При разработке функциональной архитектуры были выделены функциональные модули и плоскости, представленные на рис. 2. Рисунок также отражает распределение функциональных модулей между функциональными плоскостями. Наиболее функционально насыщенными являются модули терминал и пограничный контроллер сессий, которые участвуют в работе всех представленных функциональных плоскостей. Среди функциональных плоскостей наиболее насыщенной является плоскость сигнализации, которая объединяет функции всех без исключения функциональных модулей.

Плоскость сигнализации описывает процессы взаимодействия функциональных модулей в части управления вызовами (рис. 3). Сеть сигнализации должна обеспечивать управление предоставлением базовой услуги аудио- и видеосвязи, а также реализовывать управление дополнительными услугами.

Ведомственные сети связи целесообразно разделять на домены, каждый из которых мо-



Рис. 2. Распределение функциональных модулей по функциональным плоскостям

жет иметь независимое управление и администрирование. В этом случае в рамках плоскости сигнализации необходимо выделить два вида взаимодействия функциональных модулей: внутридоменное и междоменное. Внутридоменное взаимодействие обеспечивается контроллером сигнализации. Междоменное взаимодействие осуществляется при помощи пограничного контроллера сессий.

За управление вызовами в рамках сессии конференцсвязи отвечают терминал, сервер управления сессиями конференцсвязи и фокус конференцсвязи. Терминал является окончательным узлом сети сигнализации и выполняет функции подключения и отключения от сессии наряду с обработкой управляющей информации о медиапотоках участника в сессии, а также обработкой дополнительной информации о сессии, включая список участников.

Сервер управления сессиями конференцсвязи отвечает за запуск и завершение работы фокусов конференцсвязи. Последние отвечают за управление медиапотоками в сессиях. Возможны два режима по управлению фокусом конференцсвязи: централизованный и децентрализованный.

В централизованном режиме за запуск и завершение сессий отвечает сервер управления сессиями, который производит запуск и завершение фокуса конференцсвязи. В децентрализованном режиме фокус конференцсвязи является частью терминала и управляется пользователем.

Наличие дополнительных услуг в сети обеспечивается функциональными модулями сервера присутствия и сервера передачи файлов и мгновенных сообщений.

Аналогичным образом можно рассмотреть взаимодействие функциональных модулей в других плоскостях.

Анализ представленной функциональной модели показывает, что подавляющее большинство связей между функциональными модулями основано на использовании сетевых протоколов (SIP, RTP, MSRP и т. д.). Это означает, что физические реализации этих модулей могут быть при необходимости распределены между различными компьютерами сети. Исключение составляет только фокус конференцсвязи, который должен быть встроен в сервер управления сессиями или в терминал.

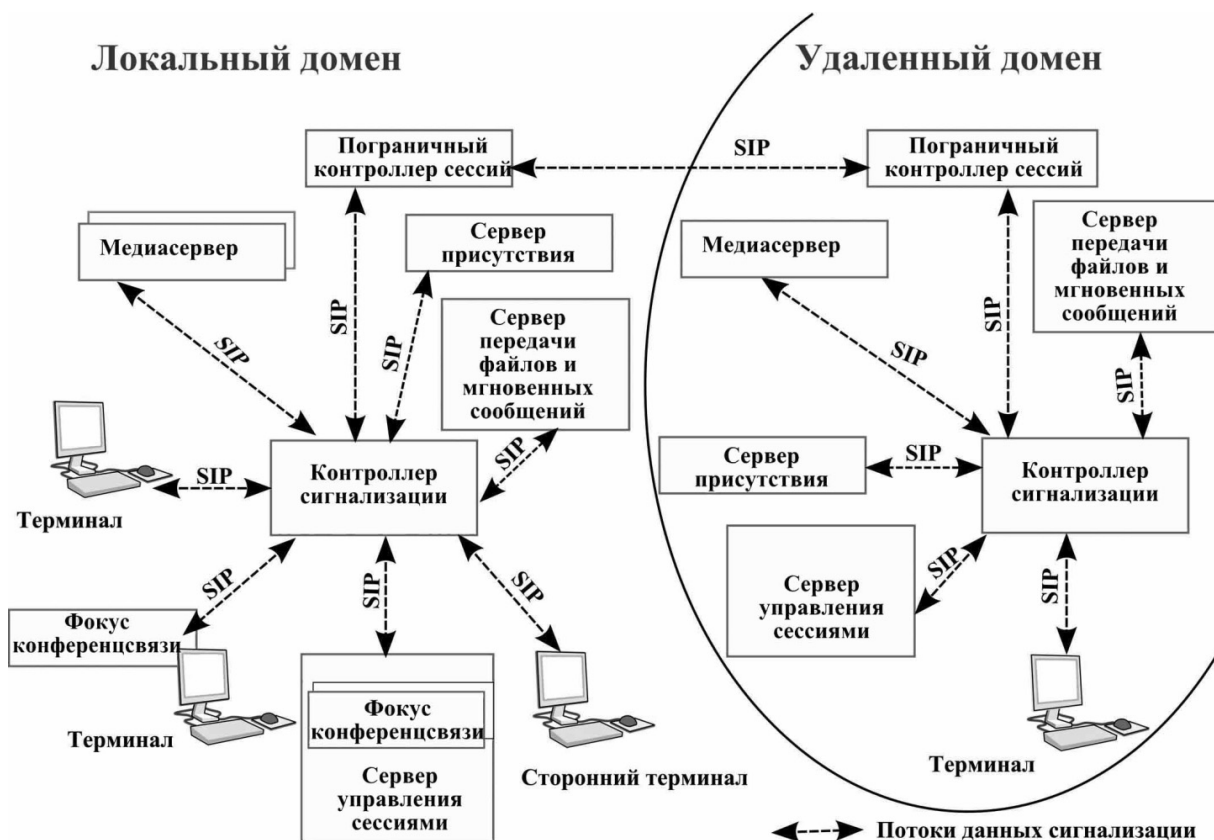


Рис. 3. Взаимодействие функциональных модулей в плоскости сигнализации

Предлагаемая функциональная модель ориентирована в первую очередь на создание перспективных распределенных телекоммуникационных служб. При этом она допускает различные степени децентрализации, которые условно можно разделить на четыре уровня:

- полностью распределенная система, серверное оборудование любой функциональности отсутствует в ее составе, терминалы могут самостоятельно организовывать многоточечные сессии конференцсвязи, при этом один из них управляет сессией, медиаданные передаются по принципу «от каждого к каждому»;
- централизация по обработке сигнализационных сообщений, в системе появляется сервер сигнализации (SIP-прокси), что обеспечивает поддержку функций авторизации, аутентификации, мобильности пользователей, а также возможность работы с символическими адресами;
- централизация по управлению сессиями конференцсвязи, система наращивается сервером конференции, что обеспечивает возможность ведения централизованных адресных книг и сце-

нариев, позволяет запускать сессии конференцсвязи в автоматическом режиме по расписанию, позволяет централизованно осуществлять сбор статистики (предбиллинг) по сессиям конференцсвязи;

- централизация по обработке медиаданных, система дополнительно наращивается медиа сервером, что позволяет понизить требования к производительности терминального оборудования и пропускной способности каналов связи, а также в некоторых случаях более гибко учитывать топологические особенности сети связи.

Методология разработки специального ПО

При реализации интегрированных телекоммуникационных служб для ведомственных сетей связи наибольшая трудоемкость приходится на разработку специального ПО. При этом большое значение имеет правильно выбранная архитектура программных средств и организация процесса разработки.

Если модульные принципы построения аппаратно-программных средств хорошо извест-

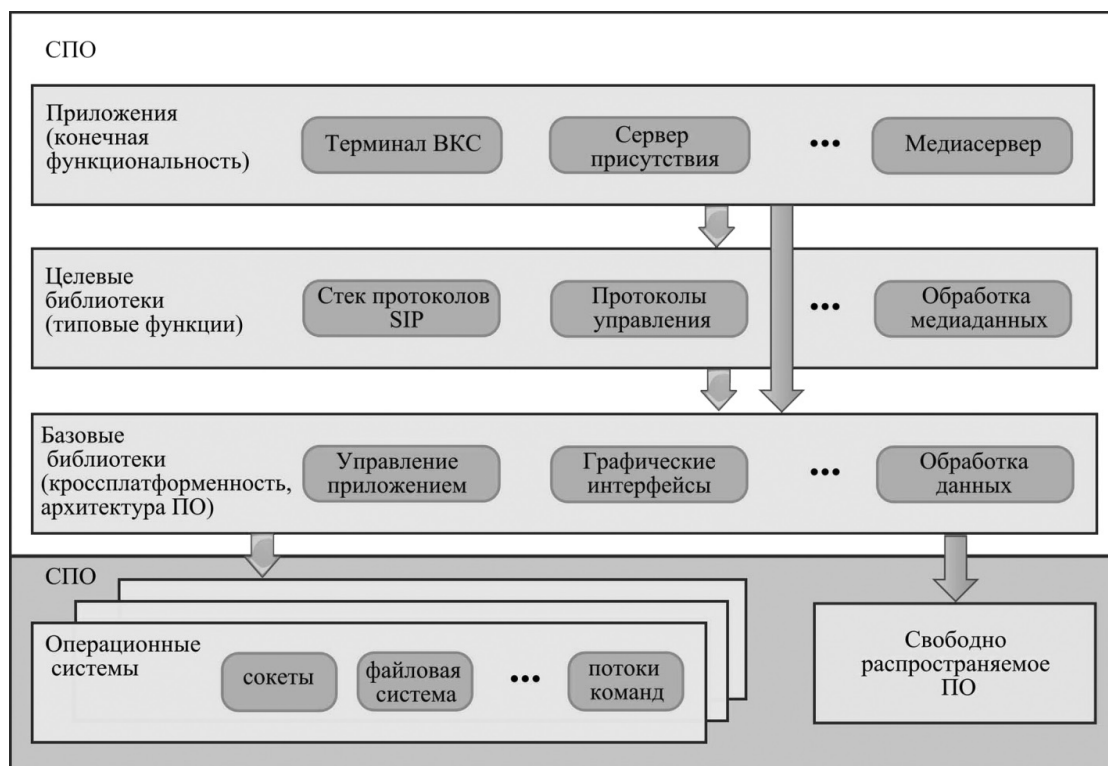


Рис. 4. Структура модульного специального программного обеспечения

ны и проработаны на уровне промышленных стандартов, то для ПО на сегодняшний день такие стандарты практически отсутствуют. Существуют лишь отдельные частные стандарты на построение программных интерфейсов (application programming interface, API) в рамках конкретных операционных систем. Стандарты на кроссплатформенные API для построения широкого спектра приложений на сегодняшний день существуют лишь в рамках конкретных разработок. Кроссплатформенность является одним из основных требований при создании ПО для ведомственных сетей связи, это связано в первую очередь с тем, что различные ведомства сертифицируют самые разные программные платформы для использования в своих автоматизированных системах управления. Более того, для организации межведомственных взаимодействий требуется возможность построения гетерогенных программных систем.

При разработке ПО, реализующего интегрированные телекоммуникационные службы, изначально заложены модульные подходы к проектированию. При этом создана иерархия модулей различной степени интеграции, на нижнем уровне которой находятся базовые программные библи-

отеки, а на верхнем – приложения, реализующие конечную функциональность телекоммуникационных служб. Модули верхнего уровня для обеспечения своей работоспособности используют функциональность модулей нижних уровней.

Структура программного комплекса разработана в соответствии с функциональной моделью, при этом, как правило, функциональные плоскости реализуются набором модулей программных библиотек, а функциональные модули реализуются использующими функциональность этих библиотек приложениями.

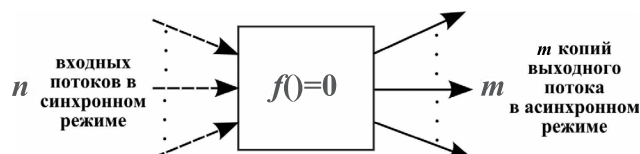
При создании иерархии модулей с учетом требований кроссплатформенного функционирования реализованы следующие основные принципы:

только нижний слой базовых библиотек может зависеть от используемых компонентов общего ПО;

вся основная функциональность должна быть сосредоточена в модулях специализированных библиотек.

Структура ПО, реализованного с учетом этих принципов, представлена на рис. 4.

Соблюдение первого принципа обеспечива-



- Защита для работы в многопоточном приложении
- Оптимизация по производительности за счет исключения копирования данных

Рис. 5. Базовый класс модулей обработки медиаданных

ет возможность сравнительно несложной портации всего программного комплекса на новые аппаратно-программные платформы, включая адаптацию к различным операционным системам. В любом случае потребуются модификация только небольшого слоя базовых библиотек, все остальные модули ПО даже не почувствуют смены платформы. Очевидно, что внешние API базовых библиотек при этом должны строиться с учетом возможности кроссплатформенного функционирования и также не зависеть от смены платформы.

Используемое из открытых источников ПО также представляется целесообразным скрывать за прослойкой базовых библиотек, чтобы минимизировать риски зависимости от внешних обстоятельств и иметь возможность в перспективе безболезненно отказываться от использования отдельных внешних программных продуктов.

Соблюдение второго принципа позволяет в максимальной степени реализовать многократное использование одного и того же кода при реализации конечных приложений. Во многих случаях при таком подходе целевые приложения оказываются практически вырождены до нескольких сотен строк программного кода, обеспечивающего необходимую инициализацию и контроль за выполнением приложения.

При этом значительно возрастает надежность и производительность функционирования СПО за счет возможности концентрации критических частей кода в компактных программных модулях, многократно используемых модулями верхнего уровня иерархии.

Выполнение представленных выше принципов создает определенные трудности при разработке структуры ПО, однако с лихвой окупается

при дальнейшем его развитии и сопровождении.

Рассмотрим реализацию модульных подходов и использование программных модулей для построения приложений на примере решения задач ввода/вывода, обработки и передачи по сети потоков медиаданных, содержащих аудио- или видеoinформацию.

Базовый класс, условно представленный на рис. 5, определяет правила построения программных интерфейсов и минимальную общую функциональность всех модулей этого типа, которые будут являться производными от него. Требования к организации входных и выходных потоков данных являются результатом анализа процессов в функциональной плоскости передачи медиапотоков реального времени.

Функциональность специализированных библиотечных модулей реализуется на этапе создания и компиляции производных классов, представленных на рис. 6. Их состав является результатом анализа общей функциональности плоскости передачи медиапотоков реального времени. Важно, что все программные модули обработки медиаданных имеют унифицированные программные интерфейсы, позволяющие связывать объекты соответствующих классов на этапе исполнения для оптимизированной и надежной передачи потоков данных при их последовательной обработке в конечном приложении.

Функциональность конечных приложений реализуется на этапе исполнения путем организации связей между предварительно созданными объектами представленных выше классов.

Таким образом, использование технологии создания модульного ПО позволяет в сжатые сроки создавать надежные и производительные кроссплатформенные программные комплексы, которые могут гибко приспосабливаться как к

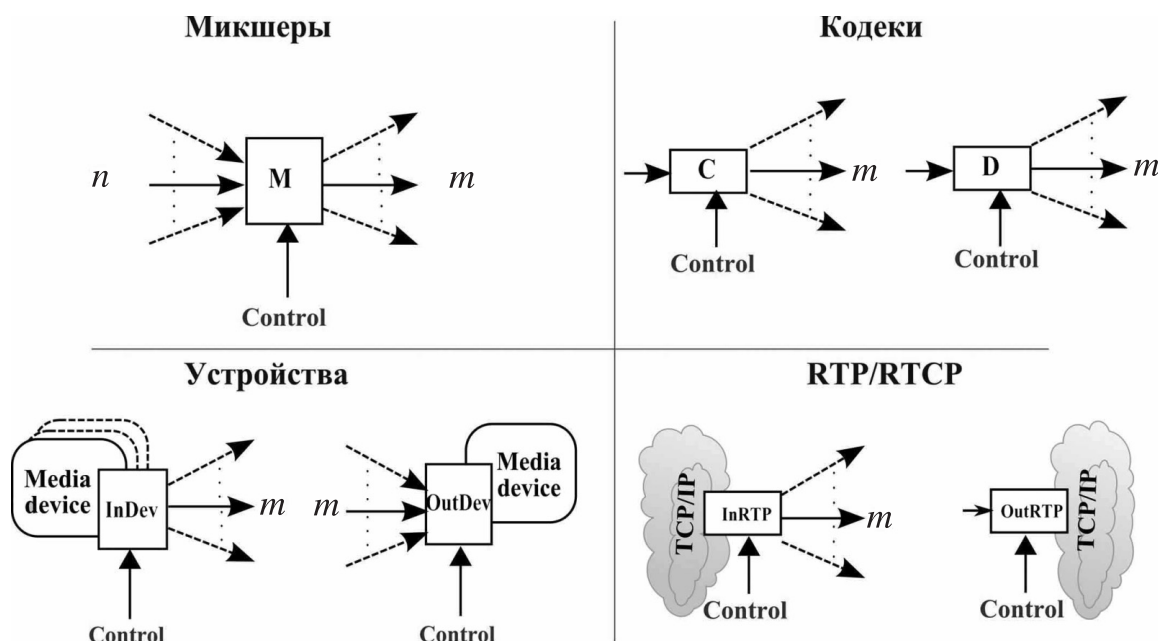


Рис. 6. Состав и интерфейсы модулей обработки медиаданных

различным условиям применения, так и к изменяющимся функциональным требованиям.

Реализация программного комплекса

Представленная выше методология апробирована при создании программного комплекса (ПК) «Видеоконференцсвязь», основным назначением которого является предоставление интегрированных телекоммуникационных услуг пользователям сетей, функционирующих на базе стека протоколов TCP/IP. ПК реализует функциональность персональных систем видеоконференцсвязи с возможностью применения в сетях связи различного масштаба, а также обеспечивает предоставление ряда дополнительных телекоммуникационных услуг. В состав комплекса входит более 20 программных библиотек различной степени интеграции, а также ряд терминальных и серверных приложений. Выбор в качестве протокола сигнализации SIP позволил решить весь спектр задач в едином ключе.

ПК позволяет реализовать все ступени децентрализации системы вплоть до полностью распределенной. Основным режимом работы является децентрализованная обработка медиаданных, передаваемых по протоколу RTP, непосредственно на терминальном оборудовании. Именно это и определяет большинство представленных ниже экономических и технических преимуществ и от-

личительных особенностей.

В централизованных системах видеоконференцсвязи подавляющую часть затрат на внедрение составляет стоимость сервера MCU. Бессерверная конфигурация позволяет при одних и тех же вложениях обеспечить увеличение числа портов системы в два и более раз. Возможность поставки чисто программных кроссплатформенных решений позволяет в максимальной степени использовать уже имеющиеся на объектах заказчика аппаратно-программные комплексы.

Отсутствие единой точки отказа, которой в централизованных системах является сервер MCU, позволяет обеспечить высокую надежность системы в целом. Кроме того, система является более устойчивой к отказам каналов связи. Даже если в результате отказов сеть распадается на ряд отдельных несвязанных сегментов, организация сессий возможна в каждом из сегментов в отдельности.

Протокол SIP базируется на принципах сетевой мобильности пользователей, что позволяет использовать ПК для предоставления интегрированных телекоммуникационных услуг пользователям стационарных и подвижных узлов полевых систем. Количество пользователей в системе и количество одновременно проводимых сессий может наращиваться неограниченно, что обеспечивается бессерверным режимом работы и свойствами масштабируемости протокола SIP.



Качественные и количественные характеристики сессий также могут наращиваться простым повышением производительности терминального оборудования.

Кроссплатформенное, построенное в соответствии с модульными подходами, ПО собственной разработки может быть легко адаптировано к различным операционным средам и аппаратным платформам. Наличие полного комплекта исходных кодов позволяет адаптировать ПО практически под любые требования заказчика, обеспечивая интеграцию с уже существующими на объектах управляющими и инфокоммуникационными системами.

Система обеспечивает возможность гибкой настройки каждого терминала и предоставляет средства управления занимаемой полосой в канале связи, изменения конфигурации устройств захвата и воспроизведения аудио- и видеoinформации, настройки звуковых оповещений. Механизм управления масштабированием и позиционированием видеоокон позволяет гибко управлять качеством отображаемых видеоданных в зависимости от производительности терминального оборудования. ПК имеет многоязычный пользовательский интерфейс.

Функциональные возможности системы в целом определяются комплексированием серверных модулей. Настройки терминала позволяют обеспечить его работу практически при любой конфигурации серверного оборудования, вплоть до полного его отсутствия. Рассмотрим основные функциональные возможности системы, которые обеспечиваются в полностью децентрализованном режиме.

Видеовывод может использоваться как для общения двух пользователей, так и для групповых конференций с участием удаленных залов для совещаний. Групповые конференции возможны за счет функции «второй экран» и возможности подключения периферийного оборудования коллективного пользования. В режиме «точка-точка» система может обеспечивать передачу видео с наилучшим качеством, а при использовании HD-камер позволяет передавать видео высокой четкости.

Максимальное количество участников одной сессии может достигать тридцати двух человек. Децентрализованный режим обработки медиаданных позволяет каждому пользователю индивидуально на своем терминале в реальном мас-

штабе времени управлять способом отображения каждого из участников сессии. Многоточечные конференции также могут отображаться на дополнительном экране коллективного пользования.

Участники сессии видеоконференцсвязи имеют возможность передавать друг другу файлы или короткие текстовые сообщения с использованием гарантированного режима протокола MSRP. При этом доставка файлов может производиться как одному, так и группе участников сессии. Кроме этого, поддерживается передача файлов и текстовых сообщений вне сессии видеоконференцсвязи. Эта функция может использоваться для передачи презентационных материалов до начала сессии или в случае невозможности организовать сессию видеосвязи.

Наличие серверов позволяет расширить возможности для представленных выше функций, а также добавить новые, такие, как регистрация, авторизация и аутентификация пользователей в сети, услуга присутствия, ведение централизованных адресных книг и т. д.

Программный комплекс функционирует на базе операционных систем, имеющих сертификаты соответствия в системе сертификации силовых ведомств (МО, ФСБ и ФСТЭК): MCBC 3.0, Astra-Linux, ALT-Linux, Windows XP, а также под управлением других ОС платформы Windows, включая Windows 2000 и Windows 7, а также ОС платформы Linux, включая Ubuntu, Debian, Red Hat и др.

Для работы терминального ПО достаточно наличия современного персонального компьютера. Возможно также функционирование на мобильных компьютерах (ноутбуках, нетбуках), а также моноблоках и других терминалах с сенсорными экранами.

Серверное ПО не предъявляет жестких требований к характеристикам компьютера, т. к. не занимается обработкой потоков медиаданных, а выполняет только функции управления и обработки сигнализации.

Отечественный программный комплекс «Видеоконференцсвязь» рекомендован к применению межведомственной комиссией из представителей силовых и коммерческих структур, имеет литеру «О₁», а также сертификат соответствия в системе сертификации средств защиты информации по требованиям безопасности информации.

В статье рассмотрена методология создания и реализации интегрированных телекоммуникационных служб, таких, как аудио- и видеосвязь, аудио- и видеоконференцсвязь, передача файлов, передача текстовых сообщений, индикация присутствия и др. Основополагающие принципы предложенной методологии – унификация аппаратного и программного обеспечения, а также основных системотехнических решений, включая организацию единых механизмов управления сеансами, единых технологий сетевого взаимодействия и единых подходов по управлению и обслуживанию комплексов связи.

Описанные подходы основываются на по-

требности в создании современных отечественных систем, которые могут использоваться для оснащения различных уровней ведомственных АСУ. Затрагивая основные требования, предъявляемые к подобным системам, рассмотрены вопросы децентрализации функций управления и передачи мультимедийной информации.

Представленная методология апробирована при создании программного комплекса «Видеоконференцсвязь», основным назначением которого является предоставление интегрированных телекоммуникационных услуг пользователям сетей, функционирующих на базе стека протоколов TCP/IP.

СПИСОК ЛИТЕРАТУРЫ

1. **Rosenberg, J.** RFC 3261. SIP: Session Initiation Protocol [Электронный ресурс] / J. Rosenberg, H. Schulzrinne // IETF. 2002.
2. **Синепол, В.С.** Системы компьютерной видеоконференцсвязи [Текст] / В.С. Синепол, И.А. Цикин. –М.: Связь и бизнес, 1999.
3. **Schulzrinne, H.** RFC 3550. RTP A Transport Protocol for Real-Time Applications [Электронный ре-

сурс] / H. Schulzrinne, S. Casner // IETF. 2003.

4. **Campbell, B.** RFC 3428. Session Initiation Protocol Extension for Instant Messaging [Электронный ресурс] / B. Campbell, J. Rosenberg // IETF. 2002.

5. **Campbell, B.** RFC 4975. The Message Session Relay Protocol [Электронный ресурс] / B. Campbell, R. Mahy // IETF. 2007.

СИСТЕМА ЦИФРОВЫХ «ВОДЯНЫХ» ЗНАКОВ ДЛЯ АУДИОСИГНАЛОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ЯВЛЕНИЯ РЕВЕРБЕРАЦИИ

Техника цифровых «водяных» знаков (ЦВЗ) является эффективным средством защиты прав собственности для различных продуктов, распространяемых в мультимедийных средах, в частности, звуковых (музыкальных) файлов. При этом дополнительная информация, вкладываемая в эти файлы, может содержать не только идентификатор собственника, но и данные легального покупателя, что позволяет в дальнейшем обнаруживать и привлекать к ответственности тех нечестных покупателей, которые копируют и затем нелегально распространяют купленные ими продукты. Вложение дополнительной информации не должно увеличивать объем файлов и изменять их формат. Более того, качество звуковых (и, особенно, музыкальных) произведений практически не должно меняться по сравнению с оригиналом.

Нелегальные распространители копий («пираты») могут попытаться удалить вложенный ЦВЗ, причем под удалением понимается ситуация, когда исходная информация сохраняет приемлемое качество, а ЦВЗ тем не менее не может быть выделен.

Можно также ожидать, что файл с вложением будет подвергаться сжатию с потерями, изменению частоты дискретизации и микшированию. ЦВЗ должен надежно извлекаться после этих преобразований.

Наиболее популярными методами вложения ЦВЗ в аудиосигнал являются метод модуляции мгновенной фазы [1] и добавления эхосигналов [2–8]. Первый использует слабую чувствительность человеческого уха к фазовому спектру аудиосигнала. Результаты исследования этого метода опубликованы в работе [3]. Однако дальнейшие наши исследования показали, что для вы-

деления ЦВЗ требуется обеспечить очень точную синхронизацию на приемной стороне, что фактически невозможно в реальных условиях.

Второй метод основывается на свойстве человеческого слухового аппарата не замечать наличие эхосигнала, если задержка между основным сигналом и эхосигналом не превышает определенной величины. Однако этот метод оказывается уязвимым к преднамеренным атакам, основывающимся на подавлении эха. Использование явления реверберации для построения системы ЦВЗ также рассматривалось ранее в научной литературе [4]. Однако предложенная система не удовлетворяла всем перечисленным выше требованиям. Потому мы представляем в данной статье более полное исследование данного вопроса.

Описание методов вложения и извлечения дополнительной информации на основе использования явления реверберации

Явление реверберации состоит в суперпозиции различных эхосигналов, полученных от одного источника звука. Эффект реверберации можно наблюдать в закрытых помещениях после выключения источника звука [1]. Реверберация присутствует изначально в любой звукозаписи, иногда ее добавляют искусственно с целью придания музыкальным произведениям определенных оттенков звучания.

Для искусственного добавления реверберации нужно произвести свертку исходного сигнала с «комнатным импульсом». Комнатный импульс – это реакция моделируемого помещения на единичный звуковой импульс [2].

Временем реверберации обычно называется время, за которое уровень звука уменьшается на 60 дБ.

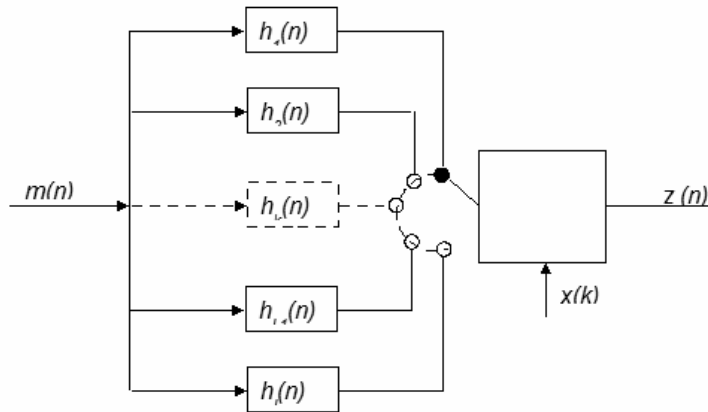


Рис. 1. Схема погружения ЦВЗ на основе явления реверберации

В ранних исследованиях погружения информации в аудиосигнал с помощью добавления реверберации [2] предлагалось использовать два комнатных импульса для вложения, соответственно «0» и «1». Данный способ имеет, однако, существенный недостаток – присутствие некоторой разности в звучании сегментов с разными реверберациями.

В настоящей статье предлагается использовать один комнатный импульс, незначительно модулируя время задержки между основным сигналом и временем начала реверберации.

В этом случае схема погружения выглядит так, как показано на рис. 1, где $m(n)$ – исходный звуковой сигнал; $x(k)$ – погружаемая двоичная последовательность; $h_1(n) - h_L(n)$ – импульсные отклики фильтров, применяемых для имитации реверберации; $z(n)$ – сигнал с вложением.

Различие между фильтрами $h_1(n) - h_L(n)$ заключается в разном времени задержки между исходным сигналом и реверберацией.

Так как возможно применение нескольких величин задержки, то в промежуток времени, в течение которого действует один фильтр, можно вложить L бит, где L – число используемых вариантов задержки.

Если бы исходный звуковой сигнал $m(t)$ (для упрощения мы перешли к непрерывному време-

ни) можно было аппроксимировать гауссовским процессом с известной корреляционной функцией, то, казалось бы, можно было построить оптимальный приемник для выделения одного бита вложенной информации (при использовании двух фильтров) по схеме, представленной на рис. 2 [5].

Здесь $h_\Delta(t, u)$ – решение интегрального уравнения:

$$\int_0^T \int_0^T R_0(t, x) h_\Delta(x, y) R_1(y, u) dx dy = R_1(t, u) - R_0(t, u), \quad (1)$$

где $R_{0(1)}(t, x)$ – корреляционные функции сигналов на выходе первого и второго фильтра соответственно; T – временной интервал, на котором вкладывается один бит.

Однако модель с гауссовской аппроксимацией входного сигнала $m(t)$, которая приводит к оптимальной решающей схеме, показанной на рис. 2, отличается от нашей исходной модели следующим:

звуковой сигнал плохо аппроксимируется гауссовским процессом;

корреляционная функция входного звукового сигнала в точности не известна;

сигналы на входах первого и второго фильтров нельзя полагать взаимонезависимыми, что

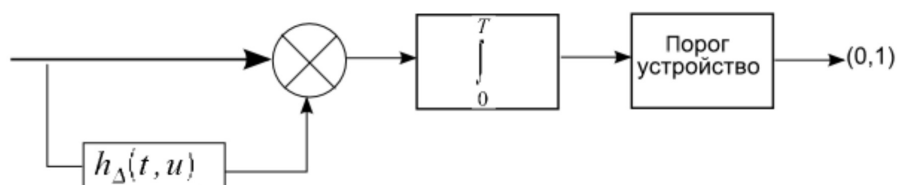


Рис. 2. Оптимальный прием гауссовских сигналов



существенно при доказательстве оптимальности схемы, показанной на рис. 2 [5].

Потому используем т. н. *кепстральное преобразование сигнала* с вложением, которое позволяет свести задачу выделения бита к различению двух гипотез на фоне аддитивного шума [8].

Известны два кепстральных преобразования (комплексное S_c и вещественное S_R), для которых справедливы следующие представления: $S_c = IFT(\log(|FT(S)|)) + 2\pi l$ и $S_R = IFT(\log(|FT(S)|))$, где l – целое число, необходимое для установки фазы; FT – преобразование Фурье; IFT – обратное преобразование Фурье.

Известно следующее свойство кепструма (как комплексного, так и вещественного) [6]:

$$(S_1 * S_2)' = S_1' + S_2', \quad (2)$$

где «*» – означает свертку сигналов; «'» – преобразование кепструма. Тогда, казалось бы, для модели, изображенной на рис. 2, должно выполняться соотношение:

$$Z'(n) = m'(n) + h'_i(n). \quad (3)$$

Поскольку $Z'(n)$ может быть получено из звука с вложением, а $h'_i(n)$, $i = 1, 2, \dots, l$ рассчитывается заранее по известным импульсным реакциям фильтров, то задача извлечения вложенной информации сводится к обнаружению присутствия одного из L (в частном случае $L = 2$) сигналов $h'_i(n)$ на фоне помех $m'(n)$. Если аппроксимировать помеху гауссовским белым шумом, то оптимально решающее правило будет иметь вид:

$$i = \text{Arg}_{i=1,2,\dots,l}^{\max} \sum_{n \in T} (Z'(n) - h'_i(n))^2. \quad (4)$$

Если выполнено условие

$$\sum_{n \in T} h_i'^2(n) = \text{const}(i), \quad (5)$$

то (4) сводится к известному правилу корреляционного приема [7]:

$$i = \text{Arg}_{i=1,2,\dots,l}^{\max} \sum_{n \in T} Z'(n) \cdot h'_i(n). \quad (6)$$

Однако в нашем случае правило (4) оказывается неоптимальным по следующим причинам:

1) соотношение (3) не выполняется в точности на конечных промежутках « T », соответствующих вложению одного бита;

2) аддитивная помеха $m'(n)$ не является гауссовской последовательностью;

3) аддитивная помеха $m'(n)$ не является стационарной по « n ».

Более того, правила решения (4) или (6) оказываются неоптимальными даже при хорошем приближении помехи $m'(n)$ моделью гауссовской независимой последовательности.

Действительно, предположим, что один вкладываемый бит соответствует N отсчетам сигнала. Выберем минимальный параметр $N_0 < N$, на котором практически реализуется полная энергия кепструмов импульсной реакции фильтров, т. е. $\sum_{n=1}^{\infty} h_i'^2(n) \approx \sum_{n=1}^{N_0} h_i'^2(n)$, и рассмотрим следующее правило решения вместо (6):

$$i = \text{Arg}_{i=1,2,\dots,l}^{\max} \sum_{\hat{e}=1}^{N/N_0} \sum_{n \in N_k} Z'(n) \cdot h'_i(n), \quad (7)$$

где кепструмы $Z'(n)$ и $h'_i(n)$ вычисляются на интервалах N_k , $k = 1, 2, \dots, N/N_0$.

На первый взгляд это правило выглядит парадоксально, поскольку противоречит правилу (6) и фактически означает, что разделение интервала принятия решения на подынтервалы и вычисление на каждом из них кепструмов $Z'(n)$ $h'_i(n)$ с последующим накоплением на них взаимных корреляций может дать лучший результат, чем вычисление взаимных корреляций от кепструма на полном интервале.

Действительно, если бы $Z'(n)$ представляла собой отсчеты гауссовского белого шума, а $h'_i(n)$ – отсчеты, скажем, гармонического сигнала частоты ω_i , то переход от (6) к (7) был бы бессмысленным, поскольку энергия полезного сигнала $h(n)$ на N_0 отсчетах была бы в N/N_0 раз меньше, чем энергия на N отсчетах.

Однако в случае, когда $h'_i(n)$ представляет собой кепструм импульсной реакции, это не так, поскольку выполняется соотношение:

$$\sum_{n \in N_k} h_i'^2(n) \approx \sum_{n=1}^{\infty} h_i'^2(n). \quad (8)$$

Рассмотрим далее случай с двумя фильтрами, имеющими одинаковую импульсную характеристику $h(n)$ за исключением разной задержки, выбранной так, что кепструм импульсных реакций этих фильтров оказывается ортогональным.

Тогда, следуя методике общей теории связи [7], можно показать, что вероятность ошибки при детектировании одного бита имеет вид:

$$P = 1 - F \left(\frac{\sqrt{\frac{N}{N_0}} \cdot \sqrt{\sum_{n=1}^{N_0} h_i'^2(n)}}{\delta^2} \right), \quad (9)$$

где $h'(n)$ – кепстрем импульсной реакции фильтра, рассчитанный на интервале N_0 ; $\delta^2 = \text{Var}\{m'(n)\}$ – дисперсия отсчетов гауссовского шума $m'(n)$.

Из соотношения (9) видно, что чем на большее число интервалов N / N_0 разбито все множество отсчетов N , соответствующих вложению одного бита, тем меньше будет вероятность ошибки. Однако минимальная длина интервала N_0 всегда будет ограничена протяженностью импульсной реакции фильтра, т. е. должно выполняться приближенное равенство (8).

На рис. 3 а показан кепстрем типичного му-

зыкального произведения. Видно, что основная энергия кепструма аудиосигнала сосредоточена в окрестностях «нуля», т. е. на т. н. малых «кепсепсу». На рис. 3 б, в показаны кепструмы импульсных реакций фильтров, взятых с различными задержками. Видно, что основная энергия кепструмов этих сигналов сосредоточена на интервалах, соответствующих величинам задержек. Поэтому для ослабления воздействия аддитивной помехи $m'(n)$ в виде кепструма аудиосигнала на обнаружение сигналов $h'_i(n)$, соответствующих кепструмам фильтров, целесообразно выбрать задержки n_0 и n_1 таким образом, чтобы основ-

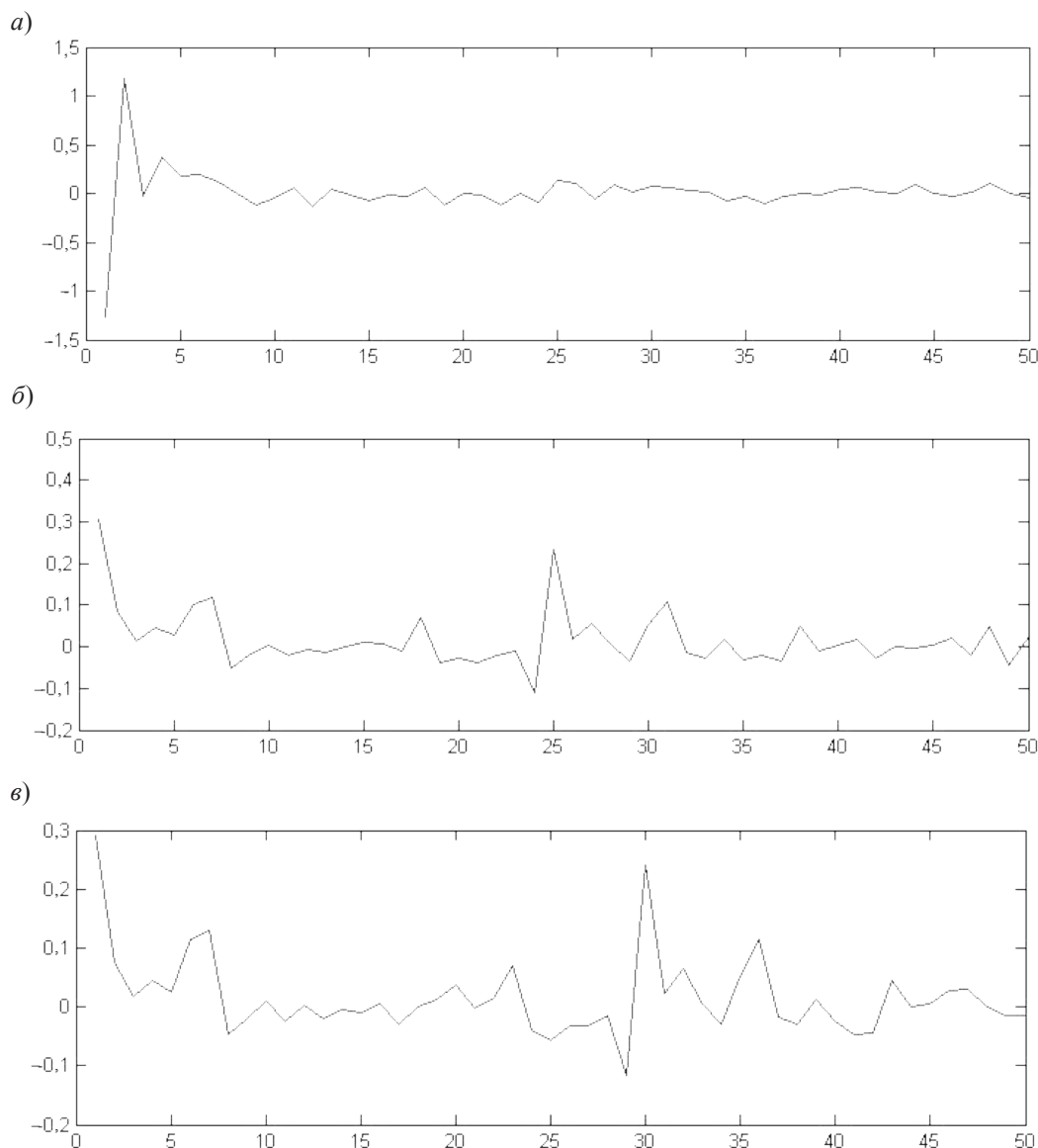


Рис. 3. Кепстрем типичного аудиосигнала (а) и импульсных реакций фильтров, отличающихся задержками в 25 (б) и 30 (в) отсчетов

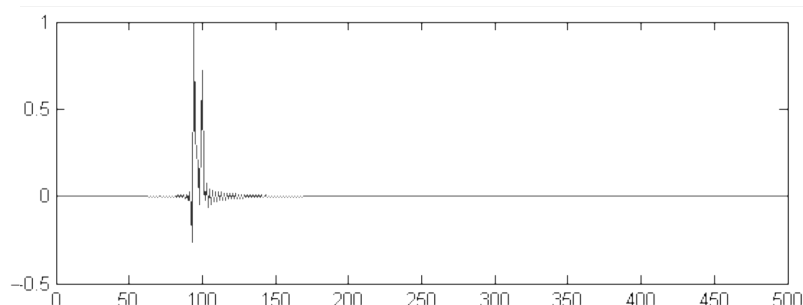


Рис. 4. Импульсная характеристика исследуемого фильтра

ная энергия кепструмов $h'_i(n)$ располагалась вне интервала отсчетов, где присутствует основная энергия аудиосигналов.

Исследование предложенного метода вложения и извлечения ЦВЗ при помощи моделирования

Для проведения эксперимента был выбран фильтр, импульсная реакция которого показана на рис. 4.

Для вложения «0» и «1» используются два фильтра, отличающихся только временем задержки между основным импульсом и реверберацией. В табл. 1 приведены результаты моделирования для различных выбранных параметров, позволяющих оценить скорость вложения и количество битовых ошибок при извлечении. В качестве покрывающего сообщения (ПС) использовались отрывки музыкального произведения Юты «Ревность» длительностью 29 с. Формат файла WAV, частота выборок 44,1 кГц, разрядность 16 бит на отсчет. Все N отсчетов на интервале вложения одного бита разбивались на l подынтервалов, со-

стоящих из $N_0 = \frac{N}{l}$ отсчетов в каждом, а извлечение битов производилось по правилу (7). Как и следовало ожидать, частота ошибок уменьшается при уменьшении скорости вложения и при увеличении (до некоторого значения) величины задержки импульсных реакций относительно аудиосигнала. Также частота ошибок зависит от максимальной амплитуды реверберации. Временем задержки между основным импульсом и реверберацией было выбрано 29 и 25 отсчетов.

Видно, что увеличение количества подынтервалов позволяет уменьшить число ошибок при неизменной амплитуде реверберации или уменьшить амплитуду реверберации, что позволяет уменьшить искажения, вносимые ЦВЗ.

Важнейшим требованием к системе ЦВЗ является ее устойчивость к различным естественным и преднамеренным преобразованиям аудиофайлов с вложением, при которых сохраняется высокое качество основного аудиосигнала (например, музыкального файла).

К естественным преобразованиям относят, прежде всего, сжатие с использованием стандарт-

Таблица 1

Зависимость количества ошибок от параметров вложения

Амплитуда реверберации	Количество подынтервалов	Вложено бит	Ошибок
1	1	144	1
0,5	1	144	1
0,3	1	144	4
0,15	1	144	20
0,15	2	144	16
0,15	10	144	5
0,15	15	144	3
0,1	15	144	11
0,05	15	144	35

Таблица 2

Количество ошибок после сжатия с использованием mp3

Амплитуда реверберации	Количество подынтервалов	Вложено бит	Ошибок
1	1	144	2
0,5	1	144	2
0,3	1	144	4
0,15	1	144	30
0,15	2	144	22
0,15	10	144	6
0,15	15	144	4
0,1	15	144	15
0,05	15	144	48

ных методов. В табл. 2 представлены результаты эксперимента для популярного формата mp3. В качестве ПС использовался аудиофайл Юта «Ревность» с параметрами $N_0 = 6000$, $N_0 = 29$, $N_1 = 25$. Сжатие производилось с «битрейтом» 128 Кбит/с.

Видно, что, хотя количество битовых ошибок возрастет, это не приведет к полному «обрыву» канала ЦВЗ, как можно было бы ожидать при таких методах вложения, как, например, вложение в наименьшие значащие биты или квантованной индексной модуляции.

Были проведены также экспериментальные

исследования по влиянию параметров вложения на качество основного сообщения (музыкального произведения). К сожалению, ввиду особенностей слуховой системы человека, некорректно использовать для оценки изменения качества звучания какие-либо известные из теории связи числовые значения, такие, как соотношение сигнал/шум, ввиду того, что разные по форме и даже по спектру сигналы могут иметь одинаковое для человека звучание. Поэтому анализ качества проводился методом «экспертной оценки». Для этого группе из 5–7 человек предлагалось прослушать музыкальное произведение как в оригинале, так

Таблица 3

Результаты эксперимента для различных произведений

Название композиции	Длина фрагмента, с	Количество отсчетов для вложения одного бита	Количество вложенных бит	Количество ошибок
«Священная война»	20	10000	86	0
«Есть только миг»	20	10000	86	0
Oldfield. «Moonlight Shadow»	20	10000	86	0
Notre-Dame de Paris. «Belle» (Русская версия)	20	10000	86	0
Mozart, fantasia D-moll, KV 397 (Yudina)	20	10000	86	10
Metallica. «Fuel»	20	10000	86	0
Алиса. «Трасса Е95»	20	10000	86	0
Maria Maifestati	20	10000	86	0
Высоцкий. «Песенка sentimentalного боксера»	20	10000	86	0
Юта. «Ревность»	20	10000	86	6



и с вложением и оценить их идентичность. Исследования показали, что при выборе амплитуды реверберации 0,3 большинство «экспертов» в группе не смогли заметить разницу между оригиналом и сигналом с вложением; при выборе амплитуды реверберации 0,15 ни один из экспертов не отличил оригинальный сигнал от сигнала с вложением.

Проведенные экспериментальные исследования позволяют выбрать следующие оптимальные параметры при использовании систем ЦВЗ для аудиосигналов, основанных на реверберации:

количество отсчетов импульсных реакций фильтров 300–600;

количество отсчетов, используемых для вложения одного бита 6000–10000;

количество подынтервалов для принятия решения 10–15;

задержка реверберации 20–40 отсчетов.

В табл. 3 представлены результаты эксперимента по количеству ошибок для десяти различных фрагментов музыкальных файлов в формате WAV с частотой выборок 44,1 кГц, разрядностью 16 бит, при использовании набора параметров, приведенных выше.

Что же касается устойчивости ЦВЗ к преднамеренным преобразованиям, то защиту от них обеспечивает выбор секретной импульсной реакции фильтров, которая может вычисляться по специальному, разделяемому между пользователями, секретному стежоключу. Конечно, нельзя считать эту атаку полностью исследованной, тем более, что существует еще один метод удаления вложенных ЦВЗ, который в известной литературе [6] носит название «deconvolution» и заключает-

ся в том, что для звукового сигнала производится компенсация фильтрации дереверберационным фильтром. Однако решение задачи по защите системы от подобной атаки требует проведения значительных дополнительных исследований и поэтому мы предполагаем представить их в наших последующих публикациях.

В статье исследован метод вложения ЦВЗ для аудиосигналов, где для вложения бита «0» или «1» использовались фильтры, имитирующие некоторую акустическую среду звукозаписи. Фильтры, соответствующие «0» и «1», отличаются друг от друга только величиной задержки реверберации относительно основного аудиосигнала. Исследования данной системы ЦВЗ позволили сделать следующие выводы.

Извлечение информации целесообразно производить при помощи корреляционного приема кепструмов аудиосигналов, выполненных на интервалах, соответствующих протяженности кепструмов импульсных реакций фильтров, с последующим накоплением результата на интервале одного бита.

Типичный музыкальный файл позволяет вложить около 350 бит в одну минуту звучания при низкой вероятности ошибок.

Применение популярных алгоритмов сжатия с потерями, таких, как mp3, к сигналам с вложением ухудшает качество извлечения, но, тем не менее оно остается удовлетворительным.

Вид импульсной реакции реверберационных фильтров определяется допустимыми искажениями аудиосигналов и возможностью их нахождения при выполнении «атак» на ЦВЗ.

СПИСОК ЛИТЕРАТУРЫ

1. **Nishimura, R.** Audio watermark based on periodical phase shift [Text] / R. Nishimura, Y. Suzuki // J. Acoust. Soc. Japan. –2004. –Vol. 60. –№ 5. –P. 268–272.
2. **Kim, H.J.** A novel echo-hiding scheme with backward and forward kernels [Text] / H.J. Kim, Y.H. Choi // IEEE Transactions on Circuits and Systems for Video Technology. –2003. –№ 13 (8). –P. 885–889.
3. **Alekseyev, V.** Design of «Robust Audio Watermark System» [Text] / V. Alekseyev, A. Grudin, V. Korzhik // Proc. of the XI international symp. on problems of redundancy in information and control systems. –2007. –P. 163–165.
4. **Ansari, R.** Datahiding in audio using frequency-selective phase alteration [Text] / R. Ansari, H. Malik, A. Khokhar // Paper presented at the IEEE International Conf. on Acoustics Speech and Signal Proc. –Montreal, Canada, 2004, May.
5. **Ван Трис, Г.** Теория обнаружения, оценок и модуляции [Текст] / Г. Ван Трис. –М.: Сов. радио, 1972. – Т. 2.
6. **Childers, D.G.** The Cepstrum: A Guide to Processing Proceedings [Text] / D.G. Childers, D.P. Skinner, R.C. Kemerait // IEEE. –Oct. 1977. –Vol. 65. –№ 10. –P. 1428–1443.
7. **Финк, Л.М.** Теория передачи дискретных сообщений [Текст] / Л.М. Финк. –Сов. радио, 1970.
8. **Korzhik, V.** The Use of Wet Paper Codes With Audio Watermarking Based on Echo Hiding [Text] / V. Korzhik, G. Morales-Luna, I. Fedyanin // Proc. of the Federated Conf. on Computer Science and Information Systems. –2012. –P.727–732.

АВТОМАТИЗИРОВАННОЕ ДЕТЕКТИРОВАНИЕ ДЕФЕКТОВ ПРИ РЕКОНСТРУКЦИИ АРХИВНЫХ ФОТОДОКУМЕНТОВ

В различных архивах мира хранится большое количество архивных фото- и видеодокументов. Многие фотографии являются уникальными и в некоторых случаях единственными сведениями об отдельных элементах прошлого. Оценивание истинных значений пикселей изображений в той или иной степени необходимо в большинстве задач цифровой обработки изображений. Методы восстановления двумерных сигналов находят свое применение в задачах обработки архивных документов в виде изображений, имеющих различные искажения, например: царапины, пятна, пыль, ненужные надписи, линии сгиба [1–3].

Восстановление архивных фотодокументов состоит из двух главных этапов: обнаружение дефектов и их устранение. В настоящее время первый и второй этап в большинстве случаев делает человек-реставратор, при этом цифровая реставрация изображений является трудоемкой и требует много времени [3]. Особенно остро это ощущается при необходимости реставрации большого количества фотодокументов в короткий срок. При этом многие операции, выполняемые человеком в процессе реставрации, можно автоматизировать.

Упрощенная математическая модель изображения может быть представлена в следующем виде:

$$Y_{i,j} = (1 - M_{i,j})S_{i,j} + M_{i,j}R_{i,j}, \quad i = \overline{1, I}, \quad j = \overline{1, J},$$

где $Y_{i,j}$ – исходное изображение; $S_{i,j}$ – оригинальное (неискаженное) изображение; $M_{i,j}$ – бинарная маска области с искаженными значениями (единица соответствует искаженным пикселям, нуль – неискаженным пикселям); $R_{i,j}$ – искаженные значения пикселей.

Дефекты на изображении можно классифицировать следующим образом: пятна, волокна, искажения фотографического слоя, зернистость, искусственные надписи. Все эти искажения можно отнести к локальным особенностям изображений и анализировать их с помощью текстурного анализа, а задачу детектирования таких особенностей свести к задаче выделения локальных дескрипторов на изображениях. Среди таких методов за

основу выбран метод Speeded Up Robust Features (SURF), как наиболее распространенный и эффективный метод выделения локальных особенностей. Метод SURF применяется для выделения локальных особенностей, создания дескрипторов, сравнения изображений, 3D реконструкции и поиска объектов на изображении [4].

Обнаружение особых точек в SURF основано на вычислении детерминанта матрицы Гессе (гессиана), который достигает экстремума в точках максимального изменения градиента яркости и позволяет детектировать пятна, углы и края линий.

Матрица Гессе для двумерной функции и ее детерминант определяется следующим образом:

$$H(Y(x, y)) = \begin{bmatrix} \frac{\partial^2 Y}{\partial x^2} & \frac{\partial^2 Y}{\partial x \partial y} \\ \frac{\partial^2 Y}{\partial x \partial y} & \frac{\partial^2 Y}{\partial y^2} \end{bmatrix},$$
$$\det(H) = \frac{\partial^2 Y}{\partial x^2} \frac{\partial^2 Y}{\partial y^2} - \left(\frac{\partial^2 Y}{\partial x \partial y} \right)^2.$$

Значение гессиана используется для нахождения локального минимума или максимума яркости изображения. Метод SURF использует бинаризованную аппроксимацию лапласиана гауссиана (Fast-Hessian). Для нахождения особых точек с помощью SURF анализируются все пиксели изображения и определяется максимум гессиана [4].

В данной статье предлагается модифицировать метод SURF с предварительной обработкой изображения энергетическими характеристиками Лавса.

Лавс (Law) разработал энергетический подход, в котором оценивается изменение содержания текстуры в пределах окна фиксированного размера [5]. Для вычисления энергетических характеристик используется набор из 22 масок размерами 5×5 . Затем энергетические характеристики каждого пикселя анализируемого изображения представляются в виде вектора и девяти чисел.

Для вычисления масок используются следующие векторы: $L5$ (Level) = [1 4 6 4 1], $E5$ (Edge) = [-1 -2 0 2 1], $S5$ (Spot) = [-1 0 2 0 -1], $R5$ (Ripple) = [1 -4 6 -4 1], $W5$ (Wave) = [-1 2 0 -2 1].

Название вектора описывает их назначение. Вектор $L5$ предназначен для вычисления симметричного взвешенного локального среднего значения. Вектор $E5$ предназначен для обнаружения краев, $S5$ – для обнаружения пятен, $R5$ – для обнаружения образа в виде ряби, $W5$ – для обнаружения волн [5]. Двумерные маски вычисляются умножением пар векторов.

Предлагаемый подход позволяет находить особые точки на различных типах текстур, а также определяет для каждой точки направление

максимального изменения яркости (градиент) и масштаб, взятый из масштабного коэффициента матрицы Гессе. Такой подход позволяет анализировать различные типы дефектов на поверхности фотографий как один из типов текстуры, таких, как края, пятна, рябь, волны и т. д.

На последнем этапе обработки в качестве дефектов анализируются особые точки, для которых масштабный коэффициент матрицы Гессе больше порогового уровня. Такой подход позволяет учитывать только те дефекты, которые значительно выделяются на фоне остальных пикселей. На следующем этапе изображения, полученные после обработки фильтром Лавса, бинаризируются. Дефектными считаются только те области пиксе-

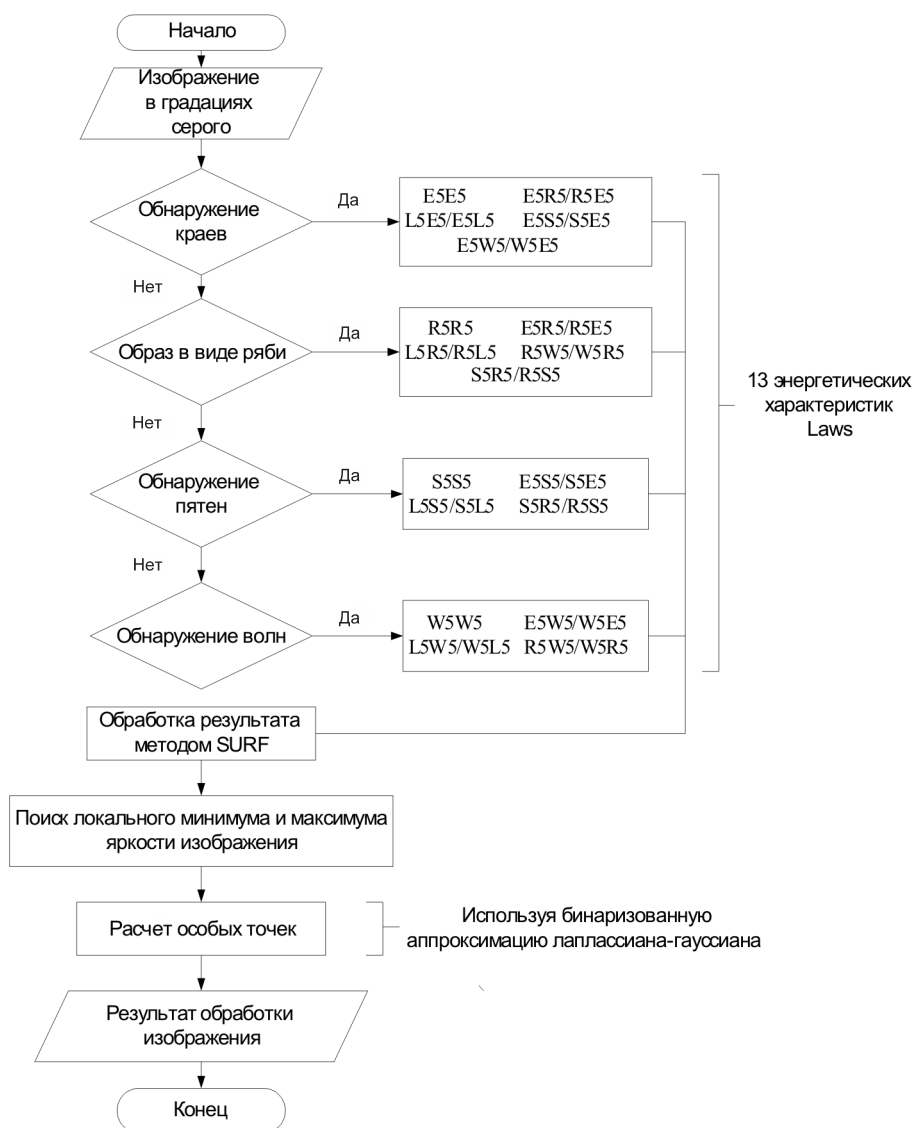


Рис. 1. Блок-схема алгоритма модифицированного метода SURF

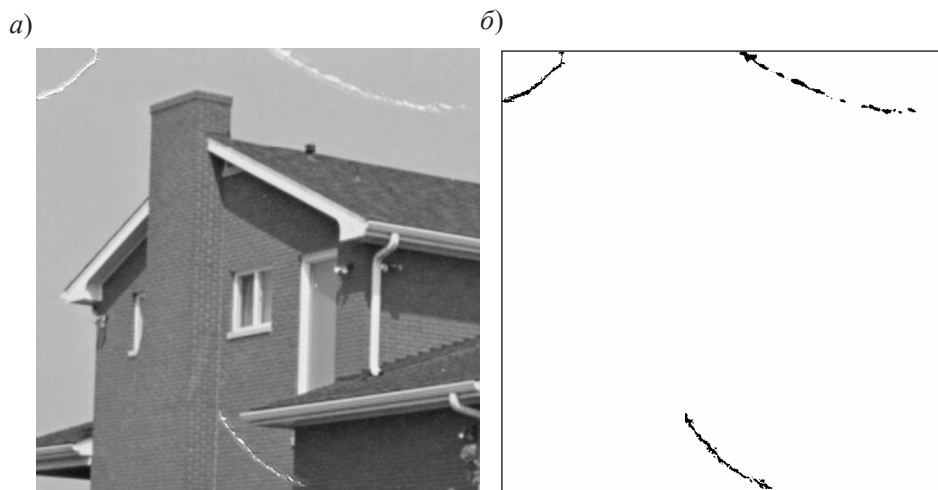


Рис. 2. Тестовое изображение 1

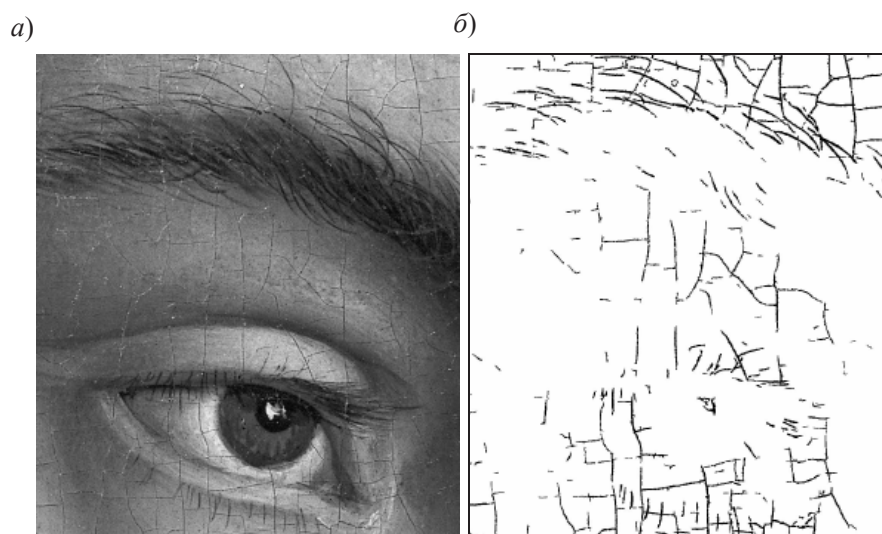


Рис. 3. Тестовое изображение 2

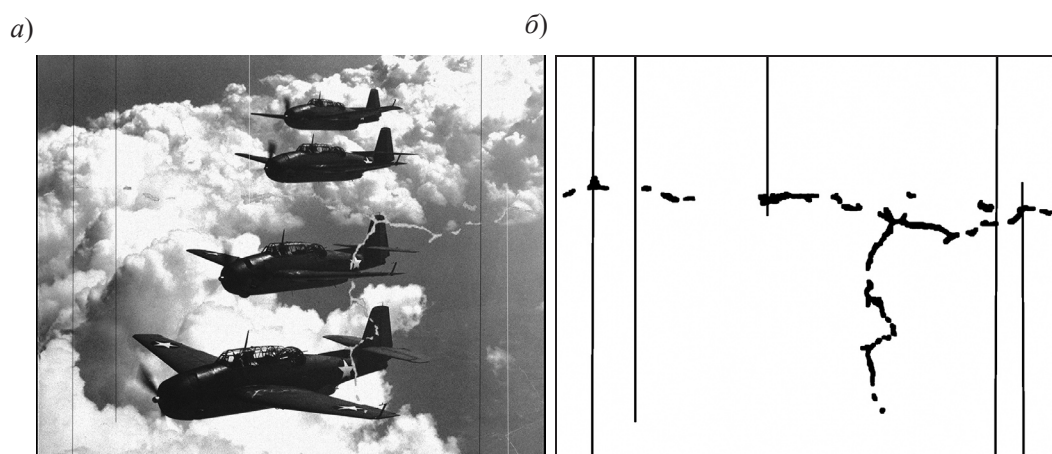


Рис. 4. Тестовое изображение 3



лей на бинаризованном изображении, для которых найдены особые точки с масштабным коэффициентом больше порогового уровня [6].

На рис. 1 представлена блок-схема алгоритма предложенного метода. Алгоритм состоит из следующих шагов.

1. Выбор интересующей локальной особенности (типа дефекта): царапина, рябь, пятна или волны.

2. Изображение $Y_{i,j}$ в зависимости от типа локальной особенности обрабатывается энергетическими масками Лавса.

3. Полученное изображение обрабатывается методом SURF.

4. Поиск локального минимума и максимума яркости изображения.

5. Расчет особых точек с помощью бинаризованной аппроксимации лапласиана гауссиана.

6. Дефектными считаются только те области пикселей на бинаризованном изображении, для которых найдены особые точки с масштабным коэффициентом больше порогового уровня.

На рис. 2–4 представлены примеры обработки фотодокументов предлагаемым методом (a – исходное изображение, b – маска дефектов, обнаруженная предложенным способом).

Использование известных методов детектирования дефектов изображений требует априорных знаний о типе дефектов. Уменьшение объема априорной информации существенно усложняет процесс обработки изображений и в ряде случаев делает обязательным визуальный контроль и использование экспертных оценок.

Анализ результатов обработки показывает, что предлагаемый метод позволяет детектировать различные типы искажений, такие, как линии, отслоение фотослоя, надрывы, горизонтальные и вертикальные линии. Изображение 3 имеет сложный тип дефекта: отслоение фотослоя в виде светлых пикселей и вертикальные царапины в виде темных пикселей. Предлагаемый метод позволяет обнаружить оба типа дефектов.

Искажения архивных фотодокументов можно отнести к локальным особенностям изображений и анализировать их с помощью текстурного анализа, а задачу детектирования таких особенностей отнести к задаче выделения локальных дескрипторов на изображениях. Для решения данного вопроса разработан модифицированный метод SURF, состоящий из нескольких этапов. Первый – преобразование изображения в градации серого и предварительная обработка энергетическими характеристиками Лавса, в результате чего получается набор отфильтрованных изображений. Следующий этап – непосредственно обработка методом SURF.

Анализ результатов обработки показал, что алгоритм модифицированного метода SURF позволяет выделять различные типы локальных особенностей с помощью предварительной обработки изображения энергетическими характеристиками Лавса.

Работа выполнена в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг.

СПИСОК ЛИТЕРАТУРЫ

1. Шапиро, Л. Компьютерное зрение [Текст] / Л. Шапиро, Дж. Стокман; пер с англ. – М.: БИНОМ. Лаборатория знаний, 2006. – 413 с.
2. Варламов, А.Д. Точечные дефекты фотодокументов и методы их устранения [Текст] / А.Д. Варламов, С.С. Садыков // Тр. Междунар. симпоз. Надежность и качество; под ред. Н.К. Юркова. – Пенза: Изд-во ПГУ, 2004. – С. 411–413.
3. Варламов, А.Д. Технология реставрации фотоизображений [Текст] / А.Д. Варламов // Сб. матер. VII Междунар. конф. Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации. Рас-

познавание. – Курск: Изд-во Курского гос. техн. ун-та, 2005. – С. 34–35.

4. Bay, H. Speeded-Up Robust Features (SURF) [Электронный ресурс] / H. Bay, A. Ess // LNCS 3951 (ECCV'06). – 2006. – Vol. 1. – P. 404–417.

5. Laws, K. Textured image segmentation: Ph.D. diss. [Text] / K. Laws. – University of Southern California, 1980.

6. Морозова, Т.В. Алгоритм детектирования локальных особенностей на изображении [Текст] / Т.В. Морозова, В.В. Воронин // Сб. науч. тр. Современные проблемы радиоэлектроники. – Ростов-на-Дону: РИО РГИСТ ФГБОУ ВПО «ЮРГУЭС», 2012. – С. 151–153.

МНОГОУРОВНЕВАЯ МОДЕЛЬ С ДИФФЕРЕНЦИАЛЬНЫМ УРОВНЕМ ДОВЕРИЯ К СУБЪЕКТАМ ДЛЯ ПОВЫШЕНИЯ ЗАЩИЩЕННОСТИ РАСПРЕДЕЛЕННЫХ БАЗ ДАННЫХ

Распределенные базы данных (РБД), стали доминирующим инструментом хранения больших массивов информации. Современные информационные приложения полагаются не на файловые структуры операционных систем, а на многопользовательские РБД, реализованные по технологии клиент/сервер. В этой связи обеспечение информационной безопасности РБД и, в первую очередь, их серверных компонентов приобретает решающее значение.

Для РБД важны все три основных аспекта информационной безопасности: конфиденциальность, целостность и доступность [1, 2]. Общая идея защиты баз данных состоит в следовании рекомендациям, сформулированным для класса безопасности С2 в «Критериях оценки надежных компьютерных систем». В принципе для ряда РБД реализуются средства защиты, характерные для класса В1, однако практическое применение подобных средств эффективно лишь в том случае, когда все компоненты информационной структуры организации соответствуют категории безопасности В. Следует учитывать два обстоятельства. Во-первых, для подавляющего большинства коммерческих организаций класс безопасности С2 достаточен. Во-вторых, более защищенные версии средств защиты отстают по своим возможностям от обычных и представляют собой несколько устаревшие (хотя и тщательно проверенные) продукты [3, 4].

На рис. 1 представлена обобщенная схема средств поддержки безопасной обработки данных в распределенных базах данных в клиент-серверной технологии. Она включает в себя следующие элементы:

- компьютерную сеть – среду передачи данных;
- сервер базы данных;
- базу данных;
- рабочие станции клиентов-субъектов, которые формируют запросы на сервер базы данных;
- сервер безопасности данных, в состав которо-

го входят блоки аутентификации, разграничения доступа к данным, шифрования и управления мониторингом [5, 6].

Клиент в данной модели через компьютерную сеть выполняет запросы к базе данных, которые независимо от типа сети (локальная, корпоративная, глобальная) попадают на сервер безопасности данных [7, 8].

Сервер безопасности данных состоит из четырех блоков, отвечающих за определенное направление защиты данных.

Первый блок выполняет внутреннюю аутентификацию, проводимую средствами СУБД, и внешнюю, которая проводится средствами операционной системы. Процедура аутентификации реализуется путем проверки идентификатора пользователя и его пароля в зависимости от раздела БД, к которому он обращается.

Второй блок сервера безопасности данных – механизм разграничения доступа. В клиент-серверных системах обычно используются три модели: дискреционная, мандатная и ролевая. Первая и вторая модели используются для построения традиционных средств защиты СУБД, а ролевая разработана сравнительно недавно, она является более сложной и базируется на принципах, заложенных как в дискреционной, так и в мандатной моделях.

Третий блок – блок шифрования. В СУБД данные шифруются прозрачным и непрозрачным способом, каждый из которых имеет свои преимущества и недостатки. Для шифрования обычно используются алгоритмы шифрования, в частности, алгоритмы группы AES, RSA. Шифрование запросов можно проводить на уровне как самой операционной системы, так и СУБД.

Четвертый блок сервера безопасности данных – блок управления мониторингом. Важным является фактор оперативного реагирования на угрозы, связанный с тем, что процедура мониторинга проводится по собранным статистическим данным в течение определенного периода времени. С другой стороны, этого недостаточно для

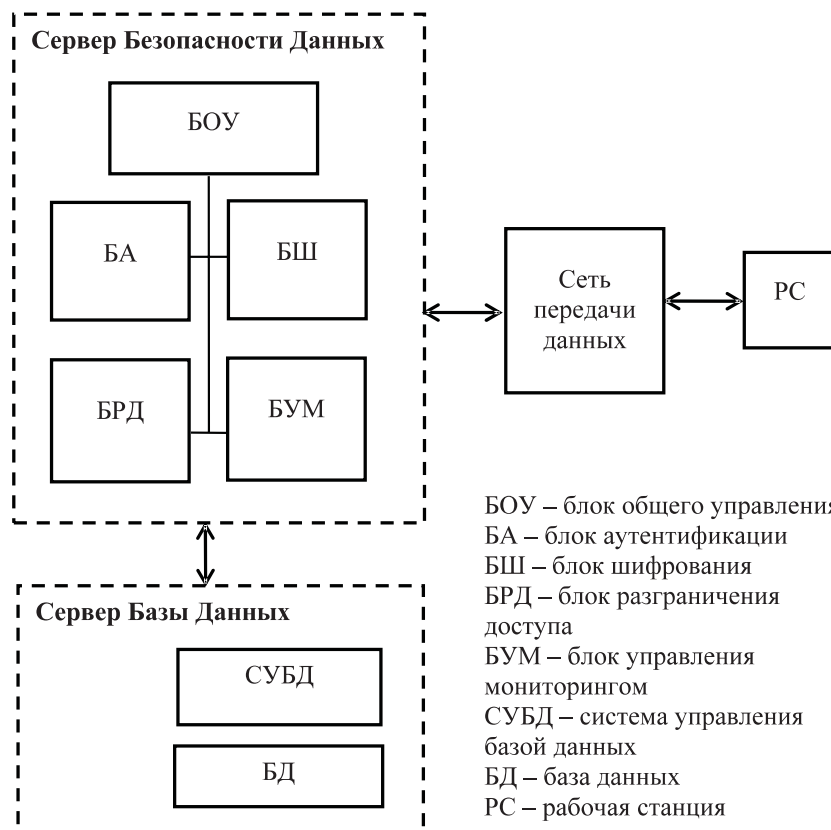


Рис. 1. Средства поддержки безопасной обработки данных в распределенных базах данных

поддержания комплексной защиты СУБД, следовательно, требуется разработка таких средств мониторинга, которые позволят проанализировать все запросы клиентов-рабочих станций на сервер базы данных с точки зрения вероятности реализации несанкционированного доступа

Безопасная среда распределенной базы данных

Вариант архитектуры безопасной СУБД для распределенной БД (рис. 2) предусматривает наличие нескольких уровней секретности в пределах одной СУБД.

Операции низкого уровня обрабатываются традиционно. Операции высокого уровня и многоуровневые операции обрабатываются СУБД, в которой хранятся высокосекретные данные и копии данных низших степеней секретности. Отметим, что такая архитектура подразумевает использование средств тиражирования, в ней должны присутствовать менеджеры тиражирования и необходима соответствующая политика обеспечения целостности данных.

Многоуровневая модель безопасности баз данных

Комбинация средств проверки полномочий и проверки подлинности – эффективный подход к обеспечению безопасности баз данных. Однако он оказывается неудовлетворительным в том случае, если в учреждении необходимо организовать реальную многоуровневую среду защиты информации.

Концепция многоуровневой защиты подразумевает, что

в компьютерной системе хранится информация, относящаяся к разным классам секретности; часть пользователей не имеет доступа к некоторым секретным классам информации.

Классический пример подобной среды – компьютерная система закрытого учреждения, где в логически единой распределенной базе данных содержится информация от полностью открытой до особо секретной, при этом степень благонадежности (доверия) пользователей также варьируется от допуска только к несекретной информации до допуска к совершенно секретным данным.

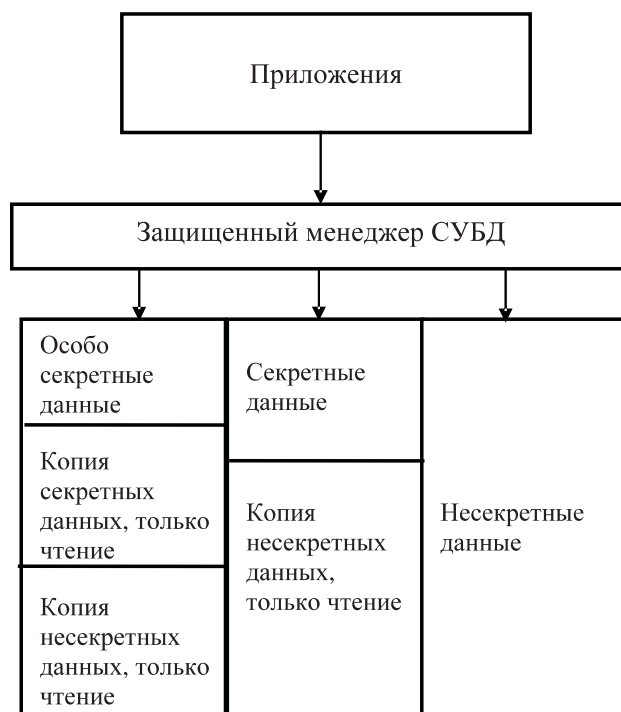


Рис. 2. Архитектура безопасной СУБД с многоуровневым компонентом

Таким образом, пользователь, имеющий низкий уровень доверия, может выполнять свою работу в системе, содержащей сверхсекретную информацию, но ни при каких обстоятельствах не должен быть допущен к ней.

Предлагается построение многоуровневой защиты баз данных на основе модели Белла–ЛаПадула (Bell–LaPadula). Объекты классифицируются, а субъектам назначается один из уровней доверия. Классы и уровни доверия представляются классами доступа, которые характеризуются двумя компонентами: первый компонент определяет иерархическое положение класса, второй – представляет собой множество элементов из неиерархического набора категорий, которые могут относиться к любому уровню иерархии.

Так, на предприятии возможны такие уровни иерархии: секретно; для ограниченного распространения; конфиденциально; для служебного пользования; несекретно. Второй компонент в такой организации может принадлежать к набору категорий: недоступно для служащих по контракту; финансовая информация компании; информация об окладах.

Далее формируется матрица соотношений между иерархическими и неиерархическими компонентами. Пусть некоторый объект классифици-

рован как особо секретный, но ему не приписана ни одна из категорий неиерархического набора, тогда он может предоставляться определенным лицам, в то время как менее секретный объект может иметь категорию «недоступно для посторонних лиц» и, следовательно, не должен им предоставляться. В матричной модели создается «решетка», где неиерархические компоненты каждого уровня иерархии автоматически приписываются и всем более высоким уровням («обратное наследование»). Эта модель отображена на рис. 3.

Следует отметить, что субъект имеет право на запись в объект только в том случае, если класс субъекта такой же или ниже, чем класс записываемого объекта. Это означает, что информация, принадлежащая к какому-либо уровню секретности, никогда не может быть записана в какой-либо объект, имеющий более низкий уровень секретности, поскольку это может привести к деградации защиты классифицированной информации.

С учетом двух принятых в матричной модели ограничений (простое свойство секретности и право записи) формируется база данных с многоуровневой защитой. Заметим, что субъект 1 (пользователь или процесс), имеющий высший уровень доверия, может читать информацию всех кортежей, но в силу права записи он может за-

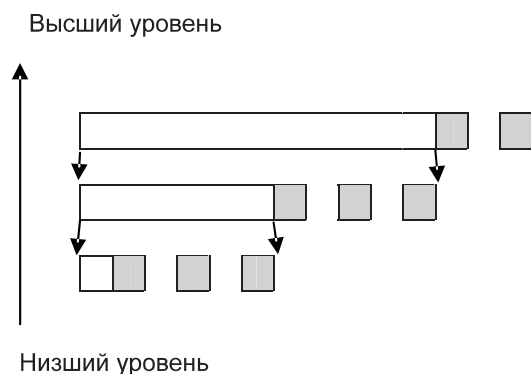


Рис. 3. Решетка классов доступа в матричной модели
 (■) иерархическая категория, «базовые» привилегии;
 (□) неиерархическая категория

писать данные только в кортеж наивысшей классификации (субъект может записывать информацию только «вверх по иерархии», но не «вниз по иерархии»). Субъект 2, принадлежащий к классу «секретно», не может прочитать «особо секретную» информацию, но может прочитать остальную. В то же время, субъекту 2 разрешена запись информации класса «секретно» и класса «особо секретно» (запись «вверх»).

Формирование уровня доверия к субъектам распределенных баз данных

Для поддержки безопасности распределенных БД на основе многоуровневого подхода требуются специальные механизмы формирования уровня доверия к субъектам.

Суть предлагаемого подхода заключается в следующем. В процессе функционирования РБД каждому новому субъекту изначально присваивается начальный, в общем случае минимальный (возможно, нулевой) уровень доверия со стороны механизмов безопасности РБД, ввиду того, что поведение данного субъекта не определено, т. к. отсутствует соответствующая статистическая информация. Таким образом, если другое не определено, новый субъект в начальный период его взаимодействия с РБД не имеет достаточного уровня доверия для обработки им конфиденциальной информации (определенного уровня секретности) и, следовательно, он может обрабатывать лишь открытые данные. В процессе функционирования РБД постоянно выполняется мониторинг действий субъекта и передаваемых им данных.

По умолчанию новый субъект автоматически принимает установленную в системе политику безопасности. В общем случае подразумевается, что новый субъект доверяет, как минимум, администратору безопасности РБД или домену, поддерживающим базовые функции безопасности РБД. Далее, в процессе работы с БД уровень доверия к субъекту изменяется, тем самым формируется показатель рейтинга субъекта на основе соответствующей статистической информации.

Для формирования рейтинга (уровня доверия) субъекта используются механизмы мониторинга действий субъекта, а также статистические данные, преобразуемые в рекомендации, получаемые от механизмов безопасности РБД, интегрированных в систему.

Уровень доверия $Tn_i(t)$ к i -му субъекту рассчитывается как

$$Tn_i(t) = Tn_0 * \left(\frac{N_{lim}}{N_i(t) + 1} \right), \quad (1)$$

где Tn_0 – начальный уровень доверия к субъекту; $N_i(t)$ – число случаев нарушения безопасности, инициированных или связанных с i -м субъектом на интервале времени $(0, t)$; N_{lim} – критическое число случаев нарушения безопасности на том же интервале времени.

События нарушения безопасности РБД разделяются на преднамеренные и случайные действия субъектов. В общем случае субъекты РБД и соответственно узлы, с которых они взаимодействуют с РКС, могут случайным образом совершать непреднамеренные ошибочные действия, формально связанные с попыткой нарушения безопасности. Управление параметром N_{lim} , который фактически определяет предельно допустимое число случаев нарушения безопасности, вызванных случайными действиями субъектов, позволяет учесть данные случаи.

В общем случае уровень доверия Tn_i к i -му субъекту в процессе его взаимодействия с РБД может повышаться, снижаться или оставаться постоянным. Так, если число случаев нарушения безопасности $N_i(t)$, связанных с i -м субъектом на интервале наблюдения $(0, t)$ превышает установленное критическое число N_{lim} ($N_i(t) > N_{lim}$), то уровень доверия к данному субъекту снижается; если число случаев нарушения безопасности меньше значения N_{lim} ($N_i(t) < N_{lim}$), то уровень доверия к субъекту, наоборот, повышается.

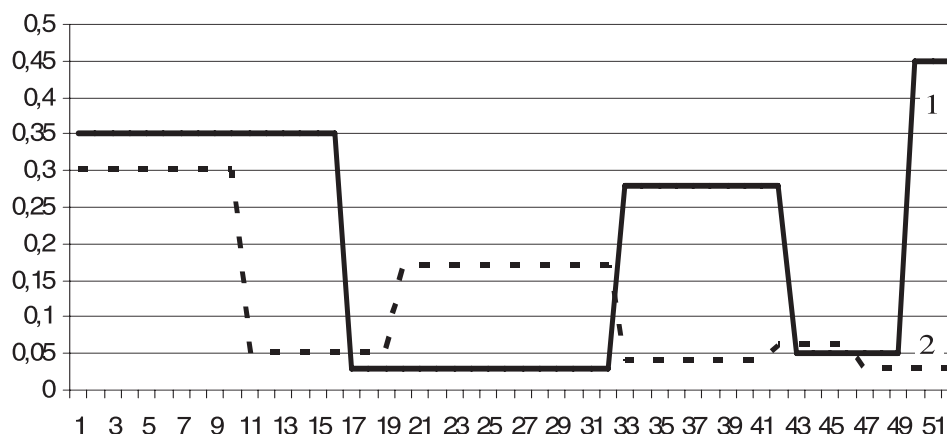


Рис. 4. Изменение рейтинга (уровень доверия) субъектов в защищенных РБД

Предлагается нормировать уровень доверия к субъекту на интервале $(0, \dots, 1)$. Для этого определяется максимально возможный уровень доверия $Tn_{\max}$ к субъектам при фиксированном начальном уровне доверия:

$$Tn_{\max} = Tn_0 * N_{\lim}. \quad (2)$$

Далее, для нормирования уровня доверия $Tn_{\text{норм}}(t)$ полученные значения $Tn_i(t)$ делятся на $Tn_{\max}$

$$Tn_{\text{норм}}(t) = \frac{Tn_i(t)}{Tn_{\max}}. \quad (3)$$

Весь период взаимодействия субъекта с РБД $(0, t)$ представим в виде конечных интервалов наблюдения (t_k, t_l) , на которых в зависимости от результатов действий субъекта на данном интервале уровень доверия к нему снижается или повышается. В результате динамика изменения уровня

доверия $Tn_{\text{норм}i}$ к i -му субъекту представляет собой кусочную функцию (рис. 4).

Для повышения безопасности РБД предложена многоуровневая модель безопасности баз данных и подход к формированию уровня доверия к субъектам на основе оперативного учета числа случаев нарушения безопасности с их стороны. Реализация средств защиты РБД на основе предложенных механизмов обеспечивает поддержку высокого уровня защищенности РБД в условиях динамического изменения числа пользователей-субъектов распределенных баз данных.

Предложенные средства соответствуют основным требованиям к построению системы безопасности РБД, а также имеют возможность расширения путем включения дополнительных механизмов.

СПИСОК ЛИТЕРАТУРЫ

1. **Смирнов, С.Н.** Безопасность систем баз данных [Текст] / С.Н. Смирнов. —М.: Гелиус АРБ, 2007. —352 с.
2. **Мельников, В.П.** Информационная безопасность и защита информации [Текст] / В.П. Мельников. —М.: Academia, 2007.
3. **Lunt, T.F.** Research Directions in Database Security [Text] / T.F. Lunt. —NY: Springer-Verlag, 1998.
4. **Ладыженский, Г.М.** Системы управления базами данных [Текст] / Г.М. Ладыженский. —Jet Infosystems. —2007.
5. **Поляков, А.** Безопасность Oracle глазами аудитора: нападение и защита [Текст] / А. Поляков. —М.: ДМК Пресс, 2009. —336 с.
6. **Гринченко, Н.Н.** Проектирование баз данных [Текст] / Н.Н. Гринченко, Е.В. Гусев, Н.П. Макаров. —2007. —468 с.
7. **Шварц, Б.** Базы данных. Оптимизация производительности [Текст] / Б. Шварц, П. Зайцев, В. Ткаченко. —М.: Символ-Плюс, 2010. —832 с.
8. **Бородина, А.И.** Технологии баз данных и знаний [Текст] / А.И. Бородина. —Мн.: Изд-во БГЭУ, 2008. —505 с.



УДК 004.5+004.77

*В.С. Заборовский, А.С. Кондратьев, В.А. Мулюха,
А.В. Силиненко, А.С. Ильяшенко*

УПРАВЛЕНИЕ ГРУППИРОВКАМИ НАПЛАНЕТНЫХ РОБОТОВ С БОРТА ПИЛОТИРУЕМОЙ ОРБИТАЛЬНОЙ СТАНЦИИ В РАМКАХ ПРОЕКТА «METERON»

Решение задач исследования космического пространства и планет Солнечной системы требует создания новых технологий управления удаленными объектами. В качестве таких объектов управления могут выступать как космические аппараты (пилотируемые и непилотируемые), так и роботы, а также группировки роботов, согласованно выполняющих технологические операции, связанные с манипуляциями, перемещениями и наблюдениями под управлением оператора [1]. В статье рассматривается использование сетецентрического подхода к задачам управления группировками робототехнических объектов (РО), управляемых с борта пилотируемой орбитальной станции.

В настоящее время существуют различные подходы к реализации сетецентрических принципов управления сложными системами, изложенные, в частности, в [2–4]. В настоящей статье под *сетецентричностью* будем понимать способ самоорганизации сложных технических систем на основе использования единого информационного пространства, обеспечивающий:

- режим ситуационной осведомленности объектов системы, основанный на непрерывной актуализации данных о состоянии компонентов и окружающей среды и формирующий целостное представление о системе для выработки управляющих воздействий;
- системную целостность и масштабируемость в условиях реализации многоцелевых операций за счет использования всеми объектами-участниками операций единых спецификаций для прикладных интерфейсов информационного взаимодействия;
- использование для целей управления как локальных, так и пространственно-удаленных информационно-вычислительных ресурсов, а также возможностей оператора.

Таким образом, носителями свойства сетецентричности можно считать методы идентифика-

ции состояния объектов-участников выполнения операций, например РО, и средства реализации механизмов обмена данными в сетевой среде, включая протоколы и интерфейсы межсетевого взаимодействия, обеспечивающие заданные характеристики качества информационного сервиса. В этих условиях сетецентричность обеспечивает возможность управления процессами информационного взаимодействия РО по интегральным критериям, используя для этого пространственные, временные, масштабные, тематические и контекстные характеристики.

1. Сетецентрический подход к управлению группировками роботов

Космическая робототехника – одно из важнейших направлений развития космонавтики, ее современный уровень развития позволяет решать широкий круг практических задач [5, 6], среди которых выделяют:

- проведение технического обслуживания спутников;
- сборку космических объектов на орбите;
- удаление техногенного мусора из околоземного космического пространства;
- исследование и освоение планет Солнечной системы.

Использование сетецентрического подхода при управлении космическими роботами представляется наиболее целесообразным применительно к исследовательским задачам, связанным с нахождением РО на поверхности планеты, а оператора – на борту пилотируемой орбитальной станции. Специфика такого рода космических операций заключается в постоянном движении орбитальной станции относительно поверхности планеты, что делает невозможным обеспечение непрерывной связи с географически локализованной группировкой напланетных роботов без использования дополнительного оборудования в виде спутников-ретрансляторов и/или развитой

сетевой инфраструктуры, включающей приемопередающие станции космической связи и объединяющие их телематические каналы. Это влечет за собой необходимость реализации в общем случае двух режимов управления группировкой РО:

супервизорного при наличии устойчивой связи с управляемым объектом, когда функционирование роботов происходит под контролем оператора, находящегося на борту орбитальной станции;

автономного при неустойчивой связи с управляемым объектом или ее отсутствии, когда группировка роботов должна обеспечивать выполнение заданной операции, полагаясь исключительно на собственный интеллект.

Наличие сетецентрических свойств у группировки напланетных роботов, функционирующих в автономном режиме, позволит обеспечить совместное выполнение поставленной задачи без непосредственного контроля со стороны оператора, что существенно увеличит эффективность выполнения технологических операций. В рамках применения данной идеи к задачам космической робототехники РО представляется удаленному клиенту (оператору или другому РО) в виде набора функций, доступных через известный интерфейс взаимодействия и позволяющих реализовать различные сценарии управления и координации совместных действий. В предложенной модели робот является частью сетецентрической инфраструктуры и представляется в виде сетевого сервиса в едином информационном пространстве, функционирующего по известным протоколам через универсальный интерфейс прикладного программирования (Application Programming Interface – API).

В свою очередь сетевая инфраструктура позволит создавать группировки – «кластеры» роботов. Группировка роботов – это совокупность РО, предназначенных для совместного выполнения заданий, функционирующих в едином информационном пространстве под управлением удаленного оператора или группы операторов, имеющих возможность формировать и динамически корректировать целевую функцию через предназначенные для этого человеко-машинные интерфейсы. Такие группировки взаимодействующих роботов являются перспективными техническими системами, повышающими безопасность и эффективность выполнения полетных заданий в условиях космического пространства [5].

Специфика решаемых в этом случае задач требует рассмотрения таких объектов управления, как автоматизированных систем, находящихся под контролем космонавта-оператора, который имеет возможность оперативно корректировать целевую функцию системы с учетом недетерминированности среды ее исполнения.

Одним из основополагающих принципов сетецентрического подхода в управлении группировкой роботов является представление РО как *сетевого сервиса*, предоставляемого оператору при помощи телематических каналов связи. Данный подход доказал свою эффективность в одноранговых сетях, когда удаленный клиент (пользователь или процесс) не нуждается в информации о специфике реализации конкретного сервиса, используя стандартные API, устройства ввода-вывода, сетевые интерфейсы и протоколы информационного обмена. При сетецентрическом подходе оператор или группа операторов, осуществляющих управление группировкой робототехнических объектов, должны использовать минимальное число устройств управления, стандартные сетевые интерфейсы и универсальные протоколы передачи данных и API. При этом каждый РО характеризуется своими функциональными возможностями и по-своему интерпретирует получаемые управляющие воздействия, а выбор форматов передаваемых сообщений должен быть прозрачным для пользователя и осуществляться программно-математическим обеспечением на борту орбитальной станции на основании данных, предоставленных РО, без непосредственного участия космонавта-оператора.

На базе группировки роботов может быть реализована *сеть робототехнических операций* (СРО). СРО является виртуальной средой функционирования РО, отвечающей за формирование последовательностей операций, реализующих заданную целевую функцию сети. Основным принципом организации СРО на базе группировки РО является ее реконфигурируемость, т. е. возможность изменения состава группировки РО в процессе ее функционирования для повышения эффективности работы СРО. При осуществлении напланетных операций должна существовать возможность добавлять РО в группировку для реализации целевой функции СРО и исключать неэффективные РО из группировки. Каждая напланетная группировка роботов должна включать в себя специализированный *планировщик*



операций, получающий команды управления непосредственно от оператора и осуществляющий планирование последовательностей простейших операций, которые необходимо совершить РО, входящим в состав СРО, для реализации целевой функции. При добавлении в группировку нового РО планировщик должен проанализировать его функциональные возможности на основании полученных от него данных и модифицировать последовательность простейших операций с целью оптимизации выполнения целевой функции.

При объединении РО в группировку при помощи СРО, проявляется свойство *эмерджентности* (системный эффект), когда у такой робототехнической системы возникают новые свойства, не присущие отдельным РО. При организации операций планировщик учитывает возможности роботов, входящих в состав СРО, а также анализирует вновь появляющиеся свойства объединенной группировки робототехнических объектов.

Для организации эффективного информационного обмена между компонентами СРО, внутри каждой напланетной группировки РО должна существовать возможность *приоритезации* процессов их информационного взаимодействия. Это необходимо в случае, когда в группировке имеется «ведущий» РО, отвечающий за управление всей группой, команды которого являются командами наивысшего приоритета. Доставка таких команд будет являться важнейшей задачей управления потоками данных внутри СРО. При создании сети робототехнических операций должна быть реализована возможность настройки приоритета потокам данных, которые создают РО в процессе функционирования в СРО. Например в большинстве режимов работы СРО видеосигнал должен иметь более низкий приоритет, чем команды управления, поэтому можно «пожертвовать» частью видеоданных для сохранения стабильности управляющих сигналов. С другой стороны, в процессе выполнения операций может возникнуть ситуация, при которой визуальная информация от камер наблюдения окажется наиболее важной для пользователя. Таким образом, в СРО должно осуществляться динамическое приоритетное управление потоками данных как на уровне устройств, так и на уровне сетевых виртуальных соединений.

Для математического описания такого подхода предлагается использовать модель теории приоритетных систем массового обслуживания

$\bar{M}_2 / M / 1 / k / f_2^1$, рассмотренную и подробно исследованную в работах [6, 7].

2. Технические возможности удаленного управления напланетными группировками РО с борта орбитальной станции

Одним из основных принципов управления группировки РО является использование комбинированного управления РО и оптимальное сочетание возможностей РО и оператора при выполнении операций. Для реализации данного принципа предлагается использовать комбинированные алгоритмы управления, распределенные между оператором и планировщиком операций, а также гетерогенные каналы связи, обеспечивающие информационный обмен между оператором на борту орбитальной станции и РО с определенными характеристиками. Для отработки данной технологии предлагается использовать следующие телематические каналы связи:

- канал связи *реального времени*, характеризующийся низкими значениями математического ожидания и дисперсии задержки в передаче данных, сравнительно невысокой пропускной способностью и небольшой продолжительностью сеанса связи, используемый преимущественно для телеуправления напланетными РО, а также широкополосный канал связи, характеризующийся большими, но стабильными задержками и высокой пропускной способностью, используемый, в основном, для передачи видео-, аудио- и других потоковых данных между оператором и группировкой напланетных роботов;

- канал связи, нечувствительный к задержкам и *гарантирующий факт доставки* сообщения, но имеющий высокие и нестабильные значения задержек передачи данных, а также сравнительно невысокую пропускную способность, используемый для передачи заданий на выполнение многоцелевых операций и результатов их выполнения между оператором и группировкой РО.

Организация группировок РО и устройств управления позволяет более эффективно использовать ресурсы РО и оператора, предоставляя возможности программного управления группой роботов, тем самым перекладывая выполнение простейших операций на планировщика и освобождая оператора для выполнения более важных задач. При этом необходимо отметить, что управляющая программа может быть как полностью

создана непосредственно самим космонавтом-оператором, так и заполнена на базе подготовленного заранее шаблона. Помимо программного управления существует возможность прямого телеуправления каждым из РО для решения сложных задач и выполнения ответственных операций в недетерминированных условиях. При этом для управления одной или несколькими группировками напланетных робототехнических устройств могут быть задействованы несколько территориально удаленных операторов. Данный подход позволяет оптимально сочетать возможности РО и оператора при реализации заданной операции.

Для отработки технологии выполнения многоцелевых операций группировкой напланетных роботов, управляемых с борта пилотируемой орбитальной станции, планируется организация специализированного космического эксперимента (КЭ).

3. Подход к реализации экспериментального решения по отработке технологии выполнения многоцелевых операций группировкой РО

Для изучения предлагаемого подхода предлагается проведение научно-прикладных исследова-

ний по отработке технологий выполнения многоцелевых операций напланетными роботами и устойчивыми группировками роботов, управляемых космонавтом-оператором с борта пилотируемой орбитальной станции [10, 11]. В рамках данного эксперимента должна осуществляться проработка одного из перспективных сценариев исследования планет Солнечной системы (Луна, Марс) при помощи напланетных интеллектуальных роботов, контролируемых группой удаленных операторов с орбитальной станции.

В соответствии с рис. 1 представлены основные компоненты и связи в системе управления группировками напланетных роботов в планируемом КЭ. В системе выделяется три функциональных уровня в соответствии с рис. 2:

- уровень *формирования заданий на выполнение многоцелевых операций*, находящийся в составе бортового сегмента и включающий оператора и автоматизированное рабочее место (АРМ) с набором задающих устройств;
- уровень *гетерогенной инфраструктуры* доступа, объединяющий сетевое и радиооборудование передачи данных бортового и напланетного сегментов;

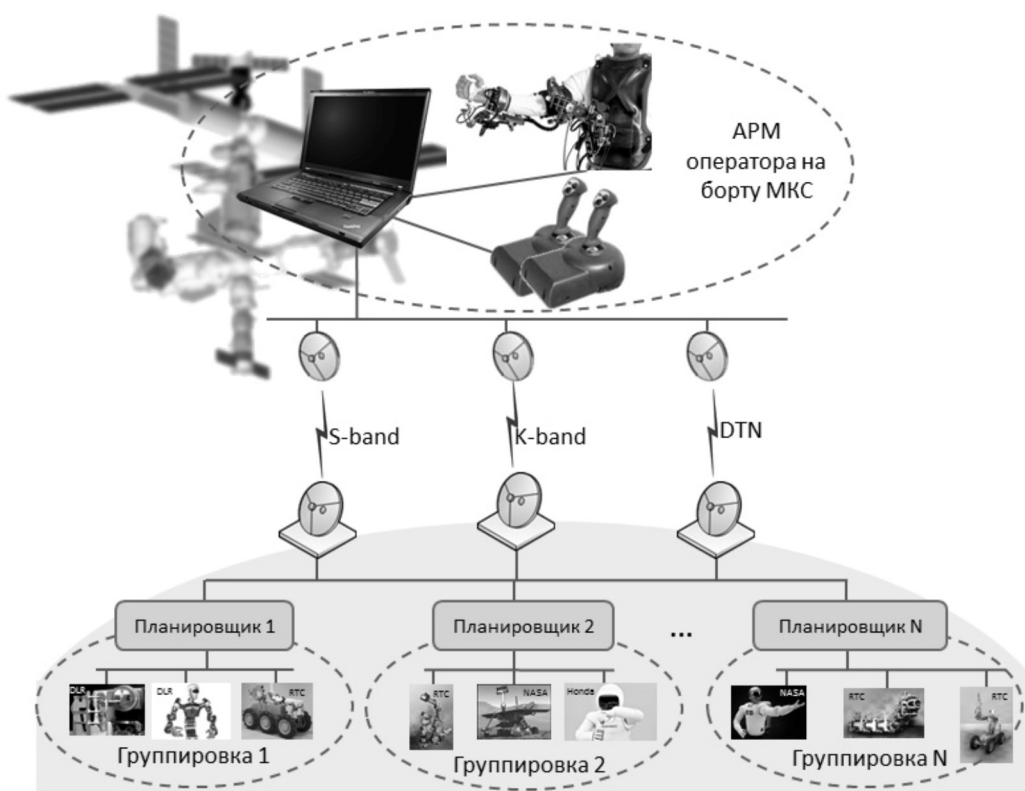


Рис. 1. Организация системы управления группировками напланетных роботов

- уровень *выполнения заданий*, включающий планировщиков и исполнительные устройства, объединенные в группировки роботов напланетного сегмента.

Задание для группировки напланетных роботов (многоцелевая операция) рассматривается как сложная, функционально завершенная операция, состоящая из комбинации целей, достижение которых связано с перемещением в пространстве, манипуляцией предметами и наблюдением за окружающими объектами. Сетецентрические свойства, реализованные в напланетных роботах, позволят объединять эти устройства в группировки, совместно выполняющие заданные технологические операции.

Операции могут выполняться как под контролем группы операторов в реальном масштабе времени, так и автономно под управлением пла-

нировщика заданий в зависимости от доступных в данный момент каналов связи между бортом орбитальной станции и напланетной сетевой инфраструктурой. Предполагается наличие нескольких таких каналов, отличающихся используемой технологией передачи данных, а также своими количественными и качественными характеристиками: пропускной способностью, задержками, вероятностью потерь пакетов данных, стоимостью и доступностью в заданное время.

Для управления напланетными роботами на борту орбитальной станции должен присутствовать набор задающих устройств, каждое из которых обеспечивает выполнение заданного перечня робототехнических операций с помощью роботов, входящих в группировку. Состав набора задающих устройств должен обеспечивать поддержку максимально широкого перечня робототехниче-

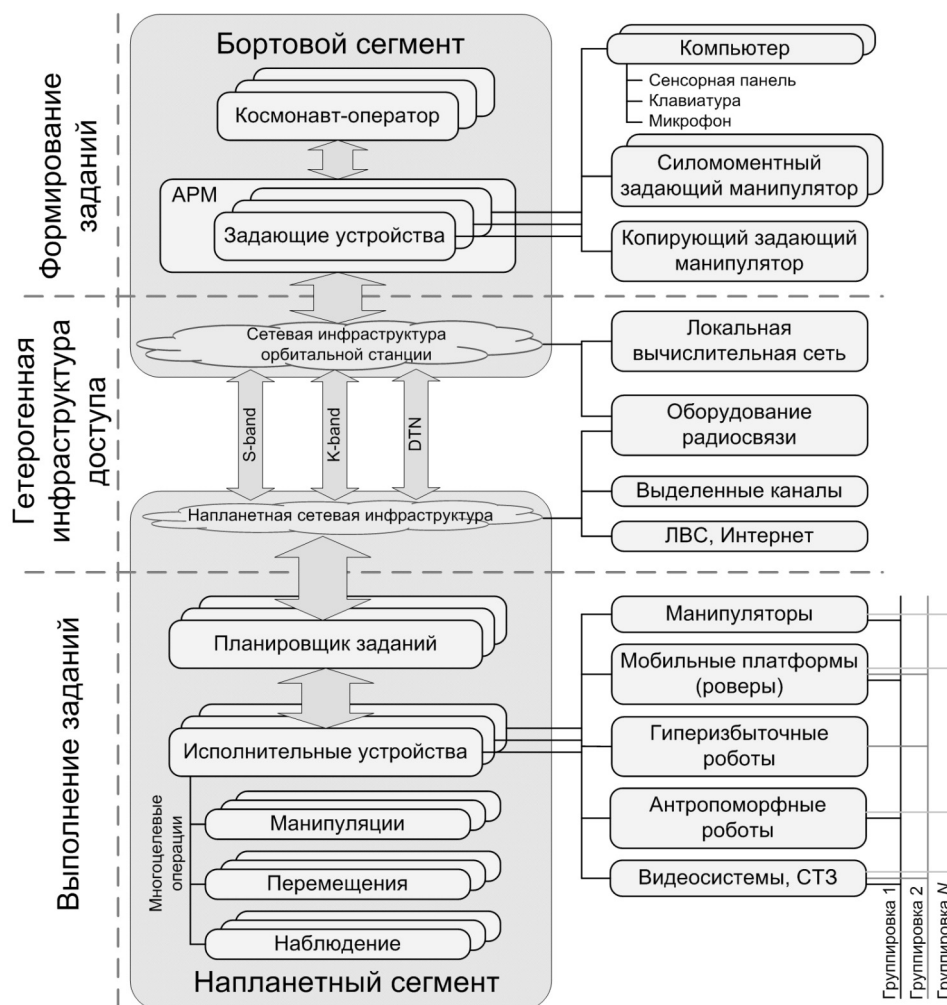


Рис. 2. Общая структура информационных и функциональных связей системы управления группировками напланетных роботов

ских операций, которые могут быть выполнены роботами. Целесообразно включение в этот набор оборудования, позволяющего с наибольшей реалистичностью «погрузить» оператора в среду функционирования исполнительных устройств, создав у оператора эффект присутствия на исследуемой планете.

Таким образом, актуальными задачами в настоящее время являются отработка и выбор оптимального состава задающих устройств, технологий и каналов связи, а также роботов для формирования группировок, позволяющих эффективно реализовывать сценарии исследовательских миссий в условиях недетерминированной среды малоизученных планет [11].

4. Реализация предложенных принципов управления: космический эксперимент «МЕТЕРОН-Р»

Для реализации предложенных принципов управления планируется проведение международного космического эксперимента (КЭ) «МЕТЕРОН-Р», целью которого является отработка технологии выполнения многоцелевых операций группировкой напланетных роботов, управляемых с борта пилотируемой орбитальной станции для решения задач исследования и освоения планет Солнечной системы. Инициаторами проведения эксперимента являются со стороны Российской Федерации:

Санкт-Петербургский государственный политехнический университет;

Центральный научно-исследовательский институт робототехники и технической кибернетики (Санкт-Петербург);

со стороны Евросоюза:

Делфтский технический университет (г. Делфт, Нидерланды);

Европейское космическое агентство (ЕКА).

С 2009 г. Европейское космическое агентство совместно с NASA проводит наземную отработку и подготовку к реализации проекта «METERON» (Multi-purpose End-To-End Robotic Operations Network), моделирующего сценарии исследования и освоения Луны и Марса с помощью напланетных роботов [12]. Космический эксперимент «МЕТЕРОН-Р» будет представлять собой российскую часть данного проекта, позволяющего консолидировать усилия групп европейских, американских и российских исследователей, со-

вместно использовать доступные ресурсы (каналы связи, задающие устройства, робототехнические объекты, разработанные протоколы), равно как и результаты, полученные при реализации данного КЭ.

Основным объектом исследований в КЭ «МЕТЕРОН-Р» являются технологии выполнения множества операций группировкой напланетных роботов (ГНР), управляемых с МКС. Под управлением ГНР понимается как телеуправление отдельными роботами из состава ГНР в случаях, когда это необходимо, так и реализация заданной целевой функции в режиме автономного функционирования ГНР под локальным управлением планировщика. В качестве объекта управления выступает ГНР, объединяющая группу робототехнических объектов, информационное взаимодействие которых обеспечивается при помощи планировщика.

Предметом исследования КЭ являются:

технологии обеспечения безопасности и эффективности информационного обмена между МКС и напланетными объектами с использованием существующих телекоммуникационных протоколов и каналов космической связи с различными характеристиками качества;

группировки напланетных роботов, предназначенных для решения задач исследования планет Солнечной системы;

комплексные универсальные интерактивные задающие устройства (в т. ч. с силомоментным очувствлением) и их использование в условиях микрогравитации.

Один из возможных сценариев управления в настоящем КЭ предполагает оперативное формирование и использование напланетной телекоммуникационной инфраструктуры на основе группировки мобильных и/или стационарных приемопередающих станций, обеспечивающих связь между роботами внутри ГНР. При этом связь ГНР с АРМ оператора на космической станции обеспечивается через специализированный планировщик, реализующий интерфейс между роботами, входящими в группировку, и каналами космической связи. Такая телекоммуникационная инфраструктура характеризуется рядом параметров (пропускная способность, задержка, длительность сеанса связи, вероятность потери пакетов данных и др.), значения которых могут существенно варьироваться во время сеанса управления в зависимости от типа канала и условий проведения сеансов.



В качестве задающих устройств предполагается использование в т. ч. устройств, изменяющих свои тактико-технические характеристики при их применении в условиях микрогравитации по сравнению с земными условиями, например, задающих манипуляторов с силовой связью.

В рамках КЭ предполагается исследовать возможность совместного использования трех различных типов каналов связи между АРМ оператора на борту МКС и ГНР:

канал связи в диапазоне S-band (канал реального времени), который характеризуется низкими значениями математического ожидания и дисперсии задержки в передаче данных (5–15 мс) и небольшой продолжительностью сеанса управления (8–10 мин);

канал связи на основе технологии Delay&Disruption Tolerant Network (DTN), который гарантирует факт доставки сообщения, но при этом имеет высокие и нестабильные значения задержек передачи данных;

канал связи в диапазоне K-band, который характеризуется большими, но стабильными значениями задержек и высокой пропускной способностью.

С учетом различного качества используемых каналов связи актуальными становятся задачи исследования влияния характеристик телекоммуникационных каналов на возможность эффективного управления ГНР с борта МКС, а также разработки методов адаптации управляющих сообщений к параметрам каналов связи на основе управления пропускной способностью и приоритетами виртуальных соединений.

Сетецентричность как способ самоорганизации представляется закономерным этапом эво-

люции сложных технических систем в условиях современного уровня развития информационных технологий. Использование сетецентрической парадигмы в робототехнике, когда робот является сервисом в едином информационном пространстве, позволит повысить эффективность выполнения операций в тех случаях, когда не представляется возможным обеспечить непрерывный контроль оператора над системой.

Целесообразно развитие сетецентрического подхода применительно к космическим роботам, решающим исследовательские задачи на поверхностях планет под управлением операторов, находящихся на борту орбитальной станции. Реализация данного подхода в планируемом космическом эксперименте «МЕТЕРОН-Р» позволит сделать следующее:

отработать стратегии управления, обеспечения информационной безопасности, человеко-машинные и прикладные программные интерфейсы для взаимодействия оператора на борту орбитальной станции и группировки напланетных роботов, а также роботов внутри группировки;

создать универсальный масштабируемый испытательный стенд для отработки технологий телеуправления роботами с борта МКС;

разработать архитектуру сети информационного взаимодействия, включающую комбинацию проводных и беспроводных телекоммуникационных каналов связи;

сформулировать требования к характеристикам интерфейсов управления, включая интерактивные механизмы формирования управляющих воздействий, учитывающие свойства мехатронных робототехнических объектов и среды передачи управляющей информации.

Работа выполнена при финансовой поддержке Министерства образования и науки РФ.

СПИСОК ЛИТЕРАТУРЫ

1. Интеллектуальные роботы [Текст] / Под ред. Е.И. Юревича. – М.: Машиностроение, 2007. – 360 с.
2. Wilson, C. Network Centric Operations: Background and Oversight Issues for Congress [Электронный ресурс] / C. Wilson // CRS Report for Congress. – Updated March 15, 2007. – 55 p.
3. Затуливетер, Ю.С. Компьютерный базис сетецентрического управления [Текст] / Ю.С. Затуливетер // Тр. конф. Технические и программные средства систем управления, контроля и измерения. – М., 2010. – С. 492–511.
4. Каляев, И.А. Самоорганизующиеся распре-

ленные системы управления группами интеллектуальных роботов, построенные на основе сетевой модели [Текст] / И.А. Каляев, С.Г. Капустян, А.Р. Гайдук // Управление большими системами. Спецвыпуск 30.1 Сетевые модели в управлении. – М.: ИПУ РАН, 2010. – С. 605–639.

5. Юревич, Е.И. Космическая робототехника: состояние и перспективы развития [Текст] / Е.И. Юревич // Rational Enterprise Management. – 04.2011. – С. 24–26.

6. Минаков, Е.П. Концепция развития робототехнических систем в интересах пилотируемой космонавтики, исследования Луны и планет Солнечной систе-

мы [Текст] / Е.П. Минаков, В.А. Лопота, Е.И. Юревич, А.С. Кондратьев // Матер. XX Междунар. науч.-техн. конф. Экстремальная робототехника. Нано- микро- и макророботы (ЭР-2009). – Таганрог: Изд-во ТТИ ЮФУ, 2009. – С. 103–106.

7. **Бабкин, Е.В.** Супервизорное управление космическим роботом на Международной космической станции (МКС) с использованием сети Интернет [Текст] / Е.В. Бабкин, М.Ю. Беляев, И.А. Васильев, В.С. Заборовский, В.П. Макарычев, А.В. Силиненко, К. Ландцеттель // Матер. Всерос. молодежной конф. Экстремальная робототехника. – 2011. – С. 32–42.

8. **Заяц, О.И.** Управление пакетными коммутациями в телематических устройствах с ограниченным буфером при использовании абсолютного приоритета и вероятностного выталкивающего механизма. Ч. 1 [Текст] / О.И. Заяц, В.С. Заборовский, В.А. Мулюха, А.С. Вербенко // Новые технологии. Программная инженерия. – 2012. – № 2. – С. 22–29.

9. **Заяц, О.И.** Управление пакетными коммутациями в телематических устройствах с ограниченным буфером при использовании абсолютного приоритета

и вероятностного выталкивающего механизма. Ч. 2 [Текст] / О.И. Заяц, В.С. Заборовский, В.А. Мулюха, А.С. Вербенко // Новые технологии. Программная инженерия. – 2012. – № 3. – С. 21–29.

10. **Заборовский, В.С.** Проведение космических экспериментов с использованием сети многоцелевых робототехнических операций [Текст] / В.С. Заборовский, В.А. Мулюха, А.В. Силиненко, А.С. Ильяшенко // Матер IV Всерос. мультikonф. по проблемам управления. – Таганрог: Изд-во ТТИ ЮФУ, 2011. – Т. 1. – 370 с.

11. **Заборовский, В.С.** Перспективные задачи космических экспериментов на борту РС МКС в области создания многоцелевой сети робототехнических операций [Текст] / В.С. Заборовский, В.А. Мулюха, А.В. Силиненко, А.С. Ильяшенко, А. Шиле // К.Э. Циолковский и будущее космонавтики: Матер. XLVI Научных чтений памяти К.Э. Циолковского. – Калуга: Изд-во «Эйдос», – 2011. – 97 с.

12. **Nergaard, Kim.** METERON. Conducting robotic operations and experiments from orbit [Text] / Kim Nergaard [et al.] // ESA Bulletin. – Aug. 2011. – № 147. – P. 22–29.

УДК 621.391.037.3

А.Б. Кислицын, А.В. Рашич

ПЕРЕДАЧА И ПРИЕМ СИГНАЛОВ С DOQPSK И CDOQPSK С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА ВИТЕРБИ

При передаче сообщений с помощью сигналов с DOQPSK (Differential Offset Quadrature Shift Keying) обеспечивается сравнительно малое значение пик-фактора излучаемых колебаний. Кроме того, при использовании таких сигналов появляется возможность осуществлять прием без фазовой и частотной синхронизации в широком диапазоне величин отстройки. Применение вида манипуляции CDOQPSK (Constant envelope Differential Offset Quadrature Shift Keying) обеспечивает значение пик-фактора колебаний 0 дБ, что позволяет использовать усилители мощности передатчиков с эффективностью, близкой к максимально достижимой. Сигнал CDOQPSK формируется в результате амплитудного клиппирования сигнала с DOQPSK после его цифровой фильтрации на выходе передающего устройства.

Процедура формирования сигналов с CDOQPSK (устранение девиаций вещественной

огибающей) приводит к двукратному расширению спектра (по сравнению со спектром сигнала с DOQPSK) для уровня излучения менее –40 дБ. В то же время данный эффект не увеличивает шумовую полосу сигнала вследствие применения одинаковых входных фильтров как в процессе приема сигналов с DOQPSK, так и в процессе приема сигналов с CDOQPSK.

Недостатком сигналов с DOQPSK и CDOQPSK является их малая энергетическая эффективность, что обусловлено интерференцией, вносимой выходным фильтром, относительной задержкой модуляции синфазной и квадратурной компонент сигнала, а также выполнением дифференцирования точек максимального эффекта сигнала в процессе его приема. При некогерентной поэлементной демодуляции сигналов с DOQPSK в канале с аддитивным белым гауссовским шумом (АБГШ) вероятность ошибки на бит 10^{-5} достигается для

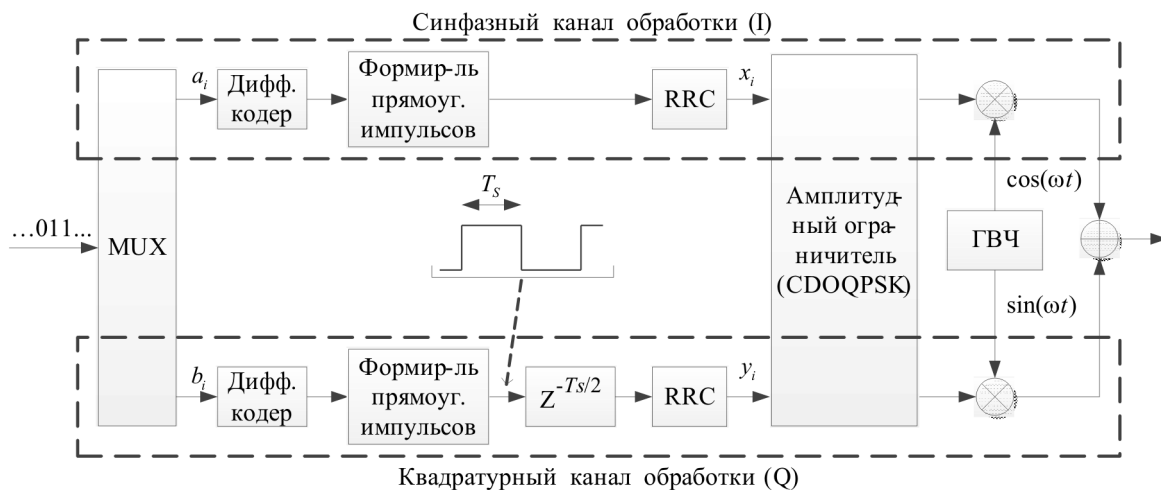


Рис. 1. Структурная схема модулятора сигналов DOQPSK и CDOQPSK

величины энергии на бит не менее 19 дБ.

Цель работы – повышение помехоустойчивости приема сигналов с DOQPSK путем использования приема «в целом», эффективная реализация которого обеспечивается применением алгоритма Витерби [1].

Схемы модулятора и устройства цифровой обработки сигналов

Модулятор сигнала (рис. 1) строится на основе формирователя сигналов с OQPSK с дополнительными блоками дифференциальных кодеров в каждом из квадратурных каналов, а также амплитудного ограничителя для устранения девиаций вещественной огибающей. Амплитудный ограничитель используется только при формировании сигналов с видом манипуляции CDOQPSK. Заданное значение ширины спектра и уровня внеполосных излучений обеспечивают фильтры низких частот (ФНЧ) с характеристикой корня из приподнятого косинуса. Сигналы с видами манипуляции DOQPSK и CDOQPSK формируются посредством одной и той же схемы в зависимости от режима работы блока амплитудного ограничения.

Процедура формирования сигналов с видами манипуляции DOQPSK или CDOQPSK включает несколько этапов. Поток канальных символов, состоящий из двоичных элементов 0 и 1, подается на мультиплексор (блок «MUX» на рис. 1). Данное устройство осуществляет распределение четных канальных символов a_i в синфазный канал обработки I, а нечетных b_i – в противофазный канал обработки Q. После распределения двоичные элементы поступают на блоки дифференциального кодирования (блоки «Дифф. кодер» на рис. 1), осуществляющие операцию бинарного интегрирования потока (рис. 2).

Канальные символы с выхода блоков дифференциального кодирования поступают на вход формирователей прямоугольных импульсов (блоки «Форм-ль прямоуг. импульсов» на рис. 1). При этом элементу 0 ставится в соответствие прямоугольный импульс длительностью T_s положительной полярности, а элементу 1 – прямоугольный импульс длительностью T_s отрицательной полярности. Сформированный видеосигнал в квадратурном канале обработки Q задерживается относительно видеосигнала в синфазном канале I на половину длительности импульса $T_s/2$.

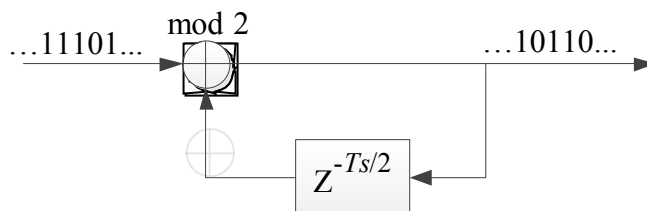


Рис. 2. Схема дифференциального кодера канальных символов

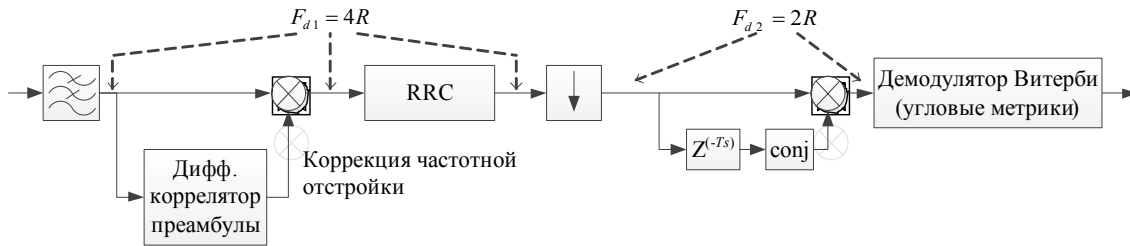


Рис. 3. Структурная схема устройства цифровой обработки сигналов на приемной стороне

Каждый из видеосигналов поступает на цифровой фильтр, имеющий передаточную характеристику вида «корень из приподнятого косинуса» (Root-Raised Cosine – RRC). Амплитудный ограничитель сигнала (блок «Амплитудный ограничитель (CDOQPSK)» на рис. 1) применяется для осуществления перехода от сигнала с DOQPSK к сигналу с CDOQPSK посредством осуществления клиппирования в соответствии с выражением $\hat{z}_i = \frac{z_i}{|z_i|}$, где $z_i = x_i + j \cdot y_i$ – i -й комплексный отсчет сигнала на выходе демодулятора, образуемый x_i – i -м отсчетом с выхода фильтра в синфазном канале обработки и y_i – i -м отсчетом с выхода фильтра в квадратурном канале обработки; $\hat{z}_i$ – i -й комплексный отсчет сигнала на выходе амплитудного ограничителя.

Последним этапом при формировании сигнала является перенос его спектра на несущую. Данная операция схематически изображена на рис. 1 модуляцией квадратурными низкочастотными колебаниями выхода генератора высоких частот (ГВЧ) и последующим сложением полученных сигналов.

На рис. 3 представлена структурная схема устройства цифровой обработки сигналов на приемной стороне.

Частота дискретизации F_{d1} сигнала, поступающего с входного ФНЧ на схему дифференциальной корреляции преамбулы (блок «Дифф. коррелятор преамбулы»), равна учетверенной бодовой скорости: $F_{d1} = 4R$. С помощью данной схемы решаются задачи обнаружения сигнала, тактовой

синхронизации, а также грубой оценки и коррекции частотной отстройки.

Перед поступлением на демодулятор Витерби поток комплексных отсчетов сигнала с частотой дискретизации, равной удвоенной бодовой скорости $F_{d2} = 2R$, подается на схему через тактового дифференцирования (рис. 4). Необходимость в применении данного блока обусловлена выбором некогерентного алгоритма приема.

Сигнальные отсчеты, получаемые с выхода блока дифференцирования на интервале $2k$ -го такта $\hat{x}_{2k}^{dif} + j \hat{y}_{2k}^{dif}$, связаны с подаваемыми на него отсчетами $\hat{x}_{2k} + j \hat{y}_{2k}$ следующим соотношением:

$$\hat{x}_{2k}^{dif} + j \hat{y}_{2k}^{dif} = (\hat{x}_{2k} + j \hat{y}_{2k}) \cdot (\hat{x}_{2k} - j \hat{y}_{2k}).$$

Конфигурация сигнального созвездия (рис. 5), формируемого отсчетами на выходе устройства дифференцирования, не зависит от начальной фазы обрабатываемого сигнала, а также от величины остаточной (после грубой коррекции) частотной отстройки. Как показано на рисунке, каждой точке дифференциального созвездия соответствует от двух до четырех комбинаций канальных битов, состоящих из четырех элементов. Данное обстоятельство обусловлено несколькими факторами:

применением видов манипуляции DOQPSK и CDOQPSK;

применением на передающей и приемной стороне RRC-фильтров, имеющих полосу пропускания, равную бодовой скорости;

выполнением через-тактового дифференцирования последовательности сигнальных отсчетов.

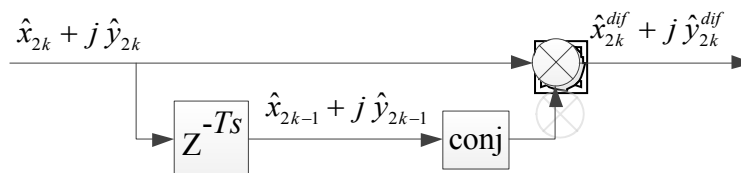


Рис. 4. Структурная схема блока через-тактового дифференцирования сигнальных отсчетов

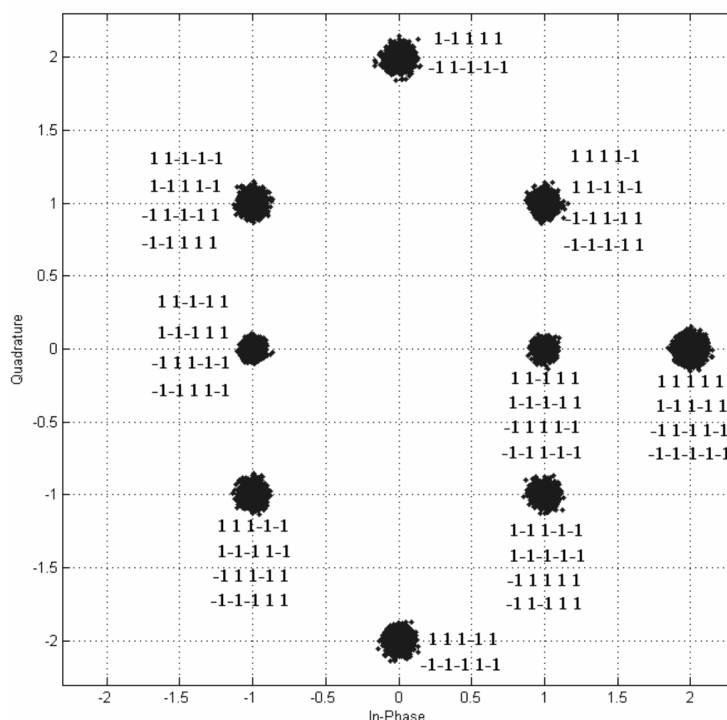


Рис. 5. Конфигурация сигнального созвездия на выходе блока дифференцирования при коэффициенте сглаживания применяемых RRC фильтров, равном единице

Демодулятор Витерби

В процессе обработки сигнала временной переход от одной точки созвездия к другой не означает смену комбинации полностью [2, 3]. Последовательность, соответствующая следующей точке, содержит три из четырех канальных символов предыдущей точки.

Таким образом, переход от одной точки дифференциального сигнального созвездия к другой в процессе приема определяется одним канальным символом. При этом схема, включающая модулятор DOQPSK (CDOQPSK), RRC-фильтр приемника, а также блок через-тактового дифференцирования, может рассматриваться в качестве единого устройства – эквивалентного кодера со скоростью $r = 1/1$, кодовым ограничением $K = 5$ и кодовыми символами на выходе, представляющими собой точки на сигнальной плоскости (рис. 6).

Область значений кодовых символов на выходе эквивалентной схемы при отсутствии АБГШ и воздействия замираний определяется набором сигнальных точек дифференциального сигнального созвездия DOQPSK. Решетка эквивалентного кода (рис. 7) включает на каждом шаге кодирования (декодирования) $2^{K-1} = 16$ узлов и два ребра, исходящих из каждого узла. Общее количество возможных переходов на каждом шаге равно 32.

Схема демодулятора Витерби для сигналов DOQPSK и CDOQPSK реализуется в соответствии с принципами построения декодеров Витерби. Алгоритм строится на основе описанной решетки эквивалентного кода. В качестве метрики Δ_{2k} используется угловое расстояние по модулю 2π , вычисляемое как разность между фазой идеальных точек дифференциального созвездия $s^{i, dif}$ и фазой сигнальных отсчетов, поступающих на вход демодулятора $\hat{z}_{2k}^{dif} = \hat{x}_{2k}^{dif} + j \cdot \hat{y}_{2k}^{dif}$:

$$\Delta_{2k} = \begin{cases} |\angle(s^{i, dif}) - \angle(\hat{z}_{2k}^{dif})|, & |\angle(s^{i, dif}) - \angle(\hat{z}_{2k}^{dif})| < \pi \\ 2\pi - |\angle(s^{i, dif}) - \angle(\hat{z}_{2k}^{dif})|, & |\angle(s^{i, dif}) - \angle(\hat{z}_{2k}^{dif})| \geq \pi \end{cases}$$

Выбор угловых расстояний вместо евклидовых обеспечивает независимость результатов

работы демодулятора от девиаций вещественной огибающей сигнала, что позволяет отказаться от

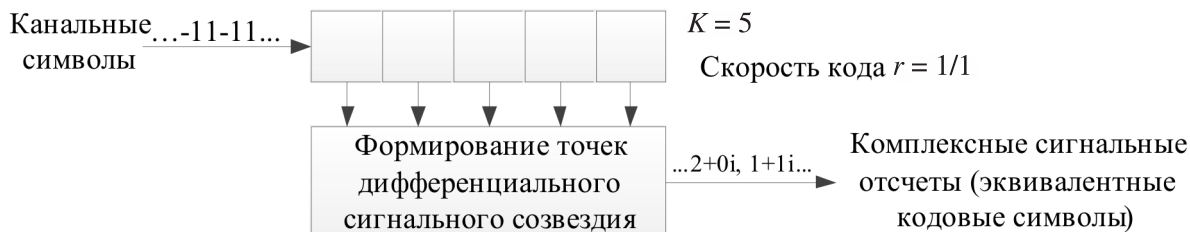


Рис. 6. Схема эквивалентного кодера

реализации вычислительно сложных схем автоматической регулировки усиления (АРУ).

В целях повышения помехоустойчивости приема с выхода демодулятора поступают мягкие решения, вычисляемые в соответствии с классическим алгоритмом SOVA (Soft Output Viterbi Algorithm).

На рис. 8 представлены полученные в результате имитационного моделирования кривые помехоустойчивости. Из характеристик видно, что осуществление приема «в целом» для сигналов DOQPSK с использованием разработанного демодулятора Витерби позволяет получить энергетический выигрыш по сравнению с поэлементным приемом до 7,5 дБ по уровню вероятности ошибки на бит 10^{-5} . При этом достигается потенциальная помехоустойчивость приема сигналов с $\pi/4$ -DQPSK. Учитывая значения пик-фактора, переход от вида манипуляции $\pi/4$ -DQPSK к DOQPSK в сочетании с предложенной схемой приема позволяет получить суммарный энергетический выигрыш до 3,4 дБ.

Характеристики помехоустойчивости некогерентного приема сигналов с CDOQPSK, а также когерентного и некогерентного приема сигналов с DOQPSK представлены на рис. 9. При

обработке сигналов с DOQPSK и CDOQPSK применяется одна и та же схема приема, что значительно упрощает реализацию радиоэлектронных устройств. Как видно из представленных графиков, применение одного алгоритма для обработки всех излучений в системе связи приводит к относительно небольшому (не более 0,5 дБ) энергетическому проигрышу приема сигналов с CDOQPSK по сравнению с приемом сигналов с DOQPSK при вероятности ошибки на бит 10^{-5} . Однако, учитывая значения пик-фактора, суммарный энергетический выигрыш, обеспечиваемый переходом от вида манипуляции CDOQPSK к DOQPSK, составляет не менее 1,6 дБ. Энергетический проигрыш при некогерентном приеме сигналов DOQPSK по сравнению с когерентным приемом не превышает 1 дБ (по уровню вероятности ошибки 10^{-5}), что также показывает эффективность разработанной схемы демодулятора Витерби.

Таким образом, используемые в разработанной сигнально-кодовой конструкции виды манипуляции DOQPSK и CDOQPSK обеспечивают относительно небольшое значение пик-фактора (от 0 до 2,1 дБ). Для приема данных сигналов не требуется осуществление фазовой синхронизации, а также частотной синхронизации в широком

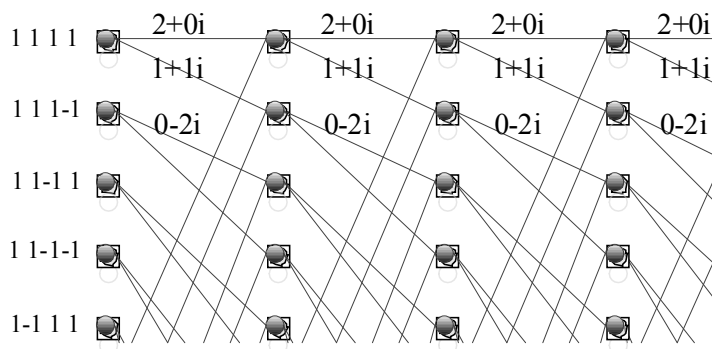


Рис. 7. Фрагмент решетки эквивалентного кода

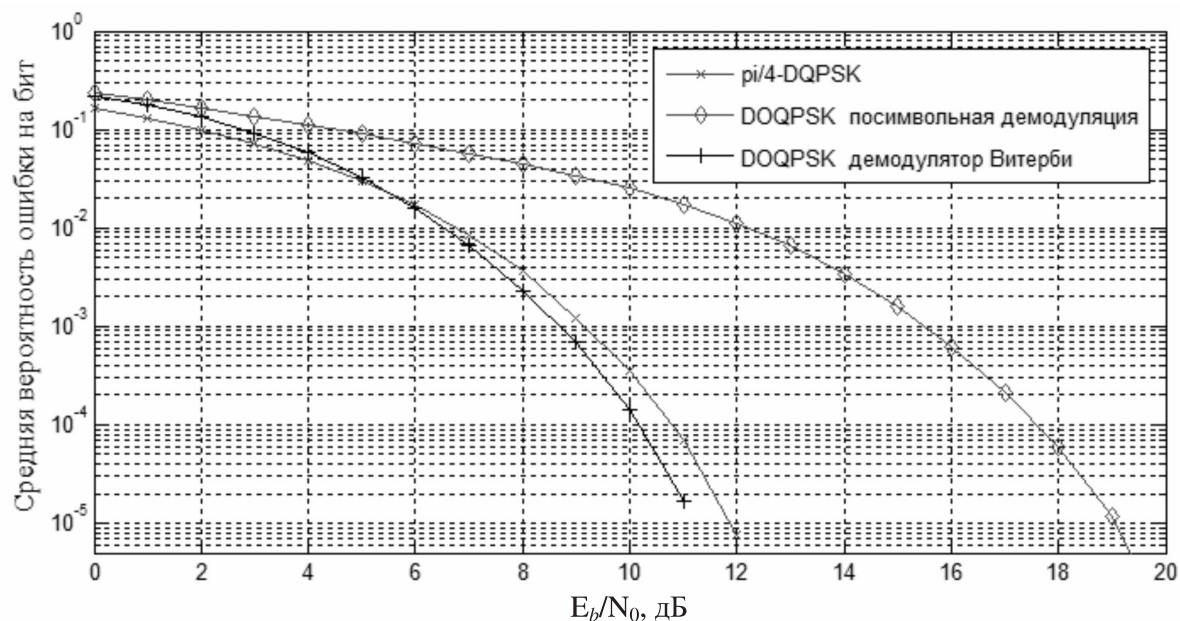


Рис. 8. Характеристики помехоустойчивости некогерентного приема сигналов с видами манипуляции $\pi/4$ -DQPSK и DOQPSK в случае осуществления посимвольной демодуляции и демодуляции по Витерби

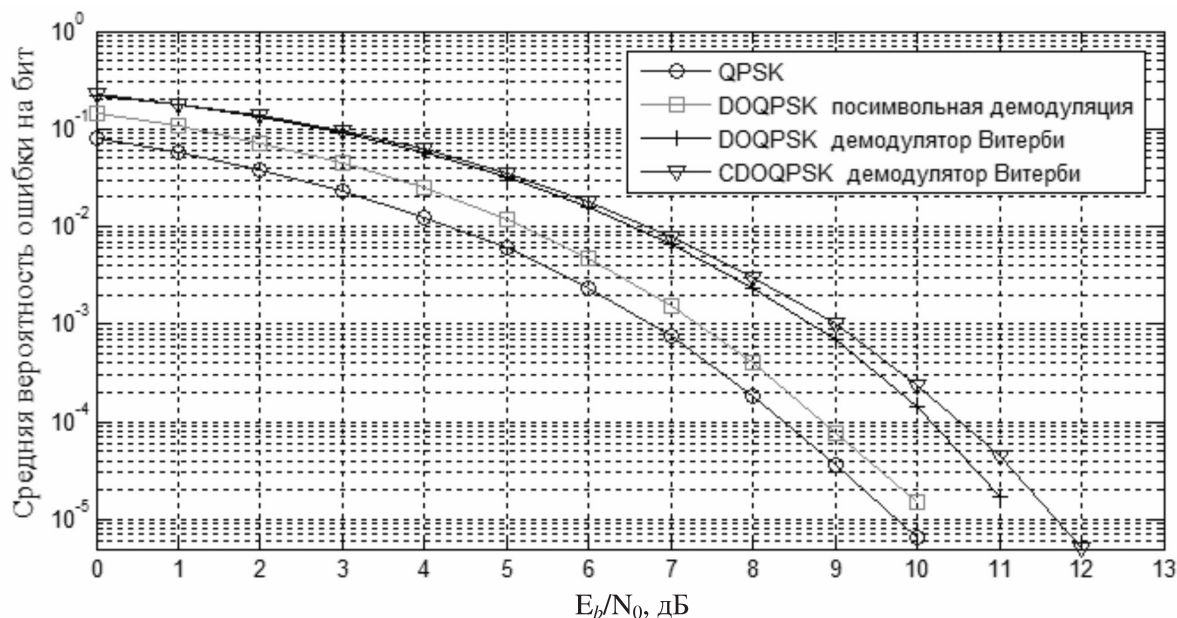


Рис. 9. Характеристики помехоустойчивости некогерентного приема сигналов с видом манипуляции CDOQPSK, а также когерентного и некогерентного приема сигналов DOQPSK

диапазоне величин отстройки, что обеспечивает работоспособность системы при небольших значениях отношения сигнал-шум на отсчет (менее 5 дБ). Помехоустойчивость приема данных сигналов в условиях АБГШ при использовании разработанного демодулятора Витерби соответствует

потенциальной помехоустойчивости сигналов с видом манипуляции $\pi/4$ -DQPSK. В таблице представлены значения энергии на бит, требуемые для достижения средней вероятности ошибки на бит 10^{-5} , при использовании различных алгоритмов некогерентного приема сигналов с четырехпо-

Значения пик-фактора и энергии на бит при использовании различных алгоритмов некогерентного приема сигналов с четырехпозиционными видами манипуляции ($BER = 10^{-5}$)

Метод манипуляции	DQPSK	$\pi/4$ -DQPSK	DOQPSK	DOQPSK	CDOQPSK
Алгоритм приема	Некогерентный поэлементный, дБ	Некогерентный поэлементный, дБ	Некогерентный поэлементный, дБ	Некогерентный прием «в целом», дБ	Некогерентный прием «в целом», дБ
E_b/N_0 ($BER = 10^{-5}$)	12,0	12,0	19,0	11,2	11,7
Пик-фактор	3,6	3,4	2,1	2,1	0
Суммарное значение	15,6	15,4	21,1	13,3	11,7

зиционными видами манипуляции. Суммарные значения энергии на бит и пик-фактора, представленные в таблице, могут рассматриваться в качестве критерия для грубой оценки эффективности применения различных видов манипуляции и схем приема. Как видно, значения данного

показателя для методов манипуляции DOQPSK и CDOQPSK при использовании разработанного демодулятора являются самыми низкими среди всех рассмотренных.

Работа выполнена при финансовой поддержке гранта Президента Российской Федерации МК-2612.2012.9 за 2012 г.

СПИСОК ЛИТЕРАТУРЫ

1. **Макаров, С.Б.** Передача дискретных сообщений по радиоканалам с ограниченной полосой пропускания [Текст] / С.Б. Макаров, И.А. Цикин. – М.: Радио и связь, 1988. – 304 с.
2. **Рашич, А.В.** Снижение пик-фактора сигналов с ортогональным частотным уплотнением [Текст] / А.В. Рашич, С.Б. Макаров // Научно-технические ведомости СПбГПУ. – СПб.: Изд-во Политехнического университета. – 2010. – № 3 (101). – С. 32–39.

мости СПбГПУ. – СПб.: Изд-во Политехнического университета. – 2008. – № 2 (55). – С. 79–84.

3. **Рашич, А.В.** Алгоритм приема спектрально-эффективных сигналов с OFDM [Текст] / А.В. Рашич, С.Б. Макаров, С. Атик // Научно-технические ведомости СПбГПУ. – СПб.: Изд-во Политехнического университета. – 2010. – № 3 (101). – С. 32–39.

О РАСЧЕТЕ ВЕРХНЕЙ ГРАНИЦЫ ЧУВСТВИТЕЛЬНОСТИ ИНТЕРФЕРЕНЦИОННЫХ ВОЛОКОННО-ОПТИЧЕСКИХ ГИДРОФОНОВ ПРИЕМНОЙ ГИДРОАКУСТИЧЕСКОЙ АНТЕННЫ

В течение двух последних десятилетий в качестве перспективного направления проектирования компонентов гидроакустических антенн рассматриваются волоконно-оптические интерферометрические гидрофоны (ВОИГ) [1–3]. Принцип действия ВОИГ основан на том, что падающие на них акустические сигналы изменяют фазу света, распространяющегося в волоконном световоде, который размещен в чувствительном элементе. Порожденные акустическим сигналом флуктуации фазы в интерферометрической схеме преобразуются в фазомодулированные (ФМ) колебания интенсивности, которые регистрируются фотодетектором и обрабатываются с целью получения сигнала, пропорционального принимаемому акустическому сигналу [4].

Применение волоконно-оптических интерферометрических технологий для построения гидроакустических антенн связано с целым рядом преимуществ (низкая отражательная способность антенны, ее малый вес, в перспективе дешевизна и др.). С другой стороны, при использовании таких технологий необходимо пересмотреть и заново проанализировать систему взаимосвязи параметров элементов конструкции, обрабатываемых сигналов и характеристик гидроакустической антенны.

Несмотря на то, что в публикациях по данной тематике приведен серьезный теоретический и эмпирический анализ многих важных характеристик ВОИГ для гидроакустических антенн [3, 5], некоторые проблемы, актуальные для разработки и изготовления практических устройств, рассмотрены не в полной мере. Один из таких вопросов связан с ограничениями на чувствительность ВОИГ, обусловленными необходимостью обе-

спечения требуемого динамического диапазона антенны.

Эффективность функционирования гидроакустических комплексов (ГАК) в значительной степени определяется корректностью выбора совокупности параметров гидрофонов антенны и тракта предварительной обработки сигналов (ПОС). Далее для краткости совокупность последовательно соединенных гидрофонов и тракта ПОС называем *трактом*. Под трактом ПОС понимается совокупность аппаратных средств, которая разделяется на следующие две части:

оптическую, представляющую собой совокупность лазера, модуляторов, волоконно-оптической линии, опорного интерферометра и фотоприемника, реализующих функции формирования оптических сигналов, интерференционной схемы обработки оптических сигналов, прошедших через группу гидрофонов, и преобразования оптического сигнала в высокочастотный фазомодулированный многоканальный сигнал;

цифровую, представляющую собой совокупность программно-аппаратных средств, решающих задачу преобразования аналоговых высокочастотных фазомодулированных многоканальных сигналов с выхода фотоприемника в цифровые низкочастотные сигналы, аналогичные сигналам от традиционных (пьезокерамических) гидрофонов.

Цель настоящей статьи – решение задачи (разработка методики) расчета верхней границы чувствительности ВОИГ по критерию достижения максимального динамического диапазона тракта. Очевидно, что при повышении чувствительности ВОИГ при фиксированной требуемой величине динамического диапазона снижаются требова-

ния к отношению сигнал/шум на входе фотоприемника, т. е. к качеству оптической компоненты тракта ПОС. Эта чувствительность ограничена сверху эффектом нелинейности фазовой модуляции. Превышение чувствительностью гидрофона (произвольного, а не только ВОИГ) допустимой верхней границы приводит к соответствующему снижению верхней границы динамического диапазона тракта. Таким образом, в качестве оптимальной, в свете указанного критерия, корректен выбор чувствительности, являющейся предельной при указанном ограничении.

1. Состояние вопросов построения интерферометрических схем с мультиплексированием и выбора верхней границы чувствительности ВОИГ

Для возможности размещения группы гидрофонов на одном оптоволокне реализуется их мультиплексирование. Для ВОИГ востребованной является, в частности, технология временно-го мультиплексирования, подразумевающая, что информация от каждого гидрофона группы позиционируется в специально выделенном интервале времени внутри периода опроса гидрофонов [1–3, 5]. Следует заметить, что частота опроса гидрофонов (f_o) может быть многократно меньше частоты дискретизации (f_d) упомянутых выше высокочастотных ФМ многоканальных сигналов в АЦП. Соотношение указанных частот влияет на требования к уровню спектральной плотности фазовых шумов тракта и количеству корректных разрядов АЦП. Анализ этих требований выходит за рамки настоящей статьи.

Для обеспечения работы приемного тракта ГАК при заданных требованиях к минимальному уровню принимаемого акустического сигнала и его динамическому диапазону необходимо рассчитать оптимальное сочетание таких параметров, как чувствительность гидрофона γ (радиан/Па) [4] и частота опроса каждого гидрофона f_o (Гц). Как правило, в публикациях, связанных с разработкой чувствительных элементов для волоконно-оптических интерферометрических акустических устройств, анализируются возможности увеличения чувствительности γ (что улучшает минимальное регистрируемое звуковое давление), границы и неравномерность рабочего диапазона частотной зависимости $\gamma(f)$, паразитная чувствительность к вибрациям и ускорениям и т. п. [4, 6–10]. Однако при заданной верхней границе динамического

диапазона величина чувствительности гидрофона ограничена сверху. В связи с этим актуальна задача разработки методики расчета соответствующей верхней границы чувствительности.

В работе [5] вопрос корректного выбора чувствительности ВОИГ решен следующим образом. Отмечено, что ширина спектра, поступающего на демодулятор ФМ сигнала в системе с дискретным временем, не должна превышать величину частоты Найквиста (заметим, что в данном случае частота Найквиста равна половине частоты опроса ВОИГ при их временном мультиплексировании; заметим также, что для узкополосных сигналов в этом условии следовало бы вместо частоты Найквиста рассматривать окно Найквиста). Использовано оценочное соотношение для *эффективной* ширины полосы ФМ сигнала при монохроматической модулирующей функции. На основе этих данных установлены соотношения между частотой Найквиста и предельным значением индекса модуляции и соответствующим ему предельным значением чувствительности ВОИГ.

В связи с указанными материалами работы [5] следует заметить следующее. В общем случае подвергнутый ограничению по полосе частот фазомодулированный сигнал и его спектр являются сложными в аналитическом описании, особенно при дискретизации этого сигнала по времени. Поэтому описание нелинейной операции фазового детектирования такого сигнала и аналитическое рассмотрение результирующих искажений представляется проблематичным. Ширина полосы частот ФМ сигнала, строго говоря, бесконечна даже при сколь угодно малом индексе модуляции. Поэтому как бы мы не определили *эффективную* ширину полосы частот ФМ сигнала, описанное в [5] решение задачи является заведомо приближенным, причем с неизвестной степенью приближения (ведь эффект нелинейности тракта, обусловленной отличием *эффективной* ширины полосы ФМ сигнала от фактически имеющей место *бесконечной*, нигде не исследован). Кроме того, вопрос динамического диапазона тракта актуален для широкополосного акустического сигнала в большей степени, чем для монохроматического. Используемая в [5] методика расчетов корректного соотношения частоты опроса ВОИГ и их предельной чувствительности на случай приема широкополосного сигнала (тем более с заданным энергетическим спектром) не может быть обобщена.



Для получения ответов на поставленные вопросы в данной статье предлагается использовать методику, основанную на прямых расчетах уровней, получаемых в результате детектирования сигналов и их сопоставления с уровнями исходных модулирующих сигналов.

2. Описание алгоритма фазовой демодуляции сигнала

Интерференционные колебания интенсивности, регистрируемые фотоприемником, задаются квазигармонической функцией от акустического входного воздействия $S_{\text{ак}}(t)$, модулирующего фазу света. Поэтому при построении практических волоконно-оптических интерферометрических измерительных устройств необходимо предпринимать специальные меры, позволяющие сформировать акустическое входное воздействие $S_{\text{ак}}(t)$ в результате обработки сигнала фотоприемного устройства. В случае разработки ВОИГ, особенно систем с мультиплексированием чувствительных элементов, как правило, используют методы с фазомодулированной несущей [1–3]. При этом за счет вспомогательной модуляции (манипуляции) разности фаз интерферирующих лучей формируют интерференционный сигнал вида

$$y(t) = \sin(2\pi f_n t + \gamma S_{\text{ак}}(t)), \quad (1)$$

где f_n – частота несущего колебания (неинформативная начальная фаза колебания опущена).

Дальнейшая обработка сводится к тем или иным стандартным способам фазового детектирования, выделяющего сигнал $S_{\text{ак}}(t)$. Уточним далее вид этой процедуры. В литературе рассматривается много алгоритмов фазового детектирования [11, 12]. Для узкополосного сигнала вида (1) все эти алгоритмы имеют общий физический смысл и могут быть рассмотрены как процедура дополнительной фильтрации и преобразования Фурье со скользящим окном. В связи с этим мы выбрали приведенный ниже относительно универсальный

подход и считаем, что такой выбор не ограничивает общности полученных результатов или, другими словами, что модификации алгоритма фазового детектирования не изменят приведенных ниже результатов анализа.

Итак, сигнал (1) с выхода фотоприемника дискретизируется в АЦП с частотой $f_d = q \cdot f_o$, где q – целое число. В свете решаемой в статье задачи величина q не существенна, поэтому далее для простоты изложения полагаем $q = 1$. Данный сигнал подвергается полосовой фильтрации с формированием результата в виде аналитического сигнала; обозначим последний как $z(n)$, где n – аргумент дискретного времени. Указанная процедура фильтрации реализуется как быстрая свертка [13] сигнала с импульсной реакцией (ИР) полосового фильтра $h(n)$ вида

$$h(n) = \cos\left(\pi \frac{(n - N/2)}{N}\right) \cdot \frac{\sin\left(\pi \Delta F \left(\frac{(n - N/2)}{f_d}\right)\right)}{\Delta F \left(\frac{(n - N/2)}{f_d}\right)} \times \\ \times \cos\left(2\pi \cdot f_{\text{cp}} \left(\frac{(n - N/2)}{f_d}\right)\right),$$

где $n = 0 \dots N - 1$; N – количество ненулевых отсчетов ИР; $\Delta F, f_{\text{cp}}$ – соответственно ширина и средняя частота полосы пропускания полосового фильтра.

При вычислении упомянутой свертки операция обратного дискретного преобразования Фурье выполняется над половиной спектра (результата перемножения спектров фильтруемого сигнала и ИР фильтра). В итоге формируется результат фильтрации $z(n)$ при частоте его дискретизации $f_{\phi} = 0,5f_o$.

Процедура фазовой демодуляции (фазового детектирования) осуществляется в три этапа. На первом этапе вычисляется фаза каждого временного отсчета $z(n)$ на интервале $[-\pi; \pi]$:

$$\varphi(n) = \begin{cases} \arctg \left[\frac{\text{Im } z(n)}{\text{Re } z(n)} \right], & \text{при } \text{Re } z(n) > 0; \\ \arctg \left[\frac{\text{Im } z(n)}{\text{Re } z(n)} \right] + \pi \cdot \text{sign}(\text{Im } z(n)), & \text{при } \text{Re } z(n) < 0; \end{cases} \quad (2)$$

Далее (на втором этапе) реализуется компенсация линейно нарастающей компоненты фазы, т. е. текущей фазы несущего колебания на частоте f_n . При $f_n = 0,5f_{\phi}$ (ниже будет проиллюстрирована опти-

мальность именно такого соотношения данных частот) эта операция предусматривает вычисление в каждый n -й момент времени (отсчет времени начинается с $n = 0$ -го момента), скорректированных фаз:

$$\varphi_k(n) = \begin{cases} \varphi(n) & \text{для четных } n \\ \varphi(n) - \pi & \text{для нечетных } n \end{cases} \quad (3)$$

Следует заметить, что ширина диапазона возможного изменения фаз отсчетов превышает размер интервала 2π в разы. В связи с этим далее реализуется третий этап фазового детектирования, называемый «раскрутка» фазы (т. е. приведение ее значений к интервалу, ширина которого многократно превышает 2π) [11, 12]. Корректное выполнение данной операции возможно в связи с тем, что частота дискретизации отсчетов фаз f_φ многократно (на практике более чем на порядок) превышает верхнюю частоту принимаемого гидрофоном акустического сигнала f_v . При этом приращение фазы сигнала (1) за период частоты f_φ гарантированно не превышает величины π . Операция «раскрутка» фазы реализуется следующим образом. Отсчеты $\psi(n)$ так называемой «раскрученной» фазы вычисляются рекуррентным образом в соответствии с выражением

$$\psi(n) = \psi(n-1) + W[\varphi_k(n) - \varphi_k(n-1)], \quad (4)$$

где $W(x)$ – функция «сворачивания» или приведения фазового аргумента к интервалу шириной 2π (один конец интервала должен быть «выколот»), определяемая как

$$W(x) = \begin{cases} x, & \text{при } -\pi < x \leq \pi; \\ x + 2\pi, & \text{при } x \leq -\pi; \\ x - 2\pi, & \text{при } x > \pi. \end{cases} \quad (5)$$

В итоге сформирована временная последовательность отсчетов фазы $\psi(n)$ на практически неограниченном интервале ее значений.

3. Анализ предельной чувствительности гидрофона, обеспечивающей линейность тракта по отношению к тональному сигналу

Приемный тракт ГАК рассчитан на прием широкополосного сигнала либо аддитивной смеси широкополосного и монохроматического (узкополосного) сигналов. Однако для иллюстрации имеющих место в тракте с ВОИГ эффектов, предопределяющих нелинейность этого тракта, сначала рассмотрим его реакцию на узкополосный (тональный) сигнал.

Будем считать, что чувствительность гидрофона γ (изменение фазы сигнала (1) при единичном уровне принимаемого акустического давления, рад/Па) частотно независима.

Далее, если это не оговорено особо, под уровнем или спектром шумов (акустических и внутренних шумов тракта) понимаем значения корня квадратного из спектральных плотностей их мощности как функции частоты.

Анализ предельной чувствительности гидрофона на заданной частоте проводится путем расчета по приведенному в предыдущем разделе алгоритму уровня отклика тракта (цепочки «гидрофон – тракт ПОС») $Y(\gamma \cdot P_s)$ на тональный сигнал:

$$S_{\text{ак}}(t) = \sqrt{2} \cdot P_s \cdot \sin(2\pi f_s \cdot t), \quad (6)$$

где f_s – частота тонального сигнала, Гц; P_s – уровень принимаемого сигнала, Па.

Результаты расчета *нормированного (относительного)* уровня отклика тракта на входное воздействие (6) при значениях частоты не-

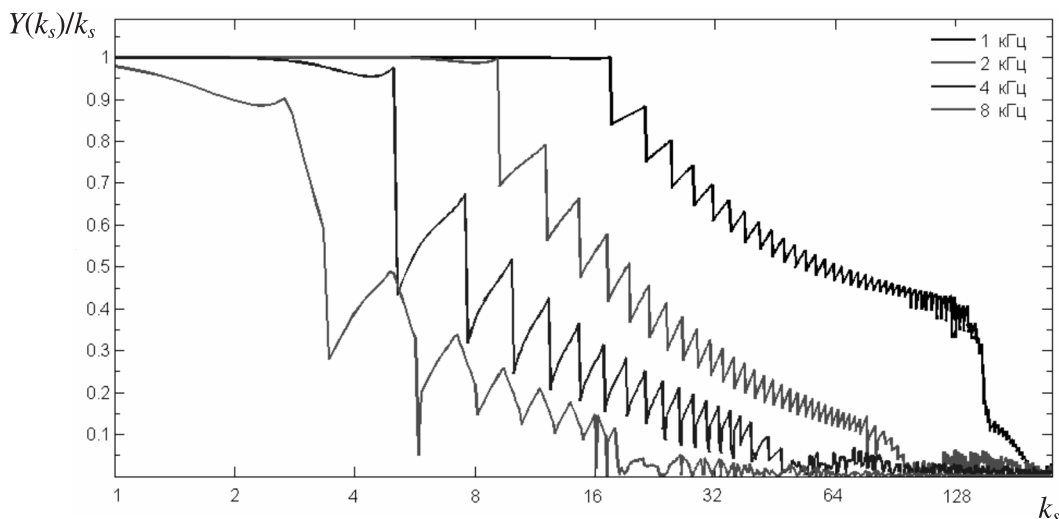


Рис. 1. Нормированный уровень отклика на тональный сигнал при полосе пропускания фильтра $\Delta F = 40$ кГц

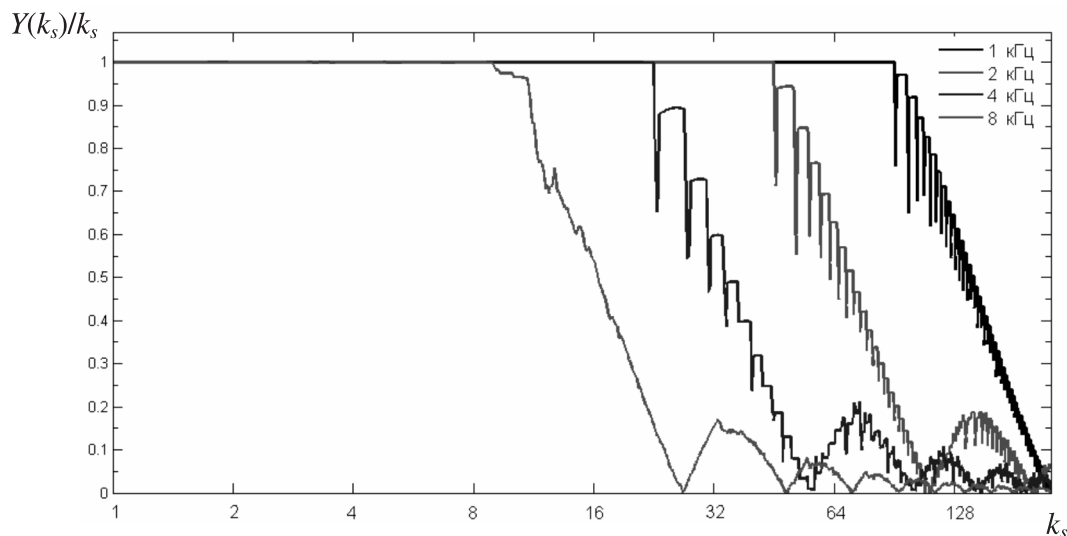


Рис. 2. Нормированный уровень отклика на тональный сигнал при полосе пропускания фильтра $\Delta F = 250$ кГц

сушего колебания $f_n = 125$ кГц, частоты опроса $f_o = 500$ кГц (при этом $f_\phi = 250$ кГц) и значениях частоты сигнала $f_s = 1, 2, 4$ и 8 кГц приведены на рис. 1 и 2 (соответственно для случаев ширины полосы пропускания полосового фильтра $\Delta F = 40$ кГц и $\Delta F = 250$ кГц; в последнем случае операция фильтрации сводится только к формированию аналитического сигнала с двукратной децимацией частоты дискретизации; $k_s = \gamma P_s$).

Величина нормированного уровня отклика $Y(k_s)/k_s$ есть арктангенс наклона его амплитудной характеристики к оси абсцисс (по ней отложен уровень входного воздействия). Заметим также, что тракт линеен в тех пределах изменения уровня входного воздействия, в которых наклон его амплитудной характеристики к оси абсцисс (практически) постоянен. Отсюда с учетом кривых на рис. 1 и 2 получаем критические значения

Таблица 1

Узкополосный сигнал. Предельное значение индекса модуляции ($\Delta F = 40$ кГц)

f_s , кГц	1	2	4	8
k_{sk} , рад	18	10	5	2,5

Таблица 2

Узкополосный сигнал. Предельное значение индекса модуляции ($\Delta F = 250$ кГц)

f_s , кГц	1	2	4	8
k_{sk} , рад	90	45	22	9,5

k_{sk} величины уровня k_s , при которых тракт практически линеен, приведенные в табл. 1 и 2 (соответственно для случаев $\Delta F = 40$ кГц и 250 кГц).

Важно отметить, что приведенные в табл. 2 критические значения не ограничены известным пределом на приращения фазы между соседними отсчетами, необходимым для корректной работы процедуры «раскрутки» фазы [11, 12], который для случая тонального сигнала имеет вид $k_s < 0,5f_\phi / f_s$.

При проведении подобных расчетов применительно к удвоенным значениям частот несущего колебания $f_n = 250$ кГц и опроса $f_o = 1000$ кГц (при этом $f_\phi = 500$ кГц) критические значения чувствительности гидрофонов k_{sk} при $\Delta F = 1000$ кГц (в сравнении с приведенными в табл. 2) удваиваются.

Расширение полосы пропускания фильтра приводит к повышению уровня шумов, мешающих фазовой демодуляции. Так, в рассмотренной ситуации при увеличении параметра ΔF в 6,25 раза уровень (напряжение) шумов повысится только в $\sqrt{6,25} = 2,5$ раза. При этом указанное расширение полосы пропускания фильтра приводит к повышению границы k_{sk} примерно в 4 ... 5 раз. В связи с этим далее при анализе предельной чувствительности гидрофона на заданной частоте при $f_i = 125$ кГц остановимся на варианте ширины полосы пропускания фильтра $\Delta F = 250$ кГц.

Пусть, например, необходимый динамический диапазон составляет 90 дБ, а минимальные уровни сигнала на указанных выше частотах, со-

Таблица 3

Широкополосный сигнал. Предельное значение чувствительности ($\Delta F = 250$ кГц)

f_s , кГц	1	2	4	8
$P_{s \max}$, Па/√Гц	14,2	8,0	4,5	2,5
$\gamma_{\max} = k_{sk} / P_{s \max}$, рад/Па	6,3	5,6	4,9	3,8

ответствующие волнению моря, например, 1 балл, составляют соответственно $4,47 \cdot 10^{-4}$, $2,5 \cdot 10^{-4}$, $1,42 \cdot 10^{-4}$ и $8 \cdot 10^{-5}$ Па/√Гц [14]. Тогда максимальные уровни сигнала ($P_{s \max}$) на рассмотренных частотах составляют соответственно 14,2, 8,0, 4,5 и 2,5 Па/√Гц. При этом с учетом приведенных выше значений k_{sk} приходим к следующим максимально допустимым значениям чувствительности гидрофона $\gamma_{\max}$, приведенным в табл. 3

Имеющие место в тракте эффекты, определяющие его линейность по отношению к входному воздействию в виде узкополосного сигнала, иллюстрируются рис. 3, на котором приведены спектры сигнала (1) и относительной функции отклика $Y(k_s)/k_s$ для частоты $f_s = 8$ кГц при различных значениях параметра k_s . Из графиков на рис. 3 следует, что одним из факторов, существен-

но влияющих на величину предельной чувствительности (и в конечном счете – на достигаемый динамический диапазон), является попадание основной части энергии фазомодулированного сигнала в окно Найквиста (в рассматриваемом примере это диапазон частот 0 ... 250 кГц). Этот эффект отмечался (без анализа последствий расширения спектра, т. е. количественных эффектов нелинейности), в частности, в [5].

Как видно из рис. 3, при значениях $k_s = 20$ и 50 спектр сигнала существенно шире окна Найквиста, что и приводит к нелинейности анализируемого тракта. При этом заметим, что в силу симметричности спектра ФМ сигнала относительно частоты несущего колебания наилучшие условия расположения спектра сигнала в окне Найквиста при прочих равных условиях имеют место при частоте несущего колебания, равной половине частоты Найквиста или четверти частоты опроса ВОИГ.

4. Анализ предельной чувствительности гидрофона, обеспечивающей линейность тракта по отношению к широкополосному сигналу

При анализе предельной чувствительности гидрофона по методике, аналогичной использо-

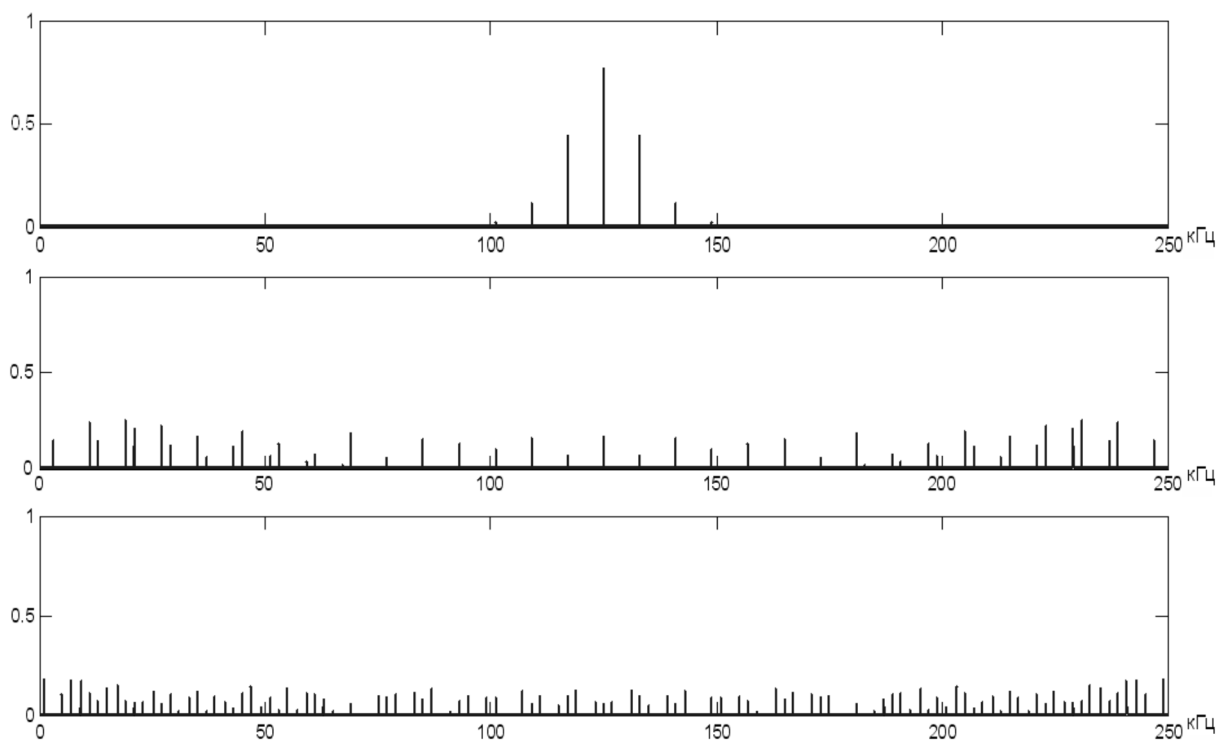


Рис. 3. Спектр фазомодулированного сигнала при величинах параметра $k_s = 1, 20, 50$, сверху вниз соответственно



ванной в разделе 3, в данном случае в качестве входного воздействия $S_{\text{ак}}(t)$ интересен широкополосный шумоподобный сигнал, спектр которого соответствует типичному спектру акустических сигналов (шумов) моря [14]. Однако при применении упомянутой методики в случае использования исходных сигналов с непрерывным спектром возникает сложность с оценкой эффекта нелинейных искажений этих сигналов. В связи с этим далее в качестве широкополосного сигнала используется политональный сигнал с определенными соотношениями частот и амплитуд.

Анализ предельной чувствительности гидрофона по широкополосному сигналу проводится путем расчета функции отклика $Y(k_s)$ тракта на сигнал, представляющий собой суперпозицию 14 тональных сигналов, соответствующих спектру шумов моря при волнении 1 балл:

$$S_{\text{ак}}(t) = \sqrt{2} \sum_{i=1}^{14} \sqrt{\Delta f_i} \alpha_i \sin(2\pi f_i \cdot t), \quad (7)$$

где f_i – частоты спектральных отсчетов, составляющие геометрическую прогрессию со знаменателем $\nu \approx 1,3$ Гц; $\Delta f_i = f_i(\sqrt{\nu} - 1/\sqrt{\nu})$ – ширина полосы частот спектральной компоненты (7), определяемая таким образом, чтобы среднегеометрическое крайних частот равнялось f_p Гц; α_i – амплитуды спектральных отсчетов, соответствующие уровню шума моря при волнении 1 балл на частоте f_i [9], Па/√Гц (мы исходим из того, что указанный уровень шума моря оправданно считать минимальным).

Выбор входного воздействия (7) при анализе предельной чувствительности гидрофона по широкополосному сигналу обусловлен следующим.

Сигнал (7) характеризуется дискретным спектром с некратными частотами, что позволяет проанализировать амплитудную характеристику тракта (характеризующую степень его линейности) на разных частотах, при этом гармоники одних частот не мешают анализу амплитудной характеристики тракта на других частотах. Используемые веса тональных составляющих сигнала (7) обеспечивают имитацию реального входного сигнала как в части формы его спектра, так и в части его интегрального уровня.

Результаты расчетов зависимости функции отклика $Y(\gamma)/\gamma$ при значениях частоты несущего колебания $f_n = 125$ кГц для спектральных компонент сигнала на частотах $f_i = 0,26, 0,98, 4,73$ и 8 кГц приведены на рис. 4. Из рисунка следует, что критические значения γ_k величины чувствительности, при которых тракт практически линеен в рабочем диапазоне частот при минимальном уровне принимаемого сигнала, составляет $\approx 10^3$.

Если указанное значение величины γ_k является критическим при минимальном уровне принимаемого сигнала (шума), то, поскольку индекс модуляции, согласно (7), определяется произведением $\gamma \alpha_i$, при возрастании уровня сигнала (т. е. всего массива уровней α_i) в δ раз критическое значение чувствительности уменьшится в те же δ раз. Поэтому при необходимости обеспечения динамического диапазона (δ) по широкополосному сигналу, например, $\delta = 60$ или 80 дБ (или в размах соответственно 10^3 или 10^4) предельно допустимое значение чувствительности гидрофонов в рассматриваемой ситуации (т. е. при фиксированных частотах опроса гидро-

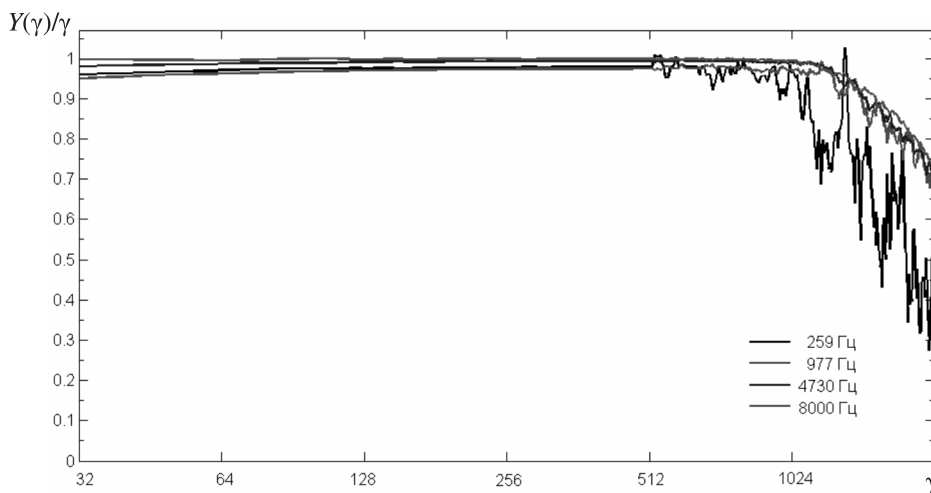


Рис. 4. Нормированный уровень отклика на широкополосный сигнал

фонов и несущего колебания) составляет соответственно $\gamma_{\max} = \gamma_{\kappa} / \delta = 1$ или 0,1.

Разработана методика расчета верхней границы чувствительности ВОИГ по критерию достижения максимального динамического диапазона. Эта методика, в отличие от известной [10], характеризуется высокой точностью расчета и применима при произвольных спектрах акустического сигнала.

Приведены примеры расчета верхней границы чувствительности, исходя из требований обеспечения заданного динамического диапазона, для случаев тонального и широкополосного сигналов. При этом для сигналов с энергетическим спектром произвольного вида количественно подтверждено правило: величина предельной чувствительности при заданном уровне принимаемого сигнала определяется из условия попа-

дания основной части энергии ФМ сигнала в окно Найквиста, ширина которого определяется частотой опроса гидрофонов.

Использованная методика решения поставленной задачи изложена в контексте проблемы выбора оптимальной чувствительности ВОИГ, актуальной при разработке гидроакустических антенн. Однако основные рассмотренные в статье вопросы касались допустимого индекса фазовой модуляции для корректного детектирования фазомодулированного сигнала, в цифровом представлении. Поэтому примененный метод анализа и ряд полученных результатов могут быть актуальны и в других задачах, касающихся разработки интерферометрических систем и любых других систем с обработкой и детектированием фазомодулированных сигналов, в особенности, работающих в дискретном времени.

Статья подготовлена при финансовой поддержке РФФИ, проект № 11-08-01097.

СПИСОК ЛИТЕРАТУРЫ

1. Удда, Э. Волоконно-оптические датчики. Вводный курс для инженеров и научных работников [Текст] / Э. Удда. – М.: Техносфера. – 2008. – 520 с.
2. Shizhuo, Y. Fiber Optic Sensor [Text] / Y. Shizhuo, P.B. Ruffin, T.S. Yu Francis. – Talor & Francis Group, 2008. – 479 p.
3. Kirkendall, C.K. Overview of high performance fiber-optic sensing [Text] / C.K. Kirkendall, A. Dandridge // J. Phys. D: Appl. Phys. – 2004. – Vol. 37. – P. 197–216.
4. Ликумович, Л.Б. Волоконно-оптические интерферометрические измерения [Текст] / Л.Б. Ликумович // Ч. 2. Волоконный интерферометрический чувствительный элемент. – СПб.: Изд-во Политехн. ун-та, 2007. – 68 с..
5. Cranch, G.A. Large-Scale Multiplexing of Interferometric Fiber-Optic Sensor Using TDM and DWDM [Text] / G.A. Cranch, P.J. Nash // J. Lightwave Tech. – 2001. – Vol. 19. – №5. – P. 687–699.
6. Zefeng, Wang. Fiber-optic hydrophone using a cylindrical Helmholtz resonator as a mechanical anti-aliasing filter [Text] / Zefeng Wang, Hu Yongming, Meng Zhou [et al.] // Optics Letters. – 2008. – Vol. 33. – № 1. – P. 37–39.
7. Кудряшов, А.В. Повышение эффективности преобразования звукового давления в волоконных интерферометрических микрофонах [Текст] / А.В. Кудряшов, Л.Б. Ликумович, А.В. Медведев // Научно-технические ведомости СПбГПУ. Сер. Физико-математические науки. – 2011. – № 3 (129). – С. 76–83.
8. Yin, Kai. Research on the acceleration responsivity of the fiber-optic air-backed mandrel hydrophones [Электронный ресурс] / Kai Yin, Min Zhang, Liwei Wang [et al.] // Proc. of SPIE. – Vol. 6830. – P. 683013-1–683013-8.
9. Lagakos, N. Linearly Configured Embedded Fiber-optic Acoustic Sensor [Text] / N. Lagakos, J.A. Bucaro // J. of lightwave Technology. – 1993. – Vol. 11. – № 4. – P. 639–642.
10. Ole Henrik Waagaard. An investigation of the pressure-to-acceleration responsivity ratio of fiber-optic mandrel hydrophones [Text] / Waagaard Ole Henrik, Geir Bjarte Havsgård, Gunnar Wang [et al.] // J. of Light wave technology. – 2001. – Vol. 19. – 994 p.
11. Васильев, В. Компьютерная обработка сигналов в приложении к интерферометрическим системам [Текст] / В. Васильев, И. Гуров. – СПб.: БХВ-Петербург, 1998. – 240 с.
12. Daniel, Malacara. Interferogram Analysis for Optical Testing [Text] / Malacara Daniel, Servin Manuel, Malacara Zacarias; 2 ed. – Taylor & Francis Group, 2005.
13. Рабинер, Л. Теория и применение цифровой обработки сигналов [Текст] / Л. Рабинер, Б. Гоулд. – М.: Мир, 1978. – 848 с.
14. Урик, Р.Дж. Основы гидроакустики [Текст] / Р.Дж. Урик. – Л.: Судостроение, 1978. – 445 с.

МОДЕЛИРОВАНИЕ ДИНАМИКИ ДВУХДВИГАТЕЛЬНОГО СЕРВОПРИВОДА С ЭФФЕКТОМ САМОТОРМОЖЕНИЯ

Сервопривод на основе электромотора с редуктором является типовым конструктивным элементом автоматических систем, широко применяется в станкостроении, приборостроении и робототехнике. В связи с этим особое значение приобретают исследования, направленные на повышение эффективности и расширение функциональных возможностей сервоприводов.

Одна из востребованных функций сервопривода – поддержание высокой точности позиционирования выходного звена (вала) при наличии внешних воздействий. Реализация этого режима работы сервопривода может обеспечиваться следящей системой, обеспечивающей необходимый для удержания момент сил на выходном валу (1), тормозом (2), либо свойством самоторможения редуктора (3).

Вариант 1 связан с высокими затратами электроэнергии, ограниченными мощностью источника питания, особенно для автономного источника.

В варианте 2 кроме дополнительного привода необходимы конструктивные решения тормоза и принятие дополнительных мер по увеличению быстродействия при переключении режимов работы сервопривода.

Свойством самоторможения (вариант 3) и малыми габаритами обладает редуктор с червячной передачей. При этом нулевая скорость обеспечивается эффектом самоторможения, возникающего в результате действия сил трения, что не требует работы двигателя и сервоконтроллера и, как следствие, затрат энергии. Существенным недостатком прямого использования червячной передачи в редукторе сервопривода является низкий КПД (меньше 0,5 [1]) в двигательном режиме. Это требует использования более мощных электродвигателей, а также решения проблемы под-

держания температурного баланса (из-за высоких потерь на трении) и повышенного износа рабочих поверхностей, либо отказа от червячной передачи в пользу более эффективных редукторов в сочетании с иными способами поддержания высокой точности позиционирования.

В [2] предложена кинематическая схема сервопривода с нетрадиционным подходом к использованию червячной передачи: червяк и червячное колесо передачи имеют собственные электроприводы. Такое решение позволяет получить сервопривод, сочетающий в себе и возможность поддержания нулевой скорости выходного звена без затрат энергии, используя свойство самоторможения, и относительно высокий КПД (значительно выше, чем у самотормозящейся червячной передачи) в двигательном режиме работы. Следует отметить, что в патенте [2] не рассмотрены вопросы управления такими системами ни в стационарных, ни в реверсивных режимах.

В настоящей статье проводится анализ и математическое моделирование динамики сервопривода с двумя двигателями, связанными внутренней связью со свойством самоторможения, и предлагается подход к управлению таким сервоприводом.

На рис. 1 приведена структурная схема объекта управления. Двигатели червяка и колеса согласованно управляются с обратной связью по скорости и по току. Роторы электродвигателей связаны через редукторы с червяком и червячным колесом. Червячная передача имеет коэффициент передачи, обеспечивающий самоторможение от колеса к червяку. Концепция рассматриваемого привода заключается в разделении его базовых функций между подсистемами приводов червяка и червячного колеса.

Функцию поддержания высокой точности по-

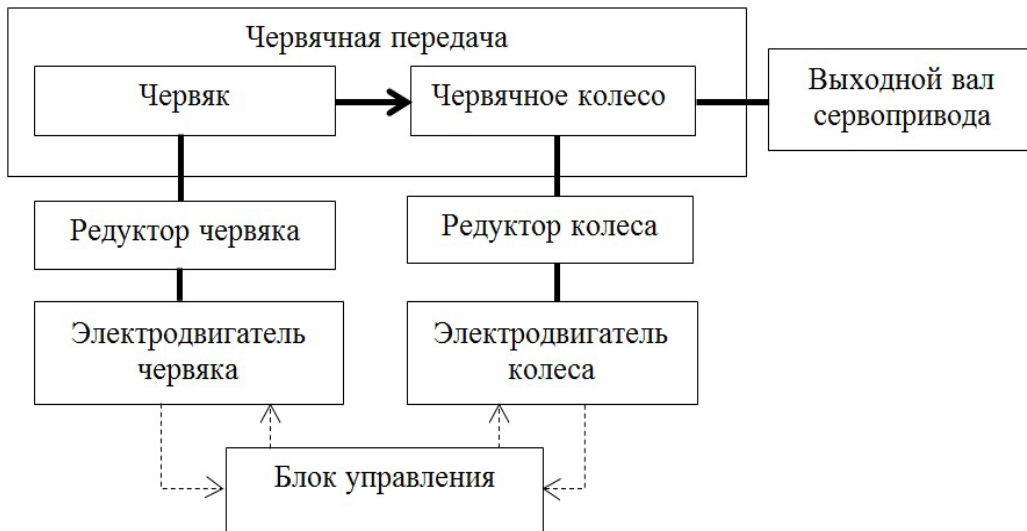


Рис. 1. Структурная схема двухдвигательного сервопривода с эффектом самоторможения

позиционирования выходного звена обеспечивает самотормозящаяся червячная передача, не требующая в этом режиме управления или затрат энергии. Выходное звено должно быть конструктивно объединено, либо присоединено через редуктор к червячному колесу, что исключает перемещение выходного звена вследствие действия внешних сил. Функции двигательного режима (позиционирование, поддержание скорости или момента сил на выходном звене) обеспечивает двигатель червячного колеса (как источник механической энергии). Двигатель червяка при этом должен поддерживать режим движения, обеспечивающий предотвращение нарушения желаемого режима движения выходного звена вследствие эффекта самоторможения.

Режим поддержания высокой точности позиционирования требует только управления обесточиванием электродвигателей привода и обеспечивается конструктивными элементами редуктора. Задача согласованного управления приводами в режиме движения осложняется возможностью возникновения эффекта самоторможения. Даже кратковременные возмущения в угловых скоростях червяка или колеса приводят к рассогласованию взаимного расположения зубьев, что может стать причиной ударной непредвиденной остановки привода из-за проявления свойства самоторможения. Поэтому традиционное управление по скорости с поддержанием соотношения скоростей [3] приводов является недостаточным для эффективного функционирования такой системы.

Кроме согласования скоростей необходимо обеспечивать особое качество переходных процессов, позволяющее избежать самоторможения.

Рассмотрим двухдвигательный сервопривод, состоящий из двух двигателей постоянного тока с редукторами, выходные ступени которых связаны червячной передачей. Выходной вал сервопривода жестко связан с червячным колесом.

Математическая модель объекта управления составлена с использованием аппарата библиотек Matlab и описывается уравнениями, которые приводятся ниже.

Математическое описание электродинамики двигателей постоянного тока с одной парой полюсов соответствуют уравнениям:

$$\begin{aligned} U_{\text{я}} + E_{\text{я}} &= R_{\text{я}} I_{\text{я}} + L_{\text{я}} \dot{I}_{\text{я}}, \\ M_{\text{эм}} + M_{\text{вн}} &= J_{\text{р}} \dot{\omega}_{\text{дв}} + \lambda_{\text{р}} \omega_{\text{дв}}, \\ M_{\text{эм}} &= k_M I_{\text{я}}, \\ E_{\text{я}} &= -k_{\omega} \omega_{\text{дв}}, \end{aligned} \quad (1)$$

где $\omega_{\text{дв}}$ и $M_{\text{вн}}$ – угловая скорость ротора и момент внешних сил, действующих на ротор; $E_{\text{я}}$ – противо-ЭДС обмотки якоря; $U_{\text{я}}$ и $I_{\text{я}}$ – напряжение и ток обмотки якоря; $M_{\text{эм}}$ – электромагнитный момент двигателя; k_{ω} и k_M – постоянные ЭДС и момента двигателя; $R_{\text{я}}$ и $L_{\text{я}}$ – сопротивление и индуктивность обмотки якоря; $J_{\text{р}}$ и $\lambda_{\text{р}}$ – момент инерции ротора и коэффициент вязкого трения в подшипниках.

Модель редукторов приводов описывается уравнениями:



$$\begin{aligned} i_p \omega_2 &= \omega_1, \\ i_p M_1 + M_2 &= i_p (1 - \eta) |M_1| \text{sign}(\omega_{\text{out}}), \\ \text{при } \omega_{\text{out}} &= \begin{cases} \omega_1, \omega_1 M_1 < \omega_2 M_2 \\ \omega_2, \omega_1 M_1 > \omega_2 M_2 \end{cases}, \end{aligned} \quad (2)$$

где i_p и η – передаточное число и КПД передачи; ω_1, ω_2, M_1 и M_2 – угловые скорости и моменты сил на выходах передачи.

Червячной передаче с люфтом $2g$ соответствует система уравнений:

$$\begin{aligned} i_{\text{чп}} \omega_k &= \omega_q, \\ i_{\text{чп}} M_q + M_k &= M_{\text{тр}} \text{sign}(\omega_k), \\ \eta_{\text{чк}} &= \frac{\text{tg}(\gamma)}{\text{tg}(\gamma + \varphi)}, \\ \eta_{\text{кч}} &= \frac{\text{tg}(\gamma - \varphi)}{\text{tg}(\gamma)}, \\ M_{\text{тр}} &= \begin{cases} (1 - \eta_{\text{кч}}) |M_k|, \omega_q M_q < \omega_k M_k \\ i_{\text{чп}} (1 - \eta_{\text{чк}}) |M_q|, \omega_q M_q > \omega_k M_k \end{cases}, \\ M_k &= \begin{cases} K(\delta - g) + \mu(i_{\text{чп}} \omega_q - \omega_k), \delta \geq g \\ 0, -g < \delta < g \\ K(\delta + g) + \mu(i_{\text{чп}} \omega_q - \omega_k), \delta \leq -g \end{cases}, \end{aligned} \quad (3)$$

где $i_{\text{чп}}$ – передаточное число передачи; γ – делительный угол подъема витка червяка; φ – угол трения; ω_q, ω_k, M_q и M_k – угловые скорости и моменты сил червяка и колеса; δ – угловое смещение колеса от положения центрирования зубьев в люфте; K и μ – коэффициенты жесткости и вязкого трения на контакте червяка и колеса.

Связь блоков двигателей и редукторов задается равенством угловых скоростей ротора двигателя и входного звена редуктора и равенством нулю суммы моментов сил, приложенных к ротору и входному валу редуктора,

$$\begin{aligned} \omega_{\text{дв}} &= \omega, \\ M_{\text{вн}} + M_1 &= 0. \end{aligned} \quad (4)$$

Связь выходов редукторов и червячной передачи описывается уравнениями:

$$\begin{aligned} \omega_2^q &= \omega_q, \\ \omega_2^k &= \omega_k, \\ -M_2^q - M_q &= J_q \dot{\omega}_q + \lambda_q \omega_q, \\ -M_2^k - M_k &= J_k \dot{\omega}_k + \lambda_k \omega_k + M_n, \end{aligned} \quad (5)$$

где J_q, J_k, λ_q и λ_k – приведенные моменты инерции, коэффициенты вязкого трения в подшипниках передач приводов червяка и колеса; M_2^q, M_2^k, ω_2^q и ω_2^k – моменты сил и угловые скорости на выходах

редукторов приводов; M_n – момент сил нагрузки на выходном валу сервопривода.

Эффект самоторможения возникает при конструктивном условии самотормозящейся передачи $-\eta_{\text{кч}} \leq 0$ и динамическом условии $-\omega_q M_q < \omega_k M_k$. Конструктивное условие является неизменным для всех режимов работы и не может быть изменено системой управления. Рассмотрим условие отсутствия самоторможения из (3):

$$\omega_q M_q > \omega_k M_k. \quad (6)$$

Без потери общности предполагается, что все передаточные числа положительны – $i_x > 0$. Выражая из (1)–(5), учитывая, что привод червяка может работать только в двигательном режиме, подставив в (6), получим развернутое динамическое условие для отсутствия самоторможения при положительных скоростях:

$$\begin{aligned} \text{при } \omega_k > 0: & \begin{cases} -M_2^k < M_n + J_k \dot{\omega}_k + \lambda_k \omega_k, \\ k_m^q I_{я}^q > \tilde{J}_q \dot{\omega}_{\text{дв}}^q + \tilde{\lambda}_q \omega_{\text{дв}}^q, \end{cases} \\ \text{при } \omega_k < 0: & \begin{cases} -M_2^k > M_n + J_k \dot{\omega}_k + \lambda_k \omega_k, \\ k_m^q I_{я}^q < \tilde{J}_q \dot{\omega}_{\text{дв}}^q + \tilde{\lambda}_q \omega_{\text{дв}}^q, \end{cases} \end{aligned} \quad (7)$$

$$\tilde{J}_q = J_p^q + \frac{J_q}{\eta_p^q (i_p^q)^2},$$

$$\tilde{\lambda}_q = \lambda_p^q + \frac{\lambda_q}{\eta_p^q (i_p^q)^2}.$$

Выполнение неравенств из (7) является необходимым и достаточным условием для двигательного режима работы самотормозящегося привода без возникновения самоторможения.

Из неравенств в (7), задающих ограничения для червячного колеса (индекс «к»), следует, что нельзя полностью исключить влияние червячной передачи на выходной вал сервопривода. С другой стороны, возможно полное исключение привода колеса из системы (система станет обычным однодвигательным приводом с самоторможением). Как указывалось выше, КПД обычного привода с самоторможением низок, поэтому для повышения КПД системы в целом необходимо минимизировать момент сил, сообщаемый колесу от червяка. Таким образом, червячная передача всегда должна быть нагружена, но ее доля в работе сервопривода должна минимизироваться, а нагрузка при этом должна ложиться на провод колеса с более эффективным редуктором, что позволяет записать критерий эффективного самотормозящегося привода:

$$M_2^k \gg M_k. \quad (8)$$

Из неравенств в (7), задающих ограничения для червяка (индекс «ч»), следует, что основной нагрузкой для привода червяка являются потери на трение в подшипниках и моменты инерции, ограниченные неравенствами:

$$\begin{aligned} \text{при } \omega_{\text{ч}} > 0: & \begin{cases} 0 < \tilde{\lambda}_{\text{ч}} \omega_{\text{дв}}^{\text{ч}} \leq \tilde{\lambda}_{\text{ч}} \omega_{\text{хх,дв}}^{\text{ч}} = k_{\text{м}}^{\text{ч}} I_{\text{хх}}^{\text{ч}} \\ -\infty < \tilde{J}_{\text{ч}} \dot{\omega}_{\text{дв}}^{\text{ч}} < k_{\text{м}}^{\text{ч}} (I_{\text{я}}^{\text{ч}} - \frac{\tilde{\lambda}_{\text{ч}} \omega_{\text{дв}}^{\text{ч}}}{k_{\text{м}}^{\text{ч}}}) \end{cases} \\ \text{при } \omega_{\text{ч}} < 0: & \begin{cases} 0 > \tilde{\lambda}_{\text{ч}} \omega_{\text{дв}}^{\text{ч}} \geq -\tilde{\lambda}_{\text{ч}} \omega_{\text{хх,дв}}^{\text{ч}} \\ \infty > \tilde{J}_{\text{ч}} \dot{\omega}_{\text{дв}}^{\text{ч}} > -k_{\text{м}}^{\text{ч}} (I_{\text{я}}^{\text{ч}} - \frac{\tilde{\lambda}_{\text{ч}} \omega_{\text{дв}}^{\text{ч}}}{k_{\text{м}}^{\text{ч}}}) \end{cases} \end{aligned} \quad (9)$$

где $\omega_{\text{хх,дв}}^{\text{ч}}$ и $I_{\text{хх}}^{\text{ч}}$ – угловая скорость и ток двигателя при холостом ходе червяка.

Из неравенств (7) и (9) видно, что чем больше значение $k_{\text{м}}^{\text{ч}} I_{\text{я}}^{\text{ч}}$, тем в более широких диапазонах скоростей и ускорений червяка выполняется условие отсутствия самоторможения. Однако чем больше величина $(k_{\text{м}}^{\text{ч}} I_{\text{я}}^{\text{ч}} - \tilde{J}_{\text{ч}} \dot{\omega}_{\text{дв}}^{\text{ч}} - \tilde{\lambda}_{\text{ч}} \omega_{\text{дв}}^{\text{ч}})$, тем больший момент сил передается через редуктор червяка и червячную передачу на нагрузку, что влечет за собой снижение КПД всей системы. В случае если изменение внешней нагрузки и соответственно ускорение $\omega_{\text{дв}}^{\text{ч}}$ известно заранее или поддается предсказанию, возможна адаптивная подстройка значения $k_{\text{м}}^{\text{ч}} I_{\text{я}}^{\text{ч}}$, позволяющая поддерживать и высокий КПД самотормозящегося привода, и широкие рабочие диапазоны скорости и ускорения. Если значение внешней нагрузки изменяется непредсказуемо, то необходимо постоянно поддерживать некоторое высокое значение $k_{\text{м}}^{\text{ч}} I_{\text{я}}^{\text{ч}}$, компенсирующее непредвиденные изменения ускорения. Здесь выявляются противоречивые критерии: с одной стороны желание повысить КПД системы, с другой – расширение рабочих диапазонов скоростей и ускорений. Так, максимизация КПД приведет к невозможности избежать самоторможения без полного отказа от червячной кинематической связи, поскольку любой непредвиденный импульс вызовет самоторможение привода. Максимальное расширение рабочих диапазонов приведет к увеличению мощности двигателя червяка и значительным потерям в червячной передаче и, как следствие, к уменьшению КПД. Также расширение вызовет

значительное изменение динамики выходного вала – нарушение (8) и, следовательно, усложнение разработки самотормозящегося привода.

Компромиссным решением здесь может стать изначальное задание необходимых рабочих диапазонов и выбор двигателя со значением $k_{\text{м}}^{\text{ч}} I_{\text{я}}^{\text{ч}}$, минимально превышающим требуемое, для удовлетворения неравенства (7) в полных диапазонах (9). Этот подход позволяет минимизировать габарит привода червяка и влияние червяка на выходной вал в режиме движения. Для управления таким двухдвигательным приводом достаточно поддерживать значение $I_{\text{я}}^{\text{ч}}$, близкое к номинальному со знаком, соответствующем желаемому направлению вращения, и управлять приводом колеса, как обычным однодвигательным приводом (без самоторможения). Для перехода из режима самоторможения в двигательный режим и смены направления вращения необходимо создавать кратковременный корректирующий импульс ускорения колеса в противоположном желаемому направлению вращения, что обеспечит начальное выполнение неравенств (7) и перебрасывание зазоров в люфте.

Следует заметить, что наличие люфта в червячной передаче также является противоречивым условием. С одной стороны, в двигательном режиме при нахождении зубьев передачи не в зацеплении, т. е. в люфте, кинематическая связь между двигателями разомкнута и, следовательно, независимо от соотношения мощностей $\omega_{\text{ч}} M_{\text{ч}}$ и $\omega_{\text{к}} M_{\text{к}}$ эффект самоторможения не проявится. С другой стороны, в режиме самоторможения с увеличением люфта уменьшается точность позиционирования, а в двигательном режиме из-за неопределенности соотношения $\omega_{\text{ч}} M_{\text{ч}}$ и $\omega_{\text{к}} M_{\text{к}}$ возможен непредвиденный переход в режим самоторможения при соответствующем зацеплении зубьев.

В программном пакете Matlab-Simulink с использованием библиотеки Simscape-SimDriveline проведено компьютерное моделирование двухдвигательного самотормозящегося сервопривода (подробная информация о библиотеке Simscape- приведена в справочной системе Matlab-Simulink). Модель соответствует математическому описанию, изложенному в настоящей статье. Параметры моделей двигателей червяка и колеса соответствуют RE-8 12V и EC-32 flat

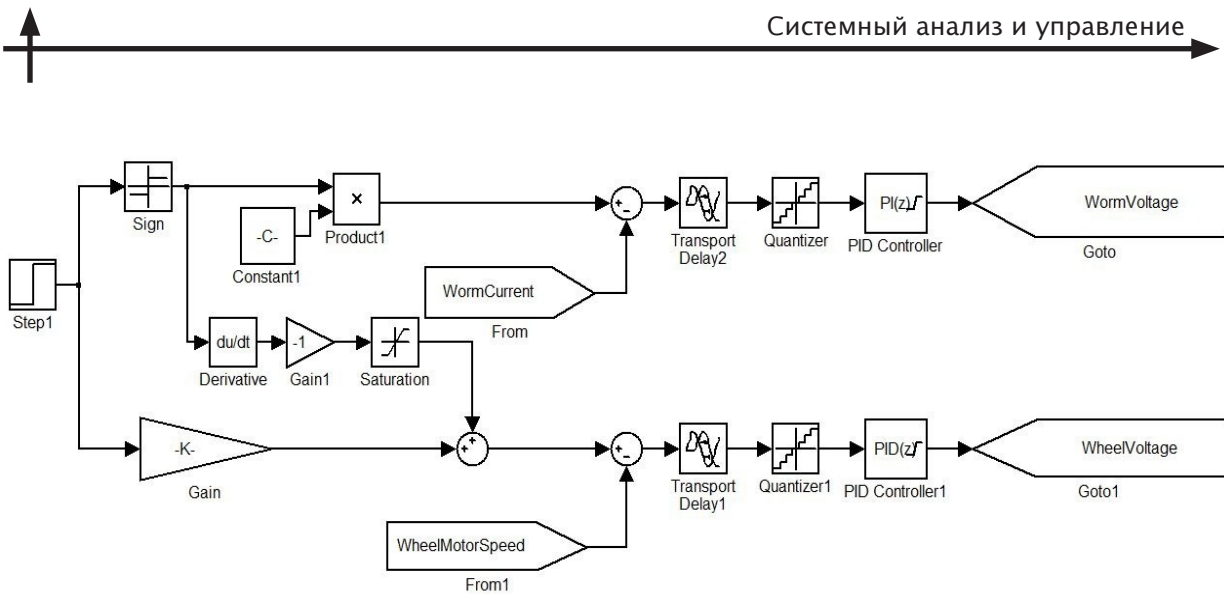


Рис. 2. Схема системы управления самотормозящегося привода

24V с планетарными редукторами GP 8A 4:1 и GP 32A 51:1 из каталога MAXON MOTOR [4]. Необходимые параметры самотормозящейся червячной передачи с передаточным числом $i_{\text{чп}} = 40$ рассчитаны в соответствии с методикой, приведенной в [1]. Схема системы управления изображена на рис. 2.

Система управления состоит из дискретного ПИ-регулятора тока привода червяка, дискретного ПИД-регулятора скорости привода колеса, блоков согласованного задания управляющих параметров (задания тока и скорости), а также из блоков корректирующего импульса скорости колеса.

На рис. 3 приведены зависимости токов и скоростей обработки изменяющейся скорости при постоянном моменте сил нагрузки. Задание скорости ротора двигателя изменяется по синусоиде с амплитудой 700 рад/с, частотой 10 рад/с, момент сил нагрузки постоянный 0,6 Нм.

Зависимости на рис. 3 демонстрируют успешное изменение направления вращения. На графике тока якоря двигателя колеса видны обратные корректирующие импульсы в моменты времени смены направлений вращения.

Анализ качественных характеристик двухдвигательного сервопривода с внутренней пассивной тормозной системой на основе самотормозящейся червячной передачи показал перспективность применения его в робототехнических манипуляционных системах. Предложены математическая модель для описания сервопривода и способ управления двухдвигательным сервоприводом с эффектом самоторможения. Проведенное компьютерное моделирование подтвердило реализуемость и эффективность таких систем в приводах робототехнических систем.

Статья подготовлена в рамках работ по выполняемому в соответствии с постановлением Правительства РФ № 218 комплексному проекту по договору № 13G25310026.

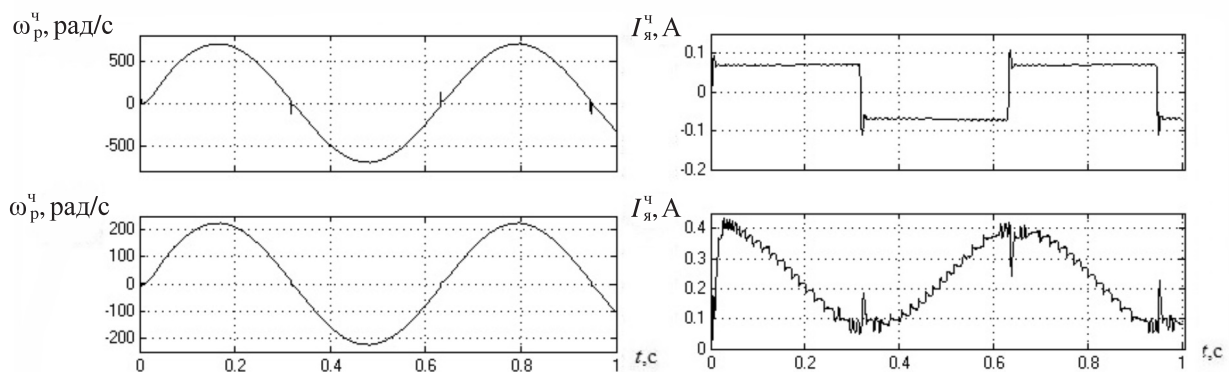


Рис. 3. Зависимости угловых скоростей ротора (слева) и токов якоря (справа) от времени в режиме отработки переменной скорости при постоянном моменте сил нагрузки

СПИСОК ЛИТЕРАТУРЫ

1. **Иванов, М.Н.** Детали машин: Учебник для машиностроительных специальностей вузов [Текст] / М.Н. Иванов, В.А. Финогенов. –М.: Высш. школа, 2008.
2. United States Patent 6.889.578 Electro-mechanical actuator [Электронный ресурс].
3. **Ковчин, С.А.** Теория электропривода: Учебник для вузов [Текст] / С.А. Ковчин, Ю.А. Сабинин. –СПб.: Энергоатомиздат. СПб. отделение, 2000.
4. Каталог продукции Maxon motor за 2011–2012 гг. [Электронный ресурс] / Режим доступа: <http://www.maxonmotor.com/maxon/view/catalog/>

АНАЛИЗ ОБЪЕКТНО-ОРИЕНТИРОВАННОЙ ПАРАДИГМЫ И ПОИСК ЛУЧШЕЙ АЛЬТЕРНАТИВЫ

Объектно-ориентированное программирование (ООП) в настоящее время является доминирующей парадигмой программирования. В данной статье произведен анализ причин этого успеха, выявлены практические задачи, решаемые ООП, и определены его недостатки. Для рассмотренных задач предложены альтернативные механизмы решения – более простые и общие и, следовательно, более подходящие в качестве основы для проектирования языков программирования.

С самого начала заметим, что термины «простой» и «сложный» будут использоваться как объективные. Простота (англ. simple, от лат. simplex) подразумевает ортогональность, отсутствие переплетения различных понятий. Противовесом является сложность (англ. complex) – сильная связанность. Легкость (англ. easy, от лат. ease → aise → adjacens), часто используемая как синоним простоты (что неправильно: легкость – субъективное понятие), относится к доступности (близости) объекта.

Цель статьи – показать недостатки ООП как парадигмы программирования и предложить более простую и гибкую альтернативу.

Характерные черты ООП

Отметим, что в дальнейшем изложении используется упрощенный взгляд на эволюцию промышленных языков программирования: структурные (C, Fortran) → гибридные (C++) → объектно-ориентированные (Java, C#).

Объектно-ориентированное программирование – общий термин, за которым могут скрываться сильно отличающиеся концепции [1]. Господствующее в настоящее время статическое ООП (C++, Java, C# и т. д.) имеет много существенных отличий от первоначального динамического (Smalltalk). В данной статье, если не говорится

иное, рассматривается именно статическое ООП. Тем не менее для ООП в целом можно выделить следующие общие характерные черты [2]:

- 1) объектно-ориентированную инкапсуляцию операций и данных;
- 2) наследование;
- 3) подтиповый полиморфизм.

Объектно-ориентированная инкапсуляция, как следует из названия, делает *объект* основным элементом управления видимостью операций (методов) и состояния (данных); объекты «владеют» методами и данными.

Наследование представляет собой механизм повторного использования кода, тесно связанный с подтиповым полиморфизмом, который, в свою очередь, основан на динамическом распределении привязанных к объекту методов (позднем связывании).

Видно, что центральное понятие объекта – *сложное* (объединяет состояние, операции, контроль видимости, реализацию полиморфизма). То же можно сказать о наследовании (объединяет механизм и гранулярность повторного использования, объявление отношения подтиповости) и подтиповом полиморфизме (объединяет иерархию классов и возможность полиморфного использования). ООП – *сложная* парадигма, основные концепции которой тесно переплетены и с трудом могут рассматриваться независимо друг от друга.

Потребности, удовлетворяемые ООП

Успех ООП объясняется большим числом факторов, среди которых весомы внешние (близость к господствовавшему структурному программированию, популяризация в «правильный» момент времени, поддержка крупными корпорациями и т.д.). В то же время, во многом он осно-

ван на том, что именно ООП было первой парадигмой, давшей относительно легкие ответы на ряд практических вопросов написания и организации программ:

- 1) инкапсуляция «на нужном уровне»;
- 2) повторное использование кода;
- 3) полиморфизм.

Разберем подробнее каждый из этих пунктов.

Инкапсуляция. Дадим инкапсуляции довольно общее определение – это механизм сокрытия деталей реализации; определение «деталей» при этом зависит от требуемого уровня абстракции. Например, функции (процедуры) в С – это механизм инкапсуляции, скрывающий детали реализации алгоритма (поток управления) за его названием. Инкапсуляция является *самым* важным инструментом при построении больших систем, не только программных. Подобные системы, как правило, организованы в виде «стопки» уровней абстракции, реализация каждого из которых основана только на интерфейсе уровня непосредственно «ниже». Хорошим примером может служить современный стек сетевых протоколов.

До распространения ООП в распоряжении программистов было только два основных уровня инкапсуляции: функция и процесс (операционной системы); наиболее популярные языки программирования того времени (С, Fortran) не поддерживали, среди прочего, концепцию модулей. Потребность группировать, например, связанные процедуры и данные очевидна.

Многие авторитетные инженеры, включая создателей Objective-C и других языков программирования (см. [3]), отмечают, что именно инкапсуляция «на нужном уровне» – основной вклад ООП в промышленное программирование.

Повторное использование. Крайне желательной является возможность повторного использования написанного кода с возможностью адаптации к новому контексту использования. Причем как и в случае с инкапсуляцией, важна правильная гранулярность единиц повторного использования: единицей повторного использования в структурном программировании является функция (с фиксированными типами аргументов), что, очевидно, недостаточно для повторного использования «поведений» (наборов связанных функций) с возможностью их настройки.

Классическое ООП в качестве механизма повторного использования предлагает наследова-

ние: функциональность, заложенная в «базовый» класс, может использоваться классами-«наследниками», которые также могут переопределить отдельные методы.

Такой подход имеет множество общепризнанных изъянов. Современные объектно-ориентированные языки и научные работы, посвященные повторному использованию (например, Scala [4, 5]), нередко предлагают полный отказ от наследования. Приведем наиболее существенные из недостатков.

- Единицей повторного использования является класс, от которого необходимо наследовать клиентов (нередко при отсутствии возможности множественного наследования); это слишком крупная единица для многих задач повторного использования поведения. Также повторное использование ограничено, т. к. не может быть расширено за пределы иерархии, образованной базовым классом.

- Внутренний интерфейс («точки расширения», методы для переопределения подклассами) и внешний интерфейс (методы для клиентов класса) не разделены, что приводит к проблеме «хрупкого базового класса» – подкласс может переопределить произвольный метод, тем самым потенциально нарушить инвариант базового класса – и анти-паттернам типа «вызов суперкласса» – требованию к подклассам об обязательном вызове переопределяемого метода.

- Проектирование хорошего базового класса крайне сложно, т. к. из-за его тесной связанности с подклассами требует априорного детального знания всех возможных будущих областей применения подклассов.

Полиморфизм. Определим полиморфизм как возможность работать с различными типами данных, используя унифицированный интерфейс. Из определения видно, что основное применение полиморфизма – написание алгоритмов, работающих с различными типами данных. Полиморфизм способствует повторному использованию кода, позволяет расширить область применения программы, не переписывая фундаментальных алгоритмов.

Отметим, что приведенное выше определение никак не связано с понятиями класса и наследования. Вместе с тем в ООП полиморфизм (который правильнее называть *подтиповым полиморфизмом*) реализуется именно путем наследования от базового класса, в терминах операций которого



написан алгоритм. Очевидно, это довольно узкий вид полиморфизма, в частности:

требуемое отношение наследования должно быть указано автором класса при его написании, что делает трудным, например, совместное использование двух независимых библиотек (т. к. у пользователя отсутствует возможность добавить в них требуемые отношения наследования);

недостаточная гибкость подхода приводит, например, к проблеме бинарного метода – невозможности в терминах ООП описать полиморфный бинарный метод [7].

Оценка ООП

Приведем еще раз список возможностей, которые были значительно усовершенствованы ООП по сравнению со структурным программированием:

- 1) инкапсуляция «на нужном уровне»;
- 2) повторное использование кода;
- 3) полиморфизм.

Отметим, как похож этот список на характерные черты ООП, приведенные ранее:

- 1) объектно-ориентированная инкапсуляция операций и данных;
- 2) наследование;
- 3) подтиповый полиморфизм.

Подобное сходство свидетельствует о том, что ООП дало конкретные и практически значимые ответы на важные вопросы организации программ, в чем его несомненное достоинство (и одна из основных причин успеха).

Но хороши ли эти ответы? Краткий анализ в предыдущих разделах показывает некоторые из их многочисленных недостатков. Если попытаться их обобщить, то можно отметить следующее: ООП – *сложная* парадигма. Центральные понятия объекта, класса и наследования перегружены назначением и обязанностями. Говоря простым языком, «когда у тебя есть только молоток, все начинает выглядеть, как гвоздь».

Кроме того, одной из основных проблем современного программирования является эффективное использование гетерогенного многопоточного аппаратного обеспечения со сложной иерархией уровней доступа к памяти. Практика показывает, что для этого необходим строгий контроль за расположением данных в памяти и их изменением. Объектно-ориентированная инкапсуляция, наоборот, поощряет зависимость от изменяемого состояния и его неконтролируемое из-

менение, при этом не давая необходимого уровня контроля за расположением данных.

Альтернативная ООП парадигма программирования

На основе проведенного анализа предложим более простую и гибкую парадигму программирования. Опять рассмотрим базовые потребности промышленного программирования:

- 1) инкапсуляция «на нужном уровне»;
- 2) повторное использование кода;
- 3) полиморфизм.

Для каждой из потребностей предложим альтернативный ООП механизм.

Инкапсуляция. В качестве «нужного уровня» инкапсуляции предложим модули. Модули были специально разработаны для решения этой задачи, но начали широко применяться только с распространением объектно-ориентированного программирования (где их функциональность дублируется классами).

Повторное использование. Для решения задачи повторного использования поведений (наборов связанных алгоритмов) будем использовать трейты [5] – специально разработанный с учетом опыта классического ООП механизм. Отметим, что трейты не требуют объектно-ориентированного языка программирования, хотя обычно описываются в таком контексте.

Полиморфизм. Обобщенное программирование («дженерики», «шаблоны» [6]) разработано для написания полиморфных алгоритмов, т. к. объектно-ориентированный подход не предоставляет необходимой гибкости. Позже объектно-ориентированные языки (Java, C# и др.) стали (частично) поддерживать этот подход, таким образом, как и в случае с модулями, предоставляя дублирующие друг друга средства.

Предлагается использовать обобщенное программирование в качестве единственного механизма полиморфизма.

В статье проведен анализ потребностей промышленного программирования и методов, предлагаемых ООП, для их удовлетворения. Выявлена объективная сложность и недостаточная гибкость объектно-ориентированного подхода. Предложена комбинация известных методов, специально разработанных для решения перечисленных потребностей, как основа более простой и гибкой парадигмы программирования.

В дальнейшем предполагается исследование построения нового языка программирования на основе тесной интеграции предло-

женных методов и проведение его детального сравнения с существующими промышленными языками.

СПИСОК ЛИТЕРАТУРЫ

1. **Pierce, B.** Types and Programming Languages [Text] / B. Pierce. –MIT Press, 2002.
2. **Mitchell, J.C.** Concepts in programming languages [Text] / J.C. Mitchell. –Cambridge University Press, 2003. –278 p.
3. **Biancuzzi, F.** Masterminds of Programming [Text] / F. Biancuzzi, S. Warden. –O'Reilly Media, 2009.
4. **Odersky, M.** An overview of the Scala programming language [Text] / M. Odersky // ACM SIGPLAN. –2006.
5. **Schärli, N.** Traits: Composable Units of Behavior [Text] / N. Schärli, S. Ducasse, O. Nierstrasz [et al.]. –Springer – Science, 2003. –P. 327–339.
6. **Musser, D.** Generic Programming [Text] / D. Musser, A. Stepanov // Symbolic and Algebraic Computation: International symp. –1988. –P. 13–25.
7. **Cardelli, L.** On Binary Methods [Text] / L. Cardelli // Theory and Practice of Object Systems. –1995. –P. 221–242.

УДК 004.413:004.414.3:004.415.53

Е.В. Пышкин

ПРОБЛЕМЫ АВТОМАТИЗАЦИИ ПРИЕМОЧНОГО ТЕСТИРОВАНИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ПРИ ИСПОЛЬЗОВАНИИ ПОДХОДА «РАЗРАБОТКА, УПРАВЛЯЕМАЯ ОПИСАНИЕМ ПОВЕДЕНИЯ»

Подход «разработка, управляемая описанием поведения» (behavior driven development, BDD [1]) развивается в рамках гибких (agile) подходов к разработке программного обеспечения давно, тем не менее в отечественных источниках эта тема освещена недостаточно. Метод BDD возник и развивается и как технология, нацеленная на автоматизацию тестирования готового программного продукта заказчиком, и как проектная практика сродни практике «тесты вначале» и основанной на ней методологии разработки через тестирование (test-driven design). Действительно, использование модульных тестов как центрального элемента подхода «тесты вначале» не эквивалентно тестированию в строгом смысле этого понятия: речь идет здесь не столько о тестировании как динамическом методе обеспечения качества, сколько о применении определенных практик тестирования в процессе реализации функциональных требований [2]. Тесты в данном случае, фактически, выступают в качестве альтернативной формы записи этих функциональных требований.

Приемочные тесты относятся к категории тестов, труднее других поддающихся формализации. Одна из задач, решаемых в рамках подхода

«разработка, управляемая описанием поведения», собственно, и состоит в создании определенной инфраструктуры, которая позволила бы осуществить переход от описания тестовых сценариев, сформулированных заказчиком, к тестовым классам и методам для модульного тестирования. При этом предполагается использование существующих библиотек и фреймворков модульного тестирования (unit-тестирования), например библиотеки JUnit для языка Java.

Разработка через описание поведения: конструктивная идея

Согласно известной V-модели организации разработки, приемочные тесты – это своего рода элемент верхнего слоя абстрактной модели процесса разработки (рис. 1).

Приемочные тесты в наибольшей степени сфокусированы на пользовательских требованиях (нередко неясных и плохо сформулированных) и нацелены на процесс верификации, вообще говоря, уже протестированной системы, с участием представителей заказчика. Ошибки, выявляемые в ходе приемочных испытаний, как правило, означают, что либо существует неадек-



Рис. 1. Приемочные тесты в V-модели разработки
(□) процесс; (→) результат процесса; (◄---) верификация

ватность в представлении требований заказчика в технической спецификации, либо обнаруживается разрыв в понимании задачи заказчиками и разработчиками. И то, и другое относится к критическим рискам, возникающим при разработке программного обеспечения, а следовательно, является серьезной проблемой управления разработкой. Поэтому развитие и совершенствование средств, предназначенных для снижения рисков возникновения подобных проблем, представляется актуальным направлением работы инженеров и системных аналитиков.

Основополагающая идея подхода BDD состоит в создании инфраструктуры, позволяющей представителям коллектива разработчиков и заказчикам двигаться навстречу друг другу: первые получают возможность реализовать приемочные тесты максимально близко к стандартной модели модульных тестов (обеспечивающих такие важные свойства agile-практик, как повторяемость, воспроизводимость, инкрементность разработки и т. д.), а вторые могут лучше видеть, каким образом представленные на естественном языке пользовательские сценарии отображаются в элементы тестового программного кода, т. е. в классы и методы на соответствующем языке программирования. Отметим, что BDD относится к развивающимся подходам, основополагающие концепции

которых являются предметом обсуждений, при этом не все ключевые определения вполне ясны и сформированы [3–5].

В чем же состоят основные трудности, с которыми сталкиваются и разработчики, и потребители?

Для того чтобы исполнить тесты, необходимо исполнить соответствующий тестовому сценарию фрагмент кода. Это означает, что нужно понимать данный код, а также уметь сформировать контекст, в котором он выполняется, определить пред- и постусловия, инварианты и т. д. Нормально, что у заказчиков и тем более пользователей обо всем этом имеются в лучшем случае весьма смутные представления. Тем не менее им важно убедиться в выполнении требований технической спецификации и должном функционировании системы.

Свою долю проблем в почти линейный процесс от требований к реализующему их коду (рис. 2) вносят неизбежные изменения, происходящие как на стороне заказчика (и ведущие к изменению требований – ведь понимание заказчиком своих интересов и потребностей тоже развивается вместе с процессом создания системы), так и на стороне разработчика. При этом очевидно, что изменения в разрабатываемом коде (например, в процессе рефакторинга) вовсе не обязательно связаны с изменениями в требованиях.

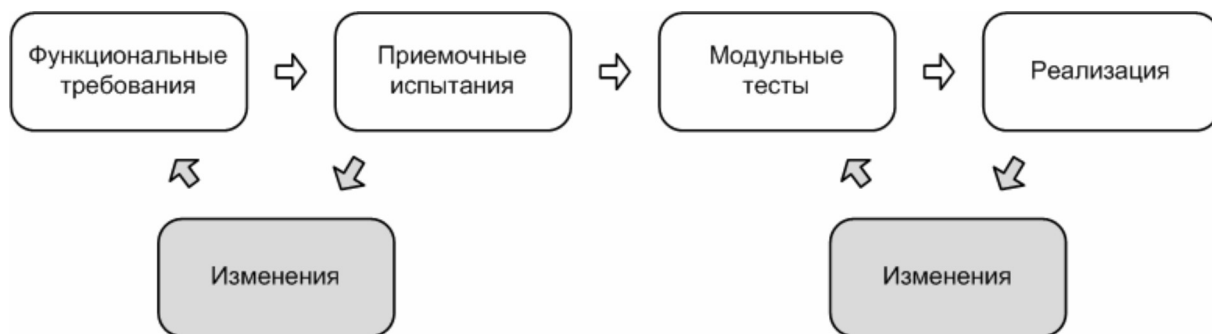


Рис. 2. От требований к модульным тестам

Разработка через описание поведения: от приемочных тестов к тестовым классам

Что же предложено «методологами» BDD? Рассмотрим небольшой пример, иллюстрирующий применение BDD в ходе разработки и тестирования небольшого приложения.

Предположим, что разрабатывается компонент информационной системы, обеспечивающий символьные и численные математические

вычисления.

Тесты правильного функционирования компонента могут быть сформулированы в виде текстовых сценариев на «упрощенном английском языке». Ниже мы рассматриваем последовательность преобразований в том виде, в каком она реализована в модели JBehave [6], хотя аналогичные этапы свойственны и другим реализациям методологии BDD (Листинг 1).

```
Given a derivative function test
When the input function is x^2+3
Then the derivative function is 2*x
Then the derivative value for x=1 is 2 where accuracy is 1e-6
```

Листинг 1. Пример текстового сценария

Очевидно, что могут существовать фрагменты текста сценария, которые неизменны от теста к тесту, в то время как отдельные элементы могут изменять свои значения, например

```
When the input function is 1/x
Then the derivative function is -1/x^2
```

Соответственно, изменяющиеся элементы тестового сценария разумно сделать параметрами теста: структура теста повторяется, параметры изменяются. Методология BDD предполагает, что на основе подобных пользовательских сценариев (narrative-сценариев) создаются соответствующие модульные тесты. Получаемые тесты обеспечивают возможность автоматического запуска, модификации, получения отчетов и интеграции со средой разработки. Иными словами, BDD создает пространство для лучшей совместной работы и лучшего взаимопонимания двух «конкурирующих» сторон (здесь термин «конкуренция» используется в том исконном смысле, в котором оно вошло в информатику: «сонаправленная работа нескольких участников»).

Основные шаги процесса BDD-разработки состоят в следующем.

1. Написание тестовых сценариев, или историй.
2. Отображение этих сценариев в unit-тесты.
3. Описание конфигурации тестовой сборки и запуска.
4. Запуск тестов (например, с использованием возможностей соответствующей среды разработки).
5. Анализ результатов, подготовка отчетов по результатам тестирования, формулирование требований по внесению изменений в программное обеспечение или по формированию новых тестовых сценариев.

Отображение представленного выше сценария проверки вычисления производной функции и значения этой функции в заданной точке в соответствующие модульные тесты (которые в рамках JBehave называются *candidate steps*), может выглядеть так, как представлено в листинге 2 (мы выделяем здесь наиболее существенный фрагмент кода тестового класса, определяемого в соответствии с правилами JBehave).

```

@Given(«a derivative test «)
public void givenADerivativeFunctionTest () {
    // Объект-вычислитель
    calc = factory.createCalculator ();
    // Объект для сравнения функций, заданных символически
    comparator = factory.createSymbolicFunctionComparator();
}

@When(«the input function is $input»)
public void theInputFunction(String input) {
    // Вычисляем производную для функции input
    calc.derivative(input);
}

@Then(«the derivative function is $result»)
public void theDerivativeFunction( String derivative ) {
    // Контролируем результат символьных вычислений
    assertTrue(comparator.equal(calc.getDerivative(),result);
}

@Then(«the derivative value for x=$x is $y where accuracy is $acr»)
public void theDerivativeValue(double x,double y,double acr) {
    // Контролируем корректность значения производной
    // в заданной точке с заданной точностью
    double derivativeVal = calc.getDerivativeValue(x);
    assertEquals(derivativeVal, y, acr );
}

```

Листинг 2. Методы тестового класса JBehave

Полученные в результате такого преобразования тесты могут быть исполнены с подстановкой в качестве аргументов значений, извлекаемых из тестовых сценариев.

Отметим, что в текущей реализации JBehave (равно, как и многих других средств) преобразование из текстового сценария в методы тестового класса, а также формирование значений соответ-

ствующих атрибутов, происходит большей частью вручную. Несмотря на то что переходы от элементов narrative-сценария к методам тестового класса происходят на основе регулярных правил, осуществление такого перехода требует участия разработчика. Актуальной представляется задача автоматизации формирования каркаса тестового класса (листинг 3).

```

@Given("a derivative test ")
public void aDerivativeFunctionTest() {
    // TODO: Write the initial condition
    throw new UnsupportedOperationException();
}

@When(«the input function is $input»)
public void theInputFunction( String input ) {
    // TODO: Write the unit test action
    throw new UnsupportedOperationException();
}

@Then(«the derivative function is $result»)
public void theDerivativeFunction( String result ) {
    // TODO: Write the assertion code
    throw new UnsupportedOperationException();
}
//...

```

Листинг 3. Каркас тестового класса

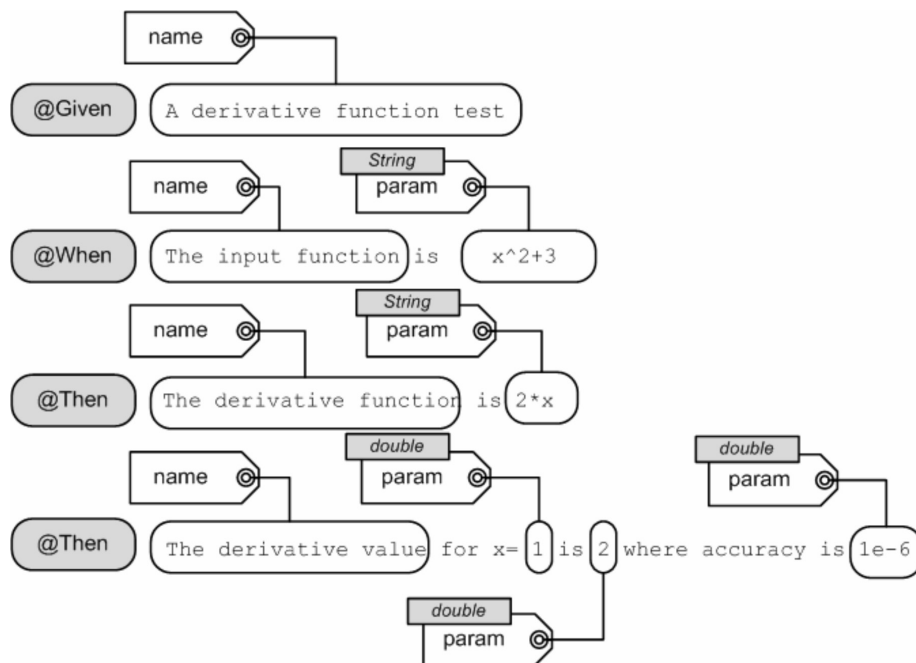


Рис. 3. Процесс разметки narrative-сценария

При постановке и решении задачи автоматизации следует иметь в виду ряд вопросов, ответы на которые неоднозначны, в частности:

- каким образом и из какой части пользовательского сценария формируется имя соответствующего метода тестового класса;
- какие элементы сценария должны быть параметрами методов;
- каковы типы этих параметров.

Таким образом, несмотря на то, что создаваемый тестовый класс для модульного тестирования напоминает тестовый сценарий с точностью до структуры кода на языке программирования, решение задачи автоматической генерации та-

кого каркаса не лишено трудностей (в т. ч. из-за сформулированных выше неоднозначностей). Сам сценарий может основываться на шаблоне (так, например, делается в SpecFlow – реализации BDD для Visual Studio [7]). Отметим, однако, что правила конструирования подобных шаблонов необязательно очевидны представителю заказчика.

Постановка задачи разметки тестовых сценариев

Если рассмотреть форму сценария, в большей степени «приближенную к неспециалисту по тестированию», то для нашего примера она могла бы выглядеть, например, как в листинге 4.

```

A derivative function test
The input function is x^2+3
The derivative function is 2*x
The derivative value for x=1 is 2 where accuracy is 1e-6
  
```

Листинг 4. Сценарий до разметки

Далее возникает задача, которую мы называем задачей разметки тестового сценария (рис. 3).

Фактически, данная разметка (для которой предлагается разработать соответствующие инструментальные средства) призвана, с одной стороны, осуществить переход от плохо структурированного сценария к структурированному описанию, ориентированному на последующую

автоматическую генерацию тестового каркаса, с другой стороны, упростить работу по «доведению» приемочных сценариев до формализованных модульных тестов.

Основные решения, которые должны быть приняты специалистом по тестированию в ходе работы с инструментарием разметки тестовых сценариев, состоят в следующем:



1) формализовать структуру приемочного теста, на которой основывается инструментарий разметки тестовых историй, в рамках предложенной модели определить назначение отдельных элементов (строк) сценария, приписав им соответствующие атрибуты, например, `@Given` – для определения контекста, `@When` – для определения выполняемого действия, `@Then` – для определения проверяемого условия выполнения теста, `@Alias` – для определения синонимов и т. д.;

2) определить названия соответствующих методов тестового класса или правила их конструирования;

3) определить, какие элементы сценария будут выступать в роли параметров тестовых методов;

4) определить типы параметров тестовых методов;

5) определить возможности расширения множества допустимых элементов разметки тестового сценария в соответствии с особенностями конкретной предметной области;

6) определить образцы (шаблоны тестовых сценариев), на основе которых автоматически генерируются или размечаются сценарии с другими значениями исходных данных.

Выполнить все эти шаги автоматически в общем случае невозможно, однако можно сформулировать постановку задачи частичной автоматизации разметки тестовых сценариев с участием специалиста по тестированию. Детальная формулировка требований к разработке инструментария разметки и правил, на которых будут основаны соответствующие преобразования тестового сценария, являются предметом дальнейшей работы.

Реализации и открытые проблемы

В статье [8] на основе обзора существующих реализаций BDD и анализа их основных возможностей, особенно в части интеграции со средами разработки, мы сформулировали список общих требований, которые необходимо учитывать при разработке средств, поддерживающих BDD. Проведенное исследование показывает, что даже те средства, которые рекламируются как поддерживающие написание приемочных тестов в виде простого текста, в действительности устанавливают существенные ограничения на тестовые истории. При этом одним из критических аспектов является обеспечение автоматизируемости отображения тестовых сценариев в модульные тесты.

В связи с этим следует выделить несколько важных моментов, определяющих основные трудности применения подхода BDD.

Несмотря на то что сами модульные тесты легко автоматизируются, задача преобразования пользовательских историй к связанным с реализацией тестовым классам остается актуальной.

Приемочные тесты суть требования. Так как в требованиях (равно как и в самом программном обеспечении) неизбежны изменения, вопрос преобразований приемочных тестов очень важен. В процессе написания «тестовой истории» мы можем обнаружить, что какие-то требования упрощены. С другой стороны, изменения могут затрагивать интерфейсный уровень приложения – в этом случае сами приемочные тесты, возможно, и не изменятся, но изменится процесс отображения. Для представленного выше примера с вычислением производной функции подобное изменение может заключаться в определении, скажем, интерфейса `RealValuedFunction` со следующими методами:

```
public interface RealValuedFunction {  
    void set( String symbolic );  
    double getValue( double x );  
}
```

Это изменение не затрагивает уровень представления приемочных тестов, но реализация тестовых методов, очевидно, изменится (листинг 5).

Еще одним моментом, требующим внимания исследователей, является то, что в существующих реализациях BDD обеспечивается формулирование сценариев ожидаемого (желательного) поведения системы. При этом учет сценариев нежелательного поведения системы также может рассматриваться как возможное расширение модели BDD, важное для практики разработки программного обеспечения.

Автоматизация перехода от плохо структурированного текста на естественном языке к упрощенному структурированному языку тестовых сценариев остается существенной проблемой (см. также [9]). Полезным может оказаться применения опыта и подходов из области NLP-технологий. Заметим, что даже более простая задача поддержания процесса разметки тестовой истории для обеспечения возможности дальнейшей автоматической генерации модульных, далека от окончательного решения.

```
@Given(«a derivative function test <»)  
public void aDerivativeFunctionTest() {  
    calc = factory.createCalculator();  
    comparator =factory.createSymbolicFunctionComparator();  
}  
  
@When(«the input function is $input»)  
public void theInputFunction(String input) {  
    // Создаем объект класса, реализующего интерфейс  
    // RealValuedFunction  
    RealValuedFunction rvf = new RealValuedFunction() {  
        @Override  
        public set( String input ) {  
            // Реализация функции...  
        }  
        @Override  
        double getValue( double x ) {  
            // Реализация вычисления значения в точке...  
        }  
        @Override  
        String toString() {  
            // Преобразование к символьному представлению...  
        }  
    };  
    rvf.set( input );  
    // Вычислитель теперь работает с функциями,  
    // реализующими интерфейс RealValuedFunction  
    calc.set( rvf );  
}  
  
@Then(«the derivative function is $result»)  
public void theDerivativeFunction( String result ) {  
    RealValuedFunction derivative = calc.derivative();  
    assertTrue(comparator.equal(derivative.toString(),result);  
}  
  
@Then(«the derivative value for x=$x is $y where accuracy is $acr»)  
public void theDerivativeValue(double x,double y,double acr) {  
    RealValuedFunction deriviative = calc.derivative();  
    assertEquals(derivative.getValue(x), y, acr );  
}
```

Листинг 5. Модификация тестового класса

Средства BDD все еще в сильной степени ориентированы на разработчиков. В то же время технологическая поддержка BDD все еще ограничена. В большинстве публикуемых примеров (включая и данную статью) тестовые сценарии и unit-тесты написаны одними и теми же специалистами. Это означает, что внедрение обсуждаемых технологий в проектную практику, фактически, означает необходимость дополнительной рабо-

ты, которая опять же ложится на плечи разработчиков. Поэтому даже небольшие улучшения существующих технологических средств могут оказать заметное влияние на востребованность BDD-инструментария как у самих разработчиков, так и у инженеров по контролю качества, которые, вероятно, и есть целевая аудитория, ответственная за формирование правильных отношений между заказчиками и разработчиками.



СПИСОК ЛИТЕРАТУРЫ

1. **North, D.** Behavior modification: the evolution of behavior-driven development [Текст] / D. North. –Better Software. –March 2006.
2. **Beck, K.** Aim, Fire [Текст] / K. Beck // IEEE Software. –2001. –Vol. 18. –№ 5. –P. 87–89.
3. **Solis, C.** A study of the characteristics of behaviour driven development [Текст] / C. Solis, X. Wang // In Proc. of 37th EUROMICRO Conf. on Software Engineering and Advanced Applications (SEAA). –Oulu, Aug 30–Sep 2. 2011.
4. **Janzen, D.** Test-driven development: concepts, taxonomy, and future directions [Text] / D. Janzen, D.H. Saiedian // Computer. –2005. –Vol. 38. –№ 9. –P. 43–50.
5. **Janzen, D.** Does test-driven development really improve software design quality? [Текст] / D. Janzen, D.H. Saiedian // IEEE Software. – March/Apr., 2008. –Vol. 25. –№ 2. –P. 77–84.
6. What is JBehave? [Электронный ресурс]. –Режим доступа: <http://jbehave.org> (Дата обращения 2012)
7. **Sanderson, S.** Behavior driven development (BDD) with SpecFlow and ASP.NET MVC [Электронный ресурс] / S. Sanderson. –Режим доступа: <http://blog.stevensanderson.com/2010/03/03/behavior-driven-development-bdd-with-specflow-and-aspnet-mvc/> (Дата обращения 2010)
8. **Pyshkin, E.** On requirements for acceptance testing automation tools in behavior driven software development [Электронный ресурс] / E. Pyshkin, M. Mozgovoy, M. Glukhikh // In Proc. of the Central & Eastern European Software Engineering Conf. Russia (CEE-SEC(R) 2012). –Nov. 1-2, Moscow, Russia. –Режим доступа: http://www.secr.ru/2012/presentations/pyshkin-mozgovoy-glukhikh_80_article.pdf (Дата обращения 2012)
9. **Borg, R.** Automated acceptance test refactoring [Электронный ресурс] / R. Borg, M. Kropp // In Proc. of WRT'11. –Waikiki, Honolulu, HI, USA. –May 22, 2011.

УДК 004.415

Л.Ю. Лабошин, А.А. Лукашин, В.Б. Семеновский

ПОДДЕРЖКА ВЫЧИСЛЕНИЙ НА ГРАФИЧЕСКИХ УСКОРИТЕЛЯХ В ВИРТУАЛЬНОЙ СРЕДЕ ХЕН ДЛЯ ОБРАБОТКИ ПОТОКОВ ДАННЫХ РОБОТОТЕХНИЧЕСКИХ СИСТЕМ

Перспективы применения графических ускорителей

Современные робототехнические системы испытывают необходимость обработки массивов потоковых данных в режиме реального времени. Актуальны задачи обработки видеопотоков, решения систем алгебраических уравнений и расчета поведения динамических систем. Перспективным направлением решения этих задач является передача потоков данных по телематическим каналам в центры обработки данных (ЦОД). Современные ЦОД базируются на технологиях виртуализации и облачных вычислений.

Для эффективной обработки потоков данных, поступающих с робототехнических систем, могут использоваться графические процессорные устройства (GPU), изначально предназначенные для обработки видеоинформации, но на данное время активно применяемые для проведения научных и инженерных расчетов. Параллельные

участки кода могут быть перенесены на GPU, где будут выполняться множеством нитей (threads). В целом работа нитей на GPU соответствует архитектуре SIMD (Single Instructions – Multiple Data), однако имеется существенное различие. Только нити в пределах одной группы выполняются физически одновременно. Нити различных групп могут находиться на разных стадиях выполнения программы. Такой метод обработки данных обозначается термином SIMT (Single Instructions – Multiple Threads) [1]. Управление работой групп производится на аппаратном уровне.

Вычислительные системы, в которых часть вычислительной нагрузки принимают на себя графические процессорные устройства, потенциально сравнимы с суперкомпьютерами в решении некоторых классов вычислительно емких задач, таких, как, например, численные методы, молекулярная динамика, квантовая химия [2]. Для решения задач на графических процессорах существу-

ет ряд средств и технологий программирования, в частности, OpenCL [3] и CUDA [4].

Несмотря на то что использование многопоточных вычислительных систем позволяет получить значительный прирост производительности на системе с графическим процессором по сравнению с процессором общего назначения, существующие системы используют GPU монополю, что препятствует использованию преимуществ технологий виртуализации как средства совместного использования и консолидации ресурсов. В статье предложена методика, позволяющая использовать графические процессорные устройства для ускорения параллельных вычислений в виртуальных машинах, что делает возможным осуществление потоковых вычислений в виртуализованных ЦОД.

Метод передачи графического ускорителя в виртуальную среду

Существует два подхода к обеспечению доступа виртуальной машины к PCI-устройствам [5]. Первый подход называется *паравиртуализация* и предполагает использование специализированных драйверов. Создание таких драйверов для современных графических ускорителей не представляется возможным ввиду их закрытой архитектуры. Второй подход – *обеспечение непосред-*

ственного доступа гостевой ОС к физическому устройству. В этом случае в виртуальную машину отображаются определенные участки памяти и прерывания физической машины. Такое отображение стало возможным с появлением технологий аппаратной виртуализации ввода-вывода под общим названием IOMMU (Input/Output Memory Management Unit). В этом случае для работы с устройством можно использовать драйверы, предоставляемые производителем. Возможность непосредственной передачи реализована в большинстве современных мониторов виртуальных машин и хорошо работает с рядом периферийных устройств на шине PCI, например сетевыми адаптерами и обеспечивает производительность, близкую к непосредственной. Однако графический контроллер, являясь PCI-устройством, имеет ряд особенностей, усложняющих организацию его передачи.

Аппаратные платформы с архитектурой x86 имеют функцию прерывания E820, использующуюся для получения информации о распределении адресного пространства, имеющейся оперативной памяти и зарезервированных адресах. Видеокарта имеет зарезервированный диапазон видеопамати, в котором размещаются данные, отображаемые на экране. Область видеопамати для кратковременного хранения одного или не-

Таблица 1

Модель памяти архитектуры x86

Диапазон адресов	Тип памяти	Назначение
00000000-000003FF	Conventional memory	Real Mode Interrupt Vector Table
00000400-000004FF		BIOS Data Area
00000500-00007BFF		Conventional Memory
00007C00-00007DFF		Boot Sector
00007E00-0009FBFF		Conventional Memory
0009FC00-0009FFFF		Extended BIOS Data Area
000A0000-000AFFFF	Upper Memory	VGA Graphics Mode Video RAM
000B0000-000B7FFF		VGA Monochrome Text Mode Video RAM
000B8000-000BFFFF		VGA Color Text Mode Video RAM
000C0000-000C7FFF		VGA Video BIOS ROM
000C8000-000CBFFF		IDE Hard Disk BIOS ROM
000CC000-000DFFFF		Optional Adapter ROM BIOS or RAM UMBs
000E0000-000EFFFF		System BIOS PnP Extended Information
000F0000-000FFFFF		System BIOS ROM
00100000-0010FFEF	High Memory	High Memory Area
0010FFF0-...	Extended Memory	Extended Memory

скольких видеокадров в цифровом виде называется *фреймбуфером* и располагается в памяти машины по фиксированному диапазону адресов 0xA0000 – 0xC0000. Для того чтобы операционная система могла адресовать видеоадаптер, этот раздел памяти должен быть отображен в адресное пространство виртуальной машины. Так как VGA адаптер не назначает адреса динамически, монитор виртуальных машин не может использовать стандартный механизм прямой передачи PCI-устройства, и логический адрес не может быть транслирован в физический средствами гипервизора, поэтому необходима специальная проверка для диапазонов VGA адресов. Адреса, которые необходимо отобразить в физическую память, представлены в табл. 1.

VGA BIOS расположен по адресам 000C0000-000C7FFF и необходим только во время загрузки операционной системы, после того как ОС получает контроль над устройством, он больше не требуется. Для того чтобы позволить виртуальной машине должным образом инициализировать графический контроллер, он должен быть выполнен, однако «простого» копирования этого диапазона в адресное пространство виртуальной машины недостаточно, потому что во время выполнения его хост-системой он может быть изменен для некоторых графических карт. Поэтому необходимо заранее извлечь VGA BIOS карты с помощью утилиты производителя непосредственно с устройства и привязать его к адресу 0xC0000 гостевой ОС.

При передаче GPU в виртуальную машину необходимо также предотвратить загрузку эмулированного графического адаптера и запретить использование адаптера хост-системой.

Таким образом, для реализации передачи графического адаптера в виртуальную машину необходимо предпринять следующие шаги:

- определить идентификатор графического контроллера;
- скрыть необходимое устройство от хостовой ОС, с использованием модуля pci-stub;
- предотвратить загрузку эмулированного графического адаптера виртуальной машины;
- разрешить доступ к портам ввода-вывода, специфичным для видеокарт 0x3B0 – 0x3BC, 0x3C0 – 0x3E0;
- обеспечить отображение диапазона MMIO видеобуфера карты (0xA0000–0xBFFFF) в виртуальную машину;
- извлечь VGA BIOS графического контроллера непосредственно с устройства и скопировать его в гостевую операционную систему по адресу 0xC000 и обеспечить его выполнение во время загрузки виртуальной машины.

Исследование эффективности применения графического ускорителя в виртуальной машине

В рамках исследования разработали стенд с характеристиками, описанными в табл. 2, осуществили необходимые шаги и модифицировали код гипервизора в соответствии с разработанной методикой для обеспечения поддержки графического ускорителя [6].

В образы протестированных виртуальных машин предустановили драйвера графического адаптера (Developer Drivers for WinXP (версия 270.81), Developer Drivers for Linux (версия 270.41.19)) и средства для производства потоковых вычислений (Cuda Toolkit 4.0).

Для оценки эффективности графического процессора в виртуальной машине провели тесты, основанные на программе генерации криптовалюты Bitcoin [7], использующей интенсивный обмен данными между оперативной памятью и памятью графического процессора. Алгоритм ра-

Таблица 2

Экспериментальный стенд

Аппаратная платформа	Монитор виртуальных машин	Хостовая операционная система	Графический контроллер	Гостевые операционные системы
IntelSR2600, с поддержкой технологии аппаратной виртуализации Intel vt-d	Гипервизор Xen версии 4.2	Ubuntu 11.10 64-бит	nVidia GT430	Windows XP SP3 32-бит
				OpenSuse 11.4 64-бит

Таблица 3

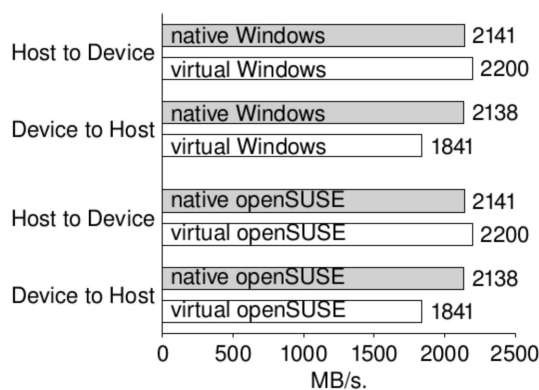
Производительность Bitcoin, млн хешей в секунду

	ОС Windows XP		ОС OpenSUSE	
	Без виртуализации	С виртуализацией	Без виртуализации	С виртуализацией
Центральный процессор	0,5	0,24	1,32	0,4
Графический процессор	14,28	14,15	14,24	14,24

боты программы – это поиск уникального хеша с определенным количеством ведущих нулей. Компьютер генерирует миллионы хешей; после вычисления хеша, удовлетворяющего условиям, результат отправляется для проверки. Программа максимально использует ресурсы компьютера для перебора хешей, поэтому хорошо иллюстрирует вычислительные возможности графического и центрального процессоров. Клиентское приложение реализовано на языке java и использует для доступа к вычислительным возможностям GPU язык OpenCL. Результаты данного теста сведены в табл. 3.

Для исследования скорости обмена данными между операционной системой и графическим ускорителем были проведены тесты пропускной способности, которые входят в набор разработчика, поставляющийся с программной оболочкой CUDA. Результаты тестов приведены на графике.

Предложен метод переноса графического ускорителя в виртуальную машину. Проведенные исследования показали, что с применением данного метода затраты на реализацию не приводят к существенной потере производительности вычислений на графическом процессорном устройстве. Это объясняется тем, что когда переданы необходимые данные, устройство работает независимо.



Скорость передачи данных между ОС и устройством

Таким образом, снижение производительности обусловлено только разницей в скорости обмена данными между памятью устройства и оперативной памятью виртуализованной и не виртуализованной систем, и составляет не более 15 %.

Результаты проведенного исследования могут применяться для решения задачи параллельной обработки интенсивного потока данных, поступающего с робототехнических систем с использованием графических процессоров в виртуальных машинах современных ЦОД, что позволит повысить вычислительные возможности робототехнических систем.

Работа выполнена при финансовой поддержке Министерства образования и науки РФ, контракт № 14.А18.21.0593.

СПИСОК ЛИТЕРАТУРЫ

1. Боресков, А.В. Основы работы с технологией CUDA [Текст] / А.В. Боресков, А.А. Харламов. – СПб.: ДМК Пресс, 2010. – 232 с.
2. GPU Technology Conference. 2012 [Электронный ресурс] / Режим доступа: <http://www.gputechconf.com> (Дата обращения 07.06.2012)
3. OpenCL – The open standard for parallel programming of heterogeneous systems [Электронный ресурс] / Режим доступа: <http://www.khronos.org/opencl> (Дата обращения 07.06.2012)
4. Параллельное программирование и вычислительная платформа CUDA [Электронный ресурс] /

Режим доступа: http://www.nvidia.ru/object/cuda_home_new_ru.html (Дата обращения 07.06.2012)

5. Williams, D.E. Virtualization with Xen [Text] / D.E. Williams; Ed. by J. Garcia. – Burlington: Syngress-Publishing, 2007. – 364 p.
6. Bitcoin P2P Digital Currency [Электронный ресурс] / Режим доступа: <http://www.bitcoin.org> (Дата обращения 07.06.2012)
7. Лабошин, Л. Методы и средства использования специализированных PCI-устройств в виртуальной среде: Дипломная работа [Текст] / Л. Лабошин – СПб.: СПбГПУ, 2012.

СТОХАСТИЧЕСКАЯ МОДЕЛЬ СТРУКТУРЫ ПУЛА ЗАЯВОК ПОСТАВЩИКОВ НА ЗАКРЫТОМ АУКЦИОНЕ СПОТОВЫХ ЦЕН

В течение продолжительного периода времени (с 2002 по 2009 г. включительно) о ходе торгов на спотовом рынке электроэнергии ЕЕХ [1] можно было судить только по двум временным рядам (в. р.): почасовым спотовым ценам Pr_{hd} и объемам сделок на каждый час Vol_{hd} . Оба процесса относят к классу нестационарных случайных процессов (с. п.) и большинство параметрических моделей предлагались как сумма нескольких слагаемых (как сумма с. п.) [2]. Традиционно слагаемые включали тренд [3, 4], явно выраженную периодическую [5], а также сингулярную составляющую [6]. Последняя описывала одно из наиболее примечательных свойств в. р. спотовых цен – выбросы. Второй в. р. с объемами сделок на каждый час, как правило, игнорировали, хотя он содержит не менее важную информацию. В данной статье предпринимается попытка привлечь и этот в.р. для построения модели.

Постановка задачи

Рассмотрим следующую задачу анализа данных, полученных в некоторый день d .

Пусть в этот день d на час h зафиксирован объем сделок в Vol_{hd} с ценой отсечения Pr_{hd} . Представим структуру пула заявок поставщиков этого закрытого аукциона в виде множества пар

$$\{pr_i^{(h)}, v_i^{(h)}\}, i \in I = \{1, \dots, N\} : \forall i < N \quad pr_i^{(h)} \leq pr_{i+1}^{(h)}, (1)$$

где $pr_i^{(h)}, v_i^{(h)}$ – цена и объем поставки, указанные в i -й заявке на час h дня d .

Другими словами, множество заявок упорядочено с помощью естественного отношения порядка первого элемента пары – упорядочение по цене заявки.

Тогда цену отсечения можно определить следующим образом:

$$Pr_{hd} := \sup_{i \in I, i < l_{\min}} \left\{ pr_i^{(h)} \mid l_{\min} = \inf_l \left\{ l : \sum_{i=1}^l v_i^{(h)} \geq Vol_{hd} \right\} \right\}. \quad (2)$$

В условиях, когда доступная информация ограничена только двумя в. р., математические модели, позволяющие идентифицировать структуру пула заявок производителей, можно построить только с привлечением некоторых допущений, справедливых не для любого участка в. р.

Предлагаемые нами построения в первоначальной формулировке можно рассматривать как правомерные только на тех временных участках, где агрегированный объем сделок за сутки ведет себя как квазистационарный с. п., и на рынке сырья не наблюдается значимых колебаний в цене. Последнее условие означает следующее: мы исходим из предположения, что цены на сырье (подчеркиваем, на сырье, а не спотовые) ведут себя на этом участке как стационарные процессы.

Рассмотрим следующую аддитивную модель в. р. со спотовыми ценами. Пусть он представлен как конечная сумма случайных стационарных в узком смысле процессов. Ключевая особенность этой суммы состоит в том, что нарушение аксиомы коммутативности или, другими словами, независимые приращения по объемам, указанным в заявке, подлежат суммированию в порядке, который задает отношение порядка по цене заявки $pr_i^{(h)}$. Аксиома коммутативности для этой суммы нарушается, что очевидно в силу связи между элементами пары $\{pr_i^{(h)}, v_i^{(h)}\}$.

В выражение для плотности вероятности как смеси распределений отдельные слагаемые входят с некоторыми случайными весами π_k :

$$f_{Pr_{hd}}(pr | Vol_{hd}) = \sum_{k=1, k_{\min} \leq K}^{\hat{k}_{\min}} \pi_k \cdot p_k(pr_{k\Sigma_{ik}} = pr | v_{k\Sigma_{ik}}^{(h)} = Vol_{hd} - \sum_{m \neq k} v_{m\Sigma_{im}}^{(h)}), \quad (3)$$

где K – число входящих в сумму слагаемых, равное количеству идентифицированных ценовых кластеров, $p_k(pr_{k\Sigma_{ik}} = pr | v_{k\Sigma_{ik}}^{(h)} = Vol_{hd} - \sum_{m \neq k} v_{m\Sigma_{im}}^{(h)})$ – вероятность того события, что для суммы объемов $v_{k\Sigma_{ik}}^{(h)}$, полученной для нескольких первых заявок из кластера k , цена последнего слагаемого в этой сумме достигает цены отсечения $pr_{k\Sigma_{ik}} = pr$.

Здесь и далее h – порядковый номер часа в пределах суток, d – порядковый номер дня во в. р.

Если предположить, что цена на каждый час формируется распределением только из одного единственного кластера, то случайные веса π_k будут играть роль индикаторных переменных. Тогда вектор π для каждого часа, составленный из π_k , содержит только одну единицу и $(K-1)$ нулей; $p_k(\bullet)$ – необязательно одномодальное распределение.

Будем предполагать, что случайная мера для отдельного k -го слагаемого в (2) определяется своим законом распределения $p_k(\bullet)$, обусловленным прежде всего себестоимостью генерируемой энергии; распределению $p_k(\bullet)$ подчиняются цены заявок производителей, генерирующих электроэнергию на основе одного и того же типа энергоносителей или различных типов энергоносителей при сопоставимой цене на них в расчете на 1 МВт.

Перед нами стоит задача кластеризации спот-цен на тех участках в. р. спотовых цен, где выполняется условие стационарности агрегированного объема сделок за сутки. Следует заметить, что за весь период с 2002 по 2008 г. удалось выделить очень мало таких участков.

В настоящей статье кластеризуемыми данными является множество $\{Pr_{hd}\}$, которое мы пытаемся представить в форме разбиения:

$$\{pr_{1\Sigma_{i1}}, \dots, pr_{K\Sigma_{iK}}\}, \quad (4)$$

где Pr_{hd} – зарегистрированная спотовая цена сделок на час h в день d . Каждому наблюдению соответствует скрытая переменная $Z_{hd} = k, k = 1, \dots, K$; Z_{hd} – случайная величина, соответствующая указанной Pr_{hd} и равная k , если она принадлежит k -му кластеру; K – количество кластеров.

Вероятность того, что случайная величина Z_{hd}

примет значение k , равна

$$p(Z_{hd} = k) = \hat{\pi}_k, k = 1, \dots, K, \text{ где } \sum_{k=1}^K \hat{\pi}_k = 1.$$

Предвосхищая будущие выводы, можем отметить, что $\hat{\pi}_k$ обеспечили бы нам параметрическую оценку для элементов π_k вектора весов π . При этом возникает принципиальная возможность рассмотреть такие распределения не только в среднем за сутки, но и по отдельности на каждый час. Другими словами, возможна еще и почасовая группировка спотовых цен, наблюдавшихся за выбранный промежуток времени. Однако ограниченность исходных данных не позволяет предложить какую-либо содержательную интерпретацию оценок в последнем случае.

Стандартным инструментом для решения задач кластеризации является ЕМ-алгоритм. Поскольку этот алгоритм предполагает предварительную параметризацию модели, предположим, что данные в каждом кластере подчиняются одному из законов распределения экспоненциального семейства. Такое предположение на практике не всегда выполняется. Тем не менее прибегнем к этой простейшей модели, обосновав свой выбор следующими рассуждениями. На формирование цены на электроэнергию оказывает влияние большое количество факторов. Например, стоимость используемого для генерации сырья, других природных ресурсов, самой электроэнергии, основных средств, трудовых ресурсов. Этот список

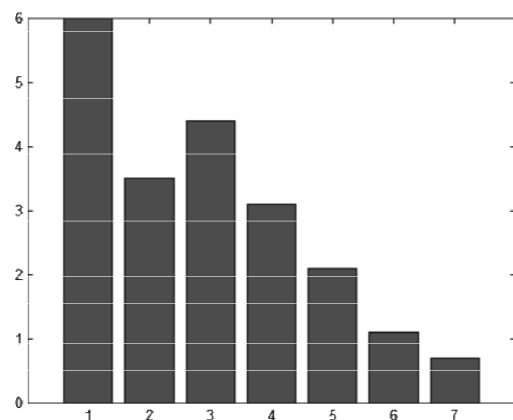


Рис. 1. Себестоимость электроэнергии, производимой с использованием различных видов топлива (средняя мировая в 2005 г.):
1 – нефть по высокой цене; 2 – нефть по низкой цене; 3 – газ по высокой цене; 4 – газ по низкой цене; 5 – уголь по высокой цене; 6 – уголь по низкой цене; 7 – атомная энергия.
По ординате центы США/кВт-ч [8]

Таблица 1

Оценки параметров распределений, входящих в смесь

	Кластер		
	1	2	3
μ , €/МВт·ч	36,15	63,76	78,39
σ , €/МВт·ч	9,11	6,35	12,93

содержит все прямые и косвенные затраты на ее производство и реализацию.

Проверке гипотезы на выбранном участке в. р. предшествовала процедура построения непараметрической оценки плотности (гистограммы) для спотовых цен. По этой гистограмме оценивалось предельно возможное количество кластеров на данном участке.

В настоящей работе для кластеризации спот-цен за выбранный промежуток времени использовался обобщенный ЕМ-алгоритм [7]. В качестве параметрической модели для спотовых цен, попадающих в один кластер, выбрали нормальное распределение. Естественно, эта гипотеза H_0 подлежала обязательной проверке.

За период с пятого октября по восьмое ноября 2005 г. (выбор этого промежутка объясняется доступностью для него независимых оценок средней себестоимости электроэнергии в зависимости от вида сырья – см. рис. 1) мы получили три кла-

стера (рис. 2) с параметрами распределения, указанными в табл. 1.

На рис. 1 для сравнения представлена средняя себестоимость электроэнергии, производимой с использованием различных видов топлива. В эти данные не включена стоимость разрешений (сертификатов) на выбросы CO_2 . Но даже с учетом этого обстоятельства можем сказать, что получили вполне сопоставимые с действительностью результаты кластеризации.

После выполнения процедуры кластеризации проверим гипотезу о том, что спотовые цены, попадающие в один кластер:

- принадлежат классу нормальных распределений (критерий Колмогорова–Смирнова; если количество наблюдений мало, то также критерий Шапиро–Уилка);
- принадлежат нормальному распределению с параметрами, полученными с помощью ЕМ-алгоритма (критерий Пирсона).

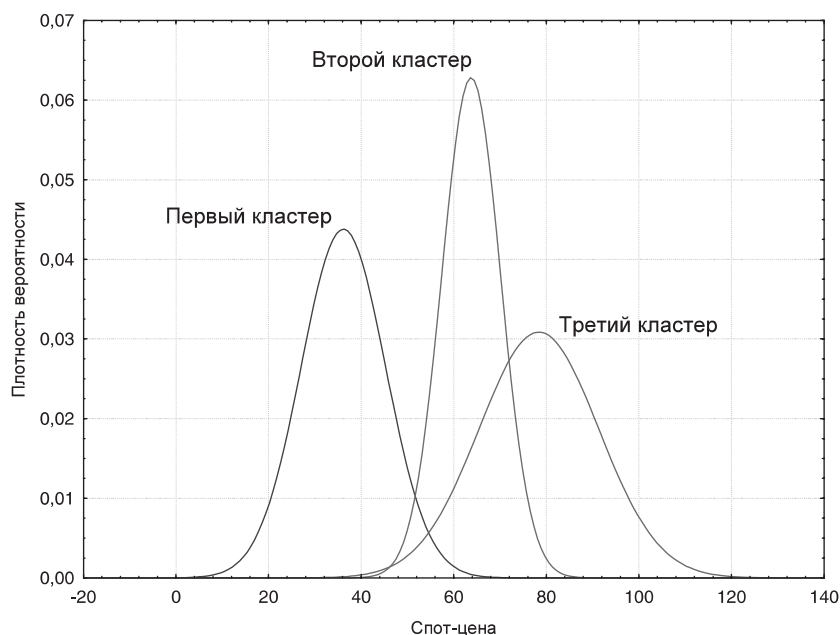


Рис. 2. График распределений спотовых цен в каждом из кластеров, полученных в результате работы ЕМ-алгоритма

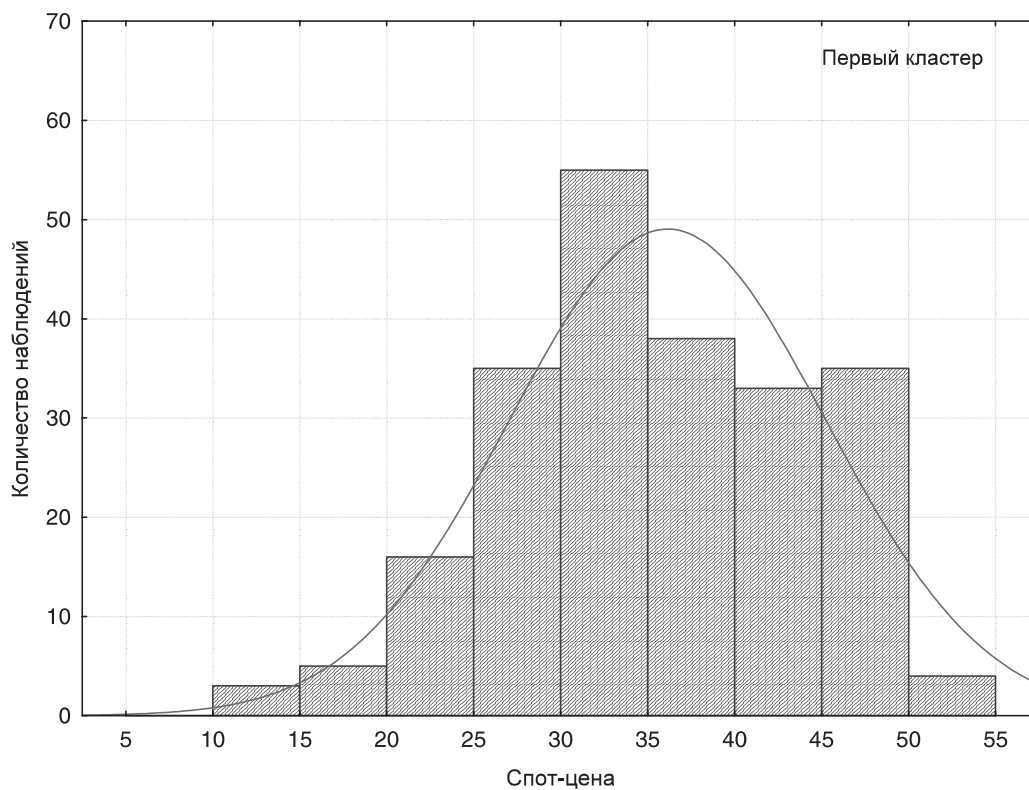


Рис. 3. Гистограмма спотовых цен, попавших в первый кластер

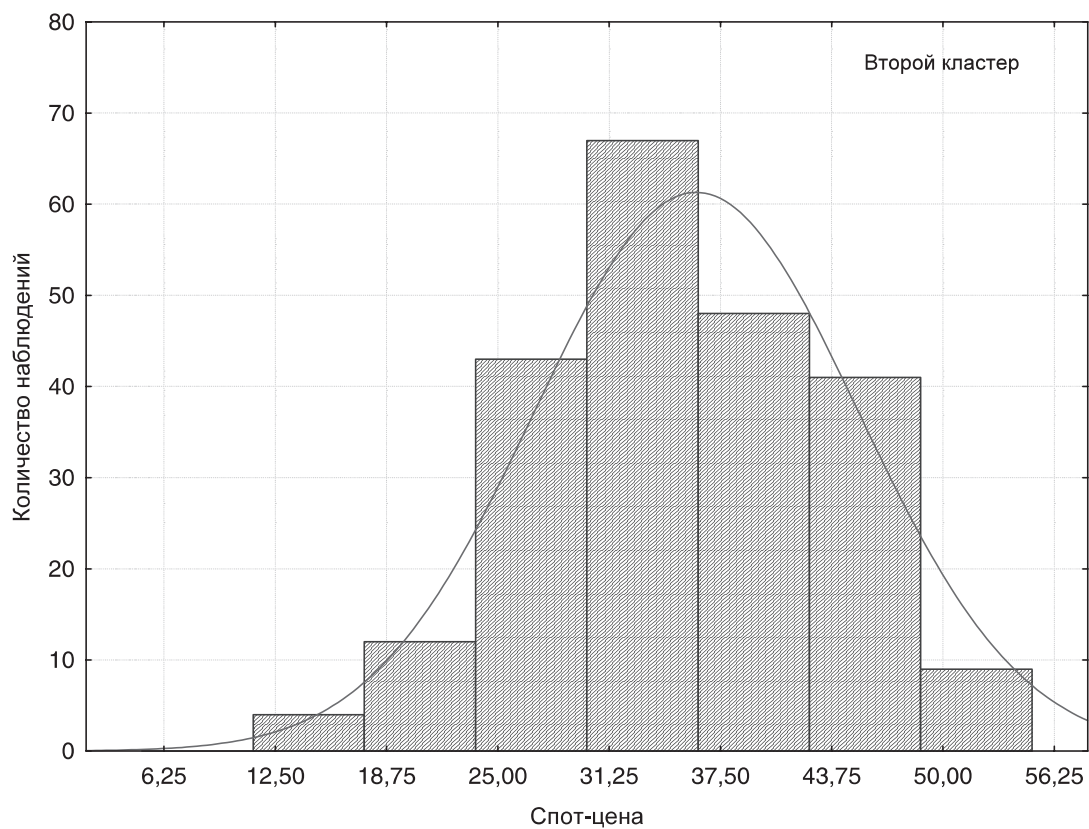


Рис. 4. Гистограмма спотовых цен, попавших во второй кластер

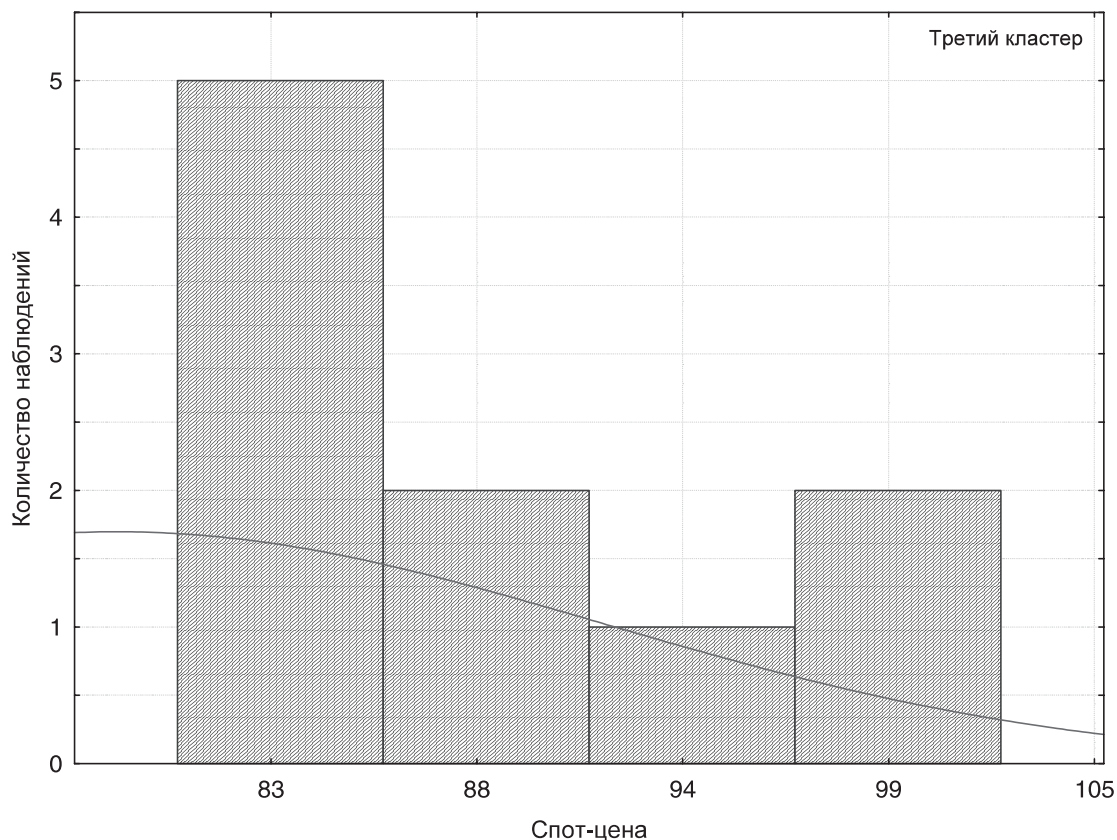


Рис. 5. Гистограмма спотовых цен, попавших в третий кластер

На рис. 3–5 изображены гистограммы спотовых цен в кластерах, а также функции плотностей нормального распределения с оценками параметров, полученными с помощью ЕМ-алгоритма.

В табл. 2 представлены расчетные и табличные

значения статистик χ^2 , Колмогорова–Смирнова и Шапиро–Уилка для каждого из кластеров.

Как видно из таблицы, простая и сложная гипотезы H_0 для первого и второго кластеров не отвергаются лишь при уровне значимости ниже 0,01.

Таблица 2

Расчетные и табличные значения статистик χ^2 , Колмогорова–Смирнова, Шапиро–Уилка для каждого из кластеров

$\alpha = 0,01$	Кластер		
	1	2	3
Количество наблюдений	224	318	10
Критерий Пирсона			
χ^2	10,583	16,679	-
Число степеней свободы	3	8	-
χ^2 табл.	11,345	20,09	-
Критерий Колмогорова–Смирнова			
D	0,072	0,041	0,624
D табл.	0,109	0,091	0,49
Критерий Шапиро–Уилка			
W	-	-	0,791
W табл.	-	-	0,781

О данных в третьем кластере можем лишь сказать, что он сформирован заявками, поступившими от поставщиков, работающих скорее всего на разных видах сырья. Очевидно, что данные, которые отнесли к первому и третьему кластерам, «засорены», т. е. для их случайной меры разумнее либо использовать бимодальные распределения, либо, в свою очередь, разделить на два кластера. Увеличение числа кластеров в данном случае осложняется ограниченным объемом данных (ограничение на длину участка, где выполняется условие квазистационарности агрегированного объема сделок за сутки).

Напомним, что выбранный для иллюстрации временной отрезок с пятого октября по восьмое ноября 2005 г., длина которого около месяца, – один из немногих, где выполняются ограничения на подобные построения. Более того, увеличение числа кластеров должно быть обосновано дополнительной информацией о структуре пула заявок поставщиков, а для того периода она недоступна.

Выявление стохастической зависимости между характером изменения почасовых объемов сделок и попаданием спотовой цены в один из кластеров

Следующий этап в нашей работе – выявление стохастической связи между характером изменения почасовых объемов сделок и попаданием спот-цены в один из кластеров. Пусть $vol_d^{\min}$ – объем, характеризующий минимальный уровень потребления электроэнергии за сутки d , необходимый для удовлетворения базовых потребностей, таких, как отопление, подогрев воды и др. Изменением почасовых объемов сделок будем считать величину:

$$\Delta vol_{hd} = vol_{hd} - vol_d^{\min}, d = 1, \dots, N, \quad (5)$$

где h – номер часа; d – номер суток; vol_{hd} – фактическое значение объема сделок на час h суток d ; $vol_d^{\min}$ – минимальное значение объема сделок, наблюдавшееся в течение суток d . Через k_{hd} бу-

Таблица 3

Оценки вероятностей попадания цены в тот или иной кластер

	Кластер		
	1	2	3
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
5	1	0	0
6	1	0	0
7	0,82609	0,17391	0
8	0	1	0
9	0	0,95652	0,04348
10	0	0,95652	0,04348
11	0	0,95652	0,04348
12	0	0,91304	0,08696
13	0,04348	0,95652	0
14	0	1	0
15	0,04348	0,95652	0
16	0,04348	0,95652	0
17	0,26087	0,73913	0
18	0,17391	0,73913	0,08696
19	0,04348	0,91304	0,04348
20	0	0,95652	0,04348
21	0,04348	0,91304	0,04348
22	0,43478	0,56522	0
23	0,82609	0,17391	0
24	1	0	0

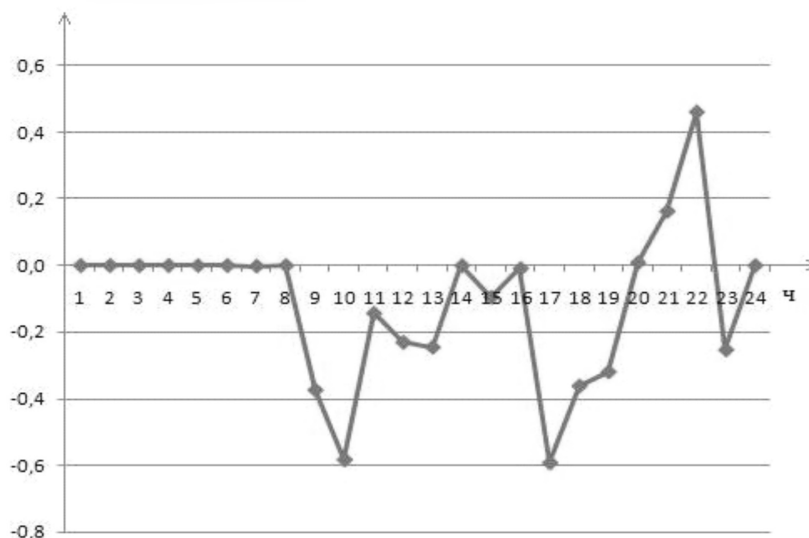


Рис. 6. Почасовые значения коэффициента корреляции между величиной изменения объемов сделок и попаданием спот-цены в тот или иной кластер
(—●—) коэффициенты корреляции

дем обозначать номер кластера, в который попала цена, соответствующая vol_{hd} . Для оценки меры зависимости двух случайных величин будем использовать коэффициент корреляции. Таким образом, выявим, существует ли взаимосвязь между характером изменения почасовых объемов сделок и попаданием спот-цены в один из кластеров. Сгруппируем пары $(\Delta vol_{hd}, k_{hd})$, $d = 1, \dots, N_d$, соответствующие h -му часу. Получим 24 группы, каждая из которых будет содержать N_d точек.

Для каждого часа получим оценки вероятности того, что цена в этом часе попадает в тот или иной кластер (табл. 3) и найдем коэффициент корреляции между Δvol_{hd} и попаданием цены в тот или иной кластер (рис. 6):

Проанализируем данные, представленные в табл. 3. В ночные часы спрос на электроэнергию покрывается самой дешевой за рассматриваемый период электроэнергией. В часы пиковых и полупиковых нагрузок, когда спрос на потребление электроэнергии значительно возрастает, для его удовлетворения привлекаются поставщики, работающие на более дорогом виде сырья. Это нетрудно видеть на отрезках 9–12 и 18–20 часов, когда цены сделок на рассматриваемых участках соответствуют самому дорогому виду электроэнергии.

Значения коэффициента корреляции (0,35–0,4) обнаруживаются в часы, когда изменяется состав поставщиков, покрывающих спрос на электроэнергию на рассматриваемый час.

Изучим, какой объем сделок на конкретный час покрывается поставщиками, цены которых попадают в один кластер:

$$vol_{hd} \leftrightarrow Pr_{hd} \cdot Pr_{hd} \leftrightarrow k_{hd} \Leftrightarrow vol_{hd} \leftrightarrow k_{hd}. \quad (6)$$

Рассмотрим участок временного ряда объемов сделок, где наблюдается монотонный его рост vol_{hd} . Пусть

$$vol_{1d} \leq \dots \leq vol_{hd} \Rightarrow k_{1d} \leq \dots \leq k_{hd} \Rightarrow \bar{k}_{1d} < \dots < \bar{k}_{ld} < \dots < \bar{k}_{md}, \quad m < h, \quad \bar{k}_{1d} \equiv k_{1d}, \quad (7)$$

где $\bar{k}_{ld}$ – уникальные номера кластеров на рассматриваемом участке в. р. объемов сделок. Тогда можем записать

$$vol_{hd} = vol_{k_{1d}} + \sum_{\bar{k}_{jd} = k_{1d}}^{\bar{k}_{md}} \left(\sum_{k_{id} = \bar{k}_{jd}} \Delta vol_{k_{id}} - \Delta vol_{k_{(i-1)d}} \right), \quad \Delta vol_{k_{1d}} = 0. \quad (8)$$

Таким образом, $\sum_{k_{id} = \bar{k}_{jd}} \Delta vol_{k_{id}} - \Delta vol_{k_{(i-1)d}}$ – это вклады поставщиков из $\bar{k}_{jd}$ -кластера в объем сделок на j -й час дня d .

Для примера возьмем один из дней анализируемого периода, в котором присутствуют цены из всех кластеров (рис. 7).

Полученная декомпозиция временного ряда объемов продаж позволила бы провести верификацию результатов кластеризации спотовых цен. Напомним, что для этих временных участков нам были доступны для анализа только два в. р.

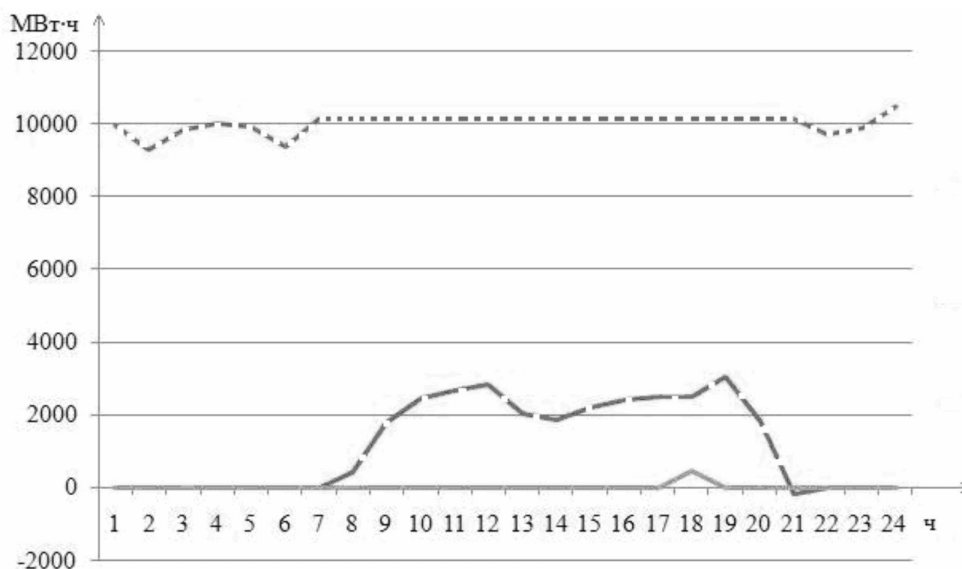


Рис. 7. Декомпозиция временного ряда объемов сделок
(---) первый кластер; (—) второй кластер; (—) третий кластер

Для верификации потребовалось привлечь информацию об объемах заявок производителей, генерирующих электроэнергию на основе энергоносителей при сопоставимой цене на них в расчете на 1 МВт. Однако эта информация недоступна. Начиная с декабря 2009 г. на сайте ЕЕХ публикуют агрегированное предложение на каждый час по каждому из 11 видов сырья (рис. 8).

Эти данные и послужили для дальнейшего развития нашей модели.

Располагая агрегированными объемами предложений, мы можем снять ограничение на квазистационарность агрегированного объема сделок за сутки, тем самым предложенная модель может применяться на любом участке в. р.

Во в. р. спотовых цен за 2010–2011 гг. мы идентифицировали 443 участка с выбросами цены в положительную полуплоскость. Выбросы в область отрицательных значений цены в данной статье не обсуждаем, трактуя их как имеющие

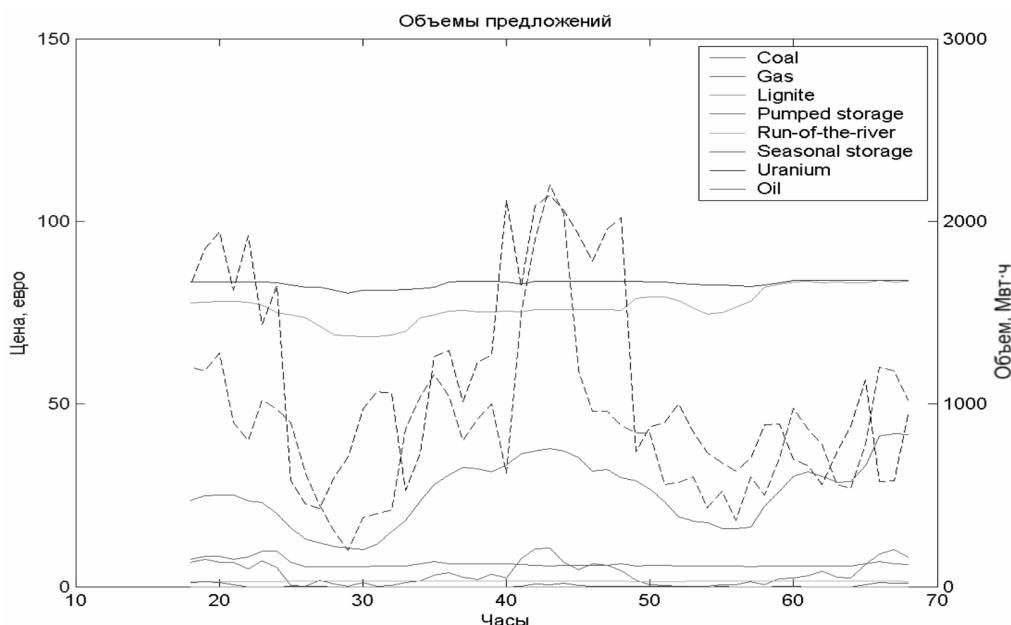


Рис. 8. Временные ряды агрегированных объемов предложений $vol_h^{(l)}$, $l = 1, \dots, K$ по видам сырья и соответствующий участок в. р. спотовых цен (---)

другую природу. Очевидно, что в каждый час h дня d цена отсечения должна попадать в один из K ценовых кластеров $Pr_{hd} \in F_j(pr, \theta)$, $j = 1, \dots, K$ согласно следующему условию:

$$Pr_{hd} \in F_j(pr, \theta) \Leftrightarrow P\left(\sum_{l=1}^{j-1} vol_{hd}^{(l)} < Vol_{hd} \mid \forall l < j \ EPr_c_l < EPr_{c_{l+1}}\right) \cdot P\left(\sum_{l=1}^j vol_{hd}^{(l)} \geq Vol_{hd} \mid \forall l < (K-1) \ EPr_c_l < EPr_{c_{l+1}}\right) = 1$$

или, эквивалентно

$$Pr_{hd} \in F_j(pr, \theta) \Leftrightarrow j = \inf_l \left\{ l : \sum_{\lambda=1}^l vol_{hd}^{(\lambda)} \geq Vol_{hd} \mid \forall \lambda < (K-1) \ EPr_{c_\lambda} < EPr_{c_{\lambda+1}} \right\}, \quad (9)$$

где EPr_c_l – математическое ожидание спотовой цены из l -го ценового кластера.

Теперь мы получим новое множество разбиений (5) для множества пар (1). Каждое разбиение будет содержать 10 подмножеств (мы практически сразу же исключили из рассмотрения агрегированные объемы по кластеру Others в силу равенства нулю подавляющего числа элементов этого в. р.) согласно мощности множества $\{vol_{hd}^{(l)}\}$.

Множество индексов из (1), I может быть кластеризовано согласно такому разбиению следующим образом:

$$I = \{1, \dots, N\} = \bigcup_{k=1}^K I_k : \forall k \in [1, \dots, K-1] I_k \cap I_{k+1} = \emptyset.$$

Можем разделить на кластеры все множество заявок на поставку энергии:

$$\{pr_i^{(h)}, v_i^{(h)}\} = \{pr_{i_1}^{(h)}, v_{i_1}^{(h)}\} \cup \dots \cup \{pr_{i_k}^{(h)}, v_{i_k}^{(h)}\} \cup \dots \cup \{pr_{i_K}^{(h)}, v_{i_K}^{(h)}\}.$$

Рассмотрим подмножество цен заявок, относящихся к первому кластеру. Оно состоит из ненаблюдаемых $Pr_{c_1}^{(h)-}$ и наблюдаемых нами $Pr_{c_1}^{(h)+}$ значений спотовой цены:

$$\begin{aligned} Pr_{c_1}^{(h)} &= \{pr_{i_1}^{(h)}\} = \\ &= \{pr_{i_1}^{(h)-}, pr_{i_2}^{(h)-}, \dots, pr_{i_j}^{(h)+}, pr_{i_{j+1}}^{(h)-}, \dots\} = \\ &= \{pr_{i_1}^{(h)-}\} \cup \{pr_{i_1}^{(h)+}\} = Pr_{c_1}^{(h)-} \cup Pr_{c_1}^{(h)+}. \end{aligned}$$

Аналогично, множества цен, относящихся к оставшимся кластерам:

$$Pr_{c_k}^{(h)} = \{pr_{i_k}^{(h)}\} = Pr_{c_k}^{(h)-} \cup Pr_{c_k}^{(h)+}$$

...

$$Pr_{c_K}^{(h)} = \{pr_{i_K}^{(h)}\} = Pr_{c_K}^{(h)-} \cup Pr_{c_K}^{(h)+}.$$

Каждое наблюдаемое значение цены отсече-

ния (спотовой цены) попадает в один и только в один из ценовых кластеров:

$$\begin{aligned} Pr_{hd} &\in \{Pr_{c_1}^{(hd)+} \cup Pr_{c_2}^{(hd)+} \cup \dots \cup Pr_{c_K}^{(hd)+}\} = \\ &= \bigcup_{l=1}^K Pr_{c_l}^{(hd)+}. \end{aligned}$$

$V_1^{(h)}$ – множество ненаблюдаемых объемов предложений из заявок поставщиков, принадлежащих первому кластеру:

$$V_1^{(h)} = \{v_{i_1}^{(h)-}\} = \{v_{i_1}^{(h)-}, \dots, v_{i_j}^{(h)-}, \dots, v_{i_{|Pr_{c_1}^{(h)}|}}^{(h)-}\}.$$

Для остальных кластеров подобные множества определяются аналогично с точностью до значения индексов и мощности множеств.

Сумма объемов предложений до некоторого индекса j доставляет нам объем сделок на час h дня d .

$$\exists Pr_{i_j}^{(h)+} \in Pr_{c_1}^{(h)+} : \sum_{m=1}^j v_{i_m}^{(h)-} = Vol_{hd} - \text{существует}$$

j -я заявка из первого подмножества упорядоченных заявок (заявок первого ценового кластера), которая определяет цену отсечения на час h дня

d в силу того, что случайный процесс $\sum_{m=1}^j v_{i_m}^{(h)-}$ достигает случайного уровня, порождаемого суммой по некоторому упорядоченному подмножеству заявок потребителей, что и доставляет нам объем сделок Vol_{hd} на час h дня d :

$$V_2^{(h)} = \{v_{i_2}^{(h)-}\} = \{v_{i_2}^{(h)-}, \dots, v_{i_j}^{(h)-}, \dots, v_{i_{|Pr_{c_2}^{(h)}|}}^{(h)-}\}$$

$$\exists Pr_{i_j}^{(h)+} \in Pr_{c_2}^{(h)+} : \sum_{m=1}^j v_{i_m}^{(h)-} = Vol_{hd} - \sum_{k=1}^{2-1} vol_{i_k}^{(h)}$$

...

$$V_K^{(h)} = \{v_{i_K}^{(h)-}\} = \{v_{i_K}^{(h)-}, \dots, v_{i_j}^{(h)-}, \dots, v_{i_{|Pr_{c_K}^{(h)}|}}^{(h)-}\}$$

$$\exists Pr_{i_j}^{(h)+} \in Pr_{c_K}^{(h)+} : \sum_{m=1}^j v_{i_m}^{(h)-} = Vol_{hd} - \sum_{k=1}^{K-1} vol_{i_k}^{(h)}.$$

В силу того, что нарушается аксиома коммутативности сложения в (2) и (3), рассмотрение вопроса адекватности стохастической модели в данном случае должно предваряться выбором некоторого конкретного упорядочивания множества кластеров. Естественный порядок может быть индуцирован отношением порядка первых моментов распределений ценовых кластеров, которые в свою очередь определяются средней себестоимостью энергии.

Для проверки гипотезы о порядке кластеров, индуцированном себестоимостью энергии, зависящей в свою очередь от вида сырья, исполь-

зуемого для ее генерации, построим следующий непараметрический аналог функции правдоподобия:

$$L = \sum_{i=1}^k \sum_{j=1}^{n_i} \ln(f_{Pr_{c_i}}(Pr_{c_{ij}})), \quad (10)$$

где сумма по i – это сумма по всем ценовым кластерам; сумма по j – сумма по наблюдаемым значениям соответствующего ценового кластера; $f_{Pr_{c_i}}(\bullet)$ – непараметрическая оценка плотности.

Если порядок, обеспечивающий максимальное значение логарифма правдоподобия (10) на большей части участков, коррелирует с априорными оценками стоимости исходя из стоимости

сырья, это означает, что наша модель адекватна, а гипотеза об отношении порядка верифицирована путем привлечения независимых данных.

По результатам вычислений согласно (10) максимум L определяет следующий порядок: Garbage, Coal, Oil, Gas, Pumped storage, Lignite, Seasonal storage, Run-of-the-river, Uranium, Wind. Действительно, из данных мировой статистики по стоимости электроэнергии известно, что такие виды сырья, как Coal, Oil и Gas находятся в верхней половине списка (т. е. являются дорогими), а Run-of-the-river, Uranium и Wind являются наиболее дешевыми.

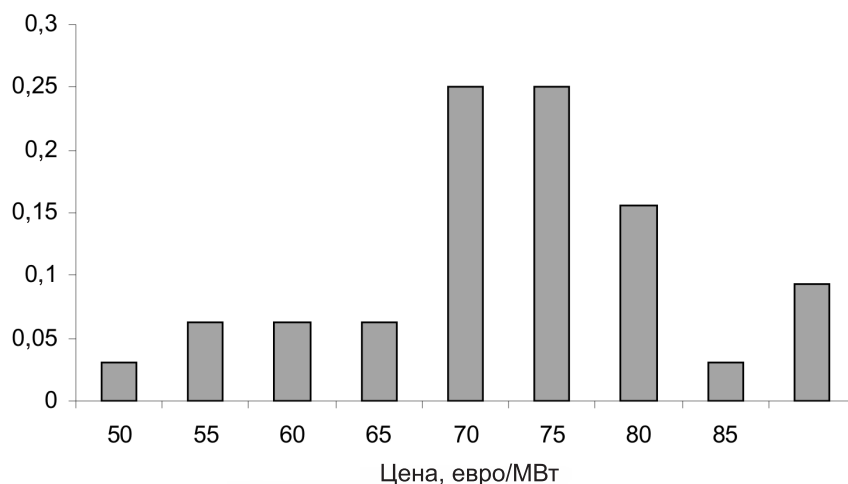


Рис. 9. Непараметрическая оценка плотности распределения для ценового кластера 2 (Coal) по всем 443 участкам с выбросами спотовой цены

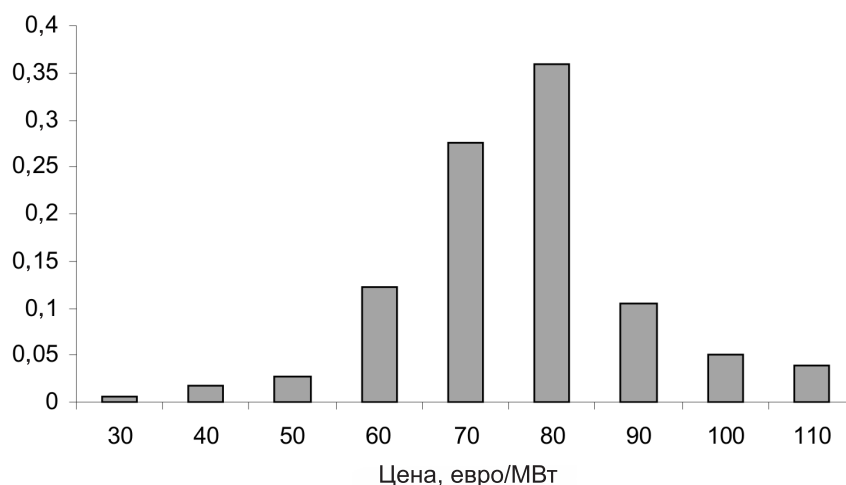


Рис. 10. Непараметрическая оценка плотности распределения для ценового кластера 3 (Oil) по всем 443 участкам с выбросами спотовой цены

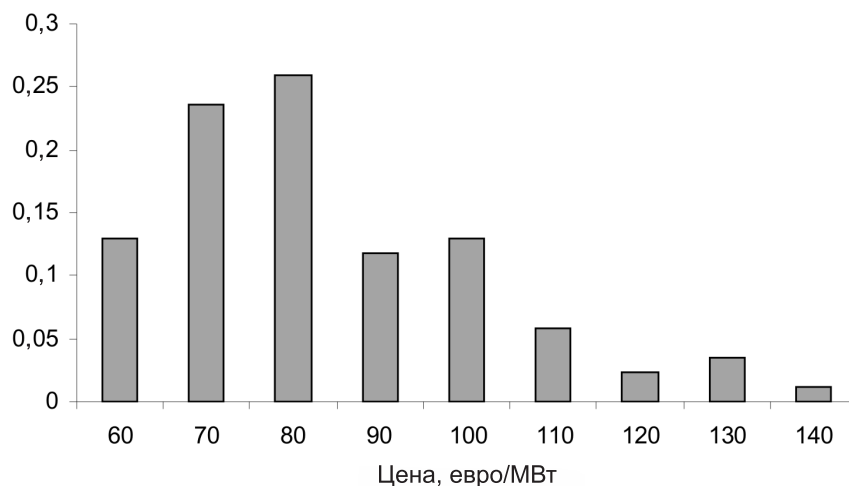


Рис. 11. Непараметрическая оценка плотности распределения для ценового кластера 4 (Gas) по всем 443 участкам с выбросами спотовой цены

На рис. 9–11 представлены непараметрические оценки плотности распределения спотовой цены для трех из упомянутых выше ценовых кластеров (Coal, Oil и Gas).

Полученные результаты дают основание утверждать, что предложенная модель адекватна исходным данным.

С ее помощью удалось получить непараметрическую оценку плотности распределения по подмножествам заявок от поставщиков, генерирующих

энергию с использованием различных видов сырья. Такие оценки позволяют параметризовать аддитивную модель в. р. спотовой цены, обоснованно выбирая вероятностные меры для ее слагаемых.

Полагаем, что наша работа – в определенном смысле обоснование перспективности выбранного направления исследования подобного рода процессов. Чем точнее должен быть прогноз и чем сложнее система, характеристики которой прогнозируются, тем больше информации следует привлекать для построения ее модели.

СПИСОК ЛИТЕРАТУРЫ

1. Европейская биржа электроэнергии (Eurex Energy Exchange) [Электронный ресурс] / Режим доступа: <http://www.eex.com/en/Download>
2. **Иванков, А.А.** Некоторые вопросы стохастического моделирования цен на электроэнергию [Текст] / А.А. Иванков // Матер. науч.-практич. конф. Научные исследования и инновационная деятельность. –СПб.: Изд-во СПбГПУ, 2007. –С. 495–501.
3. **Мельник, А.С.** О возможности построения линейных предикторов для некоторого подмножества выбросов в спотовых ценах на электроэнергию [Текст] / А.С. Мельник, А.А. Иванков // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –2008. –№ 2. –С. 179–183.
4. **Мельник, А.С.** Построение линейных предикторов для подмножества выбросов в спотовых ценах на электроэнергию [Текст] / А.С. Мельник, А.А. Иванков // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –2008. –№ 3. –С. 240–245.
5. **Валынец, Е.В.** Непараметрическая оценка сезонной компоненты в спотовых ценах на электроэнергию

- как решение вариационной задачи [Текст] / Е.В. Валынец, А.А. Иванков // Матер. Междунар. науч.-практич. конф. Высокие интеллектуальные технологии и инновации в национальных исследовательских университетах. –СПб.: Изд-во СПбГПУ, 2012. –Т. 3. –С. 9–11.
6. **Серов, А.Ю.** Фильтрация стационарной составляющей в форме процесса Леви во временных рядах, наблюдаемых на спотовых рынках [Текст] / А.Ю. Серов, А.А. Иванков // Матер. Междунар. науч.-практич. конф. Высокие интеллектуальные технологии и инновации в национальных исследовательских университетах. –СПб.: Изд-во СПбГПУ, 2012. –Т. 2. –С. 64–69.
7. **Королев, В.Ю.** ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор [Текст] / В.Ю. Королев. –М.: ИПИ РАН, 2007.
8. Ядерный рынок. Преимущества атомной энергетики [Электронный ресурс] / Атомстройэкспорт. –Режим доступа: http://www.atomstroyexport.ru/nuclear_market/advantage/
9. **Иванов, Е.Ю.** Об одной стохастической модели спотовых цен на электроэнергию [Текст] /

Е.Ю. Иванов, А.А. Иванков // Сб. науч. тр. СПбГПУ:
Формирование технической политики инновационных

наукоемких технологий. –СПб.: Изд-во СПбГПУ, 2004.
–С. 346–352.

УДК 531.133.3

А.А. Иванов

МЕТОД ПЛАНИРОВАНИЯ ФОРМЫ ГИПЕРИЗБЫТОЧНОГО МАНИПУЛЯТОРА С ОГРАНИЧЕННЫМ ДИАПАЗОНОМ ИЗМЕНЕНИЯ ШАРНИРНЫХ КООРДИНАТ

Задача управления движением кинематически гиперизбыточного манипулятора с последовательной структурой связана с необходимостью определения установок обобщенных шарнирных координат и скоростей по заданному движению конечного звена на временном интервале Δt . Классическая задача управления движением манипулятора с шестью вращательными степенями подвижности основана на решении обратной задачи кинематики и подробно рассмотрена в литературе [1, 2]. Геометрическая обратная задача для пространственного манипулятора сводится к решению системы шести нелинейных уравнений, которая может иметь неединственное решение и точки сингулярности, соответствующие нулям ее якобиана. Решение обратной задачи кинематики (ОЗК) для определения связи задаваемых линейных и угловых скоростей объекта манипуляций с обобщенными скоростями в шарнирах, либо соответственно ускорений, сводится к решению системы линейных уравнений с почти везде невырожденной квадратной матрицей [1]. При наличии ограничений на диапазон изменения шарнирных координат для решения ОЗК даже для классического манипулятора требуется привлечение аппарата квадратичного или нелинейного программирования. Обратная задача кинематики для кинематически избыточного манипулятора в любой из отмеченных постановок сводится к минимизации некоторого критерия качества в неограниченном или ограниченном пространстве переменных состояния.

В настоящей статье предложен метод решения ОЗК с ограничениями на диапазон изменения шарнирных координат, основанный на минимизации квадратичной нормы отклонения линейной и угловой скоростей конечного звена от скоростей, задаваемых пропорционально

рассогласованию векторов положения и ориентации «схвата» и «цели». Метод применен при моделировании кинематики манипулятора из однотипных модулей с шарнирами вращения ограниченной подвижности.

Постановка задачи

Задача определения по заданному шести-компонентному набору U проекций векторов линейной и угловой скорости конечного звена манипулятора N -мерного вектора-столбца $\dot{q}$ обобщенных скоростей в шарнирах гиперизбыточного манипулятора рассмотрена в работах [1, 3] и формализована записью уравнения

$$U = J\dot{q}, \quad (1)$$

где матрица Якоби J размера $(6 \times N)$ имеет столбцовый ранг не более чем 6.

Общее решение линейного уравнения (1) с использованием псевдообратной матрицы J^+ представляется суммой частного решения $\dot{q}_* = J^+U$ неоднородного уравнения (1) и общего решения однородного уравнения, содержащего произвольный вектор X_0 ,

$$\dot{q} = J^+U + (E - J^+J)X_0. \quad (2)$$

Псевдообратная матрица матрицы J – единственна, может быть определена как

$$J^+ = \lim_{\varepsilon \rightarrow 0} (J^T J + \varepsilon E)^{-1} J^T = \lim_{\varepsilon \rightarrow 0} (J^T (J J^T + \varepsilon E)^{-1}) \quad (3)$$

и позволяет получить решение системы линейных уравнений (включая несовместные системы уравнений и уравнения с плохо обусловленной матрицей), имеющее минимальную норму [4, 5].

Определение (3) позволяет записать решение уравнения (1) в одной из форм:

$$\dot{q} = y_1,$$

$$y_1 = \lim_{\varepsilon \rightarrow 0} \arg((J^T J + \varepsilon E)y_\varepsilon = J^T U) \quad (4)$$

или

$$\begin{aligned} \dot{q} &= J^T y_2, \\ y_2 &= \lim_{\varepsilon \rightarrow 0} \arg((J J^T + \varepsilon E)y_\varepsilon = U). \end{aligned} \quad (5)$$

Следует отметить, что уравнение для регулирующего решения y_ε в (4) является уравнением Эйлера для минимума квадратичного функционала

$$I_1(\dot{q}) = \frac{1}{2}(U - J\dot{q})^T(U - J\dot{q}) + \frac{1}{2}\varepsilon \dot{q}^T \dot{q}, \quad (6)$$

а уравнение для y_ε в (5) соответствует функционалу для двойственной (6) задачи нахождения седловой точки функционала:

$$I_2(\dot{q}, y_2) = \frac{1}{2}\dot{q}^T \dot{q} + y_2^T(U - J\dot{q}) - \frac{1}{2}\varepsilon y_2^T y_2. \quad (7)$$

Минимизация функционала (6) в условиях ограниченного диапазона изменения координат шарнирных координат с центром в точке $q_{i,0}$, задаваемого неравенством

$$|q_i - q_{i,0}| \leq q_{i,\max}, \quad (8)$$

требует решения на каждом промежутке времени Δt задачи квадратичного программирования с линейными ограничениями типа неравенств

$$|q_i - q_{i,0} + \dot{q}_i \Delta t| \leq q_{i,\max}. \quad (9)$$

Решение задачи (6), (9) прямыми методами минимизации требует привлечения итерационных процедур метода опорных решений или методов штрафных функций [6], существенно увеличивающих вычислительную сложность задачи.

Ограничение (8) автоматически выполняется для новых переменных θ_i , связанных с q_i зависимостью

$$q_i = q_{i,\max} f(\theta_i - \theta_{i,0}), \quad (10)$$

где $f \in [-1, 1]$ – дважды дифференцируемая монотонная функция, а $\theta_i \in (-\infty, \infty)$. Дифференцируя (10) по независимой переменной t , получаем линейное относительно производных преобразование обобщенных скоростей.

Задача (1) для новой переменной $\dot{\theta}$ имеет вид

$$U = \tilde{J}\dot{\theta}, \quad (11)$$

где $\tilde{J} = J\Lambda$, а $\Lambda = \frac{df}{d\theta_i} \text{diag}\{q_{i,\max}\}$.

Будем отыскивать решение $\dot{\theta}$ задачи (11) как решение задачи минимизации функционала

$$\tilde{I}_1(\dot{\theta}) = \frac{1}{2}(U - \tilde{J}\dot{\theta})^T(U - \tilde{J}\dot{\theta}) + \frac{1}{2}\varepsilon \dot{\theta}^T \dot{\theta}, \quad (12)$$

либо эквивалентной ей задачи (7) для функционала

$$\tilde{I}_2(\dot{\theta}, y_2) = \frac{1}{2}\dot{\theta}^T \dot{\theta} + y_2^T(U - \tilde{J}\dot{\theta}) - \frac{1}{2}\varepsilon y_2^T y_2. \quad (13)$$

Уравнения Эйлера для задачи (12) имеют вид

$$\dot{q} = \Lambda y_1, \quad (14)$$

$$y_1 = \lim_{\varepsilon \rightarrow 0} \arg((\tilde{J}^T \tilde{J} + \varepsilon E)y_1 = \tilde{J}^T U),$$

а для задачи (13)

$$\dot{q} = \tilde{J}^T y_2, \quad (15)$$

$$y_2 = \lim_{\varepsilon \rightarrow 0} \arg((\tilde{J} \tilde{J}^T + \varepsilon E)y_2 = U).$$

Решение уравнений (14) или (15) является решением обратной задачи кинематики для шестикомпонентного вектора U , составленного из проекций векторов линейной и угловой скорости, которые должны быть определены априори.

Квадратичный критерий (12), уравнения экстремали которого при $\varepsilon = 0$ совпадают с уравнениями (1), может быть изменен с целью наложения дополнительных требований на решение неопределенной системы уравнений (1). Модифицированный критерий может иметь вид

$$\begin{aligned} \tilde{I}(\dot{\theta}) &= \frac{1}{2}(U - \tilde{J}\dot{\theta})^T K^2 (U - \tilde{J}\dot{\theta}) + \\ &+ \frac{1}{2}\dot{\theta}^T \Pi^2 \dot{\theta} + \frac{1}{2}\varepsilon (\dot{\theta} - \dot{\theta}_0)^T (\dot{\theta} - \dot{\theta}_0). \end{aligned} \quad (16)$$

Здесь добавлены слагаемое взвешенной суммы квадратов переменной $\dot{\theta}$ с весами, задаваемыми диагональной матрицей Π , диагональная матрица весов K и регуляризирующее слагаемое, которое может рассматриваться как норма отклонения $\dot{\theta}$ от задаваемого набора производных $\dot{\theta}_0$. При выборе в качестве $\dot{\theta}_0$ производных на предыдущем временном интервале это слагаемое имеет смысл квадрата нормы приращения скорости. Критерий (16) может быть записан для приращений обобщенных скоростей $\dot{\theta}$ при изменении задаваемой скорости (угловой скорости). Форма критерия (16) без разделения на отдельные квадратичные слагаемые имеет вид

$$\hat{I}(\dot{\theta}) = \frac{1}{2}(\hat{U} - \hat{J}\dot{\theta})^T(\hat{U} - \hat{J}\dot{\theta}) + \frac{1}{2}\varepsilon \dot{\theta}^T \dot{\theta}, \quad (17)$$

где $\hat{J} = \begin{pmatrix} K\tilde{J} \\ \Pi \end{pmatrix}$, $\hat{U} = \begin{pmatrix} K\tilde{U} \\ 0 \end{pmatrix}$ – блочные прямоугольные матрицы размера $((6 + N) \times N)$.

Таким образом, задача предсказания обобщенных скоростей по изменяемому положению и/или ориентации «схвата» сводится к минимизации критерия вида (17) или определения седловой точки критерия вида (13), т. е. решению уравнений (14) или (15).

Метод и алгоритм решения

В работе [2] в рамках реализации распределенной процедуры нахождения обобщенных скоростей гиперизбыточного манипулятора с шарнирами вращения описан алгоритм итерационного решения уравнения (14). Для используемого метода в [2] приведено доказательство сходимости итерационной процедуры. Анализ описанного в работе [2] алгоритма показывает, что предлагаемый метод является классическим методом Зейделя, применяемым к решению уравнения (4). Сходимость метода Зейделя доказана для случая невырожденной системы линейных уравнений, а скорость сходимости определяется свойствами матрицы системы уравнений и может быть недостаточно быстрой даже при хорошем начальном приближении.

Для решения систем уравнений с симметричными матрицами наиболее устойчивым к ошибкам округления и быстро сходящимся является метод сопряженных градиентов [6] и его модификации. Свойство сходимости метода при точных вычислениях за число циклов, не превосходящее ранг матрицы без учета кратности собственных чисел, позволяет построить надежную процедуру решения сформулированной обратной задачи в форме (14), т. к. ранг обращаемой матрицы в (14) не более шести. Следует отметить, что вычислительная сложность решения системы уравнений в форме (15) при использовании немодифицированного критерия (12) – меньше, чем для уравнения (14), т. к. обращаемая матрица имеет порядок шесть. Однако при использовании параметризованного критерия (17) порядок обращаемой матрицы равен $(6 + N)$ и применение системы уравнений становится вычислительно неэффективным.

В соответствии с (14) решение ОЗК определим как $\lim_{\varepsilon \rightarrow 0} \dot{q}_\varepsilon$ решения системы уравнений

$$\begin{aligned} \dot{q}_\varepsilon &= \Lambda Y_{1\varepsilon}, \\ (\hat{J}^T \hat{J} + \varepsilon E) Y_{1\varepsilon} &= \hat{J}^T U, \end{aligned} \quad (18)$$

либо в соответствии с (15)

$$\begin{aligned} \dot{q}_\varepsilon &= \hat{J}^T Y_{2\varepsilon}, \\ (\hat{J} \hat{J}^T + \varepsilon E) Y_{2\varepsilon} &= \hat{U}. \end{aligned} \quad (19)$$

Известные методы решения некорректных задач в постановке (18) или (19) используют для нахождения регуляризирующего параметра ε , удовлетворяющего дополнительным условиям на норму невязки системы уравнений (18) или (19), трудоемкие алгоритмы оценки, требующие нахождения сингулярных чисел матрицы, либо анализа зависимости нормы невязки от регуляризирующего параметра. В связи с тем, что при движении манипулятора матрица $\hat{J}$ изменяется, а априорная оценка на каждом временном шаге чрезмерно усложняет метод решения, нами использован следующий алгоритм нахождения предельного решения.

Решение $Y_\varepsilon = Y(\varepsilon)$ отыскивается в виде интерполяционного полинома Лагранжа

$$Y(\varepsilon) = \sum_{k=1}^m Y(\varepsilon_k) \Omega_k(\varepsilon), \quad \varepsilon_k > 0, \quad (20)$$

где $Y(\varepsilon_k) \triangleq Y_k$ – решение системы уравнений

$$(\hat{J}^T \hat{J} + \varepsilon_k E) Y_k = \hat{J}^T U, \quad (21)$$

либо

$$(\hat{J} \hat{J}^T + \varepsilon_k E) Y_k = \hat{U}. \quad (22)$$

Предельное решение Y_0 находится методом аналитического продолжения в точку $\varepsilon = 0$

$$Y_0 = \sum_{k=1}^m Y_k \Omega_k(0). \quad (23)$$

Работоспособность описанной процедуры проверена для $m = 2, 3$ на примерах из [4, 5]. Получено совпадение в пределах точности представления результатов в [4, 5].

Кинематическая модель гиперизбыточного манипулятора

Разработанный метод решения использован при моделировании кинематики гиперизбыточного манипулятора, состоящего из звеньев, связанных универсальными двухстепенными ортогональными шарнирами вращения, плоскость пересечения осей которых при нулевых углах перпендикулярна отрезкам, соединяющим центры шарниров.

Функционал (17), порождающий систему уравнений, описывающих кинематику такого манипулятора из $N_{зв} = N_m/2$ одинаковых звеньев,



имеет вид

$$J = \frac{1}{2} \left(\bar{V}_t - \sum_{k=1}^{N_m} \dot{q}_k \bar{S}_k \right)^T \cdot \left(\bar{V}_t - \sum_{j=1}^{N_m} \dot{q}_j \bar{S}_j \right) + \frac{1}{2} p^2 \left(\bar{\omega}_t - \sum_{k=1}^{N_m} \dot{q}_k \bar{l}_k \right)^T \cdot \left(\bar{\omega}_t - \sum_{i=1}^{N_m} \dot{q}_i \bar{l}_i \right) + \frac{1}{2} \sum_{k=1}^{N_m} p_{1k}^2 \dot{q}_k^2 + \frac{1}{2} \varepsilon \sum_{k=1}^{N_m} (\dot{\theta}_k - \dot{\theta}_{k0})^2, \quad (24)$$

где $\bar{V}_t$ и $\bar{\omega}_t$ – скорость и угловая скорость инструмента; $\bar{l}_k$ – орт оси вращения; $\bar{S}_k = \bar{l}_k \times \bar{R}_k$, $\bar{R}_k$ – радиус-вектор положения инструмента относительно шарнира с номером k . В качестве функции $f(\theta)$ в формуле (10) выбрана функция гиперболического тангенса. Весовые коэффициенты p и $\{p_{1k}\}$ позволяют ставить задачу позиционирования «схвата» без ограничения на ориентацию и управлять «инерционностью» системы.

При выполнении движения к цели (g) скорость $\bar{V}_t$ и угловая скорость $\bar{\omega}_t$ концевой точки манипулятора (инструмента (t)) на временном промежутке Δt задаются пропорционально вектору рассогласования положения и ориентации инструмента и цели:

$$U = \alpha(Z_g - Z_t), \quad (25)$$

где $Z_g - Z_t = \{p_x, p_y, p_z, \omega_x, \omega_y, \omega_z\}$, а $\{p_x, p_y, p_z\}$ – проекции вектора линейного рассогласования, $\{\omega_x, \omega_y, \omega_z\}$ – проекции вектора конечного поворота, сопряженного тензору поворота инструмента до совмещения с целью.

Процедура определения обобщенных скоростей $\dot{\theta}_k$ повторяется каждый раз в начале временного промежутка Δt .

При моделировании перемещения концевой звена новые значения шарнирных координат $q_{i,n+1}$ определяются из (6) для $\theta_{i,n+1} = \theta_{i,n} + \dot{\theta}_{i,n+1} \Delta t$.

При использовании решения ОЗК для управления манипулятором в режиме реального времени новые значения шарнирных координат и/или скоростей в начале каждого временного интервала, для которого вычисляются обобщенные скорости, задаются как результат отработки расчетных установок шарнирных координат и/или скоростей на предыдущем временном интервале.

Как было отмечено выше, алгоритм решения ОЗК на основе поиска седловой точки сопряженного (24) функционала вида (13) связан с решением системы линейных уравнений шестого порядка только в случае, когда все p_{1k} и $\dot{\theta}_{k0}$ равны

нулю и сопряженный функционал имеет вид

$$J = y_1^T \left(\bar{V}_t - \sum_{j=1}^{N_m} \dot{q}_j \bar{S}_j \right) + p^2 y_2^T \left(\bar{\omega}_t - \sum_{i=1}^{N_m} \dot{q}_i \bar{l}_i \right) + \frac{1}{2} \sum_{k=1}^{N_m} \dot{\theta}_k^2 - \frac{1}{2} \varepsilon y_1^T \cdot y_1 - \frac{1}{2} \varepsilon y_2^T \cdot y_2. \quad (26)$$

Метод решения ОЗК, основанный на минимизации (25), позволяет накладывать дополнительные условия на форму манипулятора, однако вычислительно более трудоемок. В настоящей статье рассмотрены результаты моделирования кинематики гиперизбыточного манипулятора при решении уравнений вида (19), построенных для функционала (26).

Результаты моделирования

Для кинематической модели манипулятора, задаваемой уравнениями экстремалей функционала (26) и целевого условия (25), с помощью описанного метода поиска псевдорешения ОЗК проведены расчеты по моделированию кинематики его движения при выполнении различных заданий по перемещению концевой звена.

На рис. 1 и 2 приведены результаты расчета движения концевой точки многозвенных манипуляторов с суммарной длиной 1,2 м и различным числом двухстепенных шарниров с ограничением на поворот $\pm \pi/3$ (рис. 1) и $\pm \pi$ (рис. 2) из точки с координатами (0; 1,2; 0) в точку с координатами (0,3; -0,9; 0,3) и заданной ориентацией концевой звена параллельно и противоположно оси ординат.

Из рисунков видно, что предлагаемый метод решения ОЗК одинаково применим как к классической задаче шестистепенного манипулятора, так и к задаче с избыточным числом степеней подвижности. Все манипуляторы без шарнирных ограничений (рис. 2) обеспечивают достижение целевой конфигурации «схвата», при этом манипуляторы с избыточностью при выполнении перехода от прямолинейной конфигурации в целевую точку имеют более компактное расположение шарнирных точек. Требование достижения целевой конфигурации приводит к депланации многозвенника.

Наличие ограничений на диапазон угловых перемещений приводит к запиранию 6-степен-

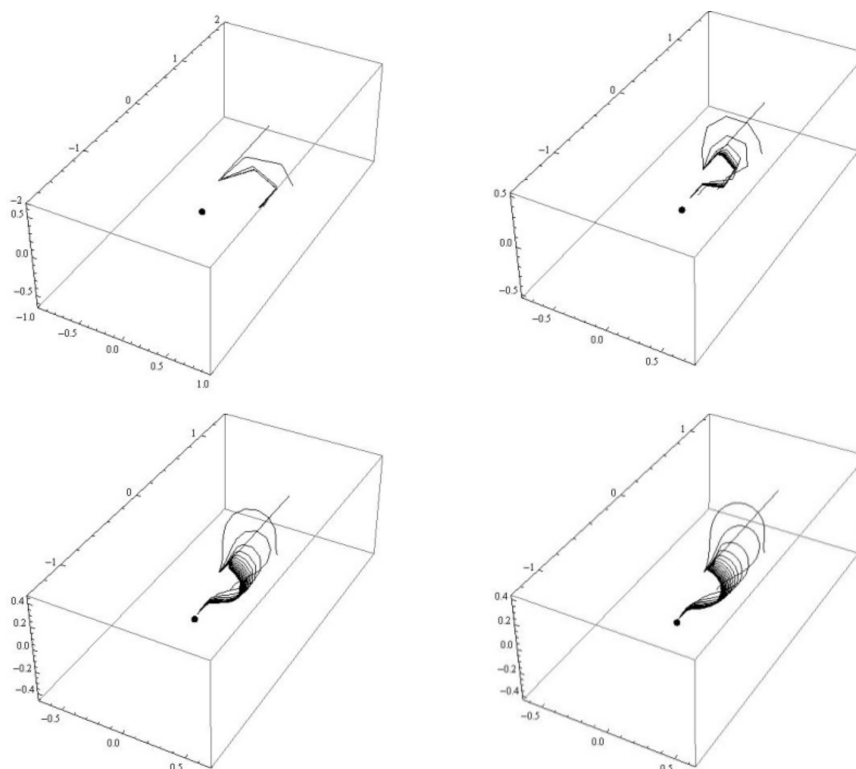


Рис. 1. Последовательные конфигурации манипуляторов с ограничением диапазона вращения в шарнире $[-\pi/3, \pi/3]$
Слева направо 3, 6, 12, 24 звеньев

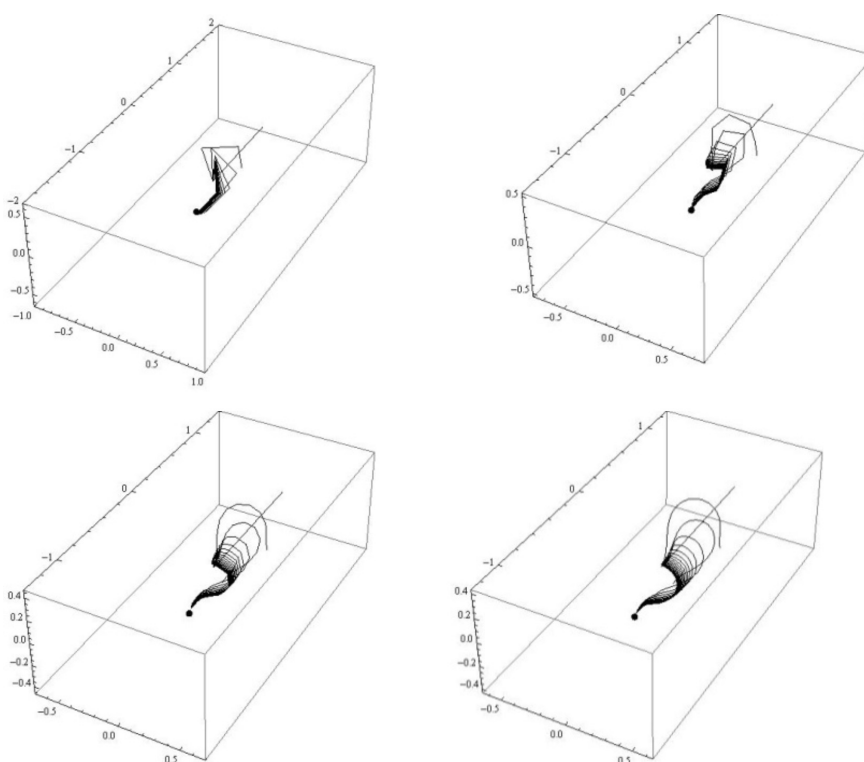


Рис. 2. Последовательные конфигурации манипуляторов с ограничением диапазона вращения в шарнире $[-\pi, \pi]$
Слева направо, сверху вниз 3, 6, 12, 24 звеньев

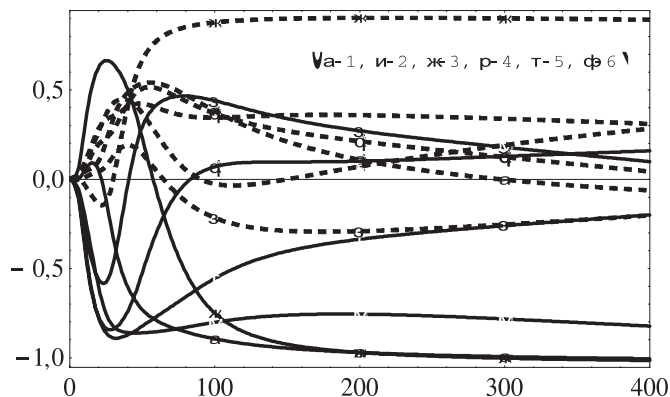


Рис. 3. Зависимости углов поворота смежных звеньев относительно курсовых (—) и тангажных (---) осей
Цифры соответствуют номерам универсальных шарниров

ного манипулятора до достижения целевой точки. Гиперизбыточные манипуляторы реализуют целевую конфигурацию, при этом в некоторых шарнирах 12-степенного манипулятора значения угловых координат выходят на границы допустимого диапазона. На рис. 3 приведены графики зависимостей углов поворота (в радианах) вокруг курсовых и тангажных осей шарниров от номера шага по времени ($\Delta t = 0,04$ с) для шестизвенного манипулятора с двенадцатью степенями подвижности. Из графиков видно, что часть шарниров выходят на ограничение $\pi/3$ и манипулятор редуцируется.

На рис. 4 приведены также результаты моделирования движения концевой точки 12-звенного манипулятора по сложной пространственной кривой, радиус-вектор которой в декартовой системе координат задается формулами

$$\begin{aligned} \overline{R}_g = \frac{L}{8} \left\{ -3 \cos\left(\frac{2\pi t}{\tau_\theta}\right) \sin\left(\frac{2\pi t}{\tau_\theta}\right) \sin\left(\frac{2\pi t}{\tau_\varphi}\right), \right. \\ \left. \left(\cos\left(\frac{2\pi t}{\tau_\theta}\right) - 1 \right), \right. \end{aligned} \quad (27)$$

$$\left. -3 \cos\left(\frac{2\pi t}{\tau_\theta}\right) \sin\left(\frac{2\pi t}{\tau_\theta}\right) \cos\left(\frac{2\pi t}{\tau_\varphi}\right) \right\},$$

где L — длина манипулятора, $\tau_\theta = 7,5$ с и $\tau_\varphi = 12,5$.

Графики на рис. 4 позволяют судить о степени повторяемости конфигурации манипулятора при выполнении циклического задания. Из рис. 5 видно, что алгоритм слежения за движущейся точкой, построенной на формуле (25), приводит к отклонению от заданной кривой (тонкая линия на рисунке) на отдельных участках траектории.

Предложенный метод решения ОЗК позволяет моделировать кинематику движения как классического, так и гиперизбыточного манипулятора с ограниченной подвижностью. Получаемое псевдорешение ОЗК дает возможность непротиворечиво описать режимы запираания, возникающие при достижении границ изменения шарнирных координат. Алгоритм эффективен как для избыточных, так и для редуцируемых в процессе моделирования манипуляторов.

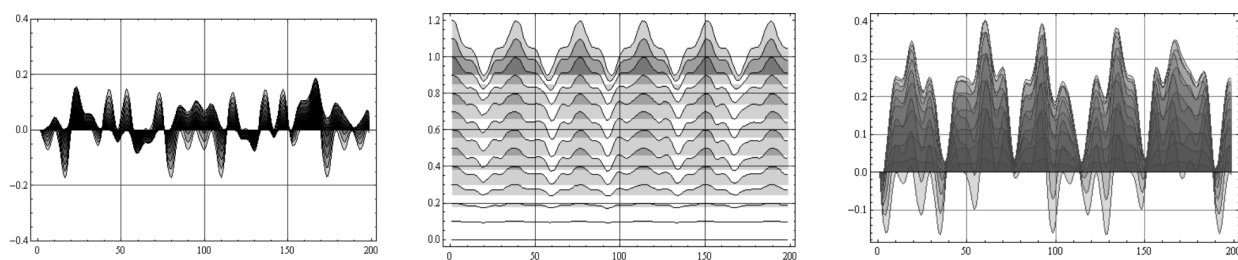


Рис. 4. Зависимости декартовых координат шарниров от номера шага

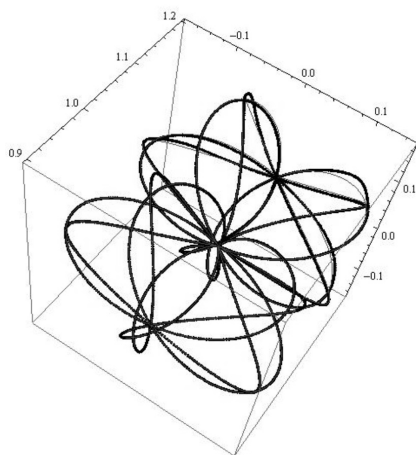


Рис. 5. Траектория движения концевой точки
и заданная траектория

Оценка быстродействия алгоритма позволяет утверждать о возможности его применения не только для моделирования кинематики, но и в режиме реального времени для задания установок углов и скоростей для управления манипулятором по заданной конфигурации схвата.

Структура алгоритма допускает распределение вычислений между вычислителями мехатронных модулей шарниров и использование его в составе встроенного программного обеспечения, позволяющего управлять многозвеновым манипулятором из унифицированных модулей.

Статья подготовлена в рамках работ по выполняемому в соответствии с постановлением Правительства РФ № 218 комплексному проекту по договору № 13 G25310026.

СПИСОК ЛИТЕРАТУРЫ

1. Springer Handbook of Robotics [Text] / Siciliano Bruno, Khatib Oussama (Eds.). –Springer, 2008. –LX. –1611 p.
2. **Фу, К.** Робототехника [Текст] / К. Фу, Р. Гонсалес, К. Ли; Пер. с англ. –М.: Мир, 1989. –624 с.
3. **Turetta, A.** Distributed Control Architecture for Self-reconfigurable Manipulators [Text] / A. Turetta, G. Casalino, A. Sorbara //International J. of Robotics Research. –2008. –Vol. 27. –Iss. 3–4. –P.481–504.
4. **Hansen, P.C.** Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems [Text] / P.C. Hansen // Numerical Algorithms. –1994. –№ 6. –P. 1–35.
5. **Кириянов, Д.В.** Вычислительная физика [Текст] / Д.В. Кириянов, Е.Н. Кириянова. –Полибук Мультимедиа, 2006. –352 с.
6. **Гилл, Ф.** Практическая оптимизация [Текст] / Ф. Гилл, У. Мюррей, М. Райт; Пер. с англ. –М.: Мир, 1985. –509 с.

УДК 004.043:004.855.5:004.852

Н.О. Кадырова, Л.В. Павлова

ЭФФЕКТИВНАЯ МЕТОДИКА ОБРАБОТКИ МНОГОРАЗМЕРНЫХ ДАННЫХ БОЛЬШОГО ОБЪЕМА

Необходимость использования современных эффективных алгоритмов обработки данных обусловлена быстрым ростом объемов корпоративных хранилищ данных, размерность которых может быть сколь угодно большой.

Методы, способные находить скрытые нетривиальные закономерности в больших объемах данных большой размерности, относятся к технологии, названной *интеллектуальным анализом данных* (ИАД, англ. data mining) [1]. Различные сочетания методов ИАД позволяют решать многие реальные бизнес-задачи.

Выделяют следующие стандартные типы закономерностей, выявляемые методами интеллектуального анализа данных: ассоциацию; последовательность; классификацию; кластеризацию и прогнозирование [1].

Выведенные из статистических данных закономерности и правила используют для описания существующих отношений, принятия решений и прогнозирования их последствий.

Классические методы восстановления зависимостей по эмпирическим данным можно применять лишь в случае небольшой размерности дан-



ных. Точность оценивания истинной зависимости катастрофически падает с ростом размерности входных векторов – признаков. Это явление называют «проклятием размерности».

Для выявления и исследования указанных типов закономерностей целесообразно использовать подходы, практически не требующие дополнительной априорной информации (например, задание распределения шума или функционального вида искомого решающего правила) и преодолевающие «проклятие размерности». Перспективным современным направлением в области извлечения знаний из хранилищ данных представляется статистическое обучение на основе ядерных машин.

Кроме того, ядерные методы устойчивы к выбросам и шуму. В настоящей статье рассмотрено применение этого подхода к закономерностям типа классификация и кластеризация, а также сопутствующие статистическому анализу задачи, касающиеся очистки и подготовки данных. Предложена эффективная методика обработки данных большого объема и большой размерности, приведены результаты ее применения к реальным данным.

Предварительная обработка данных

Рассмотрим современные методы многомерного статистического анализа, обеспечивающие возможность работы с содержащими пропуски зашумленными данными большого объема (тысячи) и большой размерности (сотни) с различными типами признаков.

Очистка данных. Корректность результатов анализа и точность построенных моделей во многом зависит от того, насколько эффективно проведена очистка данных. К первому этапу работы с данными относят обработку аномальных и фиктивных значений (выбросов), пропусков, дубликатов и противоречий, шумов и т. д. Требования к качеству данных могут быть разными для различных методов и алгоритмов анализа данных.

Заполнение пропусков. Простой способ восстановления пропусков, включенный в большинство статистических пакетов, – метод замены пропущенного значения общим средним из присутствующих значений – оказывается несостоятельным даже на простых рядах данных. Эффективным развитием этого метода в случае одиночных пропусков на гладких рядах данных является замена пропущенного значения средним из ближайших

присутствующих значений признака.

В работе [2] проведен сравнительный анализ наиболее известных алгоритмов восстановления пропущенных значений в массивах (рядах, таблицах) данных. При восстановлении групповых пропусков авторы рекомендуют использовать Zet-алгоритм, предложенный в [3]. Для разнотипных переменных рекомендуется применять алгоритм WANGA [3].

Выявление выбросов. Для больших выборок и концентрированных загрязнений особенно эффективна процедура выявления многомерных выбросов, предложенная в работе [4], – метод эксцесса.

Подготовка данных. *Нормирование, масштабирование и стандартизация.* Перед началом анализа следует провести нормирование, масштабирование или стандартизацию каждого признака. При этом существенно сужается диапазон значений признаков. Основным преимуществом указанных процедур подготовки данных является повышение численной устойчивости алгоритмов, реализующих методы статистического анализа.

Заметим, что данные для тестирования (например, построенного классификатора) должны быть подготовлены так же, как и для тренировочной последовательности, по которой проводится обучение (например, значения параметров сдвига и масштаба должны быть одинаковыми для тестовой и тренировочной последовательностей).

Работа с разнотипными признаками. При использовании методов, предусматривающих вычисление расстояний между объектами (при заполнении пропусков, при кластеризации и т. д.), в случае признаков различного типа необходимо применять корректные метрики. Метрики для разнотипных данных предложены и исследованы, например, в работах [5, 6]. Заметим, что важен не только выбор корректной метрики, но и соответствующее этой метрике масштабирование данных.

Если в исходных данных присутствует качественный признак, принимающий m возможных значений, целесообразно кодировать его значения следующим образом:

если m не велико, то каждому значению признака ставится в соответствие m -разрядное двоичное число, содержащее только одну единицу;

если m велико, то каждому качественному значению ставится в соответствие одно числовое значение.

Отбор признаков – это процесс определения самого эффективного (для кластеризации) или информативного (относительно данного отклика) подмножества из множества исходных признаков.

В работе [7] предложен метод выявления бесполезных относительно данного отклика (целевой переменной) признаков. Эта процедура рекомендуется в качестве первого этапа работы с признаками как средство очищения от шума. На втором этапе процесса отбора существенных признаков для классификации целесообразно использовать методы, предложенные в [8, 9].

Выделение признаков – преобразование исходных признаков для построения новых существенных признаков меньшей размерности. Один из самых известных методов выделения признаков – метод главных компонент (Principal Component Analysis – PCA) – не устойчив к выбросам и шуму.

В процессе решения последовательности оптимизационных задач, к которым приводит PCA, исходные объекты появляются только в составе скалярных произведений, что позволяет воспользоваться ядерным приемом. Ядерный PCA можно проводить с любым ядром, предварительно центрируя ядерную матрицу определенным образом [10]. Метод устойчив к выбросам и шуму.

Кластеризация. Типичный процесс кластеризации состоит из следующих шагов [11]:

1) *представление образов (объектов)*, включающее информацию о количестве классов, количестве имеющихся образцов, количестве, типе и масштабе имеющихся признаков, и не обязательно включающее *отбор признаков* и/или их *выделение*;

2) определение *меры близости* между образами, соответствующей данным;

3) *кластеризация* или группировка (результат кластеризации может быть четким или размытым);

4) *абстрагирование данных* (обычно это компактное описание каждого кластера в терминах прототипов или репрезентативных образов, например, таких, как центростид);

5) *оценка результата* (результатом любого алгоритма кластеризации является набор кластеров (не зависимо от того, содержат ли анализируемые данные кластеры или нет). Если данные действительно содержат кластеры, то одни алго-

ритмы могут производить кластеризацию «лучше», чем другие алгоритмы. Анализ результатов процедуры кластеризации обычно основывается на определенных критериях оптимальности: *внешнем* (сравнение полученной структуры с некоторой априорной); *внутреннем* (выяснение, соответствует ли полученная структура имеющимся данным по существу); *относительном* (сравнение двух структур и определение их достоинств друг относительно друга)).

Разные методы кластеризации могут создавать разные кластеры. Создание схожих кластеров различными методами указывает на корректность кластеризации. Заметим, что данные, которые не содержат кластеров, не следует анализировать средствами кластеризации.

Процесс кластеризации может включать множество экспериментов по выбору разнообразных параметров, например: меры расстояния; типа стандартизации переменных; количества кластеров и т. д.

Алгоритмы кластеризации, разделяющие данные на основе некоторого критерия, например, известный алгоритм K-means, можно применять для больших объемов выборок. Особенности алгоритма K-means, не всегда удобными для исследователя, являются: необходимость задания числа кластеров и начального разбиения (от которого существенно зависит окончательное разбиение на кластеры), а также определенная (гиперсферическая) форма строящихся кластеров. На практике алгоритм обычно выполняется несколько раз с различными начальными установками, и лучшая из полученных конфигураций используется как результат кластеризации.

Обобщенный метод K-means с ядром вместо метрики был бы недопустимо затратным в вычислительном отношении. Эффективной робастной процедурой является ядерная PCA-кластеризация: сначала применяют ядерный PCA к исходным данным, а затем к преобразованным данным применяют метод кластеризации K-means.

В работе [12] предложен метод кластеризации с использованием опорных векторов (SV-кластеризация). В отличие от метода K-means, SV-кластеризация не определяет кластеры посредством ряда прототипов и не требует, чтобы количество кластеров было определено заранее, а также строит кластеры произвольной формы. SV-кластеризация является одним из немногих методов, позволяющих работать с выбросами.



Классификация. Современной эффективной методологией машинного обучения с учителем является *машина опорных векторов* (Support Vector Machine – SVM) [13].

SVM-алгоритмы обладают многими привлекательными свойствами и способны обрабатывать огромные объемы информации. В основе SVM-методов лежит индукционный принцип структурной минимизации риска, сущность которого состоит в минимизации верхней границы ожидаемого риска, что обеспечивает согласованность между качеством обучения и сложностью класса аппроксимирующих функций. Принцип структурной минимизации риска оснащает SV-машины способностью к обобщению, которая и является целью статистического обучения.

Для гильбертовых пространств с репродуктивным ядром существует высокоэффективный прием вычисления скалярных произведений, который состоит в использовании ядерных функций. Функция (машина), определяющая решающее правило классификации, всегда имеет вид
$$f(x) = \sum_{i=1}^l \alpha_i k(x_i, x) + b,$$
 где $k(x, x')$ – ядерная функция, соответствующая скалярному произведению в гильбертовом пространстве (пространстве признаков), определяемом ядром k ; x_i – входные векторы (объекты), $i = 1, \dots, l$; b – вещественный параметр сдвига. Только опорным векторам соответствуют отличные от нуля значения весовых коэффициентов α_i (свойство разреженности SV-машин).

Данное множество решающих правил классификации представляет собой богатый класс нелинейных (в зависимости от вида ядра) функций. При этом они линейны по параметрам. Количество свободных параметров равно количеству опорных векторов и не зависит от размерности входных векторов. Таким образом преодолевается «проклятие размерности», и метод опорных векторов можно применять (например, для задачи классификации) при любом числе признаков. К тому же, благодаря свойству разреженности, присущему SV-машинам, число опорных векторов значительно меньше объема тренировочной последовательности (l).

Для построения SV-классификаторов разработан ряд эффективных алгоритмов: SVMLight [14], Sequential Minimal Optimization (SMO) [15], Successive overrelaxation (SOR) [16], Incremental updating (IU) [17].

При построении SV-машин целесообразно соблюдать следующие рекомендации:

- производить масштабирование данных;
- применять гауссовское ядро;
- проводить подбор параметров регуляризации (C) и ядра (σ) сначала на грубой решетке, затем уточнять значения параметров на участке, соответствующем лучшему значению оценки обобщенной точности. Для получения представления об обобщенной точности алгоритмов проводить процедуру k -кратной перекрестной проверки (k -fold cross validation).

Методика обработки данных большого объема и большой размерности

На основании современных эффективных подходов предложена следующая методика обработки данных при решении задачи бинарной классификации.

1. Исключение столбцов, в которых доля пропущенных данных выше некоторого порогового значения.
2. Заполнение одиночных пропусков методом средних из ближайших k присутствующих значений, затем заполнение групповых пропусков с применением Zet-алгоритма.
3. Кодирование качественных признаков.
4. Выявление и удаление незначимых признаков.
5. Выявление и удаление выбросов.
6. Стандартизация данных и выделение главных (новых) признаков.
7. Кластеризация данных.
8. Масштабирование данных и настройка SV-машины для бинарной классификации (подбор параметров).
9. Построение SV-классификатора.
10. Тестирование построенной SV-машины.

Результаты применения предложенной методики

В соответствии с предложенной методикой обработаны реальные web-данные, содержащие информацию о поведении пользователей сети: характеристики визитов пользователей, собранные для категорий сайтов и сети в целом. Сеть объединяет набор сайтов, доступных для фиксации поведения посетителей, категория включает группу сайтов общей тематики. Была поставлена следующая задача: по тренировочной последовательности, содержащей информацию о пове-

Таблица 1

Результаты настройки и тестирования SV-машин

Методика	Количество признаков	Параметры SV-машины, C и σ	CV_true	Test_true
1. Масштабирование данных и настройка SV-машины на основе SMO-алгоритма	20	13,5, 1	0,6640	0,6660
2. Стандартизация данных, PCA, масштабирование данных и настройка SV-машины на основе SMO-алгоритма	18	20, 0,95	0,6540	0,6700
3. Стандартизация данных, ядерный PCA, масштабирование данных и настройка SV-машины на основе SMO-алгоритма	20	10, 1,23	0,6530	0,6550

дении посетителей (истории визитов) и соответствующие гендерному признаку метки, построить решающее правило, позволяющее распознать гендерный признак нового посетителя. Исследования проводились в среде MatLab. Максимальный объем использованной в экспериментах выборки – 5000, количество признаков – 266.

Исходные данные перед классификацией предварительно подготовили согласно пунктам 1–4 приведенной методики:

на этапе 1 количество признаков сократилось до 39 при пороговом значении 0,22;

на этапе 2 пропуски были заполнены средним из ближайших $k = 3$ присутствующих значений;

на этапе 3 кодирование (трех) качественных признаков увеличило общее количество признаков до 53;

на этапе 4 после фильтрации осталось 20 признаков.

Подготовленные данные разбили на две последовательности: тренировочную (объемом $l = 4000$) и тестовую (объемом $m = 1000$). По тренировочной последовательности, согласно каж-

дой из трех выбранных методик, с использованием процедуры 10-кратной перекрестной проверки настроили SV-машины для бинарной классификации, построили SV-классификаторы и провели их тестирование. Результаты расчетов приведены в табл. 1.

Отметим, что доля правильно классифицированных образцов, полученная с использованием процедуры 10-кратной перекрестной проверки (CV_true), и доля правильно классифицированных образцов в тестовой последовательности (Test_true) при переходе к новым признакам не увеличились.

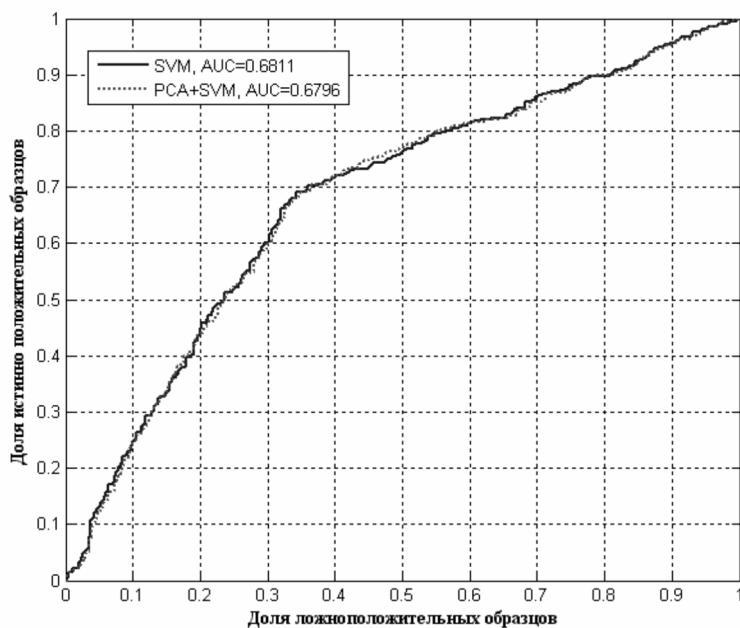
Результаты исследования влияния объема выборки на качество классификации приведены в табл. 2. Отметим, что увеличение объема тренировочной последовательности (ТП) свыше 4000 не улучшает качества SV-машины.

В процессе работы с данными выяснилось, что удаление выбросов, обнаруженных на этапе 5, не влияет на качество классификации; предварительная кластеризация данных определено не улучшает качество классификации.

Таблица 2

Влияние объема выборки на качество классификации

Объем ТП, l	Параметры SV-машины, C и σ	CV_true
2000	1, 1	0,5245
3000	0,1, 1,5	0,5723
4000	13,5, 1	0,6640
5000	12,5, 1	0,6720



ROC-кривые

На рисунке представлены ROC-кривые [18], характеризующие классификаторы, построенные по первым двум методикам из табл. 1. Площади под ROC-кривыми (AUC) практически совпадают, т. е. классификаторы обладают одинаковым качеством.

Анализ результатов проведенных экспериментов демонстрирует устойчивость SV-методов к выбросам; позволяет предположить, что увеличение количества тренировочных образцов свыше 4000 не гарантирует построения лучшего классификатора для данных той же природы и структуры; позволяет выдвинуть следующие предположения относительно предложенных к исследованию web-данных:

- данные слабоструктурированы;
- данные не содержат кластеров, поэтому

анализировать их средствами кластеризации бессмысленно;

- применение PCA или ядерного PCA не улучшает качество классификации.

Отметим, что попытка классификации этих же данных классическими методами (такими, как дискриминантный анализ) не дала полезного результата: доля правильно классифицированных образцов варьировалась в пределах 0,52–0,54. Очевидно, что природа исследуемых данных не позволяет использовать такие методы для выявления скрытых закономерностей.

Применение предложенной методики позволило улучшить качество классификации в среднем на 15 %.

СПИСОК ЛИТЕРАТУРЫ

1. **Дюк, В.** Data mining: Учебный курс [Текст] / В. Дюк, А. Самойленко. –СПб.: Питер, 2001. –368 с.
2. **Абраменкова, И.В.** Методы восстановления пропусков в массивах данных [Электронный ресурс] / И.В. Абраменкова, В.В. Крутлов // Программные продукты и системы. –2005. –№ 2. –Режим доступа: <http://www.swsys.ru/index.php?page=article&id=528>
3. **Загоруйко, Н.Г.** Прикладные методы анализа данных и знаний [Текст] / Н.Г. Загоруйко. –Новосибирск: Изд-во ин-та математики, 1999.
4. **Пена, D.** Multivariate Outlier Detection and Robust Covariance Matrix Estimation [Text] / D. Pena, F.J. Prieto // Technometrics. –2001. –Vol. 43. –№ 3. –P. 286–310.
5. **Wilson, D.R.** Improved Heterogeneous Distance Functions [Text] / D.R. Wilson, T.R. Martinez // J. of Artificial Intelligence Research. –1997. –Vol. 6. –P. 1–34.
6. **Spencer, M.S.** Heterogeneous Distance Measures and Nearest-Neighbor Classification in an Ecological Setting [Text] / M.S. Spencer [et al.] // Missouri J. of Mathematical Sciences. –2010. –Vol. 22. –№ 2. –P. 108–123.
7. **Weiss, S.M.** Predictive data mining: a practical guide [Text] / S.M. Weiss, N. Indurkha. –NY: Morgan-Kaufman, 1998.
8. **Weston, J.** Feature selection for SVM [Text] / J. Weston, S. Mukherjee, O. Chapelle [et al.] // In Neural

Information Processing Systems. –Cambridge, MA: MIT Press, 2001.

9. **Lal, T.N.** Embedded Methods [Электронный ресурс] / T.N. Lal [et al.]. –2005. –P. 1–20. –Режим доступа: <http://edoc.mpg.de/312024>

10. **Scholkopf, B.** Nonlinear Component Analysis as a Kernel Eigenvalue Problem [Text] / B. Scholkopf, A. Smola, K. Mller // Neural Computation. –1998. –Vol. 10. –P. 1299–1319.

11. **Jain, A.K.** Data Clustering: a Review [Text] / A.K. Jain, M.N. Murty, P.J. Flynn // ACM Computing Surveys. –1999. –Vol. 31. –№ 3. –P. 264–323.

12. **Ben-Hur, A.** Support Vector Clustering [Text] / A. Ben-Hur [et al.] // J. of Machine Learning Research. –2001. –Vol. 2. –P. 125–137.

13. **Vapnik, V.N.** The Nature of Statistical Learning Theory [Text] / V.N. Vapnik. –Springer-Verlag, 2000.

14. **Joachims, T.** Making Large-Scale SVM Learning

Practical [Text] / T. Joachims // Advanced Kernel Methods – Support Vector Learning, MIT Press, Cambridge, USA, 1998. –P.41–56.

15. **Platt, J.** Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines [Электронный ресурс] / J. Platt // Microsoft Research Technical Report MSR-TR-98-14, Apr. 21, 1998.

16. **Mangasarian, O.** Successive Overrelaxation for Support Vector Machines [Text] / O. Mangasarian, D. Musicant // IEEE Transactions on Neural Networks. –1999. –Vol. 10. –№ 5. –P. 1032–1037.

17. **Cauwenberghs, G.** Incremental and Decremental SVM Learning [Text] / G. Cauwenberghs, T. Poggio // Advances in Neural Information Processing Systems. –2001. –Vol. 13. –P. 409–415.

18. **Fawcett, T.** ROC Graphs: Notes and Practical Considerations for Researchers // [Text] / T. Fawcett. –Kluwer Academic Publishers, 2004.

УДК 681.3.069

И.Ю. Петрова, Т.В. Хоменко

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВЫБОРА И МЕТОД ОЦЕНКИ ВАРИАНТОВ ФИЗИЧЕСКОГО ПРИНЦИПА ДЕЙСТВИЯ ЧУВСТВИТЕЛЬНОГО ЭЛЕМЕНТА

Существующая тенденция превышения потребности в датчиковой аппаратуре над их производством, обусловленная необходимостью совершенствования (частичного обновления и/или регулярной замены) при определенном по времени жизненном цикле изделий, вызвала необходимость широкого использования автоматизированных систем поискового конструирования [1]. Автоматизация ранжирования вариантов физического принципа действия (ФПД) чувствительных элементов (ЧЭ) на начальных этапах разработки датчиковой аппаратуры позволяет сократить трудоемкость и время создания новых изделий [2].

Основной недостаток данных систем – использование классической постановки задачи выбора наиболее эффективных ФПД ЧЭ, которая в условиях даже частичного отсутствия априорных сведений об объекте проектирования неприменима [3]. Следовательно, для расширения потенциала автоматизированных систем поискового конструирования необходима разработка и ре-

ализация задачи выбора ФПД ЧЭ на этапе концептуального проектирования в условиях различных семантико-синтаксических данных, которая не может считаться решенной при отсутствии механизма оценки решения.

Используя аппарат теории множеств, представим задачу выбора коротежем:

$$ЗВ_{ФПД} := \langle ПЗ; T; R; S; F; G; P; V \rangle,$$

где ПЗ – вариант постановки задачи; T – информация о множестве альтернатив; R – информация о компонентах критерия оценивания; S – множество шкал; F – отображение множества альтернатив в множество их оценок по критериям; G – система предпочтений элемента; P – решающее правило; V – функция выбора.

В рамках различных формализаций выбора [4] информация о ФПД ЧЭ может быть охарактеризована коротежем $T = \langle T_{1i}, T_{2j} \rangle$, где T_{1i} – тип исходов ($i = 1$ – математическое описание, $i = 2$ – вербальное описание исходов); T_{2j} – тип компонент исходов ($j = 1$ – количественные,

$j = 2$ – качественные типы компонент исходов). Информация о компонентах критериев оценивания ФПД ЧЭ может быть охарактеризована кортежем: $R = \langle R_{1l}, R_{2j} \rangle$, где R_{1l} – тип критерия ($l = 1$ – независимый в совокупности, $l = 2$ – зависимый в совокупности); R_{2j} – тип компонент критерия ($j = 1$ – количественные, $j = 2$ – качественные типы компонент критерия).

Тогда математическая модель задачи выбора ФПД ЧЭ описывается кортежем $St_A = \langle \mathfrak{Z}, T, R_{OrgK}, A \rangle$, где при неоднородной входной информации о T – вариантах ФПД ЧЭ, о R_{OrgK} – критерии оценивания и составляющих его компонентах ($R; S; F \subset R_{OrgK}$) определяется $\mathfrak{Z} = \{\lambda_i\}$ – множество процедур ($G; P; V \subset \mathfrak{Z}$), позволяющих рассматривать выходную информацию об A – множестве возможных альтернатив ФПД ЧЭ ($A \subset T$).

В классическом варианте задача выбора с математически описанными исходами (T_{11}) количественными типами компонент исходов (T_{21}) при независимых в совокупности (R_{11}) количественных типах (R_{21}) компонент критерия оценивания решается классическими методами P , в частности, классическими методами ранжирования и оптимизации.

При различных сочетаниях неоднородной входной информации о T , R_{OrgK} , предлагается следующая модель задачи выбора:

$$ПЗ \equiv [\forall T_{1l}; T_{2j} \subset T; \forall R_{1l}; R_{2j} \subset R_{OrgK}$$

$$\exists \mathfrak{Z}; \exists A \quad \exists \sigma_i \subseteq \Xi := \bigvee_{m=1}^8 \langle T_{1l}; T_{2j}; R_{1l}; R_{2j} \rangle :$$

$$St_A = \mathfrak{Z}(T, R_{OrgK}, A) \rightarrow \max],$$

где для поиска решения при качественном и/или количественном варианте ПЗ определяется сценарий выбора $\sigma_i^0 := St_A$ ($i = 1, n$) такой, что отклонение оценки выбранного ФПД ЧЭ от заданных ЛПР значений минимально. Причем множество сценариев $\Xi = \{\sigma_i\}$ ($\sigma_k^0 \in \sigma_i \vee \sigma_i \subseteq \Xi$) считается заданным, если вычисляется максимальная оценка $W(\sigma_i^0)$, приводящего к цели, сценария $\sigma_i \equiv \sigma_i^0$.

Для оценки вариантов ФПД ЧЭ, отобранных в результате применения сценария $\sigma_i^0 := St_A$ на этапе концептуального проектирования, предложен новый комплексный информационный метод (КИМ), базирующийся на основных понятиях теории информации и заключающийся в определении количества информации I в управляемых

состояниях $z \in Z$ (обработка экспертной информации) для достижения максимальной оценки $W(A)$.

КИМ представляет собой вычислительную процедуру, алгоритм которой включает в себя следующую последовательность действий:

1) формируется множество операндов Δ как множество локальных целей и оценочные состояния $\tilde{z}$ как множество передаваемой информации о научно-технической результативности (НТР) варианта;

2) вычисляется минимальное $\underline{W} := I_{\min}$ и максимальное $\overline{W} := I_{\max}$ количество информации о НТР

$$\underline{W} := I_{\min \text{ НТР}} = \sum_{q=1}^w I_{\min \text{ КЗ}}^q \cdot I_{\min \text{ КД}}^q,$$

$$\overline{W} := I_{\max \text{ НТР}} = \sum_{q=1}^w I_{\max \text{ КЗ}}^q \cdot I_{\max \text{ КД}}^q,$$

где $I_{\min \text{ КЗ}}^q$, $I_{\min \text{ КД}}^q$ ($I_{\max \text{ КЗ}}^q$, $I_{\max \text{ КД}}^q$) – минимальное (максимальное) количество информации КЗ – коэффициента значимости и КД – коэффициента достигнутого уровня q -го фактора, w – количество факторов НТР [5];

3) формируется блочная матрица частот, характеризующих наступление/ненаступление события S ;

4) находится суммарное количество информации $I_{qp} = I_{\text{КЗ}} \cdot I_{\text{КД}}$ по всем q факторам (строкам блочной матрицы частот) p -го операнда Δ (столбца блочной матрицы частот) для каждой итерации ($p = 1, r$)

$$W_p := I_{qp} = \sum_{q=1}^w I_q^p;$$

5) вычисляется φ – коэффициент эмерджентности Хартли как уровень системной организации:

$$\varphi = \frac{\log_2 \sum_{p=1}^r W_p}{\log_2 \overline{W}}; \quad (\varphi \rightarrow (\log_2 \overline{W} / \log_2 \underline{W}));$$

6) вычисляется Ψ – коэффициент эмерджентности Харкевича как степень детерминированности объекта с уровнем системной организации φ , имеющего Z состояний, на переходы в которые оказывают влияние w факторов, о чем в модели накоплено N фактов:

$$\Psi = \frac{\log_2 \overline{W}^\varphi}{\log_2 N}, \quad (\Psi \rightarrow 0);$$

Рис. 1. Иллюстрация данных о выбранных вариантах ФПД ЧЭ

В рамках предложенного метода рассмотрим фрагмент оценки варианта ФПД ЧЭ терморегулятора [7], который будет использоваться в помещениях при низких температурах. В терминах автоматизированной системы «Интеллект»: природа входной величины – I_3 (электрический ток), природа выходной величины – I_3 (электрический

4) высчитывается суммарное количество информации, коэффициент эмерджентности Хартли – ϕ , эмерджентности Харкевича – Ψ ;

Рис. 2. Иллюстрация инженерной методики расчета

5) очевидно, что $\varphi \rightarrow (\log_2 \bar{W} / \log_2 \underline{W})$, $\psi \rightarrow 0$, $I \rightarrow 1$;

6) поскольку $W(\sum I) \equiv [0; 1] \rightarrow W(A_i) \equiv [1; 100]$, то решение 2 оценивается наивысшим баллом $W(A_2^0) = 17,4$ (поэтому считается наиболее эффективным); решение 4 оценивается наинизшим баллом $W(A_4^0) = 11,84$ $W(A_2^0) = 17,4$ (поэтому считается наименее эффективным). Для проверки, является ли полученное значение максимальным/минимальным, выполняется переход к шагу 2 и повтор вычислительных процедур.

В настоящее время разработана система, реализующая комплексный информационный

метод и позволяющая количественно оценить ФПД ЧЭ на этапе концептуального проектирования, что, в свою очередь, определяет для выбранного варианта ФПД ЧЭ его дальнейшую проработку на этапе технологического проектирования.

Реализация данного метода позволяет, во-первых, повышать эффективность процесса выбора решения при различных семантических и синтаксических интерпретациях данных о ФПД ЧЭ, инвариантных к различным программным средствам [9], и, во-вторых, применять разработанный метод на ранних этапах проектирования датчиковой аппаратуры.

СПИСОК ЛИТЕРАТУРЫ

1. **Лабути́н, С.А.** Доклады на научных конференциях по измерениям (Нижний Новгород, декабрь 2008 – июнь 2009) [Текст] / С.А. Лабу́тин // Датчики и системы. – 2009. – № 10. – 81 с.
2. **Коробкин, Д.М.** Автоматизированная методика извлечения структурированных физических знаний в виде физических эффектов из текстов на естественном английском языке [Текст] / Д.М. Коробкин, С.А. Фоменков // Изв. Волгоградского гос. технич. ун-та: межвуз. сб. науч. ст. Сер. Актуальные проблемы управления, вычислительной техники и информатики в технических системах. – Волгоград: Изд-во ВолгГТУ, 2011. – Вып. 10. – № 3 (76). – С. 111–116.
3. **Хоменко, Т.В.** Автоматизированные системы поискового конструирования: системный анализ и развитие системной парадигмы [Текст] / Т.В. Хоменко // Вестник АГТУ. – 2010. – № 1. – С. 136–141.
4. **Хоменко, Т.В.** Универсальное представление предметных задач поискового конструирования физического принципа действия чувствительных элементов системы управления [Текст] / Т.В. Хоменко // Информатизация и связь. – 2012. – № 5. – С. 126–129.
5. **Фатхутдинов Р.А.** Инновационный менеджмент: Учебник [Текст] / Р.А. Фатхутдинов. – 4-е изд. – СПб.: Питер, 2003. – 400 с.
6. **Хоменко, Т.В.** Формирование пространства выбора физического принципа действия чувствительных элементов при поисковом конструировании [Текст] / Т.В. Хоменко // Вестник Дагестанского гос. ун-та. – 2011. – Вып. 6. – С. 76–80.
7. **Хоменко, Т.В.** Терморегулятор. Патент № 114181 от 10.03.2012 РФ / Т.В. Хоменко, А.А. Ветрова; Заявитель и патентообладатель Астраханский гос. ун-т: заявл. № 2011140960 от 7.10.2011.
8. **Зарипов, М.Ф.** Предметно-ориентированная среда поиска новых технических решений «Интеллект» [Текст] / М.Ф. Зарипов, И.Ю. Петрова // IV Междунар. конф. РИ-95. – СПб., 1995. – С. 60–61.
9. **Хоменко, Т.В.** Анализ экспериментальных сравнений ранжирования физического принципа действия чувствительных элементов [Текст] / Т.В. Хоменко, М.А. Мурыгин // Вестник АГТУ. Сер. Управление, вычислительная техника и информатика. – 2011. – № 2. – С. 97–102.

НЕКОТОРЫЕ СТАНДАРТНЫЕ СХЕМЫ ПАРАМЕТРИЧЕСКОЙ ОПТИМИЗАЦИИ

Известны методы, позволяющие формально свести общую задачу параметрической оптимизации с ограничениями, возмущениями и векторным критерием оптимальности к последовательности концептуально более простых однокритериальных оптимизационных задач без ограничений – назовем их для определенности *каноническими* оптимизационными задачами. Мы здесь предполагаем, что такое сведение выполнено.

Возникающие при таком сведении канонические задачи обладают рядом особенностей, выделяющих их среди общих задач безусловной нелинейной оптимизации. Основные с точки зрения приложений особенности заключаются в следующем.

Особенности оптимизационных задач

1. Критерии оптимальности часто имеют характерную структуру, позволяющую строить специальные методы оптимизации второго порядка, использующие упрощенные выражения для производных.

2. Однократное вычисление функционалов задачи связано со сложным и трудоемким решением соответствующей задачи анализа. Поэтому наиболее эффективными целесообразно считать алгоритмы, которые в процессе оптимизации наименьшее число раз обращаются к вычислению значений минимизируемых функционалов и функционалов ограничений для получения решений с требуемой точностью.

3. Если оптимизируется сложная многопараметрическая система, то ее обычно можно представить как некоторую совокупность связанных подсистем меньшей размерности. Учет подобной структуры системы обычно позволяет строить более рациональные методы по сравнению с традиционными универсальными алгоритмами нелинейного программирования.

4. *Невыпуклая структура* минимизируемых функционалов существенно понижает эффективность стандартных методов нелинейной оптимизации, особенно если такой структуре сопутствует

указываемая в п. 5 овражная ситуация.

5. Как свидетельствует практика решения реальных задач параметрической оптимизации, возникающие канонические оптимизационные задачи являются, как правило, *плохо обусловленными (жесткими)*. Это определяет характерную овражную структуру поверхностей уровня минимизируемых функционалов и вызывает резкое замедление сходимости стандартных методов оптимизации [1].

Указанные характерные черты канонических оптимизационных задач определяют конкретные требования к методам параметрической оптимизации. С позиций традиционных методов нелинейной оптимизации наиболее существенными оказываются особенности, отмеченные в пунктах 4, 5. Основная трудность состоит в том, что с математической точки зрения проблема невыпуклости оказывается неразрешимой в силу сложности класса невыпуклых оптимизационных задач. Согласно результатам, представленным, например, в [2], основной вывод, сформулированный в интуитивных терминах, заключается в том, что даже для гладких одноэкстремальных функционалов в задачах с не очень малой размерностью пространства управляемых параметров скорость сходимости любого метода (равномерно по всем задачам) безнадежно мала, и попытка построить общий метод, эффективный для всех гладких невыпуклых задач, заранее обречена на неудачу.

С точки зрения практики представляют интерес методы оптимизации, вырабатывающие эффективные направления поиска в точках пространства, где стандартные процедуры оказываются неработоспособными либо неэффективными. Необходимы методы, которые как в выпуклой, так и в невыпуклой и одновременно овражной ситуации локально (с точки зрения квадратичной модели) дают существенно более удовлетворительные по скорости убывания функционала результаты по сравнению с традиционными методами. Соответствующие методы будут рассмотрены в последующих публикациях.

В зависимости от реальных ситуаций, примеры которых рассмотрены в работах [3, 4], могут формироваться различные алгоритмические схемы параметрической оптимизации. В данной статье предложена единая стратегия решения задач параметрической оптимизации на основе *среднестепенных целевых функционалов*.

По сути речь идет о методах построения целевых функционалов, обладающих определенным набором желательных свойств с позиций последующего эффективного применения специальных методов нелинейной оптимизации.

Задачи аппроксимации

Задачи аппроксимации наиболее часто возникают в практике параметрической оптимизации. К таким задачам, в частности, приводят:

- алгоритмы идентификации нелинейных детерминированных объектов по схеме с настраиваемой моделью;
- алгоритмы идентификации стохастических объектов, основанные на процедурах сглаживания;
- корреляционные алгоритмы идентификации линейных стохастических объектов;
- алгоритмы адаптации систем автоматического управления, основанные на методе обучающейся модели;
- алгоритмы оптимального проектирования устройств с заданными, например, амплитудно-частотными характеристиками;
- корреляционные алгоритмы синтеза статистически оптимальных систем автоматического управления.

Критерии оптимальности в задачах аппроксимации чаще всего формируются одним из следующих способов:

$$J_1(x) = \sum_{i=1}^N \alpha_i^2 [W(x, s_i) - W^*(s_i)]^2 \rightarrow \min; \quad (1)$$

$$J_2(x) = \max_{i=1, N} \alpha_i |W(x, s_i) - W^*(s_i)| \rightarrow \min, \quad (2)$$

где D – множество допустимых значений x ; α_i – весовые коэффициенты, определяющие необходимую точность аппроксимации в отдельных точках диапазона изменения независимой переменной s ; $s_i, i \in [1:N]$ – дискретная сетка значений s , при которых происходит сравнение заданной $W^*(s)$ и расчетной $W(x, s)$ характеристик.

Каждый из приведенных критериев имеет свои достоинства и недостатки. Функционал J_1 прост и

обладает свойством «гладкости». Если функция $W(x, s)$ является дважды непрерывно дифференцируемой функцией x , то этим же свойством будет обладать зависимость $J_1(x)$, что существенно облегчает последующую процедуру оптимизации. Недостаток J_1 заключается в возможности выбросов по точности для отдельных слагаемых. Иначе говоря, плохая точность аппроксимации в некоторых точках s_i при больших значениях N может компенсироваться хорошей точностью в других точках. Этот недостаток устранен в критерии (2), однако он не сохраняет характеристики гладкости функции $W(x, s)$, что требует привлечения специальных методов оптимизации.

Как показывает практика, простой и надежный способ решения задач аппроксимации заключается в использовании гладких среднестепенных аппроксимаций минимаксного критерия J_2 . Согласно этому подходу вместо решения задачи (2) ищется минимум функционала со *среднестепенной структурой* [4]:

$$J_3(x) = \sum_{i=1}^N \varphi_i^v(x) \rightarrow \min, v = 2, 3, \dots \quad (3)$$

где $\varphi_i(x) \triangleq \alpha_i |W(x, s_i) - W^*(s_i)|$.

При достаточно больших значениях v решения задач (2), (3) будут практически совпадать. Действительно, справедливо предельное соотношение

$$\left[\sum_{i=1}^N \varphi_i^v \right]^{1/v} \rightarrow \max_i \varphi_i, v \rightarrow \infty,$$

где $\varphi_i \geq 0, i \in [1:N]$ – произвольные числа. Кроме этого, можно показать, что операция извлечения корня v -й степени не влияет на локализацию точки минимума.

Функционал J_3 совмещает в себе достоинства J_1 и J_2 . Являясь гладким подобно J_1 , он не допускает значительных отличий по точности аппроксимации в отдельных точках.

При решении практических задач на основе критерия J_3 целесообразно пошаговое увеличение параметра v , начиная с $v = 2$. Таким способом обычно удается избежать переполнения разрядной сетки компьютера при возведении первоначально больших значений локальных ошибок аппроксимации φ_i в высокую степень v . Кроме того, проводя в интерактивном режиме оценку получаемых в процессе увеличения v решений, мы можем вовремя прервать процесс, если получены удовлетворительные результаты. Апри-

орное задание оптимального значения v обычно оказывается затруднительным. Как правило, при практических расчетах значение v не превышает 10–15. Учет ограничений $x \in D$ производится специальными методами снятия ограничений, например, методами замены переменных, методами штрафных функций, а в более сложных случаях – методами модифицированных функций Лагранжа (*точных штрафных функций*).

Рассмотренный подход очевидным образом обобщается на случай векторных функций W . При этом в качестве функций $\varphi_i(x)$ могут использоваться зависимости $\varphi_i(x) = \alpha_i \|W(x, s_i) - W^*(s_i)\|$.

Системы неравенств

Далее показывается, что изложенный выше метод построения гладких целевых функционалов можно обобщить и на задачи решения систем неравенств.

Наиболее часто задачи с неравенствами возникают при формализации основных спецификаций (требований) к оптимизируемому объекту, имеющих смысл «не хуже», «не меньше», «не больше» и т. п. В ряде случаев аппроксимационные по смыслу задачи (содержащие требования в виде равенств) также целесообразно формулировать и решать как системы неравенств. Например, если ставится задача аппроксимации заданной характеристики $W(s)$ с помощью расчетной зависимости $W(x, s)$ в точках $s_i, i \in [1:N]$, и если при этом заданы допустимые значения $\delta_i > 0$ точности аппроксимации в различных точках диапазона $W(x, s_i) \in [W(s_i) - \delta_i, W(s_i) + \delta_i], i \in [1:N]$, то требование $W(x, s_i) \approx W(s_i)$ очевидно эквивалентно двум неравенствам:

$$\begin{aligned} W(x, s_i) &\geq W(s_i) - \delta_i; \\ W(x, s_i) &\leq W(s_i) + \delta_i. \end{aligned}$$

В общем случае имеем

$$y_i(x) \leq t_i, i \in [1:m]. \quad (4)$$

С помощью задания системы (4) могут учитываться все виды ограничений на объект оптимизации: прямые, функциональные и критериальные.

Особенность рассматриваемого класса задач заключается в том, что обычно нельзя считать, что «чем меньше $y_i(x)$, тем лучше». Напротив, на одну и ту же выходную характеристику, как правило, накладываются двусторонние ограничения $\varphi_{\text{KH}} \leq \varphi_k(x) \leq \varphi_{\text{KB}}$, что приводит к двум неравенствам вида (4):

$$\begin{aligned} y_i(x) &\triangleq \varphi_k(x) \leq t_i \triangleq \varphi_{\text{KB}}; \\ y_j(x) &\triangleq -\varphi_k(x) \leq t_j \triangleq -\varphi_{\text{KH}}. \end{aligned}$$

Поэтому трактовка задачи (4) как многокритериальной вида $y_i(x) \rightarrow \min$ с применением соответствующих процедур свертки, обычно нецелесообразна. Как правило, необходимо обеспечить безусловное выполнение всех неравенств (4), не допуская больших значений $y_i(x)$ (превышающих соответствующие t_i), за счет очень малых значений $y_j(x)$ других выходных параметров.

Рациональный подход связан с применением критериев вида

$$J(x) = \min_i \gamma_i^{-1}(t_i - y_i(x)) \rightarrow \max_{x \in R^n} \quad (5)$$

или

$$J(x) = \max_i \gamma_i^{-1}(t_i - y_i(x)) \rightarrow \min_{x \in R^n}.$$

При этом параметры γ_i должны задавать «единицы измерения» разностей $t_i - y_i$ для различных i . Эти разности характеризуют *запас*, с которым выполняются неравенства (4). Согласно (5) максимизируется минимальный из запасов.

Функционал (5) так же как и (2) не является гладким, что существенно усложняет ситуацию и требует применения специальных оптимизирующих процедур [7].

Ниже излагается альтернативный подход, основанный на процедуре сглаживания исходного функционала с последующим обращением к методам гладкой оптимизации.

Пусть $z_i(x) \triangleq t_i - y_i(x)$.

Тогда, очевидно, $\min_i z_i(x) = \max_i [\exp(-z_i(x))]$.

Поэтому задача (5) эквивалентна задаче

$$\max_i [\exp(-z_i(x))] \rightarrow \min_x. \quad (6)$$

Для задачи (6) применима среднестепенная свертка (3), если положить $\varphi_i(x) \triangleq \exp[-z_i(x)], i \in [1:m]$.

В результате приходим к следующему критерию оптимальности: ...

$$J(x) = \sum_{i=1}^m \exp[-v z_i(x)] \rightarrow \min_x, v = 1, 2, \dots \quad (7)$$

Точно так же могут решаться и многокритериальные задачи вида

$$f_i(x) \rightarrow \min_x, i \in [1:l]; x \in D \subset R^n;$$

$$D = \{x \in R^n \mid g_j(x) \leq t_j, j \in [l+1:s]\},$$

где функции $g_j(x)$ задают функциональные ограничения. Сформулированная задача может быть

представлена в виде системы неравенств

$$f_i(x) \leq \hat{f}_i, g_j(x) \leq t_j; i \in [1:L]; j \in [l+1:s],$$

где $\hat{f}_i$ – некоторые предельные значения для критерийных выходных параметров f_i .

Полагая

$$z_i(x) = \alpha_i[\hat{f}_i - f_i(x)]/\gamma_i; i \in [1:L];$$

$$z_j(x) = \alpha_j[t_j - g_j(x)]/\gamma_j; j \in [l+1:s],$$

приходим к следующему целевому функционалу, аналогичному (7):

$$J(x) = \sum_{i=1}^s \exp[-vz_i(x)] \rightarrow \min_x.$$

Весовые коэффициенты α_i, α_j должны выбираться из условия $\alpha_i \ll \alpha_j$ ($i \in [1:L]; j \in [l+1:s]$). В этом случае (при достаточно больших v) критерий $J(x)$ эквивалентен минимаксному критерию (6), что в свою очередь эквивалентно задаче $\min_i z_i \rightarrow \max_x$.

Очевидно, в силу больших значений α_j ($j \in [l+1:s]$) вес соответствующих слагаемых в выражении для $J(x)$ резко возрастает при нарушении хотя бы одного из условий $g_j(x) \leq t_j$ (т. к. тогда $z_j(x) < 0$). Если указанные условия выполнены, то, по существу, производится выбор x исходя из максимальности минимального запаса по критериальным неравенствам. В данном случае параметры α_j ($j \in [1:s]$) играют роль штрафных коэффициентов при учете функциональных ограничений. Параметры α_i ($i \in [1:L]$) могут использоваться для построения на основе максиминной свертки аппроксимации множества Парето в допустимом множестве D , либо отражать интуитивные представления об относительной значимости отдельных критериальных выходных параметров.

В ряде случаев рассмотренный подход в силу своего единообразия и простоты оказывается наиболее предпочтительным при решении реальных задач параметрической оптимизации.

Как показывает вычислительная практика, трудности оптимального синтеза параметров по критериям типа максимума минимального запаса (5) в некоторых случаях вызываются исключительно негладкостью критериев, приводящей к преждевременной остановке поисковой процедуры. Целесообразно поэтому сразу обращаться к модифицированным критериям (7) с применением на первом этапе простейших алгоритмов оптимизации типа метода простого покоординатного спуска.

Использование среднестепенных критериев оптимальности в задачах параметрической оптимизации, где, по существу, необходим минимаксный подход, оправданно также с позиций явления жесткости или плохой обусловленности. Развита в настоящее время техника решения негладких оптимизационных задач достаточно сложна, и в невыпуклой овражной ситуации многие алгоритмы теряют эффективность. В то же время могут быть развиты методы, позволяющие получить удовлетворительные результаты для невыпуклых жестких (овражных) функционалов *при условии их гладкости*. При этом обычно удается использовать структурные особенности функционалов (3), (7) для увеличения эффективности соответствующих вычислительных процедур.

Решение систем неравенств в условиях неопределенности

В данном разделе рассматривается специальный случай неопределенности обстановки. Предполагается, что вектор ξ , входящий в левые части решаемых неравенств

$$y(x, \xi) \leq t, y = (y_1, \dots, y_m), t = (t_1, \dots, t_m), \quad (8)$$

имеет случайный характер. Подобные постановки задач возникают, например, при алгоритмизации процедур *управления технологическими процессами* в условиях серийного производства продукции.

В качестве критерия, глубоко отражающего конечную цель решения указанной задачи, как правило, можно выбрать вероятность P выполнения условий (8):

$$J(x) = P[y(x, \xi) \leq t] \rightarrow \max_x. \quad (9)$$

В данном случае все возможные ограничения на компоненты вектора x считаются учтенными за счет расширения системы неравенств (8). Процедура однократного вычисления функционала $J(x)$ основывается на проведении статистических испытаний по методу Монте-Карло в соответствии с заданной плотностью распределения $\Psi(x, \xi)$ случайного вектора ξ . Одновременно с расчетом значения целевого функционала можно практически без дополнительных вычислительных затрат рассчитать составляющие вектора градиента $J'(x)$ и матрицы Гессе $J''(x)$. Данное обстоятельство положено в основу построения некоторых реализаций методов параметрической оптимизации.

Альтернативный подход к решению задачи (9)

рассмотрен в [6, 7]. Он заключается в следующем. Потребуем, чтобы каждое неравенство (8) выполнялось с некоторым запасом:

$$y_i(x, \bar{\xi}) + \delta_i \leq t_i, \delta_i > 0, \quad (10)$$

где δ_i характеризует величину рассеяния i -го выходного параметра за счет статистических вариаций компонент вектора ξ относительно своих средних (как правило, нулевых) значений $\bar{\xi}_i$. Требования (10) эквивалентны неравенствам

$$z_i(x) \triangleq \left[\frac{(t_i - y_i)}{\delta_i} - 1 \right] \geq 0. \quad (11)$$

Величина z_i имеет смысл *запаса работоспособности* по i -му выходному параметру.

В вычислительной практике используется максиминная форма целевого функционала

$$J(x) = \min_i z_i(x) \rightarrow \max_{x \in R^n}, \quad (12)$$

аналогичная (5).

От представления (12) предлагается, как уже было показано выше, перейти к выражению (7), более удобному для практических расчетов. В [6, 7] и в других работах такой переход не делается, что приводит к существенным вычислительным трудностям и необходимости разработки специальных, достаточно громоздких вычислительных процедур. В наших экспериментах показано, что после перехода к гладким аппроксимациям (12) по излагаемой в данной статье методике никаких вычислительных проблем для задач из [7] не возникало. Все эти задачи были решены простейшим методом циклического покоординатного спуска и не потребовали разработки новых методов оптимизации.

В ряде случаев в выражение (12) вводятся дополнительные весовые коэффициенты α_i , позволяющие регулировать степень «перевыполнения» требований (10) по отдельным выходным параметрам.

Для определения δ_i проводится статистический анализ в окрестности текущей точки x . Значения δ_i обычно имеют смысл трехсигмовых допусков, которые периодически уточняются в процессе оптимизации. Весьма часто величины δ_i задаются как исходные данные на основе априорной информации, что значительно сокращает трудоемкость процедуры оптимизации, особенно при решении идентичных задач.

Сигномиальная оптимизация

В целом ряде практических ситуаций, связанных, например, с задачами оптимального проектирования, целевые функционалы или их аппроксимации имеют стандартную *сигномиальную* структуру:

$$J(x) = \sum_{i=1}^k s_i(x), \quad (13)$$

где $s_i(x) = c_i x_1^{\alpha_{i1}} x_2^{\alpha_{i2}} \dots x_n^{\alpha_{in}}; \alpha_{ij} \in R^1; x_j > 0; j \in [1:n]$.

В частном случае $c_i \geq 0$ выражение (13) называется *позиномом*, а соответствующая задача параметрической оптимизации – *задачей геометрического программирования* [8].

Могут быть развиты общие стандартные процедуры сигномиальной оптимизации, основанные на простоте вычисления первых и вторых производных функционалов (13). При этом предполагается, что возможные дополнительные ограничения на параметры x_i также имеют вид сигномов и учитываются методами штрафных функций или модифицированных функций Лагранжа без нарушения сигномиальной структуры расширенного целевого функционала. Методы сигномиальной оптимизации также относятся к стандартным схемам выбора параметров технических систем и устройств.

Таким образом, в статье представлена техника построения гладких критериев оптимальности, в т. ч. в задачах аппроксимации и в задачах, приводящих к системам неравенств, часто используемым для формализации технических и иных требований к оптимизируемому объекту. Традиционные минимаксные стратегии вызывают известные трудности в использовании «библиотечных» методов нелинейного программирования из-за негладкости целевых функционалов, особенно при решении систем неравенств. Эти трудности проиллюстрированы, например, в [7]. Обсуждаемые в статье гладкие аппроксимации минимаксных критериев позволяют для решения широкого класса задач (в т. ч. и представленных в [7]) ограничиваться простейшими реализациями стандартных алгоритмов нелинейного программирования.

СПИСОК ЛИТЕРАТУРЫ

1. **Ракитский, Ю.В.** Численные методы решения жестких систем [Текст] / Ю.В. Ракитский, С.М. Усти-

нов, И.Г. Черноруцкий. –М.: Наука, 1979.

2. **Немировский, А.С.** Сложность задач и эффек-



тивность методов оптимизации [Текст] / А.С. Немировский, Д.Б. Юдин. – М.: Наука, 1979.

3. **Черноруцкий, И.Г.** Методы параметрической оптимизации в задачах идентификации [Текст] / И.Г. Черноруцкий // Научно-технические ведомости СПбГПУ. – 2009. – № 2. – С. 150–155.

4. **Черноруцкий, И.Г.** Параметрические методы синтеза систем управления [Текст] / И.Г. Черноруцкий // Научно-технические ведомости СПбГПУ. – 2009. – № 2. – С. 111–115.

5. **Растрингин, Л.А.** Введение в идентифика-

цию объектов управления [Текст] / Л.А. Растрингин, Н.Е. Маджаров. – М.: Энергия, 1977.

6. **Норенков И.П.** Введение в автоматизированное проектирование технических устройств и систем [Текст] / И.П. Норенков. – М.: Высш. шк., 1986.

7. **Норенков, И.П.,** Экстремальные задачи при схемотехническом проектировании в электронике [Текст] / И.П. Норенков, С.Г. Мулярчик, С.Р. Иванов. – Минск: Изд-во БГУ, 1976.

8. **Таха, Х.** Введение в исследование операций: Кн. 1, 2 [Текст] / Х. Таха. – М.: Мир, 1985.

УДК 62.50+517.11+519.92

М.А. Марценюк, В.Б. Поляков, И.П. Селетков

МАТРИЧНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМОВ НЕЧЕТКОГО ВЫВОДА

Нечеткая логика и ее методы нашли многочисленное применение в задачах моделирования и управления [1]. Предполагается, что нечеткие логические переменные x , y могут принимать промежуточные значения между нулем и единицей. В большинстве случаев для реализации нечетких логических операций над переменными используются предложенные Л. Заде нечеткая конъюнкция $x \wedge y = \min(x, y)$ и дизъюнкция $x \vee y = \max(x, y)$ [2]. Такое определение хорошо работает при вычислениях по заданным логическим формулам, что и используется в подавляющем числе задач моделирования и управления (подробный обзор представлен в книге А. Пегата [1]).

Однако на практике большое значение имеют задачи нечеткого логического вывода. В этом случае неизвестная переменная фигурирует под знаком логической операции и требуется ее найти. Например, для правила вывода *modus ponens* $x \rightarrow y = z$ предполагается, что нам известны нечеткие переменные x , z и требуется найти нечеткость переменной y . Эту задачу затруднительно решить с помощью операций Л. Заде.

В работе [3] предложено использовать для проведения вывода матричный аппарат. В нем применяются модели логических операций, естественным образом обобщающие их «четкие» аналоги на область нечетких переменных и предикатов. Этот аппарат позволяет свести задачи нечеткого логического вывода к решению

системы линейных алгебраических уравнений с известными условиями существования и единственности решений. Цель данной статьи состоит в том, чтобы показать применимость этого подхода для решения разнообразных задач нечеткого логического вывода.

Матричное представление нечетких высказываний

Под высказыванием A в обычной (четкой) логике понимают такое предложение, которое либо истинно, либо ложно [4]. На совокупности всех высказываний можно определить функцию $\mu(A)$, принимающую значения в двухэлементном множестве $\{0, 1\}$, по следующему правилу:

$$\mu(A) = \begin{cases} 1, & \text{если высказывание } A \text{ истинно,} \\ 0, & \text{если высказывание } A \text{ ложно.} \end{cases} \quad (1)$$

Функция $\mu(A)$ называется функцией истинности, а ее значение – логическим значением или значением истинности высказывания A .

Однако не обо всех высказываниях можно однозначно утверждать истинные они или ложные. Под высказыванием в нечеткой логике понимается такое предложение, которое может принимать истинное значение с какой-то степенью уверенности. На совокупности нечетких высказываний можно определить функцию истинности $\mu(A)$, принимающую значения на отрезке $[0, 1]$, равную степени уверенности в справедливости соответ-

ствующего высказывания.

Алгебра нечетких высказываний. В матричном представлении [3] для описания степени истинности высказывания A вводится двухмерный вектор x . Если ввести базисные векторы $e^{(0)}$ и $e^{(1)}$, то вектор x может быть представлен в виде

$$x = x_0 e^{(0)} + x_1 e^{(1)}, \quad (2)$$

причем компоненты вектора x – числа x_0, x_1 – удовлетворяют условиям

$$0 \leq x_0, x_1 \leq 1, \quad x_0 + x_1 = 1. \quad (3)$$

Сами базисные векторы $e^{(0)}$ и $e^{(1)}$ описывают высказывания со степенью уверенности 0 и 1 соответственно. Операции отрицания соответствует перестановка компонент вектора x : $\neg x = x_1 e^{(0)} + x_0 e^{(1)}$.

В «классической» нечеткой логике оценка неопределенности высказывания вызывает затруднения [5]. Представление нечетких высказываний в векторном виде позволяет вычислять их неопределенность с помощью известной формулы теории информации:

$$S(x) = -x_0 \log_2 x_0 - x_1 \log_2 x_1. \quad (4)$$

Логические операции над векторами – конъюнкцию, дизъюнкцию и импликацию – можно представить в матричном виде. Составим из компонент вектора x конъюнктивную $C(x)$, дизъюнктивную $D(x)$ и импlicativeвную $I(x)$ матрицы размером 2×2 [3]:

$$\begin{aligned} C(x) &= \begin{pmatrix} 1 & x_0 \\ 0 & x_1 \end{pmatrix}; \quad D(x) = \begin{pmatrix} x_0 & 0 \\ x_1 & 1 \end{pmatrix}; \\ I(x) &= \begin{pmatrix} x_1 & 0 \\ x_0 & 1 \end{pmatrix}. \end{aligned} \quad (5)$$

С помощью этих матриц логические операции над нечеткими векторами могут быть пред-

ставлены в виде

$$x \wedge y = C(x)y; \quad x \vee y = D(x)y; \quad x \rightarrow y = I(x)y, \quad (6)$$

где вектор y следует мыслить как матричный столбец 1×2 , составленный из его компонент $y = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$, а точка обозначает обычное матричное

умножение. Результатом логических операций, сформулированных в таком виде, тоже будет нечеткий вектор. Детальное обоснование сделанного выбора матриц (4) представлено в работе [3].

Следует отметить, что введенные здесь логические операции отличаются от «классических» формул Л. Заде. Например, согласно (6) истинностная компонента конъюнкции $[x \wedge y]_1 = x_1 y_1$, в то время как «классический» результат для этой величины равен $\min(x_1, y_1)$. В частном случае «четких» (и близких к ним) значений эти формулы дают одинаковый результат (рис. 1 а), однако при $x_1 \sim y_1 \sim 0,5$ расхождение велико (рис. 1 б). На наш взгляд, из этого вовсе не следует, что одни формулы лучше, а другие хуже. Формулы Л. Заде обоснованы только их повсеместным использованием, в действительности они представляют собой только одну из возможных моделей логических операций. Наряду с ними в литературе предложено множество других вариантов реализаций нечетких операций [6]. Возвращаясь к операции конъюнкции, можно видеть, что нечеткости высказываний A («Выиграть миллион»), задаваемой вектором x , и $B = A \wedge A$ («Выиграть миллион **И** снова Выиграть миллион»), задаваемой вектором $x \wedge x$, различны, что и отражается в аппарате векторного представления нечеткой логики. В общем случае $\mu(B)$ и $\mu(A)$ не совпадают или $x \wedge x \neq x$. Равенство $x \wedge x = x$ выполняется только в том случае, когда x равно одному из базисных векторов $e^{(0)}$ или $e^{(1)}$.

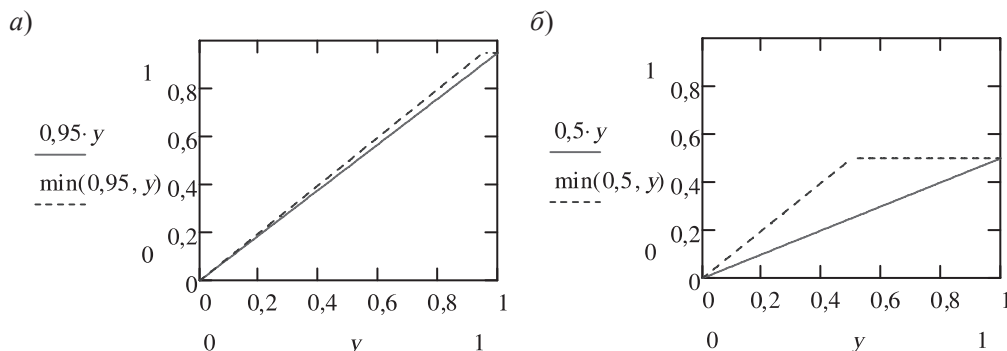


Рис. 1. Результаты нечетких конъюнкций в соответствии с (4) и Заде:

а – в приближении «четких» переменных; б – в случае наибольшей нечеткости или неопределенности

Привлекательность матричного подхода обусловлена тем, что он:

1) естественным образом обобщает логические операции на область нечетких переменных [3] и во многих случаях адекватно моделирует нечеткие рассуждения;

2) позволяет производить нечеткий логический вывод с использованием правил *modus ponens*, *modus tollens* и др. путем решения линейных алгебраических уравнений;

3) позволяет количественно оценивать неопределенность нечетких высказываний по формуле (4);

4) при выводах можно использовать нечеткое правило резолюции и вычислять его нечеткость [3].

Нечеткий вывод. Нечеткий логический вывод или приближенные рассуждения – это процесс получения значений искоемых нечетких переменных или предикатов по заданным значениям исходных переменных с использованием логических правил *modus ponens* вида «если ..., то ...», *modus tollens* и др. Предпосылки и выводы правил являются нечеткими понятиями. Совокупность всех правил представляет собой базу знаний, которая формируется человеком-экспертом в данной предметной области. Общие этапы логического вывода, а также примеры реализаций в «классической» формулировке нечеткой логики по алгоритмам Mamdani, Larsen и др. описаны во многих источниках (см. [1, 7]). В этом подходе предлагается задавать правила в базе знаний с помощью отношений $x \rightarrow y = R$, где матрица R должна быть эмпирически задана на множестве $x \times y$ (здесь знак « $\times$ » обозначает прямое произведение векторов x, y).

Наиболее точное обобщение правила *modus ponens* на нечеткую логику предложено в статье [3]. Оно имеет аналогичное четкому варианту поведение в особых точках и учитывает факт связи между предпосылкой, следствием и степенью выполнения правила.

Если известен нечеткий вектор $x = (x_0, x_1)^T$, а также нечеткий вектор z , определяющий нечеткость правила $x \rightarrow y = z$, то вывод о нечеткости вектора y можно сделать из соотношений (5), (6). Эта задача сводится к решению линейного алгебраического уравнения, относительно неизвестного вектора y , которое в случае импликативной модели имеет вид $I(x)y = z$:

$$\begin{pmatrix} x_1 & 0 \\ x_0 & 1 \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \quad \text{или} \quad \begin{cases} x_1 y_0 = z_0 \\ x_0 y_0 + y_1 = z_1 \end{cases}. \quad (7)$$

Как известно, линейная система имеет решение, если определитель матрицы коэффициентов не равен нулю. В данном случае мы должны потребовать, чтобы выполнялось условие

$$\det(I(x)) = \det \begin{pmatrix} x_1 & 0 \\ x_0 & 1 \end{pmatrix} = x_1 \neq 0. \quad (8)$$

То есть компонента истинности x_1 вектора x не должна быть равна нулю. Тогда компоненты неизвестного вектора y принимают значения

$$y_0 = \frac{z_0}{x_1}; \quad y_1 = 1 - \frac{z_0}{x_1}. \quad (9)$$

Полученное решение имеет смысл только в том случае, если обе компоненты вектора y положительны. Поэтому кроме условия $x_1 \neq 0$, получаем дополнительно, что $x_1 \geq z_0$. Отметим, что при $x_1 = 1$ и $z_1 = 1$ из (9) имеем известный результат правила *modus ponens* в четкой логике: $y_0 = 0$, $y_1 = 1$.

Таким образом, в отличие от других моделей импликации в этом подходе мы не только получаем нечеткий логический вывод, но и строго указываем область его выполнения. Аналогичным образом могут быть обобщены на область нечетких переменных и другие известные правила логического вывода [3].

Связь с алгоритмом Мамдани. В приложениях нечеткой логики большое распространение получил алгоритм Мамдани, в котором предлагается находить следствие правила y как конъюнкцию предпосылки правила x и некоторой переменной w , заданной на области определения y :

$$y = x \wedge w. \quad (10)$$

В матричном виде это выражение имеет вид:

$$y = C(x)w. \quad (11)$$

Нечеткое правило *modus ponens* может быть записано следующим образом через импликацию:

$$x \rightarrow y = z. \quad (12)$$

В матричном виде это правило выглядит как:

$$I(x)y = z. \quad (13)$$

Из формул (11) и (13) следует связь между переменными w и z :

$$z = C(x)I(x)w. \quad (14)$$

Согласно этой формуле, задавая переменную w , как того требует алгоритм Мамдани, зададим и нечеткость z правила *modus ponens* (12). В этом смысле можно строго считать, что алгоритм Мамдани является логическим выводом, хотя обычно его рассматривают только как модель вывода. При решении конкретных задач во многих случаях бывает легче определить переменную w . Матричный метод позволяет при этом указать степень нечеткости z правила (12).

Нечеткие переменные как нечеткие предикаты

Описание нечетких предикатов. В классическом подходе [2, 7] *нечеткая переменная* характеризуется нечетким подмножеством X некоторого множества S и функцией принадлежности $\mu_x(s)$, $s \in S$. Например, переменная y_1 «уровень обслуживания низкий» может характеризоваться функцией принадлежности, изображенной на рис. 2. В данном случае в качестве области определения S выступает оценка уровня обслуживания клиентом в баллах ($s \in [1...10]$) (пример из [7]).

Значение функции принадлежности может выступать как свойство некоторого объекта из множества S . Такое описание хорошо подходит под определение предикатов, применение которых для нечеткой логики введено в работе [3].

Нечеткий предикат $x(s)$ – это векторная функция, заданная на множестве определения S , принимающая значения в пространстве нечетких векторов F , элементы которого описываются формулами (2) и (3), то есть

$$x(s) = \begin{pmatrix} x_0(s) \\ x_1(s) \end{pmatrix}; \quad (15)$$

$$0 \leq x_0(s), x_1(s) \leq 1; \quad x_0(s) + x_1(s) = 1. \quad (16)$$

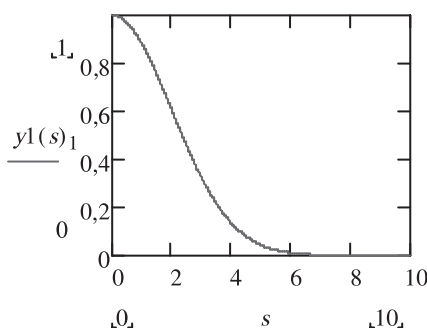


Рис. 2. Внешний вид функции принадлежности нечеткой переменной y_1 «уровень обслуживания низкий»
 s – оценка уровня обслуживания клиентом в баллах

Из определения предиката следует, что в качестве значения $x_1(s)$ используется значение функции принадлежности нечеткой переменной $\mu_x(s)$. В то время как для выполнения условий (3) необходимо $x_0(s) = 1 - \mu_x(s)$. Операции над нечеткими предикатами сводятся к операциям над нечеткими векторами.

Лингвистическая переменная – это переменная, которая может принимать значения фраз из естественного или искусственного языка. Лингвистические переменные также можно представить в терминах нечетких предикатов, как совокупность предикатов $\{x_1(s), x_2(s), x_3(s)...\}$, имеющих одну область определения S . Кроме того, лингвистическая переменная включает в себя процедуры создания новых предикатов на основе ранее определенных.

Пример. Задача о чаевых. Требуется вычислить размер чаевых в зависимости от оценки уровня обслуживания и качества пищи в ресторане в баллах [1, 10] (пример адаптирован из [8]).

Уровень обслуживания задается «лингвистической» переменной $y(s) = \{y^{(1)}(s), y^{(2)}(s), y^{(3)}(s)\}$, где $y^{(k)}(s)$ – нечеткие предикаты, описывающие низкий ($k = 1$), средний ($k = 2$) и высокий ($k = 3$) уровень обслуживания. Функции $y_1^{(k)}(s) = \mu_k(s)$ заданы в [6]. Одна из них – $y_1^{(1)}(s)$ приведена на рис. 2. Аналогично заданы лингвистические переменные «качество пищи» $z(f) = \{z^{(1)}(f), z^{(2)}(f), z^{(3)}(f)\}$ и $w(p) = \{w^{(1)}(p), w^{(2)}(p), w^{(3)}(p)\}$ – уровень оплаты, где переменная f (оценка качества пищи в баллах) принадлежит множеству $[1, ..., 10]$, а переменная p задает процент чаевых от стоимости заказа, причем $p \in [0 ... 30]\%$.

Клиент оценивает в баллах уровень обслуживания $s = s_0$ и качество пищи $f = f_0$. Необходимо найти процент чаевых p_0 , соответствующий заданным значениям s_0, f_0 .

Уровень оплаты находится с помощью алгоритма Мамдани. На первом шаге вычисляем переменные

$$v(s_0, f_0, p) = \quad (17)$$

$$= \bigvee_{k=1}^3 \{[y^{(k)}(s_0) \vee z^{(k)}(f_0)] \wedge w^{(k)}(p)\},$$

в результате дефаззификации находим процент чаевых по формуле

$$p_0 = \frac{\int_0^{30} p \cdot v_1(s_0, f_0, p) dp}{\int_0^{30} v_1(s_0, f_0, p) dp}. \quad (18)$$

Подразумевается, что все вычисления в (17) производятся в матричном виде.

Решение задачи реализовано в среде MathCAD, на языке C++ в среде C++ Builder, на языке C на микропроцессоре Silicon Laboratories Inc C8051F912.

Процессор C8051F912 имеет жестко ограниченные системные ресурсы, в частности, память (доступный объем RAM составляет 768 байт и Flash – 16 Кб). С целью сравнения, на нем был реализован алгоритм нечеткого вывода Мамдани с «классической» («по Заде») и матричной реализацией логических операций.

При реализации операций «по Заде» используются функции нахождения минимального и максимального значения, а при реализации матричного представления операций – функции нахождения истинностной компоненты матричной конъюнкции и дизъюнкции по приведенным выше формулам.

Программный код, полученный после компиляции, для функций матричных конъюнкции и дизъюнкции занимает примерно на 10 % меньше объема памяти, чем для функций нахождения максимального и минимального значения для алгоритма Заде. Учитывая, что рассматриваемые функции входят в основные программные циклы нечеткого логического вывода и выполняются многократно, выигрыш во времени исполнения программы будет тем большим, чем большее количество правил нечеткого вывода используется в алгоритме и чем больше количество точек, задающих функции принадлежности (в реализованных алгоритмах число итераций пропорционально каждому из этих параметров).

В целом реализация решения задачи показала, что матричный метод имеет более наглядный, компактный вид, удобен при разработке. Готовые программы и исходные коды можно найти на ресурсе [10].

Примеры нечеткого вывода

Люди в очках. Требуется оценить степень уверенности в высказывании о том, что человек пользуется очками (или линзами). Пусть известно, что

1) очками чаще всего пользуются люди, которые много читают;

2) очками также пользуются люди, которые имеют врожденную предрасположенность;

3) кроме этого очками пользуются люди, которые много времени проводят за компьютером;

4) много читают люди, которые любят читать;

5) много читают те люди, которые занимаются наукой;

6) если человек читает за компьютером, то он почти гарантированно станет близоруким.

Обозначим понятия, описанные в базе знаний, следующим образом:

Glas – человек пользуется очками;

Comp – человек так или иначе пользуется компьютером;

Read – человек любит читать;

MRead – человек много читает;

Congen – человек имеет врожденную предрасположенность к болезням глаз;

Scient – человек занимается наукой;

zk ($k = 0 \dots 6$) – вероятность выполнения соответствующего правила.

Тогда в математической форме записи база знаний будет выглядеть следующим образом:

$$MRead \rightarrow Glas = z1. \quad (19)$$

$$Congen \rightarrow Glas = z2. \quad (20)$$

$$Comp \rightarrow Glas = z3. \quad (21)$$

$$Scient \rightarrow MRead = z4. \quad (22)$$

$$Read \rightarrow MRead = z5. \quad (23)$$

$$Read \wedge Comp \rightarrow Glas = z6. \quad (24)$$

Исходные переменные Congen, Comp, Scient и Read заданы в виде нечетких векторов:

$$\begin{aligned} Scient &= \begin{pmatrix} 0,95 \\ 0,05 \end{pmatrix}, \quad Comp = \begin{pmatrix} 0,3 \\ 0,7 \end{pmatrix}, \\ Read &= \begin{pmatrix} 0,7 \\ 0,3 \end{pmatrix}, \quad Congen = \begin{pmatrix} 0,97 \\ 0,3 \end{pmatrix}, \end{aligned} \quad (25)$$

а переменные zk :

$$\begin{aligned} z1 = z5 &= \begin{pmatrix} 0,8 \\ 0,2 \end{pmatrix}, \quad z2 = \begin{pmatrix} 0,98 \\ 0,02 \end{pmatrix}, \quad z3 = \begin{pmatrix} 0,5 \\ 0,5 \end{pmatrix}, \\ z4 &= \begin{pmatrix} 0,96 \\ 0,04 \end{pmatrix}, \quad z6 = \begin{pmatrix} 0,4 \\ 0,6 \end{pmatrix}. \end{aligned} \quad (26)$$

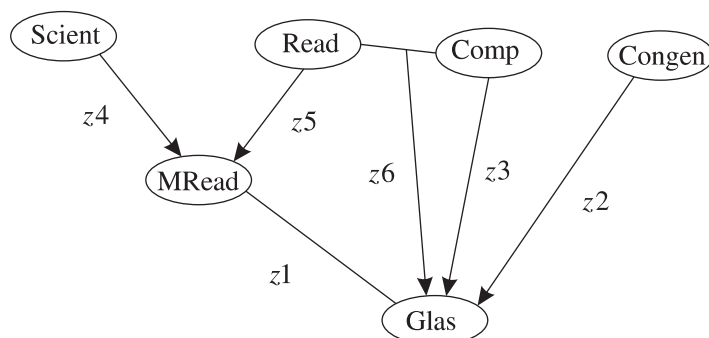


Рис. 3. Схема правил задачи «Люди в очках» в терминах исходных переменных

Схема правил задачи в терминах введенных переменных выглядит, как показано на рис. 3.

Для того чтобы сделать вывод относительно вектора Glas, нужно прежде получить вектор MRead. Перепишем уравнение (22) в матричном виде:

$$I(\text{Scient}) \cdot \text{MRead} = z4. \quad (27)$$

Решение этого уравнения в соответствии с (7) и (9) имеет вид

$$\text{MRead1} = \begin{pmatrix} -0,94 \\ 1,94 \end{pmatrix}. \quad (28)$$

Результат выходит за пределы, допустимые условием (3). Это говорит о том, что логический вывод по этому правилу в таких исходных данных невозможен. Решим аналогично уравнение для правила (25):

$$\text{MRead2} = \begin{pmatrix} 0,333 \\ 0,667 \end{pmatrix}. \quad (29)$$

Итоговое значение вектора MRead находится дизъюнкцией результатов всех правил, на которых вывод оказался успешным. В нашем случае такое правило только одно – (23), значит $\text{MRead} = \text{MRead2}$.

Далее проводим расчеты для вектора Glas. Решение уравнений (19)–(21) и (24) аналогично уравнениям (22) и (23). Результат имеет следующий вид:

$$\begin{aligned} \text{Glas1} &= \begin{pmatrix} 0,7 \\ 0,3 \end{pmatrix}, \text{Glas2} = \begin{pmatrix} 0,333 \\ 0,667 \end{pmatrix}, \\ \text{Glas3} &= \begin{pmatrix} 0,286 \\ 0,714 \end{pmatrix}, \text{Glas4} = \begin{pmatrix} -0,791 \\ 1,791 \end{pmatrix}. \end{aligned} \quad (30)$$

Решение уравнения (25) (Glas4) выходит за рамки условий (3), его необходимо отбросить.

Теперь, чтобы получить итоговое значение

вектора Glas, найдем дизъюнкцию всех результирующих векторов:

$$\text{Glas} = \bigvee_{k=1}^3 \text{Glas}_k = \begin{pmatrix} 0,067 \\ 0,933 \end{pmatrix}. \quad (31)$$

Таким образом, степень уверенности в утверждении, что человек носит очки при указанных исходных данных, равна 0,93.

Формирование управленческих решений.

Требуется определить, инвестиции в какую сферу деятельности в следующем периоде наиболее необходимы на основе некоторых ключевых показателей деятельности в области инновационного развития в текущем периоде. Имеются данные по следующим показателям деятельности [9]:

- число новых продуктов
(NewProd $\in [0..20]$);
- число новых клиентов
(NewClient $\in [0..100000]$);
- процент узнаваемости бренда
(PersBrand $\in [0..100]$);
- индекс удовлетворенности клиентов
(PersClient $\in [0..100]$);
- доля брака (PersReject $\in [0..100]$);
- затраты на НИОКР
(ValNIOKR $\in [0..100000]$);
- затраты на обучение персонала
(ValPers $\in [0..100000]$);
- затраты на проведение рекламы
(ValAdvrt $\in [0..100000]$).

Аналитики составили базу знаний следующего вида:

- 1) если доля брака высока и индекс удовлетворенности клиентов низкий, то затраты на НИОКР средние;
- 2) если доля брака средняя, то затраты на обучение персонала средние;
- 3) если число новых продуктов низкое, то затраты на НИОКР высокие;

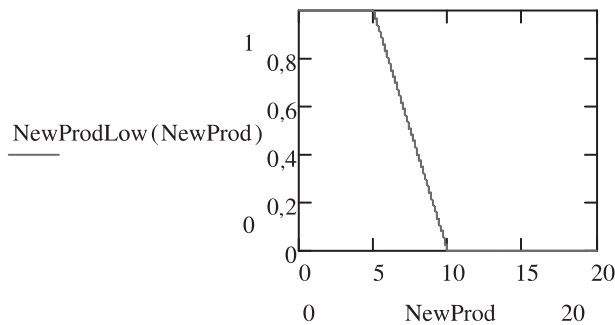


Рис. 4. Внешний вид функции принадлежности нечеткого предиката «количество новых продуктов низкое»
NewProd – количество новых продуктов предприятия в текущем периоде

4) если число новых продуктов высокое, то затраты на НИОКР низкие;

5) если процент узнаваемости бренда низкий, то затраты на проведение рекламы высокие;

6) если процент узнаваемости бренда высокий, то затраты на проведение рекламы низкие;

7) если затраты на НИОКР высокие и доля брака высокая, то затраты на обучение персонала высокие;

8) если затраты на проведение рекламы высокие и процент узнаваемости бренда низкий, то затраты на НИОКР высокие;

9) если число новых клиентов низкое, то затраты на проведение рекламы высокие.

Зададим предикаты всех нечетких понятий, используемых в базе знаний. Например, функция истинности предиката «количество новых продуктов низкое» в зависимости от количества новых продуктов может иметь вид как на рис. 4.

Аналогичным образом могут быть представлены функции истинности остальных предикатов, используемых в базе знаний (рис. 5). При подстановке конкретных исходных значений показателей деятельности можно вычислить значение всех предикатов и, соответственно, предпосылок всех правил в виде векторов. Также необходимо задать идеальные предикаты для искомых величин. Их конкретное значение в начале решения неизвестно и зависит от «четкого» значения искомой величины, которое предстоит определить.

Логический вывод проводится решением уравнения (7) для каждого правила в системе аналогично тому, как это показано в задаче «Люди в очках».

Следует отметить, что искомые предикаты (затраты на НИОКР, проведение рекламных акций, обучение персонала) вычисляются для сле-

дующего периода и не должны пересекаться с аналогичными предикатами для текущего периода. В результате вывода получатся нечеткие вектора для искомых величин, истинностная компонента которых показывает, во сколько раз реальная искомая величина отличается от идеальной:

$$\begin{aligned} \text{CoefNIOKRLow} &= \begin{pmatrix} 0,68 \\ 0,32 \end{pmatrix}; \\ \text{CoefNIOKRMid} &= \begin{pmatrix} 0,16 \\ 0,84 \end{pmatrix}; \\ \text{CoefNIOKRHigh} &= \begin{pmatrix} 0,89 \\ 0,11 \end{pmatrix}. \end{aligned} \quad (32)$$

Реальный искомый предикат находится конъюнкцией вектора-коэффициента и идеального искомого предиката:

$$\begin{aligned} \text{NIOKRLow}(p) &= \\ &= C(\text{CoefNIOKRLow}) \cdot \text{NIOKRLow}(p). \end{aligned} \quad (33)$$

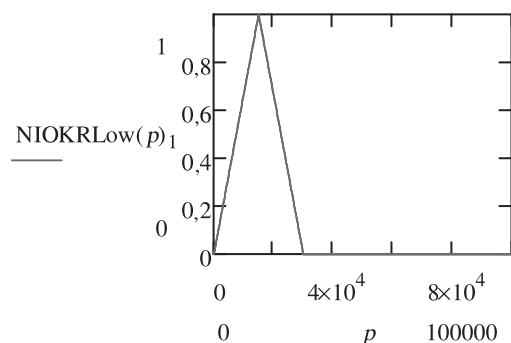


Рис. 5. Внешний вид функции принадлежности нечеткого предиката «финансирование НИОКР в следующем периоде низкое»
p – размер финансирования НИОКР в следующем периоде, у. е.

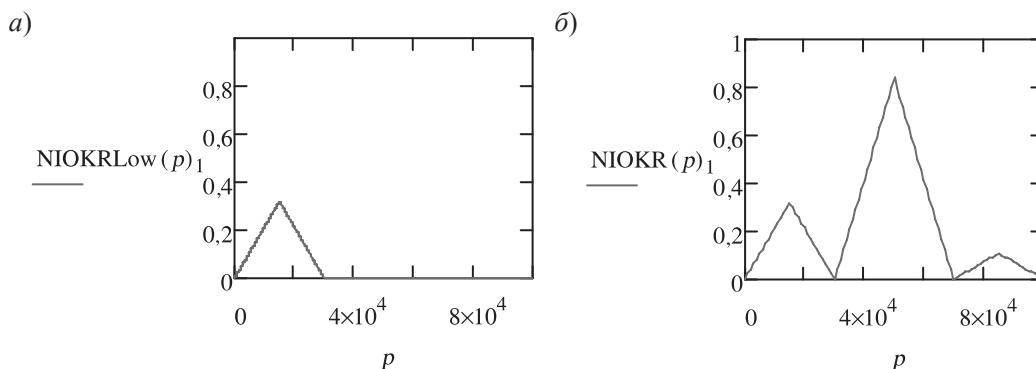


Рис. 6. Внешний вид функции принадлежности нечеткого предиката «финансирование НИОКР в следующем периоде низкое» (а); внешний вид функции принадлежности композированного предиката «финансирование НИОКР в следующем периоде» (б)
 p – размер финансирования НИОКР в следующем периоде, у. е.

После того как все искомые реальные предикаты найдены, а те из них, для которых условие (3) не выполняется, отброшены, можно объединить реальные предикаты разных правил (рис. 6 а) по одной искомой величине в один дизъюнкцией (рис. 6 б). В результате, например для НИОКР получается функция следующего вида (см. рис. 6).

Дефазгификация производится центроидным методом, подробно описанным в [7]. Результат получился в следующем виде:

$$\text{Pers} = 24900; \text{Advrt} = 43460; \text{NIOKR} = 45260. \quad (34)$$

Более подробную информацию по проекту можно найти в Интернете [10].

Использование матричного подхода позволяет свести решение задач нечеткой логики к решению системы линейных алгебраических уравнений, что дает возможность быстро и наглядно реализовывать сложные алгоритмы управления,

требующие логического вывода. В виду того что матричный подход естественным образом обобщает четкий вывод, области существования решения соответствуют информации, заложенной в базу знаний в виде правил *modus ponens*, *modus tollens* и др.

Матричный аппарат легко реализуется на микропроцессорных устройствах в условиях ограниченных системных ресурсов, в частности, оперативной памяти. В связи с этим есть возможность его применения в широком классе приборов на основе микропроцессоров.

Также данный метод применим в системах поддержки решений, где максимально точное использование имеющейся базы знаний особенно актуально.

В дальнейшем планируется применить матричный метод для реализации нечетких агентов и автоматов, экспертных систем, что позволит решать более сложные задачи управления и поддержки решений.

СПИСОК ЛИТЕРАТУРЫ

1. **Пегат, А.** Нечеткое моделирование и управление [Текст] / А. Пегат; Пер. с англ. – М.: Бином, 2009. – 798 с.
2. **Zadeh, L.A.** Fuzzy Sets [Text] / L.A. Zadeh // Information and Control. – 1965. – Vol. 8. – P. 338–353.
3. **Марценюк, М.А.** Матричное представление нечеткой логики [Текст] / М.А. Марценюк // Нечеткие системы и мягкие вычисления. Научный журнал Российской ассоциации нечетких систем и мягких вычислений. – 2007. – Т. 2. – № 3. – С. 7–35.
4. **Игошин, В.И.** Математическая логика и теория алгоритмов: Учеб. пособие для студ. вузов [Текст] / В.И. Игошин; – 2-е изд., стер. – М.: Изд. центр «Академия», 2008. – 448 с.
5. **Рыжов, А.П.** Элементы теории нечетких множеств и измерения нечеткости [Текст] / А.П. Рыжов. – М.: Диалог-МГУ, 1998.
6. **Konar, A.** Computational Intelligence: Principles [Text] / A. Konar // Techniques and Applications. – Springer, 2005. – 71 p.
7. **Круглов, В.В.** Интеллектуальные информационные системы: компьютерная поддержка систем нечеткой логики и нечеткого вывода [Текст] / В.В. Круглов, М.И. Длин. – М.: Физматлит, 2002. – 256 с.

8. Fuzzy Logic Toolbox User's Guide [Электронный ресурс] / Режим доступа: <http://www.mathworks.com/>

9. Толкач, В. КРІ в управлении: связь со стратегией [Текст] / В. Толкач // Новый менеджмент.

–2007. –№ 5.

10. Проект «матричная реализация нечеткой логики» [Электронный ресурс] / Режим доступа: <http://iseletkov.perm.ru/project/fuzzylogic>

УДК 681.3.06(075.8)

И.Г. Черноруцкий

АЛГОРИТМИЧЕСКИЕ ПРОБЛЕМЫ ЖЕСТКОЙ ОПТИМИЗАЦИИ

На практике часто возникает ситуация неудовлетворительного поведения стандартных методов нелинейной оптимизации при решении реальных прикладных задач. Как правило, это выражается в резком замедлении сходимости применяемых поисковых процедур, а в ряде случаев – в полной остановке алгоритма задолго до достижения оптимальной точки (ситуация «ложного» локального экстремума). В англоязычной литературе указанное явление носит название *джемминга* (Jamming) или *залипания*.

Возникновение подобных трудностей связано со специальной формой плохой обусловленности матрицы вторых производных минимизируемых функционалов, приводящей к характерной *жесткой* (*овражной*) структуре поверхностей уровня. В данной статье указываются причины частого появления жестких экстремальных задач в практических вычислениях.

Некоторые выводы и утверждения интерпретируются на примере стандартных задач теории управления. Однако, как правило, обсуждаемые проблемы имеют инвариантный, общесистемный характер и могут представлять интерес для специалистов из различных проблемных областей компьютерного моделирования. Рассматриваются задачи безусловной (unconstrained – UC) оптимизации.

1. Явление жесткости

Рассмотрим следующий пример критерия оптимальности, зависящего от двух управляемых параметров x_1, x_2 [1]:

$$J(x_1, x_2) = g_0^2(x_1, x_2) + \sigma g_1^2(x_1, x_2) \rightarrow \min_x, \quad (1)$$

где σ – достаточно большое положительное число.

Рассмотрим также уравнение

$$g_1(x_1, x_2) = 0, \quad (2)$$

определяющее в простейшем случае некоторую зависимость $x_2 = \varphi(x_1)$. Тогда при стремлении параметра σ к бесконечности значение функционала J в каждой точке, где $g_1(x_1, x_2) \neq 0$, будет неограниченно возрастать по абсолютной величине, оставаясь ограниченным и равным $g_0^2(x_1, x_2)$ во всех точках на кривой $x_2 = \varphi(x_1)$. То же самое будет происходить с нормой вектора градиента $J'(x) = [\partial J/\partial x_1, \partial J/\partial x_2]$, где

$$\partial J/\partial x_1 = 2g_0(x_1, x_2) \partial g_0/\partial x_1 + 2\sigma g_1(x_1, x_2) \partial g_1/\partial x_1,$$

$$\partial J/\partial x_2 = 2g_0(x_1, x_2) \partial g_0/\partial x_2 + 2\sigma g_1(x_1, x_2) \partial g_1/\partial x_2.$$

Линии уровня $J(x) = \text{const}$ для достаточно большого σ представлены на рис. 1 и имеют характерную *овражную структуру*. Там же стрелками показано векторное поле антиградиентов, определяющее локальные направления наискорейшего убывания $J(x)$.

Ясно, что при достаточно больших σ минимальные значения $J(x)$ следует искать вдоль зависимости $x_2 = \varphi(x_1)$, определяющей т. н. *дно оврага*. Из (1) следует, что изменение $J(x)$ вдоль дна зада-

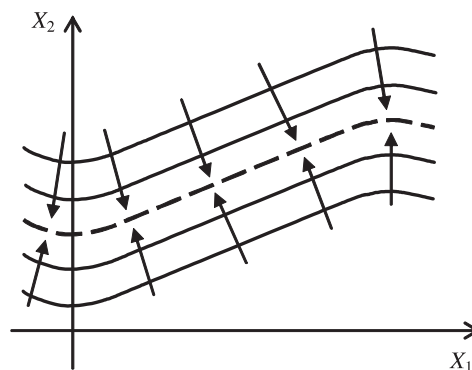


Рис. 1. Явление овражности

ется выражением $g_0^2(x_1, x_2)$ и не зависит от величины параметра σ . Таким образом, задача минимизации $J(x)$ сводится к минимизации функционала $g_0^2(x_1, \varphi(x_1))$ от одной переменной x_1 . В общем случае уравнение $x_2 = \varphi(x_1)$ обычно неизвестно.

Приведенный пример *овражной ситуации* является простым, хотя и из него уже видны принципиальные трудности, связанные, например, с применением широко распространенных методов спуска по антиградиенту. Действительно, из рис. 1 следует, что направления поиска, задаваемые антиградиентами, оказываются неэффективными. Приводя достаточно быстро процесс поиска на дно оврага, они в окрестности дна начинают осциллировать, оставаясь почти перпендикулярными направлению в точку минимума.

Возможны различные усложнения и обобщения рассматриваемой ситуации. Например, уравнение (2) может, вообще говоря, определять не одно, а несколько решений $x_2 = \varphi_i(x_1)$, каждое из которых означает свой «овраг» или «долину». С точки зрения приложений более существенным обобщением является предположение о наличии многомерного дна оврага. Чтобы проиллюстрировать это явление, обратимся к следующему примеру:

$$J(x) = g_0^2(x) + \sigma \sum_{i=1}^m g_i^2(x), x \in R^n, n > 2, m < n. \quad (3)$$

В этом случае дно оврага задается системой уравнений

$$g_i(x) = 0, i \in [1:m], \quad (4)$$

что в принципе позволяет выразить m параметров x_i через оставшиеся $n - m$ переменных. Предположим, не ограничивая общности, что уравнения (4) определяют следующие зависимости:

$$\begin{aligned} x_1 &= \varphi_1(x_{m+1}, \dots, x_n), \\ &\dots \dots \dots \\ x_m &= \varphi_m(x_{m+1}, \dots, x_n). \end{aligned}$$

Аналогично предыдущему случаю устанавливаем, что задача минимизации (3) при достаточно больших σ эквивалентна минимизации функции $g_0^2(\varphi_1, \dots, \varphi_m, x_{m+1}, \dots, x_n)$ от $r = n - m$ переменных $x_{m+1}, \dots, x_n$. Число r носит название *размерности дна оврага*. Легко представить, что, например, для $n = 3$ поверхности уровня одномерного оврага имеют характерный «сигарообразный» вид, а для двумерного оврага они будут близки к деформированным дискам.

В любом случае для овражной ситуации определяющим фактором является специальная структура поверхностей уровня $J(x)$, весьма сильно

отличающаяся от сферической. Характерно наличие некоторой области притяжения (по сути – дна оврага) $Q \subset R^n$, содержащей оптимальную точку $x^* = \arg \min J(x)$. При этом норма вектора градиента $J'(x)$ для $x \in Q$, как правило, существенно меньше, чем в остальной части пространства.

Овражную структуру могут иметь не только функционалы вида (1), (3), явно содержащие большой параметр σ . Можно привести следующий пример квадратичного функционала:

$$f(x_1, x_2) = 0,250025x_1^2 + 0,49995x_1x_2 + 0,250025x_2^2 - x_1 - x_2. \quad (5)$$

Линии уровня $f(x) = \text{const}$ функционала (5) представляют семейство подобных эллипсоидов с центром в точке (1:1). Длины полуосей эллипсоидов относятся при этом как $1:10^2$. Указать большой параметр σ в выражении (5) нельзя, хотя овражная ситуация налицо, и так же как и в предыдущих случаях явно выделяется дно оврага (прямая ab на рис. 2), имеющее уравнение $x_2 = 2 - x_1$. Подставляя выражение для x_2 в (5), снова приходим к эквивалентной задаче меньшей размерности:

$$f_1(x_1) = 10^{-4}(x_1 - 1)^2 - 1 \rightarrow \min_{x_1}.$$

Необходимость выделения овражных оптимизационных задач в отдельный класс обусловлена, с одной стороны, значительными вычислительными трудностями при их решении стандартными для компьютерного моделирования методами, а, с другой стороны, бесспорным фактом важности данного класса задач для большинства практических ситуаций. Специалисты по моделированию сталкивались с овражной ситуацией уже на заре современной компьютерной эры, когда компью-

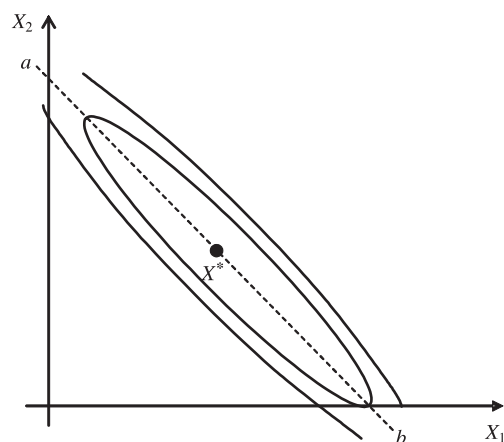


Рис. 2. Овражный квадратичный функционал



теры стали регулярно использоваться при решении реальных задач. Приведем лишь некоторые ранние свидетельства специалистов, подтверждающие тезис о типичности овражной ситуации:

[2], с. 226: «... с увеличением размерности задачи возрастает вероятность появления оврагов»;

[3], с. 353: «... большинство практических задач многопараметрической оптимизации, особенно из области оптимального проектирования, страдает обилием такого рода ловушек (то есть оврагов)»;

[4], с. 272: «... в экстремальных задачах проектирования необходимо использовать методы оптимизации, приспособленные для поиска экстремума, в овражных ситуациях»;

[5], с. 161: «... особенностью целевых функций при решении задач схемотехнического проектирования является их гребневый (овражный при поиске минимума) характер, что приводит к большим вычислительным трудностям»;

[6], с. 35: «Наибольшие трудности при поиске локального оптимума доставляют так называемые «овражные» ситуации».

Аналогичные утверждения сделаны также во многих других работах, связанных с компьютерным моделированием. Сейчас известны различные методы, ориентированные на решение рассматриваемого круга оптимизационных задач, однако и в настоящее время проблема минимизации овражных функционалов актуальна. Особенно остро стоит вопрос минимизации овражных и одновременно невыпуклых функционалов, т. к. именно в этой ситуации отказывается большинство методов поисковой оптимизации. Далее, в соответствии с основополагающими результатами работы [1], проблема овражности рассматривается с позиций явления жесткости. Последнее позволяет конструктивно вскрыть основные особенности рассматриваемого класса задач и построить теорию жесткой оптимизации.

2. Основные определения

Пусть решается задача $J(x) \rightarrow \min$, $J \in C^2(D)$, $x \in D \subset R^n$. Будем предполагать далее, что функционал $J(x)$ ограничен снизу на D .

Траектория наискорейшего спуска (ТНС) $x(\tau)$ функционала $J(x)$ задается векторным дифференциальным уравнением

$$\frac{dx}{d\tau} = -J'(x), J'(x) = \left[\frac{\partial J}{\partial x_1}, \dots, \frac{\partial J}{\partial x_n} \right]. \quad (6)$$

Эти траектории в случае овражных функционалов обладают рядом специфических черт. Например, для функционала (5) имеем

$$x(\tau) = \sum_{i=1}^2 \{ \alpha_i^* + (\alpha_i^0 - \alpha_i^*) \exp(-\lambda_i \tau) \} u_i, \quad (7)$$

где

$$x(0) = \sum_{i=1}^2 \alpha_i^0 u_i, x^* = (1, 1) = \sum_{i=1}^2 \alpha_i^* u_i,$$

$$\lambda_1 = 1, \lambda_2 = 10^{-4}, u_1 = \frac{1}{\sqrt{2}}(1, 1), u_2 = \frac{1}{\sqrt{2}}(-1, 1).$$

Из (7) видно, что из-за наличия быстро затухающей и медленно затухающей экспонент отчетливо выделяются два участка с существенно различным поведением решения. Первый, сравнительно непродолжительный, характеризуется большими значениями производных $dx_i(\tau)/d\tau$ и означает спуск на дно оврага. На дне выполняются условия типа (2) и норма вектора градиента, а с ней и производные $dx_i(\tau)/d\tau$ становятся относительно малыми. Поэтому для второго участка характерно относительно плавное изменение переменных x_1, x_2 . Таким образом, прослеживается полная аналогия с поведением решений т. н. жестких систем обыкновенных дифференциальных уравнений. В связи с этим, следуя [1], дадим следующее общее определение.

Определение 1. Функционал $J(x)$ называется жестким (овражным), если отвечающая ему система дифференциальных уравнений (6) – жесткая.

Однако с точки зрения приложений к области оптимизации более конструктивным оказывается приведенное ниже определение жесткого функционала. Это определение не содержит, в частности, таких неестественных для задач оптимизации требований, как необходимость задания промежутка интегрирования уравнения (6) [7].

Определение 2. Функционал $J(x) \in C^2(D)$, $D \subset R^n$ называется жестким в множестве $Q \in D$, если найдутся такие числа $\sigma > 0$, $\sigma \gg 1$ и множество $Q_\delta \subset D$, что

$$\begin{aligned} 1) & \forall x \in Q_\delta, \lambda_1[J''(x)] \geq \sigma |\lambda_n[J''(x)]|; \\ 2) & \forall x \in Q, \arg \min_{x' \in X_\delta(x)} J(x') \subset Q; \\ 3) & \forall x \in Q, L[X_\delta(x) \cap Q] \leq \sigma^{-1} L[X_\delta(x)], \end{aligned} \quad (8)$$

где $X_\delta(x) = \{x' \in R^n \mid \|x' - x\| \leq \delta\}$; $Q_\delta = \bigcup_{x \in Q} X_\delta(x)$; $\lambda_i[A]$ – собственные числа матрицы $A = J''(x)$, упорядоченные по убыванию.

доченные по убыванию $\lambda_1[A] \geq \lambda_2[A] \geq \dots \geq \lambda_n[A]$; $L[S]$ – константа. Липшица в соотношении $\|J'(x') - J'(x)\| \leq L[S] \|x' - x\|, \forall x', x \in S \subset R^n$. Множество Q называется *дном оврага*.

Основным в определении 2 является условие 1, констатирующее резко несимметричное расположение спектра матрицы вторых производных $J''(x)$ относительно начала координат: $\lambda_i \in [-m, M], M \gg m > 0$. Условия 2, 3 необходимы для описания свойства «устойчивости» множества Q : можно показать, что все ТНС, начинавшиеся в любой точке $x \in Q_\delta$, быстро попадают в достаточно малую окрестность Q_ε ($\varepsilon < \delta$) множества Q и остаются там до выхода из множества Q_δ .

Как правило, для приложений оказывается достаточной более грубая модель явления жесткости, когда предполагается, что собственные числа матрицы вторых производных можно отчетливо разделить на две группы, в одну из которых входят собственные числа, по модулю намного превосходящие элементы второй группы. Вводится следующее определение [7].

Пусть в $D \subset R^n$ задана r -мерная поверхность (конфигурационное пространство) $Q = \{x \in D | g_i(x) = 0; i \in [1:n-r]\}$, $g_i \in C^2(D)$.

Определение 3. Функционал $J(x) \in C^2(D)$, $D \subset R^n$ называется *жестким на множестве Q* , если найдутся такие числа $\delta > 0$, $\sigma \gg 1$ и множество $Q_\delta \subset D$, что

- 1) $\forall x \in Q_\delta, \lambda_1[J''(x)] \geq \dots \geq \lambda_{n-r}[J''(x)] \geq \sigma |\lambda_{n-r+1}[J''(x)]| \geq \dots \geq \sigma |\lambda_n[J''(x)]|$;
- 2) $\forall x \in Q, \arg \min_{x \in X_\delta(x)} J(x') \subset Q$;
- 3) $\forall x \in Q, L[X_\delta(x) \cap Q] \leq \sigma^{-1} L[X_\delta(x)]$.

Число r называется размерностью оврага (дна оврага) Q .

Пример. Рассмотрим квадратичный функционал:

$$f(x) = 1/2 \langle Ax, x \rangle - \langle b, x \rangle + c, c = \text{const.} \quad (10)$$

Пусть собственные числа $\lambda_i(A)$, $i \in [1:n]$ удовлетворяют неравенствам (9) (при $\forall x \in R^n$). Построим множество Q , являющееся в силу определения 3 дном оврага. Предположим, без существенного ограничения общности, что $\det A \neq 0$ и обозначим через x^* решение уравнения $Ax = b$. Тогда

$$f'(x) = Ax - b = A(x - x^*) = \sum_{i=1}^n \beta_i \lambda_i u_i, \quad (11)$$

где $x - x^* = \sum_{i=1}^n \beta_i u_i$.

Положим

$$Q = \{x \in R^n | \beta_i = \langle x - x^*, u_i \rangle = 0, i \in [1:n-r]\}. \quad (12)$$

Тогда, очевидно, $Q_\delta = R^n$, $\sigma \cong \lambda_1 / |\lambda_{n-r+1}|$ и условие 1 выполнено.

Обратимся к проверке выполнения условия 2. Рассмотрим сферическую окрестность $x_\delta = \{x \in R^n | \|x - \bar{x}\| \leq \delta\}$ некоторой точки $\bar{x} =$

$$= \sum_{i=n-r+1}^n \alpha_i u_i + x^* \in Q. \text{ Для произвольного } x = \sum_{i=1}^n \beta_i u_i + x^* \in R^n \text{ условие } x \in X_\delta \text{ эквивалентно тре-}$$

$$\text{бованию } \left\| \sum_{i=1}^{n-r} \beta_i u_i + \sum_{i=n-r+1}^n (\beta_i - \alpha_i) u_i \right\| \leq \delta \text{ или } \beta = (\beta_1, \dots, \beta_n) \in S_\beta = \left\{ \beta \in R^n \left| \sum_{i=1}^{n-r} \beta_i^2 + \sum_{i=n-r+1}^n (\beta_i - \alpha_i)^2 \leq \delta^2 \right. \right\}.$$

$$\text{Имеем } f(x) = \frac{1}{2} \sum_{i=1}^n \beta_i^2 \lambda_i + c_1 = \varphi(\beta), \text{ где}$$

c_1 не зависит от β . Задача поиска $x \in X_\delta$, минимизирующего $f(x)$, сводится к следующей задаче условной оптимизации

$$\varphi(\beta) \rightarrow \min_\beta, g(\beta) = \sum_{i=1}^{n-r} \beta_i^2 + \sum_{i=n-r+1}^n (\beta_i - \alpha_i)^2 - \delta^2 \leq 0.$$

Условие стационарности для функции Лагранжа $L(\mu_1, \mu_2, \beta) = \mu_1 \varphi(\beta) + \mu_2 g(\beta)$ имеет вид $L'_\beta = \mu_1 \varphi'(\beta) + \mu_2 g'(\beta) = 0$ или $\beta_i (\mu_1 \lambda_i + 2\mu_2) = 0$, $i \in [1:n-r]$, $\mu_1 \beta_j \lambda_j + 2\mu_2 (\beta_j - \alpha_j) = 0, j \in [n-r+1:n]$, причем $\mu_i \geq 0$ и не равны нулю одновременно.

Отсюда, в силу положительности «больших» собственных чисел λ_i ($i \in [1:n-r]$), с необходимостью следует $\beta_i = 0$ ($i \in [1:n-r]$). Поэтому

$$\min_{\beta \in S_\beta} \varphi(\beta) = \min_{\beta \in S_\beta} \bar{\varphi}(\bar{\beta}), \quad (13)$$

где

$$\bar{\beta} = (\beta_{n-r+1}, \dots, \beta_n); \bar{\varphi}(\bar{\beta}) = \frac{1}{2} \sum_{i=n-r+1}^n \beta_i^2 \lambda_i + c_1;$$

$$\bar{S}_\beta = \left\{ \gamma = (\gamma_1, \dots, \gamma_r) \in R^r \left| \sum_{i=1}^r (\gamma_i - \alpha_{n-r+i})^2 \leq \delta^2 \right. \right\}.$$

Из (13) следует условие 2. Для

$$x' = \sum_{i=1}^n \beta'_i u_i + x^*, x'' = \sum_{i=1}^n \beta''_i u_i + x^*, x', x'' \in X_\delta$$

$$\text{имеем } \|J'(x') - J'(x'')\| = \|A(x' - x'')\| = \left\| \sum_{i=1}^n (\beta'_i - \beta''_i) \lambda_i u_i \right\| \leq \max_i \lambda_i \left\| \sum_{i=1}^n (\beta'_i - \beta''_i) u_i \right\| = \lambda_1 \|x' - x''\|.$$

Аналогично, для $\forall \bar{x}', \bar{x}'' \in X_\delta \cap Q$ получим $\|J'(\bar{x}') - J'(\bar{x}'')\| \leq |\lambda_{n-r+1}| \cdot \|\bar{x}' - \bar{x}''\|$.



Поэтому $L(X_\delta) = \lambda_1 \gg L(X_\delta \cap Q) = |\lambda_{n-r+1}|$ и условие 3 выполнено при $\sigma = \lambda_1 / |\lambda_{n-r+1}|$.

Таким образом, для квадратичных функционалов при сдвинутом в точку x^* начале координат дно оврага Q совпадает с линейной оболочкой (12) собственных векторов, отвечающих «малым» собственным числам. Это согласуется с интуитивными представлениями, развитыми в разделе 1.

На этом же примере можно проиллюстрировать роль требований 1–3 в определениях 2 и 3. Действительно, для квадратичного функционала (10) условия 1, 2 могут выполняться для всего пространства $Q = R^n$ и любого $\sigma > 0$. Необходимая линейная оболочка собственных векторов может быть выделена только при дополнительном требовании, эквивалентном условию 3. В то же время требования 1 и 3 без 2 также оказываются недостаточными. В этом случае сдвиг линейной оболочки Q , являющейся дном оврага, вдоль любого из не вошедших в оболочку собственных векторов, не приведет к нарушению условий 1 и 3, а условие 2 при этом нарушится.

Рассмотренные выше модели явления овражности не являются исчерпывающими. Однако они описывают наиболее существенные стороны большинства практических ситуаций, связанных с решением задач параметрической оптимизации.

Определение 4. Пусть $\forall x \in Q, \det J''(x) \neq 0$. Наименьшее из чисел σ , удовлетворяющих определению 2, называется *степенью жесткости* $J(x)$ в Q и обозначается $\eta(Q)$. Отношение $\eta(x) = \lambda_1(x) / \left| \min_i \lambda_i(x) \right|^{-1}$, $x \in Q$ называется *локальной степенью жесткости* в точке x . Для вырожденных матриц $J''(x)$ величина $\eta(x)$ полагается равной ∞ .

Если $J''(x) > 0$, то $\eta(x) = \text{cond}[J''(x)] = \max_i \lambda_i(x) / \min_i \lambda_i(x)$.

В общем случае справедливо неравенство $1 \leq \eta \leq \text{cond}[J''(x)]$. При наличии больших по модулю отрицательных собственных чисел $\lambda_i(x)$, т. е. при отсутствии овражной ситуации, возможно неравенство $\eta(x) \ll \text{cond}[J''(x)]$. Следовательно, из высокой степени жесткости $J(x)$ в точке x следует плохая обусловленность матрицы $J''(x)$, а обратное неверно. Действительно, пусть спектр матрицы $J''(x)$ расположен в множестве $[-M, -m] \cup [m, M]$, $M \gg m > 0$, включая граничные значения. Тогда $\eta(x) = 1$, а $\text{cond}[J''(x)] = M/m \gg 1$. Данный функционал не будет относиться к классу

жестких, что естественно, ибо трудностей при его минимизации, например, методом наискорейшего спуска, не возникает.

Отличие между двумя характеристиками $\eta(x)$ и $\text{cond}[J''(x)]$ функционала $J(x)$ часто игнорируется, и овражными называются функционалы с большим числом $\text{cond}[J''(x)]$. Подобная точка зрения не оправдывается с позиций основных вычислительных трудностей, возникающих при решении экстремальных задач. Однако, учитывая указанную выше связь между $\eta(x)$ и $\text{cond}[J''(x)]$, жесткие задачи часто называются *плохо обусловленными экстремальными задачами*.

В каждом конкретном случае различные значения $\eta(x)$ следует считать большими. Здесь существует аналогия с понятием плохой обусловленности матрицы. В большинстве случаев все определяется точностью вычислений и типом применяемого алгоритма оптимизации. Традиционно принято классифицировать задачу как плохо обусловленную, если

$$\log_2 \eta > t, \quad (14)$$

где t – длина используемой разрядной сетки компьютера в представлении чисел с плавающей точкой. Однако и при меньших значениях η для целого ряда алгоритмов могут возникать значительные вычислительные трудности, особенно если жесткость сопровождается отсутствием выпуклости $J(x)$.

Дополнительным фактором, характеризующим степень сложности экстремальной задачи и затрудняющим применение традиционных алгоритмов минимизации, является наличие многомерных оврагов с $r > 1$. В указанной ситуации целый ряд методов, специально ориентированных на решение плохо обусловленных задач, становятся неэффективными.

3. Критерии жесткости

Рассмотрим практические методы распознавания овражной ситуации, играющие роль *критериев жесткости*. Наиболее существенной характеристикой оказывается значение показателя η в допустимой области изменения управляемых параметров.

Своеобразным индикатором может служить *метод простого градиентного спуска* (ПГС), реализуемый по схеме

$$x_{k+1} = x_k - hJ'(x_k), \quad (15)$$

с постоянным шагом $h \in R^1$.

Принадлежность $J(x)$ к классу жестких в этом случае проявляется в необходимости применения относительно малых значений шага h . Попытки увеличения h вызывают потерю свойства релаксационности (монотонного убывания) последовательности значений функционала и последние начинают резко возрастать. Если для некоторого фиксированного h (наибольшего из возможных) удалось заставить процесс (1) протекать без полной остановки, то по результатам работы метода можно количественно оценить величину η .

Для этого процесс (1) продолжается до тех пор, пока отношение $\frac{\|J'(x_{k+1})\|}{\|J'(x_k)\|}$ не стабилизируется около некоторого значения μ . Тогда справедливо следующее равенство:

$$\eta \cong 2 / |1 - \mu|. \quad (16)$$

Соотношение (16) справедливо независимо от выпуклости функционала $J(x)$ и является основным для грубой практической оценки степени жесткости решаемой задачи в окрестности текущей точки. Доказательство соотношения (16) дано в [8].

В силу изложенного выше можно рекомендовать процесс оптимизации начинать с помощью метода ПГС. Если задача простая и степень жесткости невелика, то уже этот метод достаточно быстро приведет в малую окрестность оптимума. В противном случае будет получена оценка η , что позволит правильно оценить ситуацию и выбрать наиболее рациональный алгоритм.

Другой, прямой метод оценки η сводится к вычислению матрицы Гессе функционала и решению для нее полной проблемы собственных значений. Тогда на основе непосредственной проверки выполнения неравенств (9) для вычисленных собственных чисел делается вывод о значении η . При этом определяется также размерность r дна оврага. Главный недостаток такого подхода заключается в существенных вычислительных трудностях принципиального характера, возникающих при определении малых собственных значений. Можно показать, что абсолютная погрешность $|d\lambda_i|$ представления любого собственного значения λ_i матрицы A за счет относительного искажения δ ее элементов удовлетворяет неравенству $|d\lambda_i| \leq n\delta|\lambda_i|$, где $|\lambda_i| = \max_i |\lambda_i|$. Полагая $\delta = \varepsilon_m = 2^{-t}$, где ε_m — относительная машинная точ-

ность, а t — длина разрядной сетки мантиссы числа, получим оценку для абсолютных искажений собственных чисел за счет ошибок округления:

$$|d\lambda_i| \leq n\varepsilon_m |\lambda_i|. \quad (17)$$

Параметр ε_m известен для каждой вычислительной системы.

Из последнего неравенства можно сделать следующее заключение. Если все вычисленные собственные числа матрицы $A = J''(x)$ достаточно велики, то есть $|\lambda_i| \geq n\varepsilon_m |\lambda_i|$, то параметр η может быть вычислен непосредственно. Если некоторые из вычисленных собственных чисел удовлетворяют неравенству $|\lambda_i| \leq n\varepsilon_m |\lambda_i|$, то все они должны быть отнесены к блоку «малых» собственных чисел, а для η имеем границу снизу: $\eta \geq 1/(n\varepsilon_m)$.

Качественным признаком плохой обусловленности может служить существенное различие в результатах оптимизации, например, методом ПГС, при спуске из различных начальных точек. Получаемые результирующие точки обычно расположены далеко друг от друга и не могут интерпретироваться как приближения к единственному решению или конечной совокупности решений (при наличии локальных минимумов). Описанная ситуация, как правило, означает наличие оврага, а точки остановки применяемой поисковой процедуры трактуются как элементы дна оврага Q .

4. Источники жестких оптимизационных задач

Несмотря на то, что типичность жесткой ситуации может считаться установленным экспериментальным фактом, представляет определенный интерес выяснение на качественном уровне основных причин появления оврагов в задачах управления. Рассмотрим два типа жесткости — т. н. *естественную жесткость* и *внесенную жесткость*. В первом случае подразумевается, что задачи параметрической оптимизации могут оказываться плохо обусловленными в силу естественных причин для каждого из выходных параметров объекта оптимизации, имеющих смысл частных критериев оптимальности. Напротив, внесенная жесткость возникает в связи с применением специальных методов учета ограничений типа штрафных функций и модифицированных функций Лагранжа, а также в связи с образованием обобщенных критериев оптимальности при решении задач многокритериальной оптимизации.

Естественная жесткость. Данный тип жест-

кости можно проиллюстрировать на примерах решения задач идентификации линейных динамических объектов, а также задач синтеза статистически оптимальных систем автоматического управления, приводящих к интегральным уравнениям Винера–Хопфа, имеющим вид (см. [9, 10]):

$$\int_0^T \varphi(\tau - \lambda) \omega(\lambda) d\lambda = R(\tau). \quad (18)$$

Как известно [11], задача решения уравнения Винера–Хопфа является плохо обусловленной, что проявляется в сильной чувствительности решения к малым вариациям исходных функций φ , R , имеющих смысл корреляционных функций и получаемых на практике с ограниченной точностью. Задача оказывается некорректной по Тихонову из-за потери свойства единственности решения [12]. Для построения решения уравнения (18) применяются алгебраические методы, основанные на минимизации невязки.

Пример соответствующего целевого функционала имеет вид:

$$J(x) = \sum_{j=1}^N \left\{ \sum_{i=1}^N \omega_i \varphi[q(j-i)] - R[qj] \right\}^2 \rightarrow \min_x, \quad (19)$$

где $\omega_i = \omega(qi)$, $i \in [1: N]$; $x = (\omega_1, \omega_2, \dots, \omega_N)$.

Свойство некорректности исходной задачи приводит к плохой обусловленности матрицы $J''(x)$ с резко выраженным овражным характером соответствующих поверхностей уровня. Характер решения задачи типа (19) показан на рис. 3. Как отмечается в [13], получаемые таким образом функции $\omega(t)$, как правило, имеют среднюю квадратичную погрешность, близкую к минимальной, однако из-за резко колебательного характера они сильно отличаются от точных решений.

Для ослабления отмеченного эффекта применяются регуляризованные формы целевых функционалов. Простейший критерий с регуляризацией имеет вид

$$J_1(x) = J(x) + \alpha \sum_{j=2}^N [\omega_j - \omega_{j-1}]^2 \rightarrow \min_x, \quad (20)$$

где $\alpha > 0$ – параметр регуляризации, осуществляющий сглаживание получаемого решения.

Существуют и другие формы регуляризации, связанные с расширением исходных целевых функционалов за счет сглаживающих (регуляризующих) функционалов [13]:

$$L[\omega(t)] = M\{E_2(t)\} + \alpha \Omega[\omega(t)],$$

где

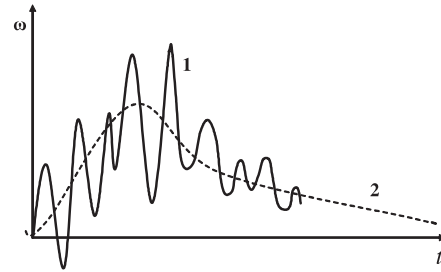


Рис. 3. Характер решения задачи
1 – решение задачи (2);
2 – точное решение уравнения Винера–Хопфа

$$\Omega[\omega(t)] = \int_0^T \left\{ k(t) \left[\frac{d\omega(t)}{dt} \right]^2 + p(t) \omega^2(t) \right\} dt. \quad (21)$$

Условие минимума $L[\omega(t)]$ при $\alpha = 0$ совпадает с уравнением Винера–Хопфа. При $\alpha \neq 0$ получаются уравнения, для решения которых могут строиться функционалы невязки и решаться экстремальные задачи, аналогичные (19) с учетом параметра регуляризации α . Если $\alpha \neq 0$, задача решения «расширенного» уравнения Винера–Хопфа является корректно поставленной и имеет единственное, непрерывное и дифференцируемое решение. Однако решение задач минимизации регуляризованных функционалов, например (20), также сопряжено с известными трудностями, вызванными некорректностью исходной задачи. Действительно, регуляризующие добавки помимо эффекта сглаживания дают искажение решения, приводя к увеличению невязки $J(x)$ [11]. Поэтому с целью получения малых невязок необходимо работать с минимально возможными α , что приводит к высокой степени «остаточной» обусловленности матрицы $J_1''(x)$. Данное обстоятельство, а также отсутствие надежных априорных методов задания α , приводят к необходимости применения специальных жестко-ориентированных алгоритмов параметрической оптимизации, позволяющих надежно получать гладкие и точные решения. Дополнительная причина, затрудняющая применение стандартного алгоритмического обеспечения, заключается в факте ухудшения обусловленности задачи с увеличением числа точек дискретности для повышения точности дискретизации при применении алгебраических методов.

В качестве другой, по-видимому, более распространенной причины возникновения жестких целевых функционалов можно указать фактор агрегированности. В большом числе случаев оказывается, что целевой функционал в действи-

тельности зависит не от n независимых переменных $x_1, x_2, \dots, x_n$, а от s агрегатов:

$$J(x) = I(z_1, z_2, \dots, z_s), \quad (22)$$

где

$$z_i = \varphi_i(x), i \in [1:s]; s < n. \quad (23)$$

Как показано, например, в [7], прямое решение задачи минимизации целевого функционала $J(x)$ в пространстве параметров x связано с наличием овражной ситуации. Действительно, значения $J(x)$ мало меняются на множестве, определяемом равенствами

$$\varphi_i(x) = z_i^*, i \in [1:s], \quad (24)$$

где z_i^* – оптимальные значения агрегатов, доставляющих минимум целевому функционалу $I(z) = I[\varphi(x)] = J(x)$. Поэтому уравнения (24) фактически являются уравнениями дна оврага.

Образование агрегатов можно пояснить на примере описания динамических свойств оптимизируемого объекта. Распространенные методы синтеза линейных оптимальных систем управления основаны на построении передаточных функций замкнутых систем, обладающих необходимыми свойствами (т. н. *этап аппроксимации*). Далее, на этапе *реализации* строятся модели реальных систем с конкретной структурой, имеющих заданные (полученные на этапе аппроксимации) передаточные характеристики.

Таким образом, процедура синтеза распадается на две стадии: *синтез в пространстве коэффициентов* передаточных функций, имеющих смысл агрегированных переменных, и *синтез в пространстве «физических» параметров* x_p , подбираемых из условия (24).

Агрегаты обычно фигурируют как коэффициенты уравнений, описывающих определенные динамические свойства объекта. При этом количество исходных переменных x_p , составляющих действительный вектор x варьируемых параметров, обычно оказывается значительно большим.

Весьма важным обстоятельством является тот факт, что в подавляющем числе случаев при оптимизации *нелинейных* систем, а также при отсутствии явных представлений для соответствующих математических моделей, мы вынуждены проводить оптимизацию непосредственно в пространстве параметров элементов оптимизируемой системы. Последнее, согласно изложенному выше, подразумевает минимизацию жестких

функционалов. Заметим, что иногда и при синтезе линейных систем необходимо осуществлять поиск в пространстве x из-за трудностей в задании множества достижимости Z при сложных функциональных ограничениях на компоненты вектора x .

В действительности обычно реализуется следующая зависимость $J(x) = I(z_1, z_2, \dots, z_s) + \varepsilon q(x)$ обобщенного показателя качества $J(x)$ от агрегатов z_i и вектора исходных переменных x . Наличие «объединяющего» функционала $g(x)$, с одной стороны, позволяет говорить лишь о приближенном агрегировании с точностью до малого параметра ε , а, с другой, при соответствующих конструкциях $g(x)$ решает проблему единственности решения в пространстве переменных x .

Высока вероятность появления жестких функционалов из-за образования т. н. «асимптотических» агрегатов при оптимизации динамических объектов, описываемых в пространстве состояний *жесткими системами* обыкновенных дифференциальных уравнений с большим разбросом значений основных постоянных времени.

Возникновение жестких дифференциальных уравнений при моделировании реальных динамических систем определяется разбросом временных характеристик, заложенным в самой их физической природе. Примерами могут служить задачи электротехники, химической кинетики, моделирования сложных систем путем формального объединения описаний различных по инерционности подсистем. Например, в задачах управления электроэнергетическими системами жесткие дифференциальные уравнения возникают из-за резко различающихся по скорости протекания быстрых электромагнитных процессов и медленных электромеханических [1].

Высокие значения чисел обусловленности матрицы Якоби дифференциальных уравнений динамики характерны для моделей гироскопических устройств, а также различных типов электронных схем.

Пример. Пусть функционирование объекта описывается вектор-функцией

$$u(t) = [u_1(t) + u_2(t)], \quad (25)$$

$$u_1(t) = t[x_1 \exp(-At) + (x_1 + x_2) \exp(-at)],$$

$$u_2(t) = t[x_1 \exp(-Bt) + (x_1 + x_2) \exp(-bt)],$$

являющейся при $A \gg a > 0, B \gg b > 0$ решением жесткой системы уравнений.

Требуется получить оптимальные значения параметров x_1, x_2 из условия наилучшего совпадения $u(t)$ с заданной вектор-функцией $\bar{u}(t)$ для $t \in [t_0, T]$. Если предположить, что $t_0 > \tau_{nc}$, где τ_{nc} — длина пограничного слоя, определяющего практически полное затухание экспонент $\exp(-At), \exp(-Bt)$, то из (9) видно, что поведение решения $u(t)$ для $t \in [t_0, T]$ будет определяться агрегатом $z = x_1 + x_2$. Если продолжать считать x_1 и x_2 независимыми параметрами, то степень жесткости, например, следующего квадратичного целевого функционала

$$J(x) = [u_1(t_1) - \bar{u}_1(t_1)]^2 + [u_2(t_1) - \bar{u}_2(t_1)]^2 \rightarrow \min_x$$

будет достаточно высока. При $t_1 = 1, a = 1, b = 1,5, A = 20, B = 15$ получим

$$\eta \cong \frac{\exp[-\min(a, b)]}{\exp[-\max(A, B)]} \cong 1,8 \cdot 10^8.$$

Таким образом, по крайней мере до тех пор пока не разработаны достаточно универсальные регулярные методы выделения агрегатов для последующей декомпозиции исходной задачи, мы вынуждены считаться с необходимостью оптимизации в пространстве переменных x_i по жестким целевым функционалам.

Внесенная жесткость. 1. Учет ограничений. Высокая степень жесткости может быть внесена в задачу оптимизации в силу используемой конструкции обобщенного критерия оптимальности. Рассмотрим эффект жесткости, возникающий при использовании методов штрафных функций и модифицированных функций Лагранжа. Изучим в качестве примера ограничения в виде равенств $g_j(x) = 0, j \in [1:p]$. Тогда согласно методу штрафных функций задача минимизации функционала $J(x)$ сводится к минимизации вспомогательных функционалов $J_0(x, \sigma) = J(x) + \sigma \sum_{j=1}^p g_j^2(x) \rightarrow \min_x$

с большим положительным коэффициентом σ . При этом структура расширенного критерия J_0 , содержащего большой параметр σ , как правило, оказывается жесткой, даже если исходный функционал $J(x)$ этим свойством не обладает.

Пример. Пусть требуется найти x_1, x_2 , минимизирующие квадратичный функционал $f(x) = x_1^2 + x_2^2$ при условии $x_1 = 2$. Задача имеет очевидное решение $x_1^* = 2, x_2^* = 0$. Поступим формально и составим вспомогательный функционал согласно общей рецептуре метода штрафных функций:

$$J_0(x) = x_1^2 + x_2^2 + \sigma(x_1 - 2)^2 = (x_1 - b_1)^2 / a_1^2 + (x_2 - b_2)^2 / a_2^2 + d,$$

где $a_1 = 1 / \sqrt{1 + \sigma}; a_2 = 1; b_1 = 2\sigma / (\sigma + 1); b_2 = 0; d = 4\sigma / (\sigma + 1)$.

Уравнение линии уровня $J_0(x) = \text{const}$ является уравнением эллипса с центром в точке (b_1, b_2) и длинами полуосей, относящимися как $a_2 / a_1 = \sqrt{1 + \sigma}$. При больших значениях σ , обеспечивающих относительно точное выполнение ограничения $x_1 = 2$, линии уровня оказываются сильно вытянутыми (рис. 4). Чем точнее выполняются ограничения, тем ярче выражен эффект овражности. В данном случае степень овражности равна $\eta = 1 + \sigma$ и стремится к бесконечности при $\sigma \rightarrow \infty$. Заметим, что линии уровня исходного функционала являются сферами и явление овражности отсутствует.

На практике метод штрафных функций широко используется на начальных этапах оптимизации с применением таких, возможно больших, значений σ , для которых удастся достигнуть относительно быстрого убывания $J_0(x)$ при точном выполнении ограничений. Для последующего улучшения решения привлекаются более тонкие стратегии, которые, как правило, оказываются и

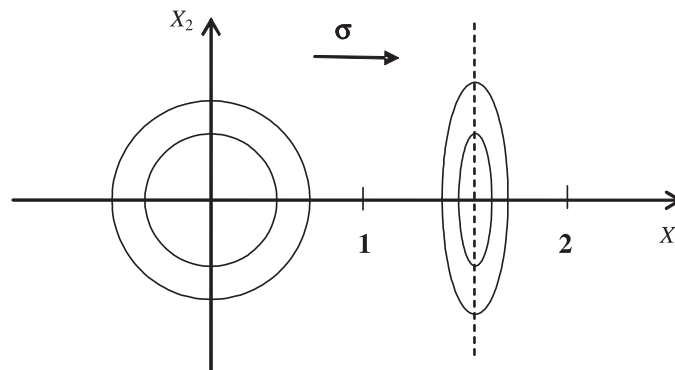


Рис. 4. Метод штрафных функций

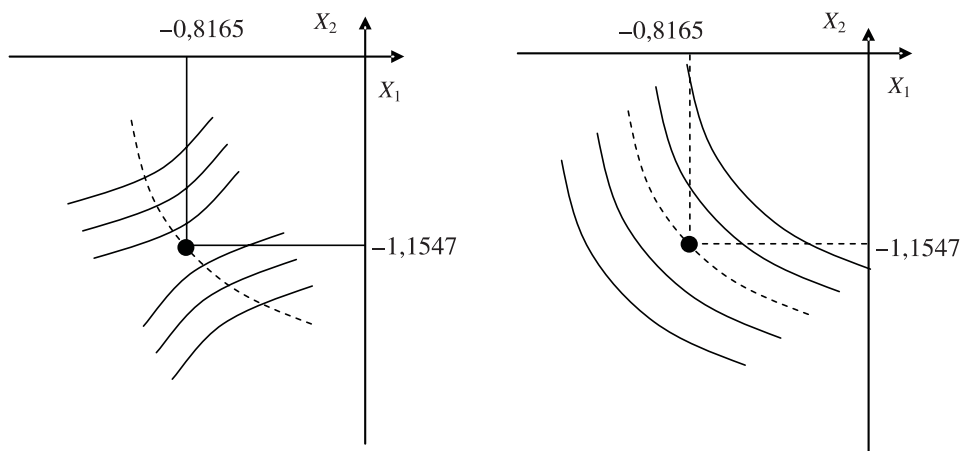


Рис. 5. Линии уровня модифицированной функции Лагранжа: $a - \sigma = 0,075$; $b - \sigma = 100$

существенно более трудоемкими. Кроме этого, метод штрафных функций до сих пор не имеет разумных альтернатив в ряде критических ситуаций, характерных для прикладных задач параметрической оптимизации.

Например, при наличии вырожденного минимума, когда не выполняется условие линейной независимости градиентов ограничений $g_i'(x)$, могут потерять работоспособность все методы учета ограничений, основанные на обычной и модифицированной функции Лагранжа, а также на линеаризации ограничений. Метод штрафных функций в указанной ситуации применим. Он оказывается наименее чувствительным ко всем формам вырождения.

Второй критической ситуацией, возникающей на практике, является несогласованность требований-спецификаций к объекту оптимизации, которая приводит к пустому множеству D допустимых значений управляемых параметров. Предположим, что решается задача с ограничениями-равенствами. В этом случае снова наиболее целесообразным образом ведет себя метод штрафных функций, позволяя получить такое решение задачи $\|g(x)\| \rightarrow \min$, для которого величина $J(x)$ минимальна. Другие методы либо теряют смысл, либо заведомо не будут сходящимися [14].

В силу изложенного выше наличие алгоритмов оптимизации, сохраняющих работоспособность при высокой степени жесткости минимизируемых функционалов оказывается чрезвычайно желательным. Этот вывод подтверждается также тем фактом, что и при использовании модифицированных функций Лагранжа мы сталкиваемся с той же проблемой, хотя и в несколько ослаблен-

ной форме. Как показано в [15], выбор параметра σ в модифицированных функциях Лагранжа весьма сильно влияет на обусловленность соответствующей канонической задачи. При этом как слишком малые, так и слишком большие σ приводят к жестким функционалам. Если учесть, что надежные рекомендации по априорному заданию σ отсутствуют, то становится совершенно ясно необходимость применения специальных жестко-ориентированных процедур для решения вспомогательных канонических задач.

Пример. Решается задача [15]

$$J(x) = x_1 x_2^2 \rightarrow \min, x \in D,$$

$$D = \{x \in R^n \mid g(x) = 2 - x_1^2 - x_2^2 = 0\}.$$

На рис. 5 показаны линии уровня вспомогательного функционала $M(x^*, \sigma) = J(x) + \lambda g(x) + \frac{\sigma}{2} g^2(x)$ при $\lambda = \lambda^* \cong 0,82$; $\sigma \cong 0,2$.

В обоих случаях линии уровня практически параллельны, что и определяет наличие высокой степени жесткости.

2. Объединение конфликтных выходных параметров. Учет многокритериальных требований к объекту оптимизации с помощью единого критерия оптимальности также является важнейшим фактором, обуславливающим возникновение овражной жестких обобщенных критериев оптимальности. Известные принципы построения Парето-оптимальных решений приводят к двум основным видам свертки: линейной и минимаксной (максиминной). Остановимся на минимаксной свертке двух частных критериев:

$$J(x) = \max \{\alpha_1 J_1(x), \alpha_2 J_2(x)\} \rightarrow \min_x,$$

$$\alpha_i > 0, \alpha_1 + \alpha_2 = 1.$$

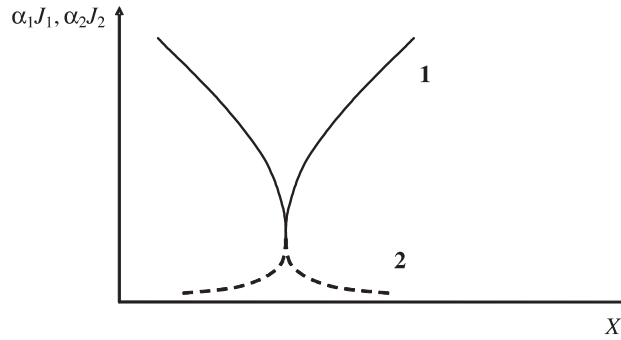


Рис. 6. Объединение частных критериев: 1 – α_{1J_1} ; 2 – α_{2J_2}

Отдельные критериальные выходные параметры, как функции от входных параметров, могут иметь нежесткий характер. Однако и в этом случае их объединение почти неизбежно приводит к возникновению высоких степеней жесткости. При этом крутые «склоны» оврага характеризуют доминирующее влияние на обобщенный критерий какого-то одного из «частных» критериев. Как следует из рис. 6, объединение критериев J_1, J_2 приводит к образованию «клювообразной» зависимости, порождающей в многомерном случае овраг с крутыми склонами. Характерно, что движение по любой поверхности $\alpha_i J_i$ в отдельности с помощью практически любых методов оптимизации может не вызывать никаких затруднений.

В качестве второго примера рассмотрим критерии оптимальности, построенные на основе функционалов вида

$$J(x) = \sum_{i=1}^N \alpha_i f_i(x), f_i(x) = \left[\sum_{j=1}^n x_j \phi_j(t_i) - y_i \right]^2, \quad (26)$$

где $y = y(t_i)$; $\{\phi_j, j \in [1:n]\}$ – заданная система функций. Задача $J(x) \rightarrow \min$ ставится при решении задач аппроксимации

$$y(t) \approx \sum_{j=1}^n x_j \phi_j(t), 0 \leq t \leq T, \quad (27)$$

характерных для многих разделов компьютерного моделирования. Например, необходимость в применении параметризации (27) возникает в задачах сглаживания экспериментальных зависимостей при идентификации стохастических объектов, а также в задачах идентификации нелинейных объектов на основе функциональных рядов Вольтер-

ра. Уже обсуждавшиеся выше алгоритмические методы решения уравнения Винера–Хопфа также приводят к задачам минимизации функционалов метода наименьших квадратов.

Функционал $J(x)$ можно рассматривать как линейную свертку функционалов $f_i(x)$, характеризующих частные критерии оптимальности, $f_i(x) \rightarrow \min_x, i \in [1:N]$. Поэтому трудности минимизации (26) в какой-то степени отражают общую ситуацию, связанную с применением линейных сверток при большом числе частных критериев. В наиболее употребительном на практике случае используется полиномиальная аппроксимация на основе системы базисных функций вида $\phi_j(t) = t^{j-1}, 1 \leq j \leq n$.

Полагая N достаточно большим и заменяя сумму интегралом

$$J(x) = \sum_{i=1}^N \alpha_i f_i(x) \approx \int_0^T \alpha(t) \left[\sum_{j=1}^n x_j \phi_j(t) - y(t) \right]^2 dt,$$

получим при $\alpha(t) \equiv 1$:

$$J''(x) = H_n, H_n = \{h_{ij}\}, h_{ij} = 1 / (i + j - 1), i, j \in [1:n],$$

где $H_n - (n \times n)$ – матрица Гильберта, являющаяся также матрицей линейной системы т. н. *нормальных уравнений* $J'(x) = 0$. Спектральное число обусловленности положительно определенной матрицы H_n , совпадающее со степенью жесткости квадратичного сильно выпуклого функционала $J(x)$, оказывается достаточно высоким: матрицы Гильберта являются стандартным примером очень плохо обусловленных матриц. При этом $\text{cond}[H_n]$ быстро растет с увеличением n :

n	2	3	4	5
$\text{cond}[H_n]$	$1,93 \cdot 10^1$	$5,24 \cdot 10^2$	$1,55 \cdot 10^4$	$4,77 \cdot 10^5$
n	6	7	8	9
$\text{cond}[H_n]$	$1,50 \cdot 10^7$	$4,75 \cdot 10^8$	$1,53 \cdot 10^{10}$	$4,93 \cdot 10^{11}$

В результате уже для небольших n минимизация (26) происходит в условиях сильно выраженной жесткости.

Таким образом, в статье обоснована актуальность задачи минимизации жестких функционалов, характерных для различных проблем компьютерного моделирования и оптимизации. Построены

модели жестких функционалов и указаны причины и механизмы их возникновения при проведении реальных вычислений. Представленные результаты могут быть положены в основу построения специальных жестко-ориентированных оптимизирующих процедур и выбора рациональных сценариев оптимизации при работе с соответствующими библиотеками стандартных программ.

СПИСОК ЛИТЕРАТУРЫ

1. **Ракитский, Ю.В.** Численные методы решения жестких систем [Текст] / Ю.В. Ракитский, С.М. Устинов, И.Г. Черноуцкий. –М.: Наука, 1979.
2. **Моисеев, Н.Н.** Элементы теории оптимальных систем [Текст] / Н.Н. Моисеев. –М.: Наука, 1975.
3. **Растрингин, Л.А.** Системы экстремального управления [Текст] / Л.А. Растрингин. –М.: Наука, 1974.
4. **Норенков, И.П.** Введение в автоматизированное проектирование технических устройств и систем [Текст] / И.П. Норенков. –М.: Высш. школа, 1986.
5. **Норенков, И.П.** Экстремальные задачи при схемотехническом проектировании в электронике [Текст] / И.П. Норенков, С.Г. Мулярчик, С.Р. Иванов. –Минск: Изд-во БГУ, 1976.
6. **Геминтерн, В.И.** Методы оптимального проектирования [Текст] / В.И. Геминтерн, Б.М. Каган. –М.: Энергия, 1980.
7. **Черноуцкий, И.Г.** Оптимальный параметрический синтез: электротехнические устройства и системы [Текст] / И.Г. Черноуцкий. –Л.: Энергоатомиздат, 1987.
8. **Черноуцкий, И.Г.** Методы оптимизации в теории управления [Текст] / И.Г. Черноуцкий. –СПб.: Питер, 2004.
9. **Черноуцкий, И.Г.** Параметрические методы синтеза систем управления [Текст] / И.Г. Черноуцкий // Научно-технические ведомости СПбГПУ. –2009. –№ 2. –С. 111–115.
10. **Черноуцкий, И.Г.** Методы параметрической оптимизации в задачах идентификации [Текст] / И.Г. Черноуцкий // Научно-технические ведомости СПбГПУ. –2009. –№ 2. –С. 150–155.
11. **Растрингин, Л.А.** Введение в идентификацию объектов управления [Текст] / Л.А. Растрингин, Н.Е. Маджаров. –М.: Энергия, 1977.
12. **Тихонов, А.Н.** Методы решения некорректных задач [Текст] / А.Н. Тихонов, В.Я. Арсенин. –М.: Наука, 1979.
13. **Дейч, А.М.** Методы идентификации динамических объектов [Текст] / А.М. Дейч. –М.: Энергия, 1979.
14. **Поляк, Б.Т.** Введение в оптимизацию [Текст] / Б.Т. Поляк. –М.: Наука, 1983.
15. **Гилл, Ф.** Практическая оптимизация [Текст] / Ф. Гилл, У. Мюррей, М. Райт. –М.: Мир, 1985.

УДК 004.932.2

А.А. Шебалов, А.Н. Баженов

ИССЛЕДОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ МЕТОДОВ ВЫЧИСЛЕНИЯ ОПТИЧЕСКОГО ПОТОКА

Оптический поток – одно из ключевых понятий технического (машинного) зрения. Алгоритмы на его основе позволяют решать большое число важных прикладных задач: оценивание и отслеживание движения, реконструкция трехмерной структуры сцены, сжатие видео (стандарты MPEG), восстановление поврежденных кадров старых видеозаписей и др.

В свете приведенных примеров актуальность задачи нахождения оптического потока, а значит, в свою очередь, сравнения и анализа производительности различных методов, не оставляет сомнений. К сожалению, среди огромного числа публикаций, появившихся за 30 лет развития дисциплины, крайне мал объем обзорно-аналитических работ.

В данной статье представлена методика для оценки и сравнения производительности методов на примере работы с набором тестовых последовательностей изображений и проводится анализ полученных результатов. В математическом пакете Matlab разработан код для сравнения восьми методов и проведены вычислительные эксперименты. Получены статистически состоятельные результаты, произведен их статистический анализ.

1. Проблема вычисления оптического потока

Рассмотрим последовательность изображений (или кадров видео), описывающих динамику некоторой сцены. Согласно определению, предложенному Б.К.П. Хорном [1], *полем движения* назовем двумерную проекцию трехмерного поля физического движения точек поверхностей объектов сцены, а под *оптическим потоком* будем понимать *видимое движение яркостной структуры* изображения.

Активные исследования проблемы вычисления оптического потока начались в конце 70-х гг., а первыми важными результатами их были, ставшие впоследствии классическими, работы Б.К.П. Хорна и Б.Г. Шунка [2], Б.Д. Лукаса и Т. Канаде [3]. В них изложены два базовых дифференциальных подхода к решению задачи и представлены результаты работы соответствующих алгоритмов. В публикации [4] был впервые предложен структурированный подход к тестированию и сравнению различных алгоритмов нахождения оптического потока. Эта наиболее цитируемая в данной области статья придала значительный импульс дальнейшему ее развитию. Следующими работами, объединяющими сведения о новых методах, были статьи [5, 6]. Многие из рассматриваемых авторами методов оказались неудачными, а само сравнение их проводилось на примере работы с очень простыми изображениями.

Точкой отсчета для современного этапа в развитии области стала работа С. Бэйкера [7]. Основываясь на идеях [4], исследователи разработали оригинальный комплексный набор тестовых последовательностей изображений, позволяющий провести объективное сравнение современных методов. Именно эти последовательности используются в данной работе.

2. Методы нахождения оптического потока

Каждый метод вычисления оптического по-

тока можно разделить на две составляющие: модель потока с соответствующим ей энергетическим функционалом и оптимизационную схему, используемую непосредственно для вычисления.

Модели оптического потока. За исключением метода Лукаса–Канаде (далее – LK-модель), все рассматриваемые в статье алгоритмы [8–11] базируются на классической модели Хорна–Шунка [2] (далее – HS-модель). В общем виде их энергетические функционалы можно записать следующим образом:

$$\varepsilon_{\text{Data}} = \sum_{x,y} \psi_1(I(x+w(x)) - I(x)) + \\ + \gamma \cdot \psi_2(\nabla I(x+w(x)) - \nabla I(x))$$

$$\varepsilon_{\text{Prior}} = \sum_{x,y} \psi_3(\nabla w)$$

$$\varepsilon_{\text{General}}(u, v) := \varepsilon_{\text{Data}} + \alpha \cdot \varepsilon_{\text{Prior}},$$

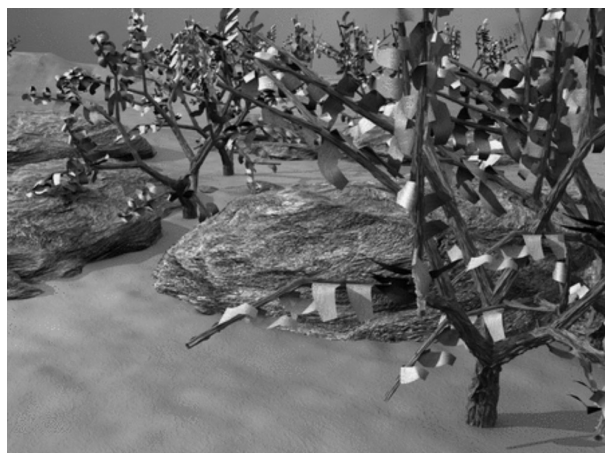
где I – функция изображения; ψ_1, ψ_2, ψ_3 – нормирующие функции; α, γ – весовые коэффициенты. Выбрав, например, $\psi_1(x) = \psi_3(x) = x^2$ и $\gamma = 0$, получим HS-модель.

Оптимизационные техники и детали реализации. Одним из основных препятствий на пути непосредственного применения существующих методов является нестабильность вычислительного процесса оптимизации функционала модели при обработке больших смещений, т. к. в этом случае существует риск того, что итерационный процесс оптимизации сойдется в точке локального минимума [3, 4]. Для борьбы с этой проблемой Лукас и Канаде [3] предложили использовать т. н. *грубо-точный подход* (coarse-to-fine approach), идея которого состоит в построении пирамиды изображений, каждое следующее из которых является субдискретизированной (downsampled) копией предыдущего. Оптимизационная схема на каждом шаге представляет собой составление, линеаризацию и решение системы уравнений Эйлера–Лагранжа [9]. Серьезный недостаток соответствующего алгоритма – низкая степень параллелизма.

Для уменьшения влияния затенения и вариаций освещенности на работу алгоритма, в [11] предложено в процессе предобработки практически полностью удалять структурную компоненту (image structure) входных изображений:

$$I_{\text{Processed}} = I_{\text{Original}} - 0,95 \cdot I_{\text{Structure}}.$$

Другим важным элементом современных ал-



Первый кадр



Эталонное поле

Рис. 1. Тестовая последовательность «Groove 3»

Цветовое кодирование векторов поля потока выполнено согласно представленной на рисунке схеме. Цветовой тон кодирует направление, а насыщенность цвета – длину вектора

горитмов является применение к полю потока различных фильтров после каждого шага грубо-точного метода [8, 11]. Это уменьшает эффект от погрешностей в решениях, полученных итерационными оптимизационными методами. Так, в алгоритме Classic+NL авторы использовали анизотропную фильтрацию, идея которой заключается в том, чтобы суммировать с большим весом значения яркости в тех точках, которые принадлежат тому же структурному элементу изображения, что и фильтруемая точка. Поскольку простой возможности точно определить, одному ли объекту принадлежат две точки, не существует, эту подзадачу решают приближенно. Можно использовать метод Собеля выделения границ, применяя к результату дилатацию с небольшим, например, 5×5 пикселей ядром. Полученную маску можно использовать для определения весов ядра анизотропного фильтра [8].

3. Экспериментальное сравнение методов

Методика сравнения. Для объективного сравнения алгоритмов требуются две компоненты: набор тестовых последовательностей с известными эталонными полями и эффективная метрика для непосредственного количественного сравнения.

Последовательности для сравнения. Сравнение алгоритмов производилось на изображениях, предложенных в [7]. В пользу репрезентативности этого набора говорит наличие в нем как реальных, так и синтетических изображений, за-

шумленных и «чистых» изображений, различных типов движения: и с малым неоднородным смещением и, напротив, со значительным и близким к линейному. Например, синтетическая последовательность «Groove 3» (рис. 1) представляет быстрое смещение камеры относительно статичной сцены, структура центрального объекта которой очень неоднородна и содержит множество мелких деталей.

Полный перечень изображений приведен в табл. 1 и 2.¹

Метрика для сравнения. На основании рекомендаций [4] и [7], выбор был сделан в пользу следующих двух естественных метрик:

средняя угловая ошибка (average angular error – AAE);

средняя оконечная ошибка (average endpoint error – AEE).

Пусть $\vec{v}_{ref}$ – вектор эталонного поля скоростей и $\vec{v}_{est}$ – расчетного. Тогда можно записать формулы для вычисления значений ошибки по введенным метрикам:

$$AAE = \frac{1}{|\Omega|} \sum_{\Omega} \arccos \left(\frac{\vec{v}_{ref} \cdot \vec{v}_{est}}{|\vec{v}_{ref}| |\vec{v}_{est}|} \right),$$

$$AEE = \frac{1}{|\Omega|} \sum_{\Omega} |\vec{v}_{ref} - \vec{v}_{est}|.$$

Алгоритмы для сравнения. Все использованные реализации методов разработаны в математи-

¹ Все они доступны на нашем сайте: <http://vision.middlebury.edu/flow/data/>

Таблица 1

Средняя угловая точность

AAE, °	LK	HS	BA	Brox	CLG	TV-L1	CI++	CI+NL
Dimetrodon	7,13	5,20	4,00	2,70	3,59	3,75	2,48	2,24
Hydrangea	7,95	2,43	2,11	2,17	2,28	3,02	1,84	1,87
Rubber Whale	14,17	4,13	3,15	4,19	5,67	5,07	2,63	2,38
Grove2	6,53	,09	2,39	2,30	2,55	3,14	2,00	1,30
Grove3	13,29	6,69	6,47	6,12	6,78	7,02	6,02	4,89
Urban2	19,96	5,12	2,97	2,89	3,13	2,93	2,55	2,03
Urban3	22,86	6,65	4,44	7,12	6,69	5,36	4,27	2,79

Таблица 2

Средняя оконечная точность

AEE, пикс	LK	HS	BA	Brox	CLG	TV-L1	CI++	CI+NL
Dimetrodon	0,36	0,25	0,19	0,13	0,20	0,19	0,12	0,11
Hydrangea	0,77	0,20	0,18	0,17	0,19	0,24	0,15	0,15
Rubber Whale	0,47	0,13	0,09	0,12	0,16	0,16	0,08	0,07
Grove2	0,50	0,22	0,16	0,15	0,17	0,23	0,13	0,09
Grove3	1,42	0,64	0,64	0,66	0,79	0,65	0,60	0,43
Urban2	4,32	0,51	0,38	0,31	0,42	0,39	0,36	0,20
Urban3	3,97	0,87	0,57	0,82	0,84	0,58	0,53	0,37

Таблица 3

Среднее время работы

	LK	HS	BA	Brox	CLG	TV-L1	CI++	CI+NL
<i>T</i> , с	23	33	103	3,9	64	7,9	116	45

ческом пакете Matlab®. Авторы и источники этих алгоритмов указан в списке литературы [16–20]. Все вычисления проводились в математическом пакете Matlab® 2011a на системе с процессором Intel® Core™ i7-2600K 4.5 GHz под управлением 64-битной ОС Linux.

Сводные результаты тестирования приведены в табл. 1–3.

Анализ результатов. При анализе результатов необходимо учитывать как точность расчетных полей, так и скорость работы соответствующих алгоритмов.

Точность расчетных полей и скорость работы. Наилучшую точность по обоим метрикам (см. п. Метрика для сравнения) на всех последовательностях показал последний в хронологическом порядке алгоритм Classic+NL. Относительно базовой вариационной модели HS ошибка

уменьшилась в среднем на 47 % (2,26°), а для более робастной модели BA – на 33 % (1,15°).

Значительный прирост в точности объясняется тем, что в вычислительную схему включены многие модификации, введенные в других методах. Однако серьезным ограничением данного метода является тот факт, что для функционала используемой модели на данный момент неизвестны способы быстрой минимизации. В частности, предложенная оптимизационная схема является плохо распараллеливаемой. Таким образом, сфера применения алгоритма Classic+NL ограничивается перечнем задач, не имеющих строгих требований по времени (например, сжатие и пост-обработка видео).

Второй алгоритм, предложенный в [8], Classic++ представляет собой модель HS, использующую обобщенную норму [12]:

$$\psi_{GC}^{\alpha}(E) = \sum_{x,y} (E_{x,y}^2 + \epsilon^2)^{\alpha},$$

где α, ϵ – положительные константы. Оптимизация соответствующего энергетического функционала с использованием современных техник позволяет добиться ошибки метода Classic++ в среднем на 18 % (0,61°) большей, чем у оптимизированной версии Classic+NL, но на 15 % (0,81°) меньшей, чем у схожей модели Brox и др. Анализируя этот результат, можно прийти к выводу, что не все используемые в алгоритме Classic++ оптимизационные приемы одинаково полезны. Так, время работы метода Brox в 30 раз меньше (116 против четырех секунд), и, на этом фоне, разница в точности выглядит совсем незначительной. Более того, для модели Brox существует т. н. *мультисеточная схема решения* [13], позволяющая дальнейшее ускорение вычислений.

Алгоритмы CLG и TV-L1 показали точность того же уровня, что и современная реализация классической робастной модели Black и Anandan, что нельзя признать успешным результатом. Тем не менее эти модели уже сыграли важную роль в развитии дисциплины вычисления оптического потока. Так, авторы алгоритма Classic+NL использовали при интегрировании медианной фильтрации в функционал модели тот же принцип, который привел А. Брун к CLG-модели. В свою очередь TV-L1 модель является базовой в одноименном классе моделей, для которых получена очень хорошо распараллеливаемая вычислительная схема.

Классический алгоритм LK показал в тестировании наихудший (и с большим отставанием) результат, но здесь следует сделать важное замечание. Локальный метод в силу своей чрезвычайной простоты применим лишь к очень простым движениям, но при этом реализуем в виде параллельного алгоритма, что позволяет проводить обработку более 100 кадров размера 512×512 в секунду [14]. Отсюда следует, что если вместе с этим методом использовать высокоскоростные камеры, дающие последовательность кадров с совсем небольшими смещениями, то в совокупности получится достаточно точный и, главное, работающий в реальном времени алгоритм.

Скорость работы современных алгоритмов, с учетом того, что во многих приложениях требуется производительность реального времени порядка $T_{RT} = 0,04$ с, следует признать удовлетво-

рительной. Действительно, как следует из табл. 3, для обработки одной пары кадров самому быстрому из рассмотренных методу Brox требуется $T_{Brox} = 3,9 \pm 0,6$ с, а самому точному – Classic+NL – необходимо $T_{Classic+NL} = 45 \pm 8$ с.

Ошибки в результатах и их причины. Для того чтобы определить, в каких случаях у современных алгоритмов возникают проблемы, обратимся к рис. 2, на котором выделены основные ошибки в расчетном поле наиболее точного из исследованных метода Classic+NL на наиболее сложной для обработки (как следует из табл. 1, 2) последовательности «Groove 3».

Во-первых, сразу видно чрезмерное сглаживание расчетного поля. При этом заметим, что резкость контура объекта сохранена, тогда как все внутренние области оказались размытыми. Причиной этого обстоятельства является тип применяемого в алгоритме анизотропного сглаживания (см. п. Свободные результаты тестирования). При использовании во время фильтрации Собелева градиента в качестве фактора, определяющего веса точек, на участках фона между «внутренними» гранями ветки (область, выделенная на рис. 2 справа) взвешивание происходит некорректно. Это связано с небольшим размером самих участков и, напротив, большим размером ядра сглаживающего фильтра – в [8] указан размер 15×15 пикселей.

Во-вторых, на рисунке в верхней части выделенной области можно видеть практически полную потерю малого движения. Эта проблема является общей для всех рассмотренных методов. Причина ее заключается в несовершенстве грубо-точного подхода. При субдискретизации малые структурные элементы «сливаются» как друг с другом, так и с фоном. Появление этих элементов при переходе на менее грубый уровень пирамиды становится причиной явных ошибок интерполяции.

Далее, следует отметить, что наличие в изображениях присущего оптическим приборам шума, а также слаботекстурированных участков не оказывает существенного влияния на результаты. Связать это можно с повсеместным использованием в алгоритмах того или иного вида сглаживания поля, в т. ч. на стадии предобработки.

Оценки точности, полученные на различных последовательностях, коррелируют не очень хорошо. Так, например, для пары последовательностей «Hydrangea» и «Urban 2» ранговый коэффициент корреляции оказался равным 0,57.

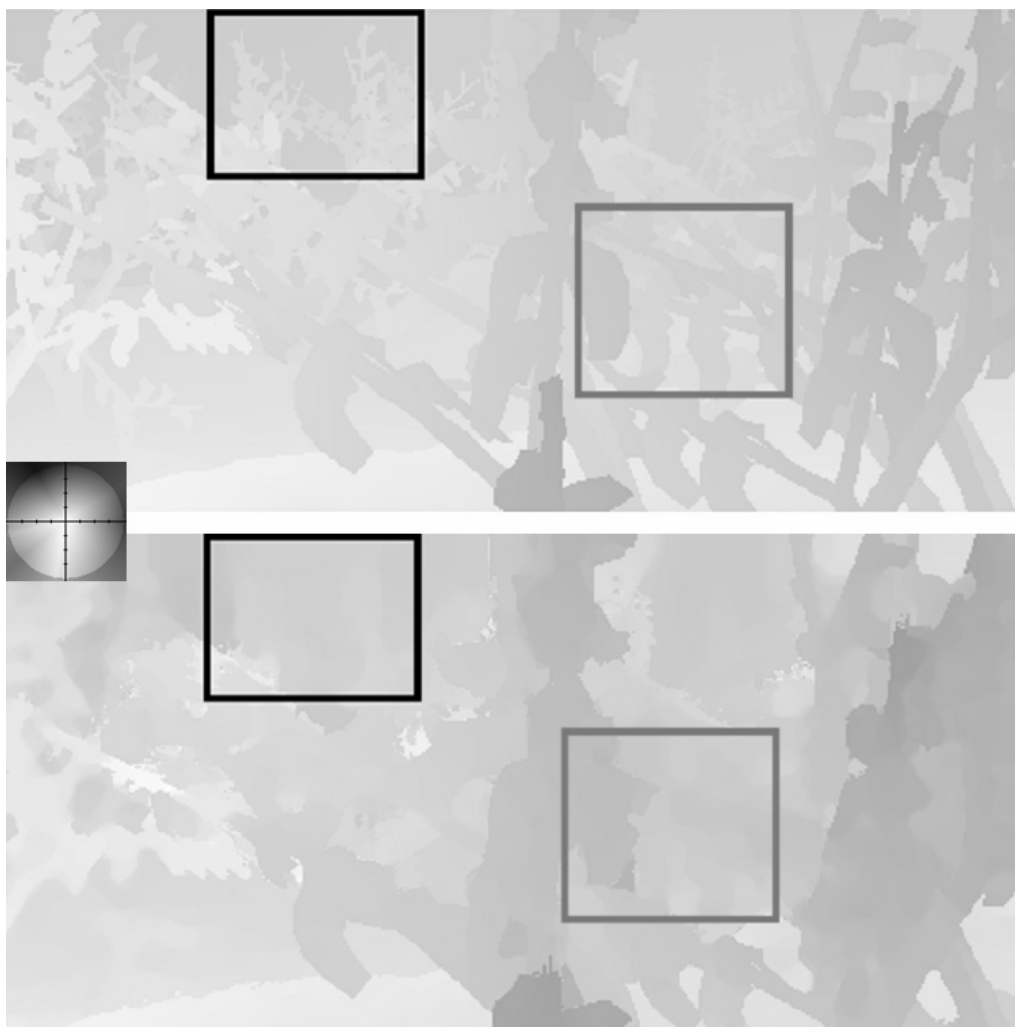


Рис. 2. Ошибки в обработке последовательности «Groove 3»
Сверху – эталонное поле последовательности «Groove 3». Снизу – рассчитанное методом Classic+NL поле.
Меры ошибки: $AAE = 4,9^\circ$, $AEE = 0,43$ пикс.

Принимая во внимание также большой разброс во времени работы среди современных алгоритмов, можно утверждать, что ни один из них не является универсальным.

Дальнейшее расширение тестового набора, вероятно, позволит выявить новые зависимости и недочеты методов.

AAE и AEE метрики. Результаты, полученные с использованием метрик AAE и AEE, хорошо коррелируют с ранговым коэффициентом $R = 0,91$. На этом основании можно сделать вывод о взаимозаменяемости двух этих метрик. Тем не менее Бэйкер и др. [15] в своей работе склоняются к использованию AEE-метрики, т. к. она дает более правдоподобные, по мнению авторов, результаты для полей больших смещений. Отметим, од-

нако, что это лишь качественный, основанный на результатах визуального анализа расчетных полей вывод. Числовые экспериментальные данные, в т. ч. и настоящей работы, такой зависимости не выявляют.

При дальнейшем развитии методологии сравнения исследователям стоит концентрировать внимание не на поиске более эффективных метрик, а на развитии техник применения существующих. Так, например, в [15] предложено усреднять ошибку лишь на границах объектов сцены, а не по всей плоскости изображения, маскируя соответствующие «простые» участки.

В данной статье представлена методика объективного количественного сравнения ал-

горитмов вычисления оптического потока. Для тестирования был разработан набор процедур в математическом пакете Matlab® 2011a.

Проведен экспериментальный анализ производительности восьми методов вычисления оптического потока, в число которых вошли как классические, так и современные алгоритмы, каждый из которых является представительным и/или основополагающим в своем классе. Сравнение их производилось на семи тестовых последовательностях изображений, покрывающих широкий спектр различных типов движения.

Проведенный анализ результатов сравнения позволил выделить достоинства и недостатки современных алгоритмов, определить степень целесообразности применения различных улучшений моделей и указать причины возникновения ошибок.

В заключение следует отметить, что вычисление оптического потока – относительно молодая, а потому интенсивно развивающаяся область компьютерного зрения. Большое число потенциальных приложений определяет актуальность дальнейших исследований и гарантирует интерес научного сообщества.

СПИСОК ЛИТЕРАТУРЫ

1. **Хорн, Б.К.П.** Зрение роботов [Текст] / Б.К.П. Хорн; Пер. с англ. под ред. Е.И. Кугушева, Ю.А. Садова. – М.: Мир, 1989. – С. 281–282.
2. **Horn, B.K.P.** Determining optical flow [Электронный ресурс] / B.K.P. Horn, B.G. Schunck // AI Memo 572, MIT. –1981.
3. **Lucas, B.D.** An iterative image registration technique with an application to stereovision [Электронный ресурс] / B.D. Lucas, T. Kanade // In Proc. of the DARPA Image Understanding Workshop. –1981.
4. **Barron, J.L.** Performance of optical flow techniques [Text] / J.L. Barron, D.J. Fleet, S. Beauchemin // IJCV. – 1994. –№ 12 (1).
5. **Stiller, C.** Estimating motion in image sequences: A tutorial on modelling and computation of 2D motion [Text] / C. Stiller, J. Konrad // IEEE Signal Proc. Mag. –1999. –Vol. 16.
6. **Galvin, B.** Recovering motion fields: An evaluation of eight optical flow algorithms [Электронный ресурс] / B. Galvin, B. McCane, K. Novins [et al.] // In Proc. BMVC. –1998.
7. **Baker, S.** A database and evaluation methodology for optical flow [Электронный ресурс] / S. Baker, D. Scharstein, J. Lewis [et al.] // In Proc. ICCV. –2007.
8. **Sun, D.** Secrets of optical flow estimation and their principles [Электронный ресурс] / D. Sun, S. Roth, M.J. Black // In Proc. IEEE CVPR. –2010.
9. **Brox, T.** High accuracy optical flow estimation based on a theory for warping [Электронный ресурс] / T. Brox, A. Bruhn, N. Papenberg [et al.] // In Proc. ECCV(4). –2004.
10. **Bruhn, A.** Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods [Text] / A. Bruhn, J. Weickert, C. Schnörr // IJCV. –2005. –№ 61(3).
11. **Wedel, A.** An improved algorithm for TV-L1 optical flow [Text] / A. Wedel, T. Pock, C. Zach, H. Bischof [et al.] // In Dagstuhl Motion Workshop. –2008.
12. **Charbonnier, P.** Two deterministic half-quadratic regularization algorithms for computed imaging [Text] / P. Charbonnier, L. Blanc-Feraud, G. Aubert [et al.] // In Proc. IEEE ICIP. –1994. –Vol. 2.
13. **Bruhn, A.** A Multigrid Platform for Real-time Motion Computation with Discontinuity-Preserving Variational Methods [Text] / A. Bruhn, J. Weickert, T. Kohlberger [et al.] // IJCV. –2006. –№ 70 (3).
14. **Le Besnerais, G.** Dense optical flow by iterative local window registration [Электронный ресурс] / G. Le Besnerais, F. Champagnat // In Proc. IEEE ICIP. –2005. –Vol. 1.
15. **Rannacher, J.** Realtime 3D motion estimation on graphics hardware [Электронный ресурс] / J. Rannacher // Undergraduate thesis, Heidelberg University. –2009.
16. Метод LK [Электронный ресурс] / Berkin Bilgic. –Режим доступа: <http://www.mathworks.com/matlabcentral/fileexchange/23142>
17. Метод TV-L1 [Электронный ресурс] / Thomas Pock. –Режим доступа: <http://gpu4vision.icg.tugraz.at/index.php?content=downloads.php>
18. Метод CLG [Электронный ресурс] / Се Liu. –Режим доступа: <http://people.csail.mit.edu/ceiliu/OpticalFlow/>
19. Метод Брок [Электронный ресурс] / Thomas Brox. –Режим доступа: <http://lmb.informatik.uni-freiburg.de/people/brox/index.en.html/code.html>
20. Методы HS, BA, Classic++ и Classic+NL [Электронный ресурс] / Deqing Sun. –Режим доступа: <http://www.cs.brown.edu/~dqsun/research/software.html>

УДК 004.4:004.932.2:615.831

М.С. Аникеева, А.Н. Софронов, С.С. Дремов, А.Л. Тер-Мартirosян

ЦИФРОВАЯ СИСТЕМА ВИЗУАЛИЗАЦИИ ДЛЯ ФОТОДИНАМИЧЕСКОЙ ТЕРАПИИ

Фотодинамическая терапия (ФДТ) представляет собой метод диагностики и лечения злокачественных раковых опухолей, основанный на способности некоторых препаратов (фотосенсибилизаторов) избирательно накапливаться в пораженных тканях и одновременно взаимодействовать с электромагнитным излучением видимого или ближнего инфракрасного диапазона. Локальная активация светом накопившегося в опухоли фотосенсибилизатора из полосы его поглощения приводит в присутствии кислорода тканей к развитию фотохимической реакции, разрушающей опухолевые клетки, за счет чего достигается лечебный эффект. При этом процесс сопровождается флюоресценцией как препарата, так и кислорода, в более длинноволновой области спектра. Картина пространственного распределения флюоресценции дает информацию о размерах опухоли и концентрации препарата, а ее изменение во времени может дать и количественную информацию о необходимой дозе облучения. Данный эффект может использоваться не только при лечении, но и при диагностике онкологических заболеваний различных локализаций на ранних стадиях [1].

Последние достижения в области полупроводниковой оптоэлектроники делают метод ФДТ доступным для широкого использования в медицинских учреждениях благодаря разработке недорогих компактных мощных источников лазерного излучения на основе полупроводниковых лазерных диодов [2]. Однако для получения и визуализации картины пространственного распределения флюоресценции фотосенсибилизатора необходима спектрально-селективная система регистрации оптического излучения, пропускающая свет во всей полосе флюоресценции и непро-

зрачная для возбуждающего флюоресценцию излучения. На настоящий момент в существующих устройствах ФДТ в качестве таких систем выступают либо очки, либо аналоговые видеокамеры с оптическими интерференционными фильтрами. Такой подход не только не позволяет количественно оценить параметры опухоли, но и не обеспечивает потенциального удобства применения систем ФДТ.

Все это говорит о необходимости разработки цифровой системы визуализации флюоресценции фотосенсибилизаторов, позволяющей в полной мере использовать все потенциальные возможности метода фотодинамической терапии, производящей визуализацию в наиболее удобной форме и осуществляющей автоматический количественный анализ изображения флюоресценции в реальном времени.

Основой такой системы является программное обеспечение, управляющее фоточувствительным элементом системы и производящее обработку поступающих от него данных.

Анализ требований к программному обеспечению систем фотодинамической терапии

Один из ключевых моментов разработки программного обеспечения для систем фотодинамической терапии – форма визуализации получаемой информации, наиболее удобная для конечного пользователя. Анализ пожеланий практикующих врачей показал, что «классическое» представление в виде изображения в оттенках серого, получаемое непосредственно с камеры системы через спектрально-селективную оптику, является недостаточно удобным для визуального восприятия, а также недостаточным

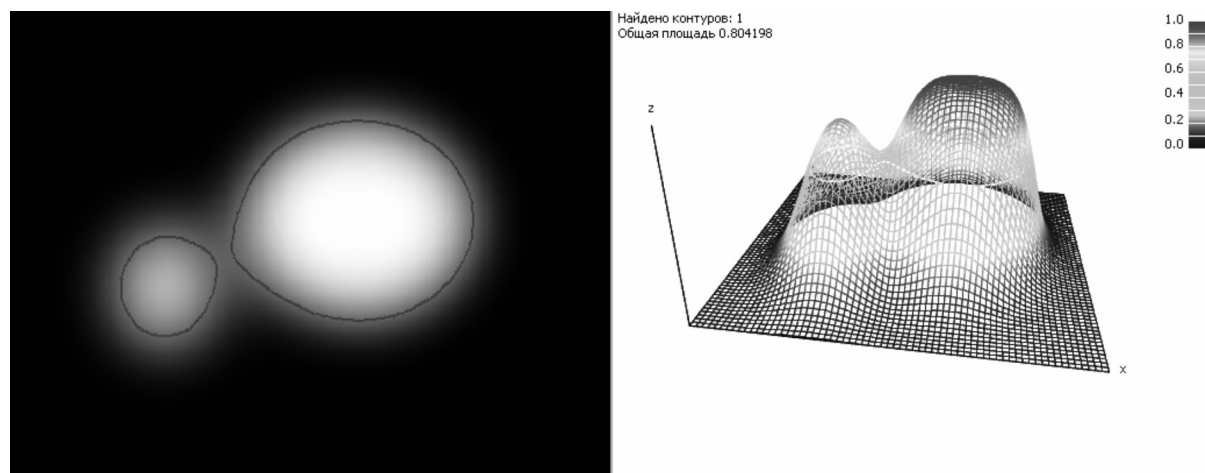


Рис. 1. Форма визуализации картины пространственного распределения флюоресценции (используется цветовой код)

для анализа визуальной информации. Наилучшей формой визуализации оказалась псевдо-трехмерная гистограмма распределения яркости по двум координатам, отображаемая рядом с реальной картиной, получаемой с камеры (рис. 1). При этом сама реальная картина может быть представлена как в исходном виде в оттенках серого, так и в псевдоцветах в соответствии с определенной принятой цветовой картой.

Для дополнительной наглядности результатов необходимо также очерчивать на полученных изображениях контуры постоянного значения яркости по определенному заданному уровню, представляющие собой, фактически, границы опухолей.

Помимо представления распределения интенсивности флюоресценции в реальном времени, необходимо также иметь возможность фиксировать изображения в определенные моменты времени для последующего сравнения и анализа.

Следующий аспект, требующий особого внимания – это графический интерфейс пользователя программного обеспечения (ПО). С учетом специфики конечных пользователей ПО, интерфейс должен быть простым, удобным, обеспечивать минимум действий пользователя и позволять легко менять вид и представление визуализируемой информации.

Аппаратная платформа

Аппаратная часть системы, на которой предполагается функционирование ПО, состоит из стандартного ПК с подключенной к нему циф-

ровой камерой по любому достаточно быстрому интерфейсу (USB2.0, IEEE1394, Ethernet). На объективе цифровой камеры установлен спектрально-селективный оптический фильтр, пропускающий только излучение в полосе флюоресценции используемого фотосенсибилизатора. Таким образом, поступающие с камеры видеокдры являются исходными данными для визуализации.

Блок-схема аппаратной части показана на рис. 2. Аналоговый видеосигнал с фоточувствительной матрицы оцифровывается блоком цифровой обработки, который также отвечает за передачу информации в ПК. Оба эти блока объединены в цифровой камере. Также в состав прибора входит излучатель, возбуждающий флюоресценцию, блок питания и некоторые вспомогательные элементы.

Объектная модель программного обеспечения

Следуя [3], построение объектно-ориентированной архитектуры разрабатываемого ПО состоит в построении абстракций для отдельных элементов системы, установления ответственностей каждой абстракции и связей между ними. Готовая объектная модель системы представлена на рис. 3 в виде совмещенной диаграммы классов и компонентов.

Программной абстракцией для источника исходных кадров является класс `frameInput`. Кроме очевидной способности возвращать очередной кадр видеопотока, в число обязанностей класса входит также предоставление информации о па-

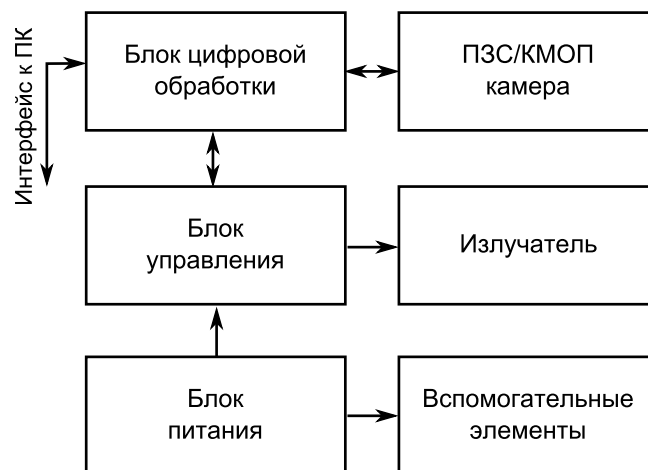


Рис. 2. Блок-схема аппаратной платформы системы

раметрах кадра (размеры, битность и т. д.). Класс прячет в себе конкретные детали аппаратного интерфейса подключения камеры к ПК, при этом конкретная реализация класса осуществляется либо с помощью фирменного SDK камеры, либо посредством DirectShow. В последнем случае может использоваться открытая библиотека videoInput [4], являющаяся объектно-ориентированной оберткой над DirectShow.

Классы VideoOutput и ChartOutput являются абстракциями для устройств вывода информации в виде битовой карты и псевдотрехмерной гистограммы соответственно. В обязанности классов входит отображение графической информации в виде, показанном на рис. 1. Конкретная реализация классов осуществлена с использованием библиотеки OpenGL, имеющей аппаратную реализацию на большинстве видеокарт, при этом использование OpenGL существенно ускоряет вывод графики по сравнению со средствами Windows GDI. Оба класса являются наследниками абстрактного класса GLWidget, предназначенного для инкапсуляции функций по созданию и обслуживанию окна графического интерфейса пользователя, в которое осуществляется вывод средствами OpenGL.

Класс Analyzer инкапсулирует функции по анализу кадра изображения, поиску контуров постоянного значения и расчету данных для построения трехмерного графика. Класс имеет методы для передачи ему параметров анализа кадра (пороги контуров и т. д.), а также основной метод обработки, принимающий в аргументах указатель на устройство ввода кадров frameInput и указатели

на устройства вывода VideoOutput и ChartOutput. Основная функция класса – определить границы пораженного участка ткани, рассчитать его площадь, а также подготовить данные к отображению на устройствах вывода и обновить последние. Конкретная реализация класса основана на использовании алгоритмов машинного зрения библиотеки OpenCV. В частности, производится топологический структурный анализ кадра для поиска полного множества контуров постоянного значения [5].

Класс mainWindow инкапсулирует функции создания пользовательского графического интерфейса, в также выполняет роль «сэмплера», вызывая по таймеру основной метод обработки класса Analyzer. Конкретная реализация основана на кросс-платформенном инструментарии Qt (версия, лицензированная по GNU LGPL). Это наиболее удобный и хорошо документированный способ быстрого создания графического интерфейса. По сравнению с набором классов MFC от Microsoft, Qt предлагает полностью правильный объектно-ориентированный подход к программированию с активным использованием иерархии классов, что упрощает и ускоряет разработку, обеспечивая более высокий уровень абстракции.

Полная модель содержит также ряд дополнительных классов. В частности, построенные по принципу связанного списка классы Contours для хранения контуров постоянного значения и ImgStack, производящий усреднение некоторого количества последовательных кадров для снижения уровня шумов, а также класс DimDialog, ответственный за вывод элементов графического

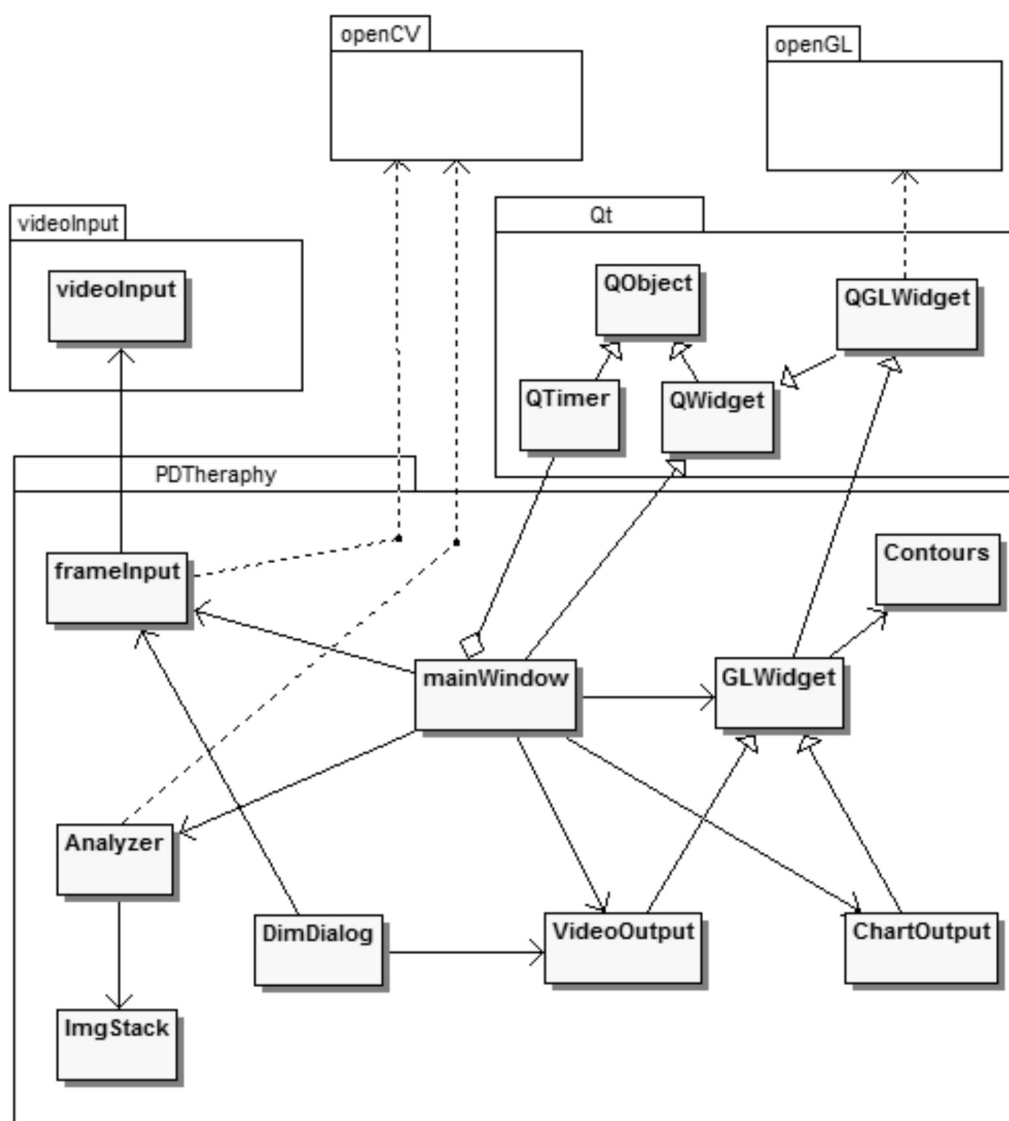


Рис. 3. Диаграмма классов и компонентов

пользовательского интерфейса, необходимых для калибровки реальных размеров объектов.

Все зависимости программы от внешних библиотек показаны на рис. 3.

Клинические испытания

На основе разработанной модели создана конкретная реализация опытного образца системы визуализации флюоресценции – прибор «Флюовизор». В приборе цифровая монохромная камера с разрешением 752×582 пикселя, использующая интерфейс USB 2.0. В качестве персонального компьютера использован ноутбук низшего ценового сегмента с двухъядерным процессором семейства Intel Celeron B800. Применение опи-

санных подходов к разработке программного обеспечения позволило добиться работы системы в режиме реального времени (фактически, в режиме обновления изображений на экране с частотой 30 fps) даже на такой аппаратной платформе.

Опытный образец прошел клинические испытания в НИИ онкологии имени Н.Н. Петрова, где использовался для визуализации при диагностике и лечении базальноклеточного рака кожи методом ФДТ при облучении опухоли полупроводниковым лазером с длиной волны излучения 665 нм. В качестве фотосенсибилизатора применялся препарат Фотодитазин.

На рис. 4 показана реальная картина распределения флюоресценции.



Рис. 4. Флюоресценция базальноклеточного рака кожи заушной области, возбуждаемая полупроводниковым лазером с длиной волны излучения 665 нм и мощностью 3 Вт

На рис. 5 представлена визуализация участка полного кадра с пораженной областью в виде псевдотрехмерного распределения интенсивности флюоресценции по координатам. Также показана найденная граница опухоли, отображаемая линиями на исходном изображении и на трехмерной картинке.

Приведенные примеры иллюстрируют возможности по интерпретации реальных изображений, которые предоставляет пользователю разработанное ПО.

Такая форма представления данных наглядно показывает, в частности, распределение концентрации фотосенсибилизатора и ее динамическое

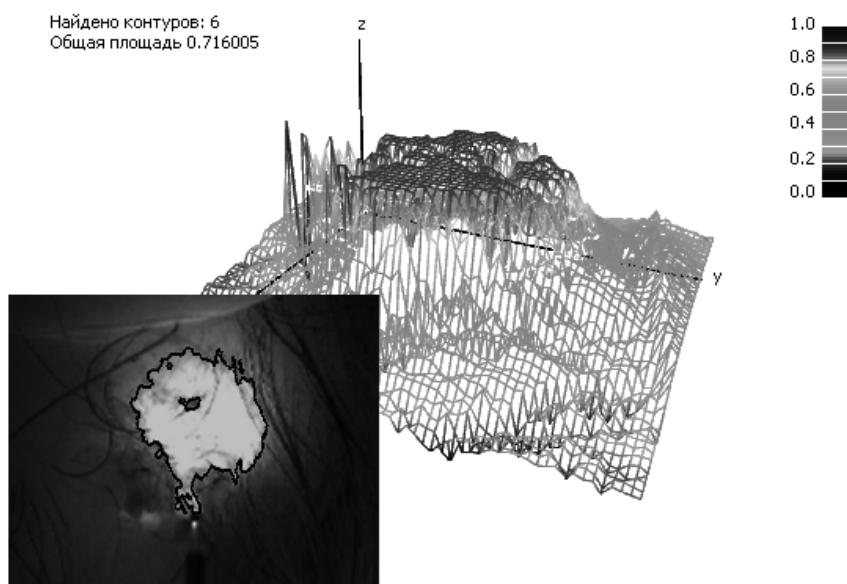


Рис. 5. Визуализация участка изображения на рис. 4 в виде псевдотрехмерной гистограммы (используется цветовой код)

изменение во времени, позволяет более дозированно и направленно обеспечивать облучение ткани лазером для оптимального проведения процедуры лечения методами ФДТ.

В результате выполнения работы создано программное обеспечение для системы визуализации пространственного распределения флюоресценции при диагностике и лечении методами фотодинамической терапии. Разработана объектно-ориентированная архитектура системы визуализации, произведен выбор оптимального варианта конкретной реализации

с активным использованием сторонних библиотек с открытым исходным кодом. Изготовлен опытный образец системы визуализации, включающий аппаратную часть и программное обеспечение, который успешно прошел клинические испытания.

Работа выполнена в рамках проекта по развитию кооперации российских вузов и производственных предприятий (договор № 13.G25.31.0055). По результатам работы направлена заявка на патент на устройство визуализации флюоресценции фотосенсибилизаторов при диагностике и лечении методом ФДТ.

Авторы выражают благодарность М.Л. Гельфонду за помощь в проведении клинических испытаний системы.

СПИСОК ЛИТЕРАТУРЫ

1. **Loschenov, V.B.** Proceeding SPIE [Text] / V.B. Loschenov, R. Steiner. –1994. –Vol. 2325. –144 p.
2. **Воробьев, Л.Е.** Лазерные диоды для фотодинамической терапии [Текст] / Л.Е. Воробьев, А.Н. Софронов, Д.А. Фирсов, Д.М. Демидов, Р.В. Леус, М.А. Свердлов, А.Л. Тер-Мартirosян // Научно-технические ведомости СПбГПУ. Сер. Физико-математические науки. –2011. –№ 2 (122). –С. 80–84.
3. **Гради, Буч.** Объектно-ориентированный анализ и проектирование с примерами приложений [Текст] / Буч Гради; 3-е изд.; Пер. с. англ. –М.: ИД «Вильямс», 2008. –720 с.
4. [Электронный ресурс] / Режим доступа: <https://github.com/ofTheo/videoInput>
5. **Bradski, Gary.** Learning OpenCV [Text] / Gary Bradski, Adrian Kaehler. –O'Reilly, 2008.

УДК 656.078: 658.5.011

Э.А. Латыпова, И.О. Бондарева

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ КАК СРЕДСТВО УПРАВЛЕНИЯ ЛОГИСТИЧЕСКИМ ПРЕДПРИЯТИЕМ НА ПРИМЕРЕ ГРУЗОВОГО ПОРТА

Для подавляющего большинства российских компаний, в т. ч. и грузовых портов, огромное значение приобретает решение проблемы повышения конкурентоспособности организации путем совершенствования менеджмента и систем качества продукции и услуг. Наиболее эффективным направлением совершенствования менеджмента предприятия является получение интегрированной системы управления предприятием.

Зарубежный опыт применения различных моделей менеджмента показывает, что интегрированную систему управления предприятием целесообразно выстраивать и совершенствовать прежде всего на принципах «Всеобщего Менеджмента Качества» (Total Quality Management – TQM) и «Бережливого Производства и Мышления» (Lean Thinking and Manufacturing – LTM), т. к. остальные подходы можно рассматривать как их частные случаи.

Достоинства идей TQM состоят в том, что согласно им в основе менеджмента должны находиться люди и бизнес-процессы, в т. ч. процессы взаимоотношений поставщиков и потребителей. Наиболее значительным достижением TQM является рассмотрение бизнеса как процесса достижения справедливо сбалансированных целей и интересов всех заинтересованных сторон (владельцев, акционеров, инвесторов, менеджеров, сотрудников, потребителей, поставщиков и общества).

Всеобщий менеджмент качества подразумевает постоянный контроль за состоянием качества всех сфер деятельности организации, а также планирование качества. Самоконтроль позволяет компаниям определять стратегическое направление своей деятельности в борьбе за потребите-

ля. Одной из базовых управленческих стратегий концепции TQM является управление на основе фактов. Такими «фактами» могут являться показатели деятельности компании или показатели качества [1].

Однако появляется сложность в выборе таких показателей из большого количества имеющихся. Для того чтобы согласно еще одной стратегии TQM всецело ориентироваться на процесс, т. е. охватывать все сферы деятельности компании, направленные на достижение общего результата, предлагается использовать концепцию Сбалансированной Системы Показателей (ССП). Данная концепция обеспечивает возможность совокупного рассмотрения финансовых и нефинансовых индикаторов, позволяет учитывать причинно-следственные связи между влияющими факторами и итоговыми показателями. Это позволяет осуществлять подробный мониторинг деятельности компании в стратегическом фокусе, повысить эффективность и оперативность принимаемых управленческих решений, а также контролировать целевые показатели деятельности, степень достижения которых определяет движение компании согласно заданной стратегии. Другими словами, суть SSP заключается в формулировании стратегии в нескольких перспективах, постановке стратегических целей и измерении степени достижения данных целей при помощи показателей [2, 3].

Придерживаясь данной последовательности, на основе стратегии развития грузового порта первоначально было разработано дерево стратегических целей, представленное на рис. 1. Таким образом, видно, что основной целью порта является повышение конкурентоспособности. Повсемест-



Рис. 1. Дерево целей грузового порта

но известно, что финансовая конкурентоспособность определяется финансовой устойчивостью (или независимостью от внешних займов), платежеспособностью, скоростью оборота активов и доходностью. Исходя из этого сформулированы

цели следующего уровня.

Стратегические цели деятельности компании также рассмотрены согласно ССП в рамках четырех перспектив (рис. 2): финансы, клиенты, внутренние бизнес-процессы, обучение и развитие.

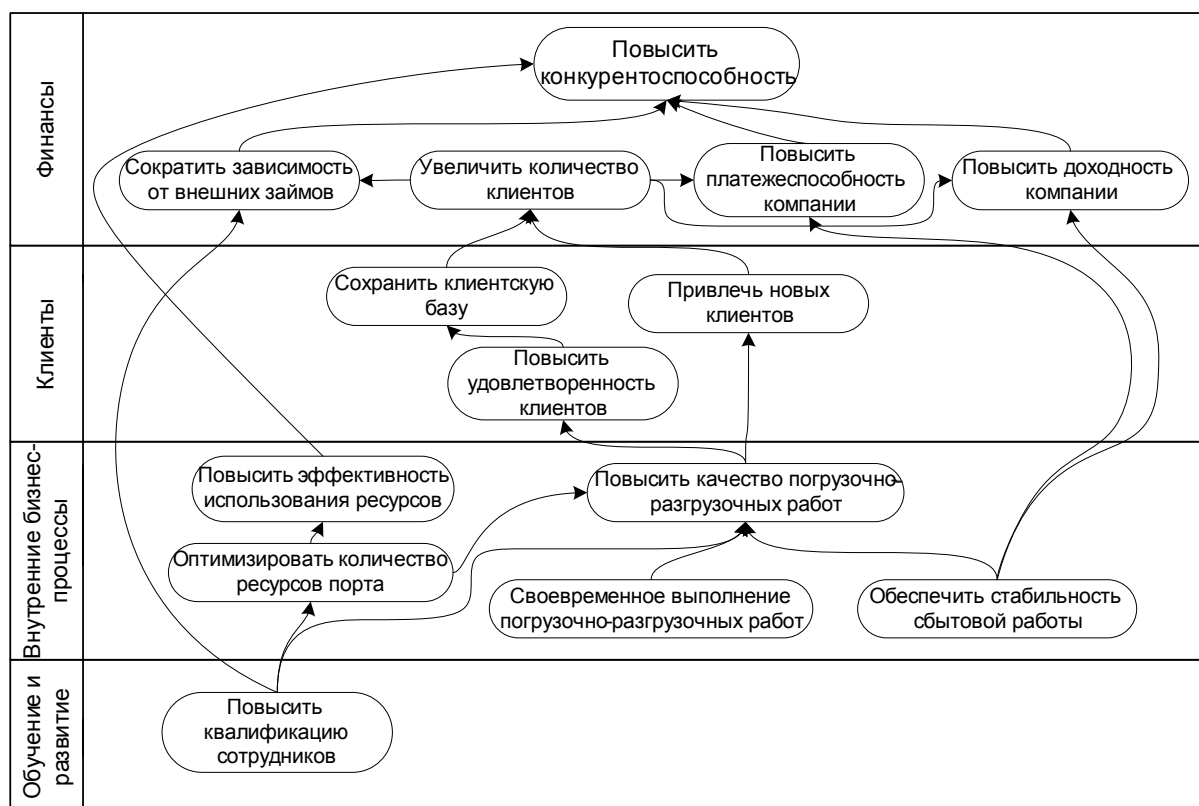


Рис. 2. Стратегическая карта целей грузового порта

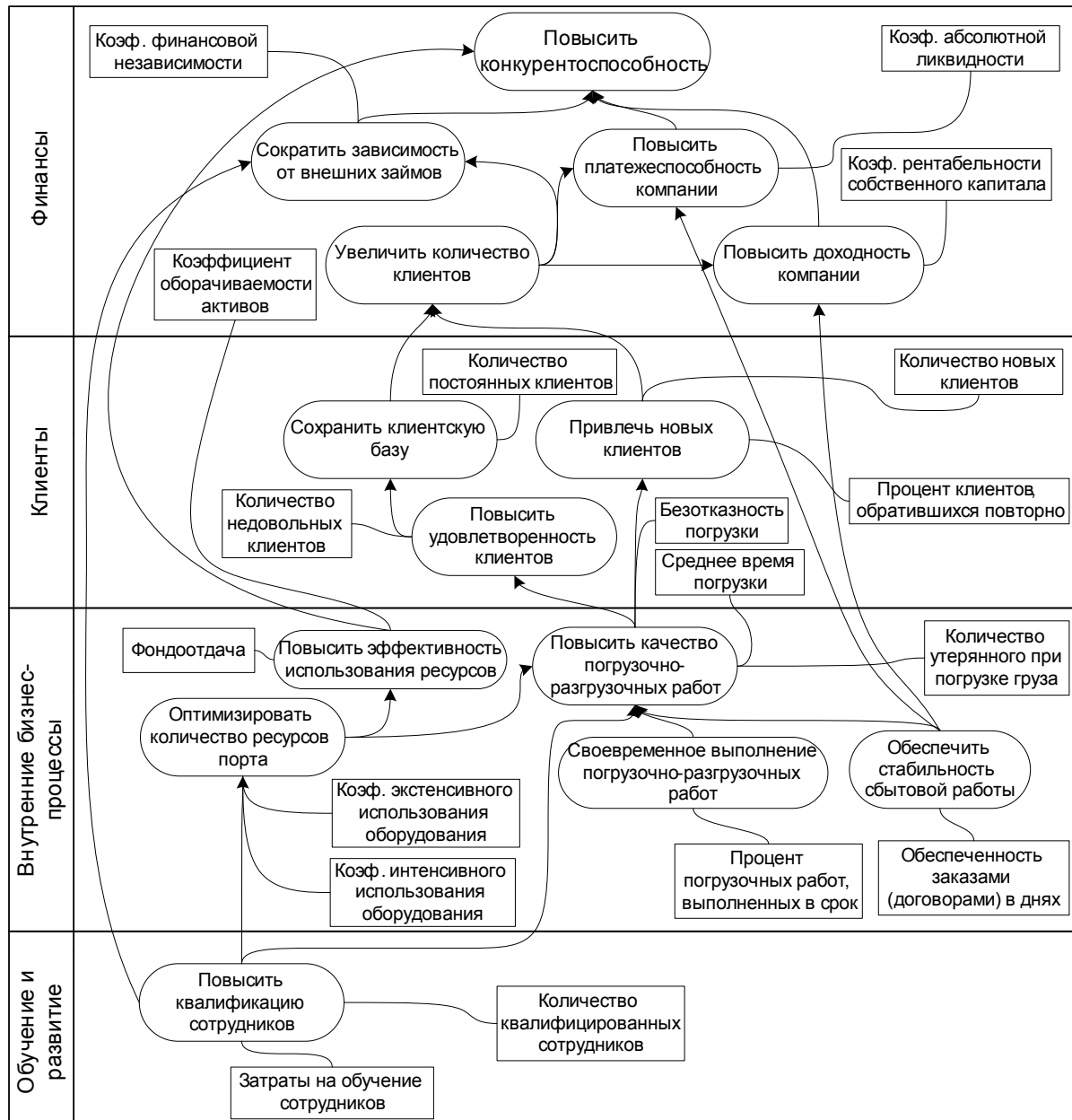


Рис. 3. Стратегическая карта грузового порта

На рисунке отражена взаимозависимость целей согласно перспективам.

Затем каждая цель в соответствии с перспективами была дополнена набором характеризующих ее показателей, что нашло отражение в разработанной стратегической карте грузового порта (рис. 3) [3, 4].

Все показатели, характеризующие деятельность порта, а также формулы и способы их расчета [5] представлены в таблице в соответствии с перспективами.

Цели и показатели определены, однако поддержание качества всех процессов на определенном уровне невозможно без информационно-компьютерной поддержки. При анализе, синтезе и оптимизации объектов и процессов на предприятиях широко используются различные модели: математические, графические, физические, имитационные и др. В данном случае предлагается использовать частный случай математического моделирования – имитационное моделирование. Замена математической модели имитатором или

Показатели деятельности грузового порта

Перспектива	Показатель	Формула/Способ расчета
Финансы	Коэффициент финансовой независимости (коэффициент автономии)	$K_a = \frac{C_c}{A_c}$, где C_c – собственные средства; A_c – совокупные активы
	Коэффициент оборачиваемости активов	$K_{oa} = \frac{B}{C_{a,cp}}$, где B – выручка от продаж, осуществленных за период времени T ; $C_{a,cp}$ – средняя стоимость активов за тот же период времени T
	Коэффициент абсолютной ликвидности	$K_{ал} = \frac{ДС + ФВ_{кр}}{O_{текущ}}$, где $ДС$ – денежные средства; $ФВ_{кр}$ – краткосрочные финансовые вложения; $O_{текущ}$ – текущие обязательства
	Коэффициент рентабельности собственного капитала	$K_{рск} = \frac{\Pi_{ч}}{C_c}$, где $\Pi_{ч}$ – чистая прибыль за период времени T ; C_c – собственные средства
Клиенты	Количество постоянных клиентов	Рассчитывается путем подсчета числа клиентов, обратившихся более двух раз
	Количество новых клиентов	Подсчитывается путем подсчета числа клиентов, вновь обратившихся за период времени T
	Безотказность погрузки	$БП = \frac{КПрЗ}{КПЗ} \times 100 \%$, где $КПрЗ$ – количество заявок, принятых за период времени T ; $КПЗ$ – количество поступивших заявок за тот же период времени T
	Среднее время погрузки	$ВП_{cp} = \frac{T_{п}}{КВЗ_{общ}}$, где $T_{п}$ – общее время погрузки; $КВЗ_{общ}$ – общее количество заявок, выполненных за период времени T
	Процент клиентов, обратившихся повторно	$Проц.Кл_{повтор} = \frac{Кл_{повтор}}{Кл_{общ}} \times 100 \%$, где $Кл_{повтор}$ – количество клиентов, повторно обратившихся за период времени T ; $Кл_{общ}$ – общее количество клиентов, обратившихся за период времени T
	Количество недовольных клиентов	Рассчитывается путем подсчета числа клиентов, оставшихся недовольными услугами, оказанными им за период времени T
Внутренние бизнес-процессы	Процент погрузочно-разгрузочных работ, выполненных в срок	$Проц.ПР_{в\ срок} = \frac{ПР_{в\ срок}}{ПР_{общ}} \times 100 \%$, где $ПР_{в\ срок}$ – количество погрузочно-разгрузочных работ, осуществленных за период времени T , без нарушения оговоренных сроков; $ПР_{общ}$ – общее количество погрузочно-разгрузочных работ, осуществленных за период времени T

Внутренние бизнес-процессы	Обеспеченность заказами (договорами) в днях	$Z_{\text{дн}} = \text{Кол}_{\text{дого}} * \overline{Дн_{\text{дого}}}$, где $\text{Кол}_{\text{дого}}$ – количество заключенных договоров; $\overline{Дн_{\text{дого}}}$ – среднее количество дней, необходимых для проведения работ по одному договору
	Количество утерянного (испорченного) при погрузке груза	$Гр_{\text{утер}} = Гр_{\text{план}} - Гр_{\text{факт}}$, где $Гр_{\text{план}}$ – общее количество (т.) груза, запланированное к погрузке за период времени T (согласно принятым заявкам); $Гр_{\text{факт}}$ – общее количество грузов (т), погруженных за период времени T
	Фондоотдача	$K_{\text{ф}} = \frac{СУ}{\text{СОС}_{\text{средн}}}$, где $СУ$ – стоимость услуг, оказанных за период времени T ; $\text{СОС}_{\text{средн}}$ – среднегодовая стоимость основных средств на начало года
	Коэффициент экстенсивного использования оборудования	$K_{\text{экт}} = \frac{T_{\text{ф}}}{T_{\text{р}}}$, где $T_{\text{ф}}$ – фактическое время работы машин и оборудования, ч; $T_{\text{р}}$ – режимный фонд времени работы машин и оборудования, ч
	Коэффициент интенсивного использования оборудования	$K_{\text{инт}} = \frac{П_{\text{ф}}}{П_{\text{р}}}$, где $П_{\text{ф}}$ – фактическая производительность основного технологического оборудования (ед. продукции/ч); $П_{\text{р}}$ – технически обоснованная производительность машин и оборудования (ед. продукции/ч)
Обучение и развитие	Количество квалифицированных сотрудников	Рассчитывается путем подсчета количества работников, прошедших профессиональную подготовку в соответствии с занимаемой должностью
	Затраты на обучение сотрудников	Рассчитывается путем суммирования затрат на квалификационную подготовку работников предприятия

имитационной моделью производится в ряде случаев, когда для того или иного (сложного в описании) объекта по различным причинам не могут быть разработаны аналитические модели, либо отсутствуют методы решения полученной модели.

В процессе имитации фиксируются определенные события и состояния или измеряются выходные воздействия, по которым вычисляются характеристики функционирования системы. Имитационная модель является универсальным методом отражения реальных процессов функционирования системы, который позволяет реагировать на все изменения в самом процессе и опытным путем (с помощью изменения первоначальных входных данных) получать необходи-

мые результаты на выходе [6]. На основе этого можно корректировать выполнение процессов на действительной модели для получения необходимого результата.

Необходимо определиться с программным обеспечением, используемым для построения модели грузового порта, позволяющей анализировать его работу, используя описанные выше показатели.

Присутствующие на сегодняшний день на рынке программные продукты имитационного моделирования можно разделить на две группы: универсальные языки программирования и пакеты имитационного моделирования. Предпочтение было отдано пакетам имитационного моделирования, т. к. созданные с их помощью

имитационные модели просты в использовании, легко модифицируются и обеспечивают более совершенные механизмы обнаружения ошибок. Из всех универсальных пакетов моделирования был выбран пакет Agena, т. к. он предоставляет возможность импорта исходных данных из MS Excel, имеет встроенный язык Visual Basic for Application (VBA), что существенно облегчает и расширяет возможности моделирования; позволяет создавать настраиваемые модули и сохранять их в качестве шаблона, использовать возможности анимации и сохранять результаты моделирования в базе данных и конвертировать их в различные форматы.

Построена имитационная модель грузового порта, включающая следующие модули: модуль поступления и обработки; погрузки; прибытия вагонов с грузом; прибытия судов с грузом; метеорологических условий.

По окончании моделирования автоматически выводится результирующая таблица, наглядно отображающая полученные значения для каждого из показателей, описанных выше.

Для проверки адекватности разработанной модели использовались результаты моделирования и статистические данные о работе грузового порта за август 2011 г. В указанный период времени на территории грузового порта функционировали все погрузчики и краны. Метеорологические факторы, такие, как туман,

ветер, ледостав на работу порта влияния не оказывали. Результаты сравнения статистических показателей работы порта за август 2011 г. и показателей, полученных в результате моделирования, позволяют говорить о том, что разработанная модель адекватна реальному экономическому объекту.

Построенная на основе стратегической карты имитационная модель грузового порта позволяет корректировать возможные мероприятия по улучшению качества работы порта путем осуществления манипуляций с исходными данными, оценивать реальное положение дел, изучая полученные значения показателей, и тем самым выявлять «узкие места» в работе порта с целью их устранения, а также принимать управленческие решения более оперативно за счет сокращения времени обработки и получения необходимых данных, повышения степени достоверности обработки информации, увеличения количества аналитических показателей, получаемых на базе исходных.

В результате изучения деятельности грузового порта выявлено, что баланс использования концепций всеобщего управления качеством, сбалансированной системы показателей и технологии имитационного моделирования позволит добиться конкурентного преимущества за счет повышения качества всех сфер деятельности порта.

СПИСОК ЛИТЕРАТУРЫ

1. **Липидус, В.А.** Всеобщее качество (TQM) в российских компаниях [Текст] / В.А. Липидус. – М.: Новости, 2000. – 432 с.
2. **Каплан, Р.С.** Сбалансированная система показателей. От стратегии к действию [Текст] / Р.С. Каплан, Д.П. Нортон; Пер. с англ. – М.: ЗАО «Олимп-Бизнес», 2003. – 304 с.
3. **Ольве, Н.-Г.** Оценка эффективности деятельности компании: Практическое руководство по использованию сбалансированной системы показателей [Текст] / Н.-Г. Ольве, Ж. Рой, М. Веттер. – М.: ИД «Вильямс», 2004. – 306 с.
4. **Гаджинский, А.М.** Логистика: Учебник для высших и средних специальных учебных заведений [Текст] / А.М. Гаджинский. – 3-е изд. – М.: Информационно-внедренческий центр «Маркетинг», 2001. – 472 с.
5. **Грищенко, О.В.** Анализ и диагностика финансово-хозяйственной деятельности предприятия: Учеб. пособие [Текст] / О.В. Грищенко. – Таганрог: Изд-во ТРТУ, 2000. – 112 с.
6. **Кельтон, В.** Имитационное моделирование. Классика CS [Текст] / В. Кельтон, А. Лоу. – СПб.: Питер; Киев: Издат. группа BHV, 2004. – 847 с.

ПЛАНИРОВАНИЕ ПОРТФЕЛЯ ПРОЕКТОВ В СТРОИТЕЛЬСТВЕ НА ОСНОВЕ МУЛЬТИАГЕНТНОГО ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

В статье рассматривается комплекс задач, возникающих при планировании деятельности холдинга, который занимается строительством нескольких объектов. Данная задача соответствует задаче планирования портфеля проектов. Задача планирования строительных работ и ее программная реализация решена в результате интеграции мультиагентного моделирования и метода критического пути.

Процессы преобразования ресурсов

Под управлением проектом будем понимать деятельность, направленную на реализацию проектов с максимально возможной эффективностью при заданных ограничениях по времени, денежным средствам, трудовым и материальным ресурсам, а также качеству конечных результатов проекта [1]. Для решения задачи планирования портфеля проектов (объектов строительства) в данной работе используется мультиагентная модель процесса преобразования ресурсов (МППР) и программы семейства BPsim [2]. Рассмотрим основные понятия процессов преобразования ресурсов (ППР).

Ресурс – количественная мера возможности выполнения какой-либо деятельности [3]. Ресурс – это, что можно использовать, тратить; возможная продолжительность эксплуатации машины [3, 4].

Под ППР понимается непрерывная или дискретная операция преобразования входа (ресурсов, необходимых для выполнения процесса) в выход (продуктов – результатов выполнения процесса). Преобразование осуществляется с помощью средства [2]. Классификация ресурсов с точки зрения их использования [5] в ППР приведена на рис. 1.

Потребляемые ресурсы (входы) – ресурсы, которые используются в процессе только один раз. *Средства* не потребляются, а используются в процессе преобразования, они не уменьшаются в процессе их использования. *Выходы* формируются в процессе преобразования [5].

Метод планирования на основе мультиагентного имитационного моделирования

Предлагаемый метод состоит из следующих этапов.

1. Построение имитационной модели (ИМ) с агентами субподряда. При построении ИМ строятся следующие подмодели: генерации объектов строительства, строительства, поставок потребляемых ресурсов (материалов, необходимых для строительства), субподряда, продаж.

При построении подмодели строительства

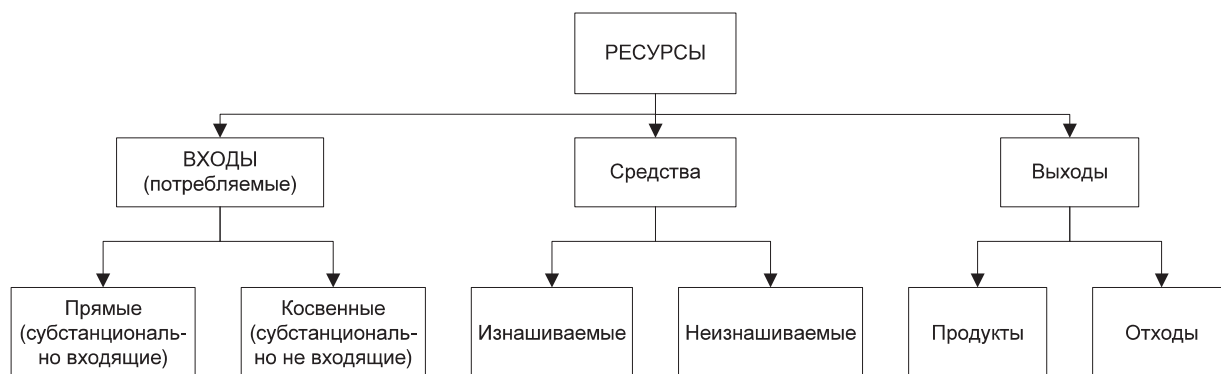


Рис. 1. Классификация ресурсов по типам использования [5]

описывается модель типового объекта строительства, которая является параметрически настраиваемой и зависит от типа и свойств сгенерированного объекта. Так, отдельные блоки модели соответствуют этапам строительства. При поступлении в блок заявки (объекта строительства), в зависимости от ее параметров, происходит вычисление времени длительности этапа, количества потребляемых ресурсов и используемых средств.

Предметная область для моделирования портфеля проектов (возводимых объектов) строительства диктует следующие специфические требования и соответствующие им правила построения ИМ:

- ограничение по объему субподрядных работ. В случае превышения ограничения на суммарные затраты данный объект строительства становится нерентабельным – проект может быть исключен из портфеля;

- должна использоваться стратегия строительства объектов «первым пришел – первым вышел», т. к. задержки при строительстве отдельного объекта приводят к ряду дополнительных затрат, связанных с обеспечением проекта (охрана, электроэнергия и т. д.). Завершение процесса строительства объекта и, соответственно, сокращение его сроков, приводит к ускорению возврата инвестиций. В связи с этим блоки модели должны применять выталкивающую стратегию для заявки «объект строительства» (приоритет блоков модели возрастает от начальных этапов строительства к завершающим);

- для процесса строительства характерно параллельное во времени выполнение этапов разных работ; работы (соответствующие блоки ИМ), относящиеся к критическому пути, должны иметь приоритет выше, чем у параллельных работ.

Данные требования предметной области и правила построения ИМ также хорошо согласуются с выводами Дэйвиса [6]: «... правило упорядочения, в соответствии с которым первой выполняется работа с наименьшим резервом (или эквивалентное правило минимизации самого позднего времени начала), в среднем дает наилучший результат».

В подмодели субподряда возможны два варианта: чужие средства привлекаются в случае нехватки своих на выполнение всей операции от начала и до окончания; на каждом последующем такте происходит пересмотр возможности отказа

от привлечения субподряда [7]. ИМ с агентами субподряда характеризуется следующим: 1) при применении второго варианта время ожидания не увеличивается в виду недостатка средств, т. е. срок проекта по критерию использования средств минимален; 2) по результатам экспериментов можно скорректировать среднее количество своего парка средств и оценить максимальный объем привлечения субподряда; 3) в случае достижения требуемой пропускной способности модели и уменьшения времен ожидания до допустимых значений можно считать план обработки портфеля проектов идеальным (минимальным по срокам). Ограничения реальных проектов строительства: не всегда допустимо применение субподряда; объем привлечения субподряда имеет ограничения. Применение агентов субподряда позволяет только решить задачу узких мест на средствах мультиагентной модели. Для решения задачи анализа узких мест, возникающих на операциях и ресурсах, необходимо выполнение последующих шагов, завершающихся реинжинирингом мультиагентной модели.

2. Проведение имитационного эксперимента с моделью субподряда.

3. Поиск узких мест модели МППР. При диагностике узких мест анализируются следующие параметры мультиагентной модели [7, 8]: коэффициент использования операции, средства, агента; среднее время заявки в очереди к операции, агенту; простой операции из-за отсутствия средств и/или входных ресурсов. Для оценки динамики работы операции и агента также анализируется средняя очередь заявок к операции, агенту и среднее состояние ресурсов.

4. Реинжиниринг модели МППР. По результатам анализа статистики экспериментов диагностируются узкие места и принимается решение о свертке/развертке модели [7]. Критерием остановки реинжиниринга модели МППР является снижение времени ожидания до допустимых значений по всем блокам модели [8]. Данный этап направлен на решение задачи распараллеливания сетевых графиков параллельно возводимых во времени объектов строительства (в блоках ИМ могут возникать ситуации с параллельной обработкой заявок).

5. Построение модели сети массового обслуживания (СМО). Применение операционного анализа вероятностных сетей к модели МППР позволяет также решить задачу уменьшения ко-

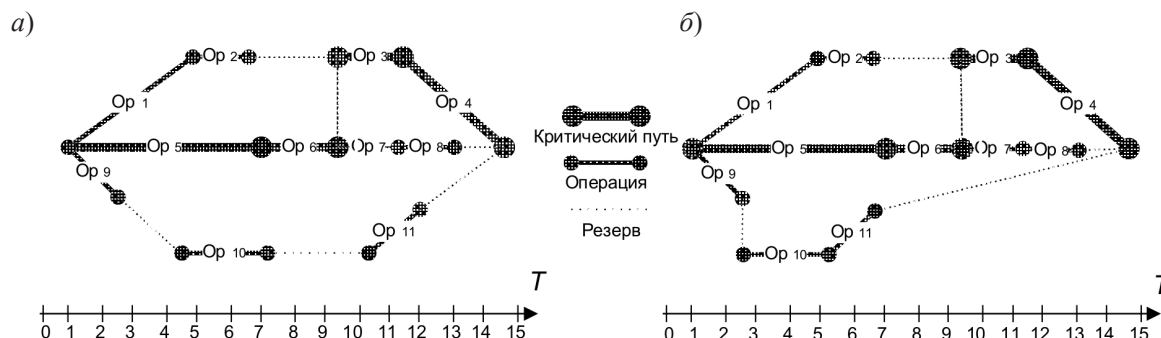


Рис. 2. Применение методов СРМ и ИМ субподряда:
а – применение метода СРМ; б – применение имитационной модели субподряда

личества экспериментов путем построения модели СМО на основе результатов экспериментов модели МППР [8], с целью быстрого нахождения среднего количества работающих устройств (средств МППР).

6. Принятие решения о допустимости применения модели с агентами субподряда. Если применение модели с агентами субподряда допустимо, то переходим на шаг 8, иначе – на шаг 7.

7. Применение алгоритмов балансировки средств модели МППР. Различные подходы балансировки средств модели МППР описаны в [1]. Эксперименты проводятся до нахождения эффективного решения.

8. Построение плана (сетевого графика). По результатам имитационного эксперимента формируется сетевой график портфеля объектов. Критический путь для каждого отдельного объекта строительства определяется как выбор из параллельных цепочек работ, цепи, не имеющей запаса времени по результатам имитационного

эксперимента. Оценка резерва времени для каждой работы выполняется автоматически. Так, для параллельных цепочек (не относящихся к критическому пути) это будет резерв времени последней работы в цепи.

9. Проверка портфеля проектов на соответствие бюджета и сроков. Если сроки и бюджет портфеля проектов (портфель объектов строительства) удовлетворяет, то завершаем, иначе возвращаемся на шаг 7.

Сравнение предлагаемого метода и классического сетевого метода

Для анализа узких мест в управлении проектами и строительстве наиболее часто применяется сетевая модель (сетевой график), которая вместе с методом критического пути (СРМ [6]) позволяет определить резервы времени выполнения отдельных работ. Применение ИМ субподряда приводит к эффекту наискорейшего «проталкивания» работ. Результаты приме-

Сравнение методов

Критерий сравнения	СРМ	МППР
1. Учет использования средств	+	+
2. Учет потребления ресурсов	НЕТ	+
3. Учет поставок ресурсов	НЕТ	+
4. Учет времени жизни потребляемого ресурса	НЕТ	+
5. Учет субподряда	+	+
6. Автоматическая генерация и параметризация объектов портфеля	НЕТ	+
7. Балансировка средств	+	+
8. Формирование сетевого графика	+	+
9. Формирование графика использования средств	+	+
10. Нахождение критического пути	+	+
11. Оценка резерва времени для каждой работы	+	+
12. Формирование графика потребления ресурсов	НЕТ	+

ния метода СРМ и МППР на примере цепочек параллельно-последовательных работ показаны на рис. 2.

Важно отметить разницу понятия «ресурс» сетевой модели и модели МППР. Так, в сетевой модели под «ресурсом» понимаются трудовые ресурсы и технические средства, что в терминологии МППР соответствует «средству». Однако сетевая модель не описывает работу с «потребляемыми» ресурсами и не позволяет описать модель управления запасами ресурсов и модель поставок, что в свою очередь оказывает существенное влияние на ход строительных работ. Новая модель позволяет кроме графика использования средств построить график потребления ресурсов для всего портфеля проекта.

Также стоит отметить существование для строительства специфических потребляемых ресурсов с очень коротким сроком полезного потребления, к которым относится бетон, обладающий высокой скоростью схватывания. С целью учета

ограниченности средств (как собственных, так и субподрядных) оба подхода поддерживают решение задачи балансировки ресурсов (сглаживания функции потребления средств). Результаты сравнения методов представлены в таблице.

С применением системы BPsim и нового метода планирования разработана модель строительного холдинга CHINA WANBAO (задача строительства одиннадцати многоэтажных зданий) [7, 8].

Задача планирования строительных работ и ее программная реализация решена с использованием гибридного подхода в результате применения мультиагентного имитационного моделирования, метода критического пути и операционного анализа вероятностных сетей. Применение метода привело к уменьшению сроков выполнения строительства на 5,94 % (81 день) и повышению загрузки средств.

Работа выполнена при поддержке гранта РФФИ № 12-07-31045.

СПИСОК ЛИТЕРАТУРЫ

1. **Аксёнов, К.А.** Применение имитационного моделирования и технологии интеллектуальных агентов для решения задачи управления проектами [Текст] / К.А. Аксёнов, А.С. Антонова // Научно-технические ведомости СПбГПУ. Сер. Информатика. Телекоммуникации. Управление. –2011. –№ 4 (128). –С. 27–36.
2. **Аксёнов, К.А.** Теория и практика средств поддержки принятия решений: Монография [Текст] / К.А. Аксёнов. –Germany, Saarbrücken: LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. –341 с.
3. Википедия – свободная энциклопедия [Электронный ресурс] / Режим доступа: <http://ru.wikipedia.org/wiki/Ресурс#> (Дата обращения 07.10.12)
4. **Ожегов, С.И.** Толковый словарь русского языка [Текст] / С.И. Ожегов, Н.Ю. Шведова. –М.: Азъ Ltd., 1992. –960 с.
5. **Пищулов, В.** Введение в теорию производства: Учеб. пособие [Текст] / В. Пищулов, К. Рихтер, Е. Дятел. –Екатеринбург: Изд-во Урал. гос. экон. ун-та, 2003. –161 с.
6. Исследование операций: Т. 2 [Текст] / Под ред. Дж. Моудера, С. Элмаграби. –М.: Мир, 1981. –677 с.
7. **Аксёнов, К.А.** Задачи свертки и развертки мультиагентной модели [Текст] / К.А. Аксёнов, О.П. Аксёнова, Ван Кай // Теория графов и приложения. –Екатеринбург: Изд-во УрФУ, 2012. –С. 4–16.
8. **Ван Кай.** Использование аппарата операционного анализа вероятностных сетей для определения среднего количества приборов обслуживания мультиагентной модели [Электронный ресурс] / Ван Кай, К.А. Аксёнов, О.П. Аксёнова, М.В. Киселёва // Современные проблемы науки и образования. –2012. –№ 3. –Режим доступа: <http://www.science-education.ru/103-6290> (Дата обращения 22.05.2012)



АНАЛИЗ И ОПТИМИЗАЦИЯ ЭКОЛОГО-ЭКОНОМИЧЕСКИХ ПАРАМЕТРОВ ИННОВАЦИОННЫХ ПРОЕКТОВ

Проблема интенсификации инновационного развития экономики в настоящее время решается во многих странах. В Республике Беларусь для этих целей реализуется Государственная программа инновационного развития (ГПИР), призванная способствовать реализации наиболее актуальных и прорывных проектов.

Ежеквартальный мониторинг ГПИР осуществляется по таким общеэкономическим параметрам, как рентабельность, степень освоенности выделенных средств, внутренние затраты на исследования и разработки и др. Вместе с тем среди индикаторов реализации полностью отсутствуют экологические показатели, хотя, по нашему мнению, проект не может называться инновационным, если его реализация наносит существенный ущерб окружающей среде (ОС) и уровень использования природных ресурсов на единицу конечной продукции выше, чем у известных и используемых технологий.

Для решения обозначенной актуальной проблемы нами проанализированы существующие

подходы к разработке и обоснованию эколого-экономических параметров и предложены новые научные подходы, основанные на уравнениях балансов.

Предлагается использовать *материально-сырьевой баланс* (отражающий материально-сырьевую обеспеченность для получения продукции, выражающийся в натуральных (весовых) единицах), *материально-энергетический* (отражающий энергоемкость и отношение материально-энергетической обеспеченности технологических процессов к количеству выпускаемой продукции) и *производственно-экономический баланс* (отражающий общие производственно-экономические (финансовые) затраты, отнесенные к стоимости полученной конечной продукции).

На основании указанных уравнений балансов нами выведены новые показатели экологичности, такие, как природоемкость, безотходность, энергоемкость и др. [1]. При этом количественный анализ природоемкости может быть осуществлен по следующей формуле:

$$\sum_{i=1}^n I_{poj} = \frac{\sum_{i=1}^n Q_{ij}^{ис} [N_{ij} - (P_{ij} + H_{nij})] + \sum_{i=1}^n V_{oij}^{yo} + \sum_{i=1}^n V_{oij}^{в.пер}}{\sum_{i=1}^n Q_{ij}^{np}}, \quad (1)$$

где $\sum_{i=1}^n Q_{ij}^{ис}$ – нормированное количество использования исходного природного сырья j -го вида в i -м технологическом процессе, необходимого для получения продукции $\sum_{i=1}^n Q_{ij}^{np}$, включающего N_{ij} – норму расхода сырья j -го вида в i -м технологическом процессе; P_{ij} – чистый (полезный) расход сырья j -го вида в i -м технологическом процессе; H_{nij} – безвозвратные потери сырья j -го вида в i -м технологическом процессе; $\sum_{i=1}^n V_{oij}^{yo}$ – количество утилизируемых нормированных отходов из сырья j -го вида в i -м технологическом процессе, потенциально пригодного для вторичной переработки

и использования; $\sum_{i=1}^n V_{oij}^{в.пер}$ – количество отходов j -го вида в i -м технологическом процессе, поступающих на вторичную переработку в качестве вторичного сырья.

В основу разработки уравнений оценки и прогнозирования ресурсообеспеченности производств рационально применить теоретические методы с использованием весовых коэффициентов b_i . Суть подхода заключается в следующем. Определяется обобщенный и объективный фактор, отражающий наиболее важный производственно-экономический показатель, например, общий объем финансирования $Q_{об}$ ГПИР, и соответствующие объемы Q_j финансирования ее j -х структур-

ных подразделений (государственных заказчиков (ГЗ)). На этом основании можно определить количественные значения весовых коэффициентов каждого из всех ГЗ:

$$b_j = Q_j / Q_{об}; \sum_{j=1}^m b_j = 1, \quad (2)$$

где m – количество структурных подразделений (ГЗ) в составе ГПИР.

Информация о наиболее природоёмких и энергоёмких предприятиях и ГЗ, включенных в ГПИР, представлена в таблице.

Для сбора, хранения, обработки и отображения распределенной информации о параметрах инновационных проектов разработан и внедрен Интернет-портал, позволяющий пользователям осуществлять обратную связь с администратором, добавлять и редактировать новости, просматривать регламент и осуществлять статистическую обработку необходимых параметров. Экономический эффект от использования разработанного инструмента достигается за счет сокращения временных и трудовых затрат на пересылку и обработку параметров экологичности инновационных проектов. Практическая значимость выражается

в т. ч. в возможности для участников сопоставления результатов и в динамике, и по сравнению с другими пользователями Интернет-портала.

Предложенные новые научные принципы и программные средства анализа параметров экологичности инновационных проектов позволяют осуществлять выбор наиболее эффективных, что представляется актуальным в условиях ограниченности финансовых средств.

Так, по показателю природоёмкости в рамках ГПИР выделены, как имеющие наиболее высокий уровень, концерн «Белнефтехим», Министерство промышленности, Министерство архитектуры и строительства Республики Беларусь.

Ранжирование по уровню энергоёмкости позволило выявить министерства и ведомства, имеющие наименьшие значения: Государственный комитет по науке и технологиям, Министерство образования, Министерство культуры.

Вместе с тем при анализе эколого-экономических параметров часто встает задача прогнозирования. В реальных условиях риска и неопределенности использование вероятностных методов не всегда представляется обоснованным.

Наиболее природоёмкие и энергоёмкие предприятия в составе государственных заказчиков, включенных в ГПИР на 2009 г. – плановое количество проектов и объемы финансирования [4]

Наиболее природо- и энергоёмкие предприятия в составе государственных заказчиков	Плановое количество проектов в 2009 г.	Объем финансирования проектов в составе ГЗ, план, млн руб.	Весовые коэффициенты, b_i
1. Брестская область	22	199705	0,02661
2. Витебская область	37	21035	0,00280
3. Гродненская область	28	27909,3	0,00372
4. Гомельская область	29	39283,2	0,00524
5. Минская область	35	104015	0,01386
6. Минский горисполком	28	53470,6	0,00713
7. Могилевская область	18	80753	0,01076
8. Министерство промышленности	126	614638,9	0,08191
9. Министерство архитектуры и строительства	29	960216	0,12796
10. Министерство сельского хозяйства	20	1358154	0,18099
11. Министерство жилищно-коммунального хозяйства	5	447,1	0,00006
12. Министерство транспорта и коммуникаций	35	437621	0,05832
13. Концерн «Белнефтехим»	29	1105893	0,14738
14. ГПО «Белэнерго»	19	1331042	0,17738
15. Концерн «Беллесбумпром»	10	194224	0,02588
Всего	470	6528407,1	0,87



Для прогнозирования уровня обеспеченности природными ресурсами реализуемых инновационных проектов и их эколого-экономических параметров предлагается следующий метод.

1. Ретроспективный числовой ряд, содержащий фактические количественные показатели за истекшие периоды (не менее 7), обрабатывается с целью построения наиболее адекватной аппроксимирующей кривой. При этом из возможных моделей (линейной, логарифмической, полиномиальной, степенной и экспоненциальной), выбирается та, у которой величина достоверности аппроксимации (R) наибольшая.

2. На основании полученного на предыдущем шаге уравнения рассчитываем прогноз для предстоящих временных интервалов.

3. Рассчитанное прогнозное значение представляем в виде нечеткого множества: при $R^2 < 1$ форма функции принадлежности будет иметь треугольный вид, при $R^2 = 1$ – вертикальная линия. При этом основание треугольника будет равно произведению рассчитанного значения на удвоенную разность единицы и величины R^2 .

4. Для учета экспертной оценки формируем нечеткое множество, отражающее наиболее ожидаемое значение прогнозируемого параметра.

5. Результирующее значение прогноза в виде нечеткого множества C находим как сумму двух нечетких множеств A и B : $C = A \cup B$. При этом функция принадлежности множества C имеет вид:

$$\mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x) = \max(\mu_A(x), \mu_B(x)) \quad (3)$$

для каждого $x \in X$ (рисунок).

6. Оценка уровня неопределенности полученного прогноза в виде нечеткого множества может быть получена на основании значения $\gamma = 1 - \frac{S_0 - S}{S_0}$ или $\gamma = \frac{S}{S_0}$, $\gamma \in [0, 1]$, где S_0 – пло-

щадь под прямоугольной функцией принадлежности, равная ширине (т. к. множество нормализованное и высота прямоугольника равна единице); S – площадь под кривой функции принадлежности, полученной в результате расчетов. При этом большим значениям γ будет соответствовать больший уровень неопределенности.

7. Для получения четкого числа используем методы дефазификации – преобразования нечеткого множества к четкой форме – методы центра тяжести и среднего центра [2]. В случае если пересечением множеств A и B является непустое множество $D = A \cap B \neq \emptyset$, предпочтительнее использовать метод центра тяжести. Искомое значение $\bar{y}$ при этом рассчитывается как центр тяжести функции принадлежности $\mu_{B^k}(y)$ (тут и далее $B^1 = A, B^2 = B, N = 2$), то есть

$$\bar{y} = \frac{\int y \mu_{B^k}(y) dy}{\int \mu_{B^k}(y) dy} = \frac{\int y \max \mu_{B^k}(y)}{\int \max \mu_{B^k}(y)}, \quad (4)$$

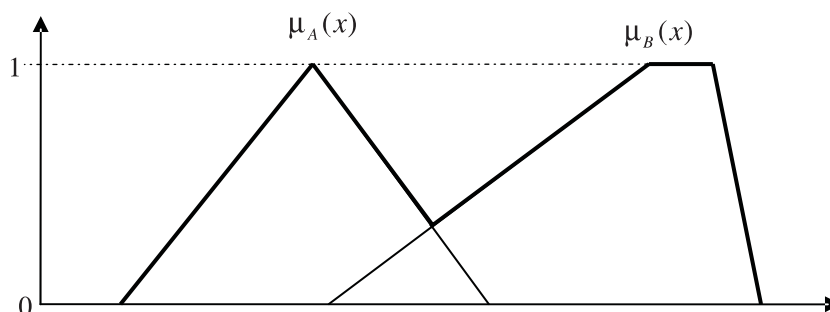
при условии, что оба интеграла в приведенном выражении существуют.

В случае непересечения множеств A и B для дефазификации используем метод по среднему центру. Значение $\bar{y}$ рассчитывается по формуле:

$$\bar{y} = \frac{\sum_{k=1}^N \mu_{B^k}(\bar{y}^k) \bar{y}^k}{\sum_{k=1}^N \mu_{B^k}(\bar{y}^k)}, \quad (5)$$

где $\bar{y}^k$ – это точка, в которой функция $\mu_{B^k}(y)$ принимает максимальное значение, то есть $\mu_{B^k}(\bar{y}^k) = \max_y \mu_{B^k}(y)$. Точка $\bar{y}^k$ называется центром нечеткого множества B^k . Обратим внимание, что значение $\bar{y}$ не зависит от формы и носителя функции принадлежности $\mu_{B^k}(y)$.

8. Для учета значимости экспертной оценки



Графическое представление операции суммирования нормализованных нечетких множеств

в результирующем прогнозном значении используем коэффициент $\lambda \in [0..1]$. Формула (3) примет вид:

$$\begin{aligned}\mu_{A \cup B}(x) &= \mu_A(x) \vee \mu_B(x) = \\ &= \max([1 - \lambda]\mu_A(x), \lambda\mu_B(x)).\end{aligned}\quad (6)$$

Для нормализации нечеткого множества используется формула:

$$\mu_{C_N} = \frac{\mu_C(x)}{h(C)}, \quad (7)$$

где h – высота нечеткого множества C .

Разработанная методика прогнозирования позволяет за счет использования теории нечетких множеств учесть фактор неопределенности и использовать экспертные оценки для получения адекватных экстраполяционных моделей. При этом разброс итоговых значений при различных функциях принадлежности исходных нечетких множеств составляет около 1 %.

Вместе с тем ТНМ в настоящее время широко применяется и для оценки экономической эффективности. Ниже перечислены основные преимущества нечетко-интервального подхода к оценке эффективности и риска инвестиционных проектов по сравнению с другими существующими методами [3].

- Данный подход позволяет формализовать в единой форме и использовать всю доступную неоднородную информацию (детерминированную, интервальную, статистическую, лингвистическую), что повышает достоверность и качество принимаемых стратегических решений.

- В отличие от интервального метода, нечетко-интервальный метод аналогично методу Монте–Карло формирует полный спектр возможных сценариев развития инвестиционных проектов (ИнП), а не только нижнюю и верхнюю границы. Таким образом, инвестиционное решение принимается не на основе двух оценок эффективности ИнП, а по всей совокупности оценок.

- Нечетко-интервальный метод позволяет получить ожидаемую эффективность ИнП как в виде точечного значения, так и в виде множества интервальных значений со своим распределением возможностей, характеризующимся функцией

принадлежности соответствующего нечеткого числа, что позволяет оценить интегральную меру возможности получения отрицательных результатов от ИнП, т. е. степень риска ИнП.

- Нечетко-интервальный метод не требует абсолютно точного задания функций принадлежности, т. к. в отличие от вероятностных методов результат, получаемый на основе нечетко-интервального метода, характеризуется низкой чувствительностью (высокой робастностью (устойчивостью)) к изменению вида функций принадлежности исходных нечетких чисел, что в реальных условиях низкого качества исходной информации делает применение данного метода более привлекательным.

- Вычисление оценок показателей ИнП на основе нечетко-интервального метода оказывается эффективным в ситуациях, когда исходная информация основана на малых статистических выборках, т. е. в случаях, когда вероятностные оценки не могут быть получены. Это всегда имеет место при предварительной оценке долгосрочных инвестиций и нередко – при последующем перспективном анализе, проводимом при отсутствии достаточной информационной базы.

- Реализация нечетко-интервального метода на основе интервальной арифметики предоставляет широкие возможности для применения данного метода в инвестиционном анализе, что обусловлено фактически отсутствием конкурентоспособных подходов к созданию надежного (в смысле гарантированности) и транспортабельного (по включению) инструментального средства для решения численных задач.

- Характеризуется простотой выявления экспертных знаний.

Таким образом, предложена новая технология анализа и оптимизации эколого-экономических параметров инновационных проектов, состоящая из разработанных методов и программных средств, что в совокупности позволяет повысить эффективность поддержки принимаемых решений в области рационального природопользования и охраны окружающей среды.

СПИСОК ЛИТЕРАТУРЫ

1. Рыбак, В.А. Антропогенная нагрузка на окружающую среду: количественная оценка, анализ, нормирование: Монография [Текст] / В.А. Рыбак. – Мн.: РИВШ, 2010. – 334 с.

2. Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы [Текст] / Д. Рутковская, М. Пилиньский, Л. Рутковский; Пер. с польск. – М.: Горячая линия – Телеком, 2006. – 452 с.



3. Деревянко, П.М. Сравнение нечеткого и имитационного подхода к моделированию деятельности предприятия в условиях неопределенности [Текст] / П.М. Деревянко // Современные проблемы экономики и управления народным хозяйством: Сб. науч. статей. –СПб.: Изд-во СПбГИЭУ, 2005. –Вып. 14. –С. 289–292.

4. Войтов, И.В. Методология развития инновационных производств на основе технологического прогнозирования и оценки использования природных ресурсов [Текст] / И.В. Войтов, М.А. Гатих, В.А. Рыбак, А.Л. Топольцев; под ред. И.В. Войтова. –Минск: Беларус. навука, 2012. –439 с.

УДК 338.45

В.И. Маслов, В.Ф. Минаков

КРИТЕРИЙ ЭФФЕКТИВНОСТИ В СИСТЕМЕ УПРАВЛЕНИЯ КАЧЕСТВОМ ПРОДУКЦИИ ПРЕДПРИЯТИЯ

На предприятиях качество продукции может быть достигнуто на основе использования принципов всеобщего менеджмента качества (TQM) [1, 2]. Его принципы ориентированы на управление качеством продукции предприятий реального сектора экономики, обеспечивающим конкурентные условия проникновения на рынок. Однако нельзя не заметить, что принципы TQM по своим целеполаганию и сущности реализуют парадигму «качество любой ценой». При этом цена качества может оказаться столь высокой, что его достижение окажется экономически неэффективным. Действительно, достаточно вспомнить «потерянное десятилетие» Японии, чтобы понять, что проблема требует разрешения. Качество продукции ее предприятий, экспортируемой на мировые рынки, безусловно, превосходит аналоги других стран-производителей. Вместе с тем в течение десяти лет повышенное качество обеспечивается непомерно высокой ценой продукции. В результате спрос на продукцию повышенного качества не окупает затрат на ее обеспечение, а экономическая эффективность деятельности предприятий оказывается низкой. Более того, проникновение на мировые рынки продукции Китая более низкого качества оказывается экономически эффективным для производителей, обеспечивает стране экономический рост. Объемы промышленного производства в этот же период показывают беспрецедентный рост. Экономика Китая становится второй после США. Японская экономика теряет позиции второй экономики мира с 2010 г. по объемам годового ВВП.

Постановка задачи. Требуется разработка критерия экономической эффективности обеспечения качества производимой предприятиями продукции. Именно на основе использования такого критерия должна быть обеспечена окупаемость и прибыльность мероприятий, направленных на повышение качества продукции.

Эластичность качества. Предлагается введение двух показателей эластичности качества: по цене (ценовая эластичность качества) и по затратам [3]. Эластичность качества по цене равна отношению относительного приращения качества как степени соответствия продукции уровню заданного (или наилучшего) значения свойства продукта, отражающего его качество, к относительному изменению цены, при которой сохраняется возможность реализации продукта на конкурентном рынке. Кроме того, эластичность качества по затратам является метрикой производства, равной отношению относительного приращения качества к относительному изменению затрат, при которых обеспечивается такое качество.

Обозначим показатель по i -й характеристике качества – Q_i . Тогда эластичность качества по цене может быть определена как

$$E_{i,D} = \frac{\Delta Q_i / Q_i}{\Delta P_i / P_i}, \quad (1)$$

а эластичность качества по затратам:

$$E_{i,S} = \frac{\Delta Q_i / Q_i}{\Delta T_i / T_i}, \quad (2)$$

где ΔQ_i – абсолютное изменение показателя каче-

ства Q_i ; $\Delta Q_i / Q_i$ – относительное изменение показателя качества ΔQ_i ; ΔP_i – абсолютное изменение цены (второй случай – затрат) при повышении показателя качества Q_i на величину ΔQ_i ; ΔT_i – абсолютное изменение затрат при повышении показателя качества Q_i на величину ΔQ_i ; $\Delta P_i / P_i$ – относительное изменение цены при повышении показателя качества Q_i на величину ΔQ_i ; $\Delta T_i / T_i$ – относительное изменение затрат при повышении показателя качества Q_i на величину ΔQ_i .

Точечная ценовая эластичность качества может быть представлена в дифференциальной форме:

$$E_{i,D,d} = \frac{dQ_i / Q_i}{dP_i / P_i}, \quad (3)$$

а точечная эластичность качества по затратам:

$$E_{i,S,d} = \frac{dQ_i / Q_i}{dT_i / T_i}. \quad (4)$$

Ценовая эластичность качества в сфере спроса (или спроса на качество) прежде всего показывает готовность потребителей платить большую цену за продукцию более высокого качества. Как правило, сама зависимость показателей качества от цены имеет нелинейный растущий характер. Например, для автомобильных свеч зажигания [4] такая зависимость приведена на рис. 1. В качестве показателя качества в рассматриваемом примере выступает такая характеристика, как степень отклонения условий работы свеч, при ко-

торых обеспечивается их работоспособность (наличие искры). В относительных единицах (о. е.) единичным значением является максимальная, достигаемая лучшими марками свеч зажигания, величина. Цена, по которой возможна реализация продукции в рыночных условиях, растет с повышением качества, однако темп роста цены по мере повышения качества снижается. Именно это свойство рынка и отражает ценовая эластичность качества в части спроса на него.

Представленная на рис. 1 зависимость цены спроса на качество характеризует спрос, при котором потребители готовы платить повышенную цену за качество. Важно заметить, что качество может быть совершенно эластичным по цене. Это в полной мере относится к качеству электроэнергии. На рынке электроэнергии заключаемые договоры не содержат повышенных по сравнению с минимальными, регламентируемыми государственным стандартом, показателями, требований. Отсутствие спроса на повышенное качество электроэнергии обуславливает и независимость ее цены от качества. Графически неэластичный спрос на качество можно отобразить горизонтальной прямой в координатах цена – качество.

Кроме того, ценовая эластичность качества при производстве более высококачественных продуктов (предложения качества) показывает степень роста затрат на повышение качества. На рис. 2 представлены кривые роста затрат и

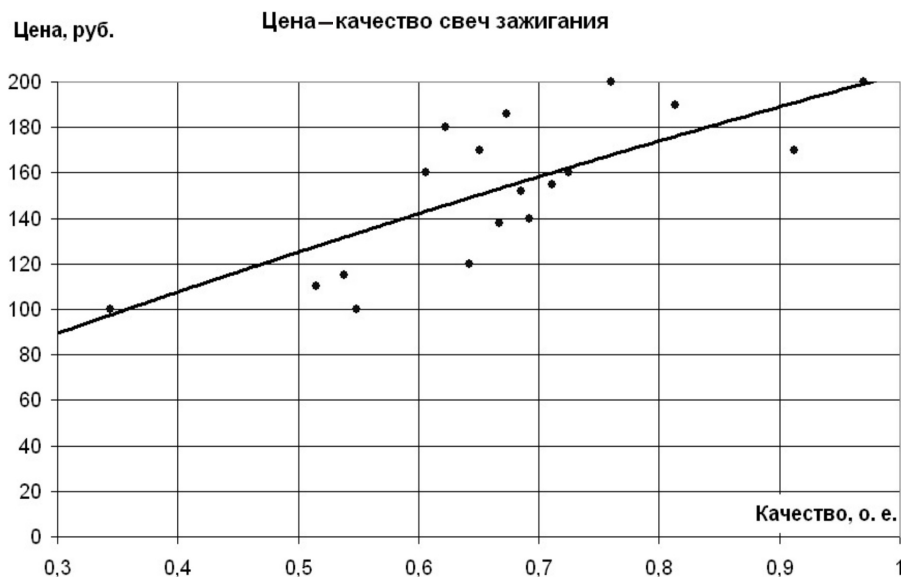


Рис. 1. Зависимость цены продукции от показателя качества

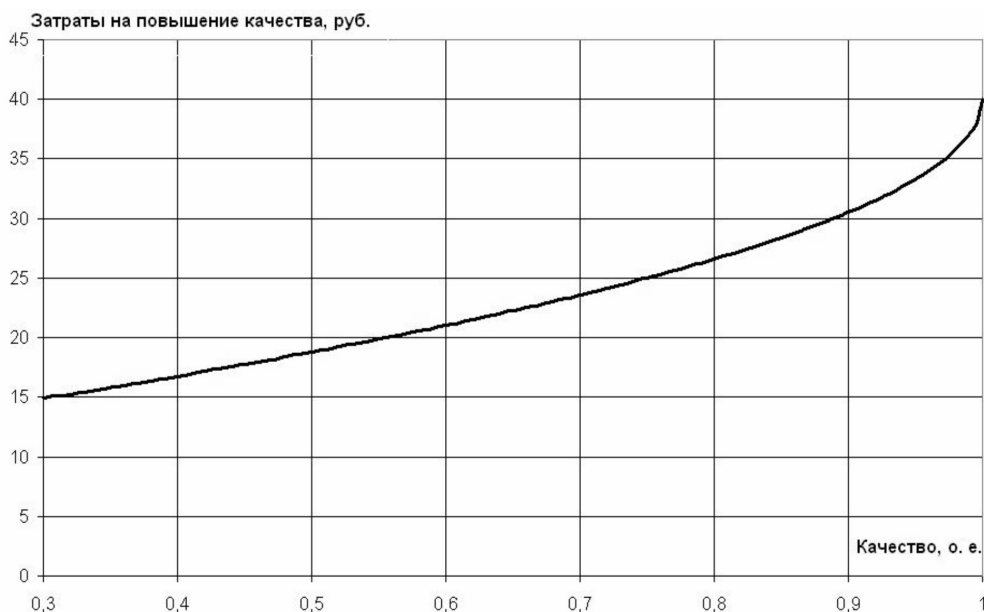


Рис. 2. Зависимость затрат от показателя качества продукции

цены изделий при повышении качества, например, машиностроительной продукции [5–8].

Из сопоставления динамики затрат и цен можно сделать следующие выводы. Во-первых, по мере приближения качества производимой продукции к максимальному уровню (достижаемому к настоящему времени либо предельно возможному) темп роста цены реализации замедляется, в то время как темп роста затрат, наоборот, повышается. Во-вторых, превышение вторых над первыми из упомянутых темпов роста делает экономически неэффективной стратегию повышения качества теми же мероприятиями, которые использовались до точки эффективности. Важно отметить, что экономическая эффективность не всегда является критерием эффективности реализуемой стратегии управления качеством.

С учетом отмеченных особенностей динамики роста затрат на производство более качественных изделий и цены их реализации, можно сформулировать математически строгий критерий экономической целесообразности перехода производителя на более высокое качество продукции: ценовая эластичность предложения качества по затратам на качество должна быть больше или равна эластичности спроса на качество

$$E_{i,S} \geq E_{i,D}. \quad (5)$$

На основе сопоставления ценовых эластичностей качества по затратам конкурирующих пред-

приятий оценивается наличие потенциала повышения качества предприятия, либо отсутствие такового.

Критическая точка равенства эластичности качества по цене и по затратам характеризует такой уровень качества продукции, при котором рост финансового обеспечения качества в точности совпадает с ростом цены на более качественную продукцию. Более высокое качество не приводит к росту доходности производителя. При качестве ниже указанной точки имеется возможность не только повышать качество, но и реализуется такое управление качеством, при котором одновременно повышается доходность более качественной продукции (рис. 3).

Очевидно, что спрос на качество (рис. 4) может быть:

- эластичным, при котором численные значения эластичности больше единицы,

$$E_{i,D} \geq 1; \quad (6)$$

- неэластичным, при котором численные значения эластичности меньше единицы (чаще всего при высоком уровне качества продуктов),

$$E_{i,D} \leq 1; \quad (7)$$

- с единичной эластичностью качества по цене, при которой повышение качества изделий пропорционально повышению цены реализации с сохранением спроса,

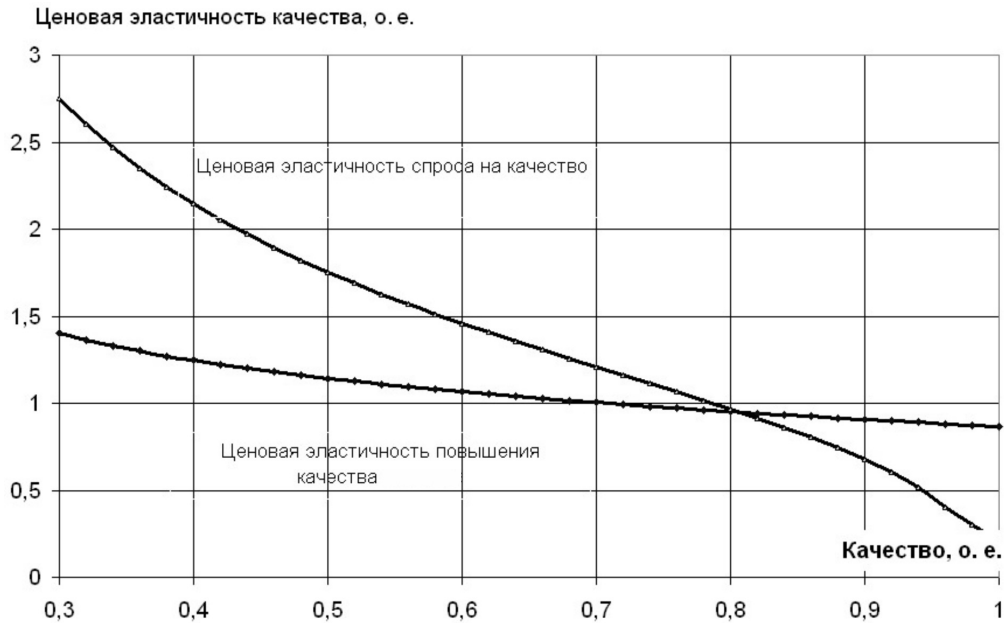


Рис. 3. Эластичности качества продукции по цене и затратам

$$E_{i,D} = 1; \quad (8)$$

• с нулевой эластичностью, при которой цена реализации не зависит от качества, например, для электроэнергии при условии соблюдения допустимых границ качества (определенных ГОСТ 13109-7 в редакции 2002 г. [9] как максимально допустимые отклонения напряжения, частоты, формы напряжения и др.)

$$E_{i,D} = 0. \quad (9)$$

Аналогичными могут быть и эластичности качества по затратам:

$$E_{i,S} \geq 1, \quad (10)$$

$$E_{i,S} \leq 1, \quad (11)$$

$$E_{i,S} = 1, \quad (12)$$

$$E_{i,S} = 0. \quad (13)$$

Графически зоны эластичного и неэластично-

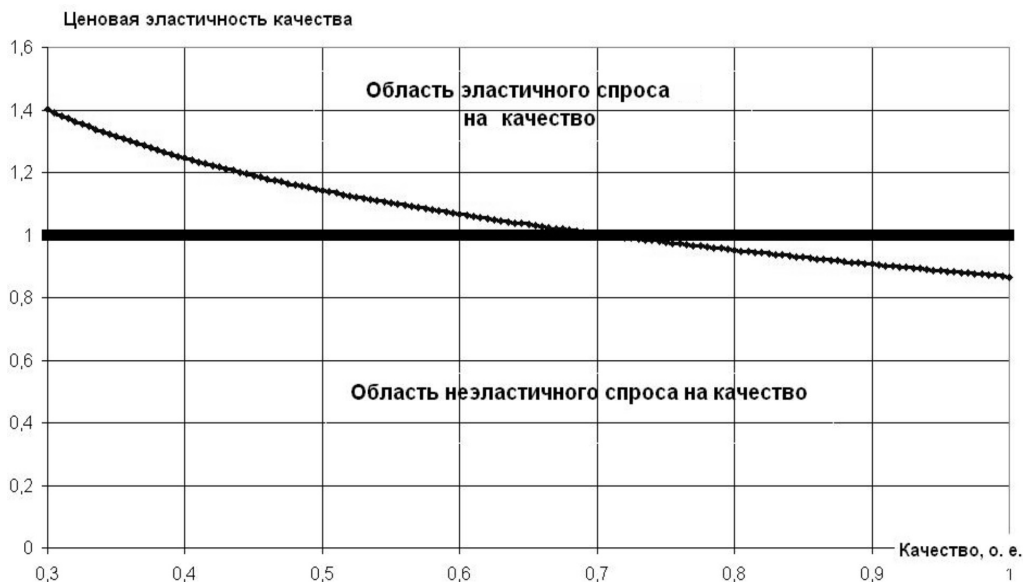


Рис. 4. Виды эластичности спроса на качество

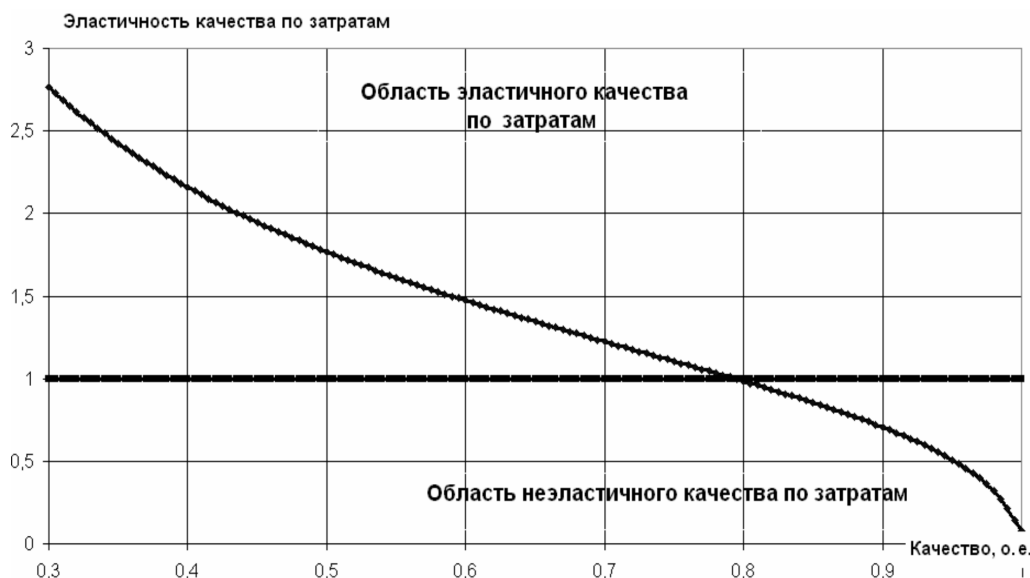


Рис. 5. Виды эластичности качества по затратам

го спроса на качество продукции приведены на рис. 4, а эластичного и неэластичного роста затрат — на рис. 5.

Из сопоставления динамики эластичности качества свеч зажигания с единичной границей видно, что один и тот же продукт может переходить из зоны эластичного спроса и предложения качества в зону неэластичного в функции уровня качества продукта. В рассматриваемом для современного рынка конкретном примере область высокого качества характеризуется снижением эластичности качества по затратам. Это является особенностью управления качеством, которую можно переформулировать в простое правило: ряд рыночных продуктов с повышением их качества требует увеличения темпов роста затрат по сравнению темпами их роста, достаточными для достижения текущего качества. Или иначе: каждый последующий процент прироста качества требует большего роста затрат, чем предыдущий. Для таких продуктов важными являются численные значения эластичности качества по затратам (рис. 4), т. к. позволяют производителям соотносить их с возможностями предприятий.

На основе сопоставления ценовых эластичностей предложения качества конкурирующими предприятиями оценивается наличие потенциала повышения качества i -го предприятия по сравнению с потенциалом j -го, что имеет место при

$$E_{i,D} < E_{j,D}, \quad (14)$$

либо отсутствие такового при

$$E_{i,D} \geq E_{j,D}. \quad (15)$$

Важно отметить, что для оценки стратегии повышения качества продукции необходимо сопоставление с аналогами продукции, но иного уровня качества. Так, традиционные лидеры производства высококачественных телевизоров Sony, Panasonic уступили свои позиции на мировых рынках корейским, а к настоящему времени — и китайским компаниям [10, 11]. Именно последние заняли позиции лидеров производства и поставок телевизоров в Азию, Европу, Америку, несмотря на отставание в качестве своей продукции. Точно так же дело обстоит на мировых рынках таких видов продукции, как радиоприемники, мобильные телефоны, стиральные и швейные машины, велосипеды, мотоциклы, часы, фотоаппараты. Китайская стратегия развития собственного рынка автомобилестроения позволила ему обойти не только Японию, но и США. Список можно продолжать еще большим числом товаров с меньшей степенью переработки. Учесть конкуренцию товаров различного качества позволяет перекрестная эластичность качества:

по цене

$$E_{ij,D} = \frac{\Delta Q_i / Q_i}{\Delta P_j / P_j}, \quad (16)$$

по затратам

$$E_{ij,S} = \frac{\Delta Q_i / Q_i}{\Delta T_j / T_j}, \quad (17)$$

где индексы i, j относятся соответственно к i -му, j -му производителям (и их продукции).

В точечном представлении:

$$E_{ij,D,d} = \frac{dQ_i / Q_i}{dP_j / P_j}, \quad (18)$$

$$E_{ij,S,d} = \frac{dQ_i / Q_i}{dT_j / T_j}. \quad (19)$$

Перекрестные эластичности качества отражают факт влияния качества, затрат и спроса на него производителей-конкурентов на рассматриваемое предприятие и наоборот.

Предложенный количественный подход к показателям качества продуктов, основанный на соотношении роста затрат на их обеспечение с ростом цены рыночной реализации продукции, переводит проблему эффективного управления качеством в плоскость новых координат: качество – экономическая эффективность его повышения.

Предложен показатель качества продукции – ценовая эластичность качества. Рыночный спрос на качество характеризуется величиной его эла-

стичности по цене продукции – соотношению относительного приращения качества и относительного приращения цены реализации товаров и услуг повышенного качества. Предложение производителями продукции повышенного качества характеризуется производной относительного приращения качества по относительному приращению затрат на обеспечение повышения качества. Ценовая эластичность качества по затратам является метрикой потенциала предприятия в части экономичности управления качеством и его обеспечения, а также оценкой его конкурентных позиций на рынке аналогов продукции. Является изменяющейся величиной в зависимости от достигнутого уровня качества выпускаемой продукции и затрат на его повышение. Превышение эластичности качества по затратам одного производителя по сравнению с другим служит количественным критерием его конкурентных преимуществ в реализации стратегии повышения качества продукции. Превышение эластичности качества по цене над эластичностью качества по затратам является критерием экономической эффективности мероприятий, обеспечивающих повышение качества продукции предприятия.

СПИСОК ЛИТЕРАТУРЫ

1. Международный стандарт ISO 9001: 2008. Системы менеджмента качества. Требования. (Quality management systems – Requirement). Для учебных целей [Текст]. – М.: Госстандарт России, 2008. – 39 с.
2. **Стивен, Дж.К.** Всеобщее управление качеством (TQM) [Текст] / Дж.К. Стивен. – СПб.: Изд-во «Санкт-Петербург», 2002. – 256 с.
3. **Маслов, В.И.** Эластичность качества по цене и затратам [Текст] / В.И. Маслов, В.Ф. Минаков // Стандарты и качество. – 2012. – № 9. – С. 88–90.
4. **Шмелева, А.Н.** Оценка и повышение операционной эффективности системы менеджмента качества предприятия: теория, методология, практика: Автореф. дис. ... д-ра экон. наук: 08.00.05 [Текст] / А.Н. Шмелева. – Тамбов: Тамбовский гос. технич. ун-т, 2011. – 41 с.
5. **Маслов, В.И.** Повышение качества производства поршневых колец [Текст] / В.И. Маслов, А.И. Арустамян // Научно-технические ведомости СПбГПУ. Наука и образование. – 2012. – № 1 (142). – С. 132–136.
6. **Маслов, В.И.** Математическая модель спектральных характеристик излучений и датчиков [Текст] / В.И. Маслов, А.И. Арустамян, В.Ф. Минаков // Матер. XII Междунар. науч.-метод. конф. Информатика: проблемы, методология, технологии. Воронеж 8-10 февраля 2012 г. – Воронеж: Издат.-полиграф. центр Воронежского гос. ун-та, 2012. – С. 241–242.
7. **Маслов, В.И.** Оптический контроль в системе управления качеством производства поршневых колец [Текст] / В.И. Маслов, А.И. Арустамян, В.Ф. Минаков // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 1 (77). – С. 16–20.
8. **Маслов, В.И.** WEB-сервисы в системе управления качеством машиностроительной продукции [Текст] / В.И. Маслов, А.И. Арустамян, В.Ф. Минаков // Современное машиностроение. Наука и образование: Матер. II Междунар. науч.-практич. конф. – СПб.: Изд-во политехн. ун-та, 2012. – С. 472 – 478.
9. ГОСТ 13109-97. Электрическая энергия. Совместимость технических средств электромагнитная. Нормы качества электроэнергии в системах электроснабжения общего назначения [Текст]. – М.: Госстандарт России. – 2002. – 33 с.
10. Экономика Японии [Электронный ресурс] / Режим доступа: http://ru.wikipedia.org/wiki/Экономика_Японии (Дата обращения 29.09.2012)
11. Экономика Китайской Народной Республики [Электронный ресурс] / Режим доступа: http://ru.wikipedia.org/wiki/Экономика_Китайской_Народной_Республики (Дата обращения 29.09.2012)



УДК 378.14.014.13, 303.433.3

С.В. Калмыкова, А.С. Соколицин

ОПТИМИЗАЦИЯ СОДЕРЖАНИЯ И СТРУКТУРЫ УЧЕБНОГО ПРОЦЕССА В ВУЗЕ С ИСПОЛЬЗОВАНИЕМ МЕТАМОДЕЛИ

Новые образовательные технологии формируются как результат научных открытий и исследований. Так, развитие кибернетики и вычислительной техники обусловило создание и развитие программированного обучения; результаты исследований закономерностей развития человеческого мышления привели к развитию проблемного обучения. Наиболее полно эти принципы могут быть реализованы посредством внедрения дистанционных технологий обучения в традиционный учебный процесс. При этом необходимо, чтобы под воздействием внешних факторов процесс образования имел гибкую структуру и был рационален с точки зрения системы управления: позволял бы достигать соответствующего результата в максимально короткие сроки и с определенными затратами. Вопрос баланса между интегративной и рациональной составляющей имеет отношение не только к образовательному процессу, но и к любому управленческому.

Внедрение дистанционных технологий обучения во все формы невозможно без следующего:

анализа информационно-профессиональной сущности задач;

моделирования междисциплинарного взаимодействия;

дифференциации дисциплин на отдельные составляющие;

решения задачи определения оптимальной последовательности изучения отдельных модулей курса (разработки алгоритма);

решения задачи эффективности усвоения учебного материала (разработки математической модели и эвристического алгоритма).

Внедрение дистанционных технологий позволяет «расширить» возможность «практики», развивать коллаборативное взаимодействие участников учебного процесса, реализуя индивидуальную и коллективную деятельность и ролевые игры.

Следует отметить, что внедрение дистанционных технологий позволяет реализовать учебный процесс без увеличения часов аудиторной работы, что весьма актуально, т. к. он регулируется рамками ФГОС.

Как уже отмечалось, тенденция сегодняшнего общества – переход от дискретно-ступенчатых форм образования к целостной системе (непрерывного) образования, обеспечивающей потребности человека в различных образовательных траекториях. Таким образом, учебный процесс определяется как процесс непрерывно-дискретного типа, поскольку функция реализации учебного процесса является непрерывной функцией от двух аргументов: времени (непрерывный аргумент) и структурной единицы (дискретный аргумент).

В моделировании данной задачи используется в основном одна и та же формализованная математическая модель. Задача выбора оптимальной реализации учебного процесса сводится к распределению заданного набора структурных единиц по заданным организационно-временным этапам в соответствии с определенными ограничениями, в т. ч. с указанием количества времени, отводимого на выполнение каждой структурной единицы конкретного этапа.

Предпосылками для реализации оптимальной модели обучения, в рамках соответствующего образовательного стандарта, являются:

1) возможность представления содержания учебного курса в виде множества отдельных элементов;

2) ранжирование этих элементов;

3) ранжирование учебных дисциплин в соответствии с учебными планами;

4) выявление междисциплинарных взаимосвязей;

5) построение модели взаимосвязи между дисциплинами на основе словесно-формульного алгоритма;

6) выявление повторяющихся или «транзитных» путей на спроектированной модели;

7) применение критериев программированного обучения в «узлах» модели для ее оптимизации;

8) формирование обобщенного алгоритма оптимизации обучения.

При построении модели следует учитывать следующие ограничения:

- в модели должны быть представлены все структурные единицы учебного процесса;
- ранжирование структурных единиц необходимо осуществлять как в пределах выделенных временных интервалов, так и в их совокупности;
- любая совокупность структурных единиц на каждом выделенном временном этапе не должна повторяться (т. е. любая совокупность структурных единиц может быть представлена на каждом временном этапе только одной подоперацией);
- в модели должны быть представлены все временные этапы учебного процесса;
- суммарное количество временных единиц должно соответствовать общему времени, отведенному на их изучение (полное выполнение данной технологической операции, а суммарная длительность изучения структурных единиц каждого этапа должна соответствовать его продолжительности).

В общем виде целевая функция решаемой задачи имеет следующий вид:

$$\text{OPT}(I) = \sum_{i=1}^n l(d_i), \dots, d_i \in D^-, \dots, n = |D^-|,$$

$$D^- = \{d \mid d \in D^{\text{SPEC1}}, \dots, D^{\text{SPEC}P}; l(d)^{\text{SPEC1}} = \dots = l(d)^{\text{SPEC}P}; k(d)^{\text{SPEC1}} = \dots = k(d)^{\text{SPEC}P}\},$$

где D^{SPEC} – множество структурных единиц соответствующего направления; $k(d)^{\text{SPEC}i}$, $l(d)^{\text{SPEC}i}$ – параметры структурных единиц; D^- – множество всех структурных единиц учебного процесса; I – задача, для которой ищется оптимальное решение.

Целевая функция в отсутствие ограничений определяет максимально возможную теоретическую оценку оптимального решения $\text{OPT}_T(I)$,

представляющую сумму длительностей общих структурных единиц, таких, для которых существует непустое пересечение множеств допустимых семестров для всех специальностей.

Логическая сеть структурных единиц представлена на рис. 1.

$$T(D_{26}^{4}) > T(D_{10}^{26})$$

$$T(D_{10}^{29}) < T(D_{26}^{34}),$$

где T – продолжительность периода времени от начала обучения; D_n^i – индекс, обозначающий, что данный период времени относится к изучению i -й темы в n -й дисциплине.

Анализ рисунка показывает, что изучение четвертой темы в 26-й дисциплине должно начаться не ранее окончания 26-й темы в десятой дисциплине.

Реализация совокупности логических цепей представляет собой граф реализуемого учебного процесса. Он определяет взаимное расположение структурных единиц, при котором обеспечивается требуемая логическая последовательность изучения всего материала. Процесс формирования такого графа подразумевает исключение значительного количества лишних связей и узлов при сохранении полного отображения требуемой логической последовательности изучения всего учебного материала.

Для получения минимизированного графа учебного процесса необходимо исключить все лишние узлы и связи (рис. 2). С этой целью вводится понятие эквивалентной связи. Эквивалентными называются связи между двумя узлами, имеющие различные пути.

Критериями формирования целостного пред-

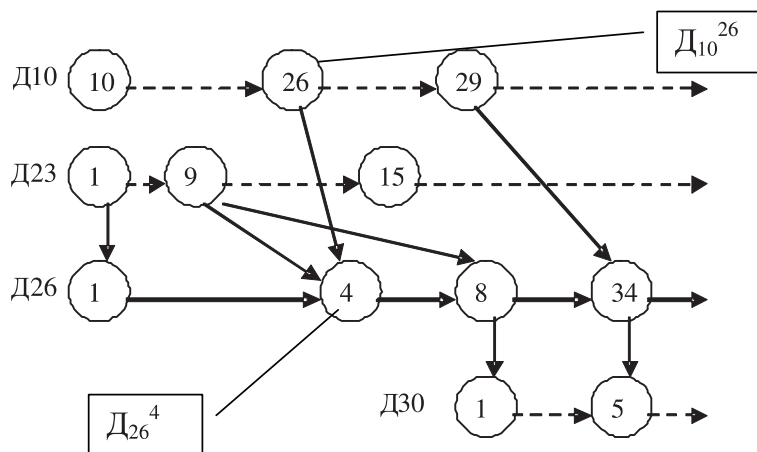


Рис. 1. Последовательность изложения материала

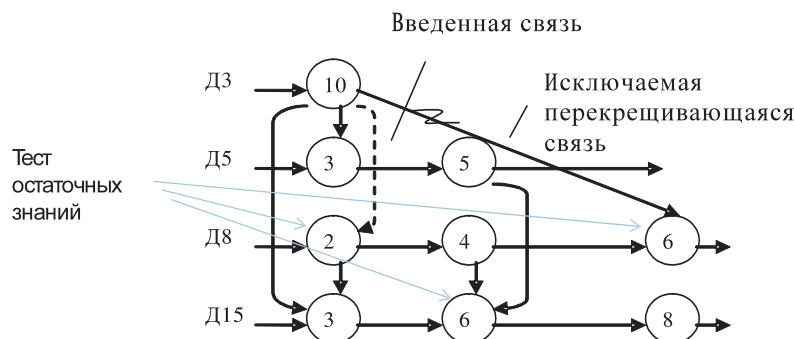


Рис. 2. Выявление исключаемых связей

ставления об оптимальной структуре образовательного процесса являются вводимые на этапе построения итогового графа показатели результатов непрерывного тестирования, формируемые как по линейной методике, так и с помощью других алгоритмов программированного обучения. Таким образом, задача построения итогового графа заключается в нахождении оптимального сочетания оптимальной задачи планирования с оценкой качества восприятия информации при слиянии узлов графов при сформированных ограничениях.

С целью выработки управленческого критерия на уровне слияния узлов графов проведен анализ некоторых алгоритмов программированного обучения [1], в т. ч.:

- линейного алгоритма Скиннера;
- разветвленного алгоритма Краудера;
- шэффилдского метода;
- блочного (варшавского) метода;
- адаптивной системы программированного обучения.

Алгоритм Скиннера включает в себя несколько принципов:

принцип малых шагов, когда учебный материал делится на мелкие части;

открытых вопросов – для проверки усвоения материала;

индивидуализации темпа обучения;

принцип дифференцированного закрепления знаний (основное повторяется несколько раз в разных разделах для лучшего усвоения).

Алгоритм Краудера представляет собой разветвленный алгоритм программированного обучения. В зависимости от реакции обучающегося могут реализовываться различные пути прохождения алгоритма. Количество набранных баллов может являться оценкой оптимальности его про-

хождения.

Шэффилдский и блочный методы являются смешанными. Основными принципами шэффилдского метода являются возможности разделения потока информации на разные порции или шаги, проведения открытых и закрытых тестов, задействования механизма обратной связи для повторения отдельных блоков теории, т. е. возможность объединения теории с практикой в учебном процессе. Аналогично рассмотренному ранее алгоритму Краудера, количество набранных баллов может являться оценкой оптимальности прохождения.

Адаптивная система предполагает переходы как на более высокий, так и на более низкий уровень исходя из критерия, основанного на количестве сделанных ошибок. Такая система нередко используется как средство выбора разделов курса, требующих дополнительного изучения, и может быть использована с целью «рейтингового» выбора оптимального уровня сложности.

Анализ различных методов программированного обучения показал, что для формирования критериев «перехода» к следующему этапу образовательной программы наиболее предпочтительным является смешанный алгоритм, т. к. в этом случае мы получаем возможность наиболее разносторонней оценки знаний.

Как частный случай программированного контроля при решении задачи оптимизации содержания учебного процесса, нами разработан алгоритм системы самоконтроля, позволяющей оценивать качество образования на различных уровнях [2]. Самотестирование является частным случаем программированного контроля, поэтому следует учитывать факторы, определяющие его эффективность: принципы подбора заданий по признакам первостепенности и сложности; фор-

му подачи вопросов и ответов, исключающую дополнительные пояснения и, наконец, схему определения уровня сложности. Использование в программе самотестирования наглядных шкал оценок и ограничений (как на время контроля, так и на количество вопросов) позволяет рассматривать алгоритм ее реализации как прототип алгоритма формирования управленческого критерия при решении поставленной задачи. Целью являлась выработка принципов построения и, соответственно, алгоритма программы самотестирования. На формируемый алгоритм были следующие ограничения: на количество вопросов, время самоконтроля, количество уровней сложности, оценку (все неправильные ответы не оцениваются).

В алгоритме использована прогрессирующая шкала оценивания в зависимости от уровня сложности вопроса. Это позволило корректно оценивать ситуацию и сформировать наивысшую оценку как оценку за наиболее эффективное решение. Для получения итоговой оценки (оценки набранного количества баллов) используются различные элементы рассмотренных ранее алгоритмов. Эффективность программы обусловлена дискретностью подаваемого материала на каждом из определенных уровней (этапов) самотестирования и предусматривает переходы к линейной системе программирования в процессе работы алгоритма. Под этапом понимается блок вопросов, уровень каждого из которых определяется алгоритмом. В каждом блоке вопросов соблюдается ряд требований, позволяющих минимизировать ошибки угадывания. К ним, например, относится рандомизация – случайный выбор вопроса из совокупности вопросов одного уровня. В предлагаемом алгоритме рандомизация не используется только при линейном алгоритме на самом низшем (нулевом) уровне. Следует отметить, что возможные повторения вопросов (при их случайном выборе из базы) на высоких уровнях не противоречат дидактическим принципам программированного контроля.

Методы формирования критерия принятия управленческого решения о переходе на следующий уровень (вопросов или блоков) определяются следующим образом:

A_j^a – состояние, подготовка к первому вопросу блока a ; j – уровень сложности вопроса (начало тестирования: A_1^1 – подготовка к вопросу первого уровня сложности); $A_{ij}(n)$ – событие, как ответ на вопрос i -го уровня; j – уровень сложно-

сти следующего вопроса; n – количество баллов, полученных за ответ.

Если ответ правильный, то $j = i + 1$ при $i = 0, 1, 2$ и $j = i$ при $i = 3$.

Как пример: $A_{33}^*(8)$ – правильный ответ (оценка «хорошо»); $A_{33}(16)$ – правильный ответ (оценка «отлично»), в скобках указана балльность.

Если ответ неверный, то $j = i - 1$ при $i = 1, 2, 3$, $j = i$ при $i = 0 (n = 0)$.

Переходы: « $\rightarrow$ » – переход от состояния к состоянию или от события к событию; « $\Leftarrow$ » – переход от события к событию с повышением уровня сложности; « $\Leftarrow$ » – аналогичный переход с понижением уровня сложности.

$\left(\sum_{k=1}^{k=4} n_k, x_1 \right)_j^a$ – состояние, достигнутое после прохождения блока a ; $k = 1, 2, 3, 4$ – номера вопросов; n_k – количество баллов, полученных за ответ на k -й вопрос; $x_1 = 0, 1, 2, 3, 4$ – количество правильных ответов; j – уровень первого вопроса следующего блока (A_j^{a+1}).

$L_j(x_1, x_2, 16 - (x_1 + x_2))$ – событие, начало линейного тестирования на уровне j ; x_1 – засчитываемое к началу тестирования количество правильных ответов; x_2 – засчитываемое к началу тестирования количество неправильных ответов; $16 - (x_1 + x_2)$ – количество вопросов линейного теста.

$A_{ij}(n)$ – событие, как ответ на вопрос j -го уровня линейного тестирования с таким же уровнем сложности следующего вопроса; n – количество баллов, полученных за ответ. Если ответ правильный, то $n = 1$, если неверный, то $n = 0$.

S – состояние подготовки к оценке результатов; $S(L)$ – состояние подготовки к оценке результатов линейного тестирования.

Одному состоянию $\left(\sum_{k=1}^{k=4} n_k, x_1 \right)_j^a$ могут соответствовать несколько вариантов последовательности событий. Например: $(30, 4)_3^1 = A_1^1 \rightarrow A_{12}(2) \Leftarrow A_{23}(4) \Leftarrow A_{33}(16) \rightarrow A_{33}^*(8) \rightarrow A_3^2$.

Это же состояние достигается при ином пути: $(30, 4)_3^1 = A_1^1 \rightarrow A_{12}(2) \Leftarrow A_{23}(4) \Leftarrow A_{33}^*(8) \rightarrow A_{33}(16) \rightarrow A_3^2$.

При прохождении линейного теста $L_j(x_1, x_2, 16 - (x_1 + x_2))$ предлагается ответить на $16 - (x_1 + x_2)$ вопросов уровня тестирования j (см. табл.). При расчете оценки результатов (правильный ответ на вопрос любого уровня – 1 балл, неверный ответ на вопрос любого уровня – 0 баллов) учитывается предыстория.

Схема возможных событий при прохождении линейного теста ($L_j(x_1, x_2, 16 - (x_1 + x_2))$)

		Вариант ответа на вопросы				Конец теста
		$x_1 + x_2 + 1$	$x_1 + x_2 + 2$		16	
Уровень вопросов j	$L_j(...)$	$A_{jj}(1) \rightarrow$ $A_{jj}^0(0) \rightarrow$	$A_{jj}(1) \rightarrow$ $A_{jj}^0(0) \rightarrow$	$A_{jj}(1) \rightarrow$ $A_{jj}^0(0) \rightarrow$	$A_{jj}(1) \rightarrow$ $A_{jj}^0(0) \rightarrow$	$S(L)$

Примечание:

1) x_1 – засчитываемое к началу линейного теста количество правильных ответов; x_2 – засчитываемое к началу линейного теста количество неверных ответов.

2) Список возможных событий: $A_{jj}(1)$, $A_{jj}^0(0)$.

Структурная схема части реализованного алгоритма приведена на рис. 3, где БАЗА ВОПРОСОВ/ОТВЕТОВ – структурированная по уровням база вопросов/ответов; БВ1-БВ3 – блоки вопросов/ответов; БА1-БА3 – блоки анализа вопросов/ответов; БЛ – блок перехода к линейному тестированию; БВЛ – блок вопросов/ответов линейного тестирования; БО – блок оценки тестирования; БВ1-БВ3 обеспечивают обращение к базе вопросов/ответов. Эти блоки производят анализ ответов на вопросы и выдачу очередных вопросов необходимого уровня с учетом временных ограничений.

Блоки анализа вопросов/ответов БА1, БА2 обеспечивают анализ состояния после прохождения соответствующего блока и разрешают переход к следующему блоку вопросов или переход к блоку линейного тестирования БЛ. Блок БА3 обеспечивает анализ состояния после прохождения третьего блока вопросов согласно таблице. Он разрешает переход или к блоку оценки тестирования БО или к блоку линейного тестирования БЛ.

БЛ обеспечивает исходные данные для линейного тестирования, т. е. определяет его уровень и количество зачтенных ранее вопросов, учитывая информацию от БА1-БА3.

Блок вопросов/ответов линейного тестирования БВЛ обеспечивает обращение к базе вопросов/ответов с целью выбора необходимых вариантов.

Блок оценки тестирования БО обеспечивает подсчет оценочных параметров, выбор и выдачу оценки-рекомендации.

Таким образом, решение поставленной задачи частично сводится к задаче сетевого планирования итоговой целью которой является установление «оптимальных взаимосвязей» входных и выходных параметров независимо от самих параметров.

Проведенный анализ показывает, что поставленная задача оптимизации представляет собой задачу построения метамодели оптимизации образовательного процесса. В литературе [3] встречается следующее определение понятия «мета-

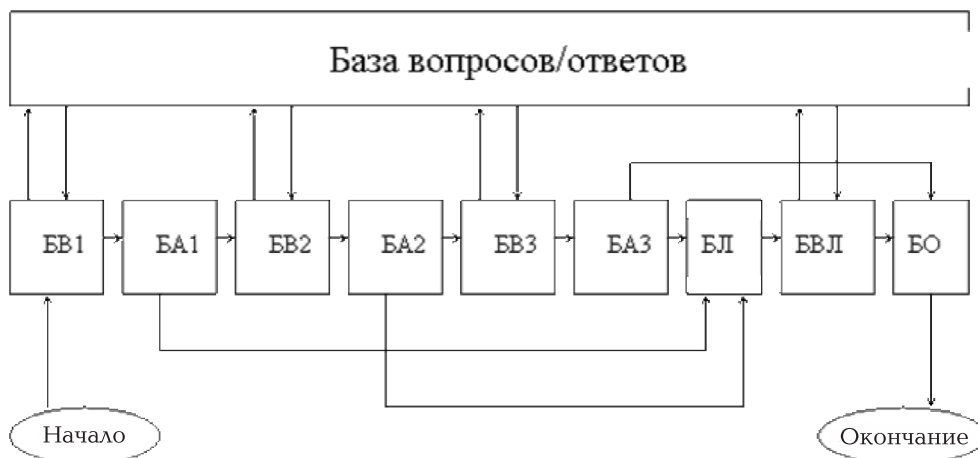


Рис. 3. Структурная схема алгоритма (блоки 1–3)

модель». *Мета модель* – модель, описывающая структуру, поведение и свойства какой-либо модели и существующая относительно нее на более высоком логическом уровне.

Принимая во внимание данное определение, алгоритмы программированного обучения и результаты проведенного исследования, можно дать

новое определение метамодели применительно к образовательному процессу. *Мета модель образовательного процесса* – это модель, формирующая управленческие критерии в соответствии с имеющимися входными и выходными параметрами для реализации алгоритма оптимизации структуры и содержания образовательного процесса.

СПИСОК ЛИТЕРАТУРЫ

1. **Аткинсон, Р.** Введение в математическую теорию обучения [Текст] / Р. Аткинсон, Г. Бауэр, Э. Кротерс. –М., 1969.

2. **Томас, К.** Перспективы программированного обучения [Текст] / К. Томас, Дж. Девис, Д. Опеншоу [и др.]; Пер. с англ. О.А. Бондина, Н.Т. Кобяковой;

Под ред. А.В. Нетушила. –М.: Мир, 1966. –247 с.

3. **Лядова, Л.Н.** Технология создания динамически адаптируемых информационных систем [Текст] / Л.Н. Лядова // Тр. Междунар. науч.-технич. конф. Интеллектуальные системы (AIS'07). –М.: Физматлит, 2007. –Т. 2. –С. 350–357.



ESEC/FSE 2013

Санкт-Петербург, Россия

19-23 августа 2013



IX объединенная Европейская конференция по программной инженерии (ESEC) и Симпозиум по основам программной инженерии ACM SIGSOFT (FSE)

Европейская конференция по программной инженерии, совмещенная с Симпозиумом по основам программной инженерии ACM SIGSOFT – это одно из важнейших и самых престижных событий мирового уровня в области разработки программного обеспечения. В 2013 году конференция пройдет в северной столице России Санкт-Петербурге. ESEC/FSE по праву завоевал международное признание среди исследователей и разработчиков ПО как платформа для демонстрации и обсуждения инноваций, тенденций и новых задач в области программной инженерии.

ESEC/FSE 2013 Приглашение к участию

Научно-исследовательские статьи

К участию приглашаются статьи, описывающие оригинальные и ранее не опубликованные результаты теоретических, концептуальных и экспериментальных исследований в области программной инженерии. Статьи должны содержать значительные и инновационные результаты, связанные с тематикой ESEC/FSE. Также будут рассматриваться работы, предлагающие революционные подходы к актуальным проблемам. В целях воспроизводимости результатов особенно приветствуется использование данных и программных средств, находящихся в открытом доступе.

Темы конференции: Мы заинтересованы в участии представителей как академического сообщества, так и индустрии разработки ПО, при условии что тема работы относится к программной инженерии. К таким темам относятся:

- ♦ Компоненты, сервисы, связующее ПО
- ♦ Развертка и управление конфигурацией систем
- ♦ Среда и инструменты разработки
- ♦ Распределенная разработка и аутсорсинг
- ♦ Распределенное, мобильное и встроенное ПО
- ♦ Встроенное ПО и системы реального времени
- ♦ Эмпирические исследования и метрики
- ♦ Разработка ПО для конечного пользователя
- ♦ Поддержка и эволюция программных систем
- ♦ Разработка, управляемая моделями
- ♦ Шаблоны, фреймворки и серии программных продуктов
- ♦ Параллельное и распределенное программирование
- ♦ Визуализация программ
- ♦ Управление требованиями
- ♦ Обратная разработка
- ♦ Архитектура ПО
- ♦ Экономика области разработки ПО
- ♦ Валидация, верификация и тестирование

Ключевые даты

- ♦ Подача статей: до 1 марта 2013 г.
- ♦ Уведомление о принятии статей: 26 мая 2013 г.

Секции

- ♦ Промышленная разработка ПО
- ♦ Новые идеи
- ♦ Демонстрации инструментов разработки
- ♦ Презентации актуальных технологий
- ♦ Симпозиум аспирантов
- ♦ Семинары

Председатель конференции

- ♦ Бертран Мэйер, Высшая политехническая школа Цюриха, Швейцария, НИУ ИТМО, Россия

Председатели программного комитета

- ♦ Лучано Барези, Миланский политехнический университет, Италия
- ♦ Мира Мезини, Дармштадский технический университет, Германия

Программный комитет

- ♦ Свет Апель, Университет Пассау, Германия
- ♦ Антония Бертолино, ISTI-CNR, Италия
- ♦ Эрик Бодден, Дармштадский технический университет, Германия
- ♦ Пауло Борба, Федеральный университет Пернамбуко, Бразилия
- ♦ Витторио Кортеллесса, Л'Аквилиский университет, Италия
- ♦ Элизабетта Ди Нитто, Миланский политехнический университет, Италия
- ♦ Массимилиано Ди Пента, университет Санньо, Италия
- ♦ Майкл Эрнст, Университет штата Вашингтон, США
- ♦ Патрик Ойгстер, университет Пердью, США
- ♦ Алессандро Гарсия, Католический университет Рио-де-Жанейро, Бразилия
- ♦ Виктор Гергель, Нижегородский государственный университет, Россия
- ♦ Пауль Грюнбахер, Линцкий университет, Австрия
- ♦ Марк Харман, Университетский колледж Лондона, Великобритания
- ♦ Патрик Хейманс, Намурский университет, Бельгия
- ♦ Владимир Ицыксон, Санкт-Петербургский государственный политехнический университет, Россия
- ♦ Наталия Юристо, Мадридский университет, Испания
- ♦ Грегор Кицзалес, Университет Британской Колумбии, Канада
- ♦ Шунгун Ким, Гонконгский университет науки и технологии, Гонконг
- ♦ Патрисия Лаго, Амстердамский свободный университет, Голландия
- ♦ Ненад Медвидович, университет Южной Калифорнии, США
- ♦ Алессандро Орсо, Технический университет Джорджии, США
- ♦ Харольд Осшер, IBM Research, США
- ♦ Корина Пасариану, NASA, США
- ♦ Мауро Пецце, Луганский университет, Швейцария
- ♦ Миланский университет Бикокка, Италия
- ♦ Авайс Рашид, Ланкастерский университет, Великобритания
- ♦ Мартин Робильярд, Университет Макгилла, Канада
- ♦ Вильгельм Шэфер, Падерборнский университет, Германия
- ♦ Паоло Тонелла, Фонд Бруно Кесслера, Италия
- ♦ Себастиан Учитель, Университет Буэнос-Айреса, Аргентина
- ♦ Имперский колледж Лондона, Великобритания
- ♦ Ари ван Дойрсен, Делфтский технический университет, Голландия
- ♦ Виллем Виссер, Стелленбосский университет, ЮАР
- ♦ Андреас Целлер, Университет Саарландеса, Германия
- ♦ Андреа Цисман, Университет Сити, Великобритания



<http://2013.esec-fse.org>



ESEC/FSE 2013

Saint Petersburg, Russia

August 19-23, 2013



9th joint meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering

The 9th joint meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering will take place in the unique Saint Petersburg, Russia, the imperial capital of Russia and a major European cultural center. Held bi-annually, ESEC/FSE is an internationally renowned forum for researchers, practitioners, and educators to present and discuss the most recent innovations, trends, experiences, and challenges in the field of software engineering.

ESEC/FSE 2013 Call for papers

Research track

We invite high quality submissions describing original and unpublished results of theoretical, empirical, conceptual, and experimental software engineering research. Contributions should describe innovative and significant original research relevant to ESEC/FSE. Submissions of papers describing groundbreaking approaches to emerging problems will also be considered. Submissions that facilitate reproducibility by using available data sets or making the described tools publicly available are especially encouraged.

Topics of Interest: We are interested in submissions from both industry and academia on all topics related to software engineering. These include, but are not limited to:

- ◆ Components, services, and middleware
- ◆ Configuration management and deployment
- ◆ Development environments and tools
- ◆ Distributed and outsourced development
- ◆ Distributed, mobile, and pervasive software
- ◆ Embedded and real-time software
- ◆ Empirical studies and Metrics
- ◆ End-user software engineering
- ◆ Maintenance and evolution
- ◆ Model-driven software engineering
- ◆ Patterns, frameworks, and product lines
- ◆ Parallel and concurrent software
- ◆ Program comprehension and visualization
- ◆ Requirements engineering
- ◆ Reverse engineering
- ◆ Software architecture
- ◆ Software economics
- ◆ Validation, verification, and testing

Important dates

- ◆ Research papers due: March 1, 2013
- ◆ Notification to authors: May 26, 2013

Further tracks

- ◆ Artifact Evaluation
- ◆ Industrial Track
- ◆ New Ideas Track
- ◆ Tool Demonstrations Track
- ◆ Technical Briefings Track
- ◆ Doctoral Symposium
- ◆ Workshops

General Chair

- ◆ Bertrand Meyer, ETH Zurich, Switzerland and ITMO, Russia

Program Chairs

- ◆ Luciano Baresi, Politecnico di Milano, Italy
- ◆ Mira Mezini, Technische Universität Darmstadt, Germany

Program Committee

- ◆ Sven Apel, Universität Passau, Germany
- ◆ Antonia Bertolino, ISTI-CNR, Italy
- ◆ Eric Bodden, Technische Universität Darmstadt, Germany
- ◆ Paulo Borba, Universidade Federal de Pernambuco, Brasil
- ◆ Vittorio Cortellessa, Università dell'Aquila, Italy
- ◆ Elisabetta Di Nitto, Politecnico di Milano, Italy
- ◆ Massimiliano Di Penta, Università del Sannio, Italy
- ◆ Michael Ernst, University of Washington, USA
- ◆ Patrick Eugster, Purdue University, USA
- ◆ Alessandro Garcia, PUC Rio, Brasil
- ◆ Victor Gergel, University of Nijny Novgorod, Russia
- ◆ Paul Grünbacher, Universität Linz, Austria
- ◆ Mark Harman, University College London, United Kingdom
- ◆ Patrick Heymans, University of Namur, Belgium
- ◆ Vladimir Itsykson, Saint-Petersburg State Polytechnical University, Russia
- ◆ Natalia Juristo, University of Madrid, Spain
- ◆ Gregor Kiczales, University of British Columbia, Canada
- ◆ Sunghun Kim, Hong Kong University of Science and Technology, Hong Kong
- ◆ Patricia Lago, Vrije Universiteit Amsterdam, The Netherlands
- ◆ Nenad Medvidovic, University of Southern California, USA
- ◆ Alessandro Orso, Georgia Tech, USA
- ◆ Harold Ossher, IBM Research, USA
- ◆ Corina Pasareanu, NASA, USA
- ◆ Mauro Pezzè, University of Lugano, Switzerland and University of Milano Bicocca, Italy
- ◆ Awais Rashid, Lancaster University, United Kingdom
- ◆ Martin Robillard, McGill University, Canada
- ◆ Wilhelm Schäfer, Universität Paderborn, Germany
- ◆ Paolo Tonella, Fondazione Bruno Kessler, Italy
- ◆ Sebastian Uchitel, Universidad de Buenos Aires, Argentina and Imperial College London, United Kingdom
- ◆ Arie van Deursen, Delft University of Technology, The Netherlands
- ◆ Willem Visser, Stellenbosch University, South Africa
- ◆ Andreas Zeller, Universität des Saarlandes, Germany
- ◆ Andrea Zisman, City University, United Kingdom



СВЕДЕНИЯ ОБ АВТОРАХ

КОНТАКТНЫЕ ДАННЫЕ

АКСЕНОВ Константин Александрович – доцент кафедры информационных технологий Уральского федерального университета имени первого Президента России Б.Н. Ельцина, кандидат технических наук.

620002, г. Екатеринбург, ул. Мира, д. 19. Тел. 8(343) 375-41-45, e-mail: Wiper99@mail.ru

АКСЕНОВА Ольга Петровна – научный сотрудник кафедры информационных технологий Уральского федерального университета имени первого Президента России Б.Н. Ельцина.

620002, г. Екатеринбург, ул. Мира, д. 19. Тел. 8(343)375-41-45, e-mail: solnceva_op@mail.ru

АЛЕКСЕЕВ Василий Геннадьевич – аспирант кафедры информационной безопасности телекоммуникационных систем Санкт-Петербургского государственного университета телекоммуникаций имени проф. М.А. Бонч-Бруевича.

191186, Санкт-Петербург, наб. р. Мойки, д. 61. E-mail: i@vasay.ru

АНИКЕЕВА Мария Сергеевна – студент кафедры физики полупроводников и наноэлектроники Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29. Тел. 8(812)552-96-71, e-mail: hegma21@yandex.ru

БАЖЕНОВ Александр Николаевич – доцент кафедры прикладной математики Санкт-Петербургского государственного политехнического университета, кандидат физико-математических наук.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: a_bazhenov@inbox.ru

БОЛДЫРЕВ Юрий Яковлевич – заведующий кафедрой математического и программного обеспечения высокопроизводительных вычислений Санкт-Петербургского государственного политехнического университета, доктор технических наук, профессор.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: yuibold@mail.ru

БОНДАРЕВА Ирина Олеговна – доцент кафедры прикладной информатики в экономике Астраханского государственного технического университета, кандидат технических наук.

414025, г. Астрахань, ул. Татищева, д.16. E-mail: i.o.bondareva@gmail.com

ВАЛЫНЕЦ Елена Владимировна – выпускница кафедры прикладной математики Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29.

ВАН Кай – аспирант кафедры информационных технологий Уральского федерального университета имени первого Президента России Б.Н. Ельцина.

620002, г. Екатеринбург, ул. Мира, д. 19. Тел. 8(343)375-41-45, e-mail: kylebjcn@gmail.com

ВОРОНИН Вячеслав Владимирович – доцент кафедры радиоэлектронных систем Южно-Российского государственного университета экономики и сервиса, кандидат технических наук.

346500, г. Шахты, ул. Шевченко, д. 147. E-mail: voronin_sl@mail.ru

ГААЗЕ Михаил Викторович – заместитель директора научно-технического центра ОАО «Информационные телекоммуникационные технологии».

197342, Санкт-Петербург, ул. Кантемировская, д. 8.

ГОЛУБЕВ Анатолий Геннадиевич – ведущий научный сотрудник ОАО «Концерн «ЦНИИ «Электроприбор», доктор технических наук.

197046, Санкт-Петербург, ул. Малая Посадская, д. 30. E-mail: agg300@mail.ru

ДРЕМОВ Сергей Сергеевич – инженер ЗАО «Полупроводниковые приборы».

19415, Санкт-Петербург, а/я 29. Тел. 8(812)294-25-32, e-mail: sales@atcsd.ru

ЗАБОРОВСКИЙ Владимир Сергеевич – заместитель главного конструктора Государственного научного центра Российской Федерации «Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики», доктор технических наук, профессор.

194064, Санкт-Петербург, Тихорецкий пр., д. 21. E-mail: vlad@neva.ru

ИВАНКОВ Алексей Александрович – доцент кафедры прикладной математики Санкт-Петербургского государственного политехнического университета, кандидат физико-математических наук.

195251, Санкт-Петербург, ул. Политехническая, д. 29. Тел. 8(812)552-75-25, e-mail: ivankov@stu.neva.ru

ИВАНОВ Александр Александрович – заведующий кафедрой мехатроники и роботостроения Санкт-Петербургского государственного политехнического университета, кандидат технических наук, доцент.

195251, Санкт-Петербург, ул. Политехническая, д. 29. Тел. 8(812)552-41-98, e-mail: al_ivanov@rtc.ru

ИЛЬЯШЕНКО Александр Сергеевич – инженер 821 лаборатории Государственного научного центра Российской Федерации «Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики».

194064, Санкт-Петербург, Тихорецкий пр., д. 21. E-mail: ilyashenko.alex@gmail.com

КАДЫРОВА Наталья Олеговна – доцент кафедры прикладной математики Санкт-Петербургского государственного политехнического университета, кандидат биологических наук, старший научный сотрудник.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: natalia.kadyrova@gmail.com

КАЛМЫКОВА Светлана Владимировна – доцент кафедры управления в социально-экономических системах Санкт-Петербургского государственного политехнического университета, кандидат педагогических наук.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: kalmykos@mail.ru

КИСЛИЦЫН Александр Борисович – аспирант кафедры радиоэлектронных средств защиты информации Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: kislitsyn@rambler.ru

КОНДРАТЬЕВ Александр Сергеевич – заместитель директора ЦНИИ РТК по научной работе, кандидат технических наук, доцент.

194064, Санкт-Петербург, Тихорецкий пр., д. 21. E-mail: kondr@rtc.ru

КОРНАГА Ярослав Игоревич – старший преподаватель кафедры математического моделирования экономических систем, соискатель кафедры вычислительной техники Национального технического университета Украины «Киевский политехнический институт».

03056, Украина, Киев, пр. Победы д. 37. Тел. +(38044)454-99-36, e-mail: slovyan_k@ukr.net

ЛАБОШИН Леонид Юрьевич – инженер Государственного научного центра Российской Федерации «Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики».

194064, Санкт-Петербург, Тихорецкий пр., д. 21. E-mail: laboshin@ya.ru

ЛАТЫПОВА Эльмира Амировна – преподаватель Волжской государственной академии водного транспорта.

414000, г. Астрахань, ул. Костина, д. 2. E-mail: el-mi-ra20@mail.ru

ЛУКАШИН Алексей Андреевич – научный сотрудник Государственного научного центра Российской Федерации «Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики».

194064, Санкт-Петербург, Тихорецкий пр., д. 21. E-mail: lukash@neva.ru



ЛЫЧАГИН Николай Иванович – заместитель генерального директора ОАО «Информационные телекоммуникационные технологии», доктор технических наук, профессор.

197342, Санкт-Петербург, ул. Кантемировская, д. 8.
E-mail: nil@ntcl.inteltech.ru

МАЛАХОВСКИЙ Артем Алексеевич – начальник научно-исследовательского отделения ОАО «Информационные телекоммуникационные технологии».

197342, Санкт-Петербург, ул. Кантемировская, д. 8.

МАРЦЕНЮК Михаил Андреевич – заведующий кафедрой компьютерных систем и телекоммуникаций Пермского государственного национального исследовательского университета, доктор физико-математических наук, профессор.

614990, г. Пермь, ул. Букирева, д. 15. E-mail: mrcn@psu.ru

МАСЛОВ Владимир Иванович – заведующий кафедрой управления конструкторско-технологическими инновациями Санкт-Петербургского государственного политехнического университета, доктор технических наук, профессор.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: vimaslov@tu.neva.ru

МИНАКОВ Владимир Федорович – профессор кафедры управления конструкторско-технологическими инновациями Санкт-Петербургского государственного политехнического университета, доктор технических наук.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: m-m-m-m-m@mail.ru

МУЛЮХА Владимир Александрович – старший научный сотрудник лаборатории 822 Государственного научного центра Российской Федерации «Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики», кандидат технических наук.

194064, Санкт-Петербург, Тихорецкий пр., д. 21.
E-mail: mva@rtc.ru

МУХИН Вадим Евгеньевич – доцент кафедры вычислительной техники Национального технического университета Украины «Киевский политехнический институт», кандидат технических наук.

03056, Украина, Киев, пр. Победы д. 37. Тел. +(38044) 454-99-36, e-mail: v_mukhin@mail.ru

ОРЖЕНОВСКИЙ Михаил Викторович – студент кафедры прикладной математики Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29.

ПАВЛОВА Людмила Владимировна – доцент кафедры прикладной математики Санкт-Петербургского государственного политехнического университета, кандидат физико-математических наук.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: lyu0510@gmail.com

ПЕТРОВА Ирина Юрьевна – проректор по науке Астраханского государственного университета, заведующий кафедрой информационных систем, доктор технических наук, профессор.

414056, г. Астрахань, ул. Татищева, д. 20А. Тел. 8(8512)61-08-43, e-mail: petrova@aspu.ru

ПЕТРОСЯН Григорий Сергеевич – аспирант кафедры прикладной математики Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: gregory.petrosyan@gmail.com

ПОЛЯКОВ Виктор Борисович – старший преподаватель кафедры компьютерных систем и телекоммуникаций Пермского государственного национального исследовательского университета.

614990, г. Пермь, ул. Букирева, д. 15. E-mail: polyakov@mpm.ru

ПЫШКИН Евгений Валерьевич – доцент кафедры компьютерных систем и программных технологий Санкт-Петербургского государственного политехнического университета, кандидат технических наук.

195251, Санкт-Петербург, ул. Политехническая, д. 21. Тел. 8(812)297-42-18, e-mail: evgeny.pyshkin@gmail.com

РАШИЧ Андрей Валерьевич – доцент кафедры радиоэлектронных средств защиты информации Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: rashich@cee.spbstu.ru

РЫБАК Виктор Александрович – директор Центрального научно-исследовательского института комплексного использования водных ресурсов Министерства природных ресурсов и охраны окружающей среды Республики Беларусь, кандидат технических наук, доцент.

220086, Республика Беларусь, г. Минск, ул. Славинского, д. 1/2. Тел. +(37517)2676522, e-mail: 6774338@tut.by

СЕЛЕТКОВ Илья Павлович – аспирант кафедры компьютерных систем и телекоммуникаций Пермского государственного национального исследовательского университета.

614990, г. Пермь, ул. Букирева, д. 15. E-mail: iseletkov@gmail.com

СЕМЕНОВСКИЙ Василий Борисович – старший преподаватель кафедры компрессорной, вакуумной и холодильной техники Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: swb@neva.ru

СИЛИНЕНКО Александр Витальевич – заведующий 822 лабораторией Государственного научного центра Российской Федерации «Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики», кандидат технических наук.

194064, Санкт-Петербург, Тихорецкий пр., д. 21. E-mail: avsv@neva.ru

СМЕТАНИН Андрей Георгиевич – аспирант кафедры информатики Московского физико-технического института.

141700, Московская обл., г. Долгопрудный, Институтский пер., д. 9. E-mail: originalnull@gmail.com

СМИРНОВ Алексей Сергеевич – ведущий инженер ОАО «Концерн «ЦНИИ «Электронприбор», соискатель.

197046, Санкт-Петербург, ул. Малая Посадская, д. 30. E-mail: smalex77@mail.ru

СОКОЛИЦИН Александр Сергеевич – профессор кафедры управления в социально-экономических системах Санкт-Петербургского государственного политехнического университета, доктор экономических наук.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: alex.sokol1951@yandex.ru

СОФРОНОВ Антон Николаевич – доцент кафедры физики полупроводников и наноэлектроники Санкт-Петербургского государственного политехнического университета, кандидат физико-математических наук.

195251, Санкт-Петербург, ул. Политехническая, д. 29. Тел. 8(812)552-9671, e-mail: sofronov@rphf.spbstu.ru

СУЛЕЙМАН Халед Манси – аспирант Белорусского государственного университета информатики и радиоэлектроники.

220013, Республика Беларусь, г. Минск, ул. П. Бровки, д. 6.

ТЕР-МАТИРОСЯН Александр Леонович – генеральный директор ЗАО «Полупроводниковые приборы», кандидат технических наук.

194156, Санкт-Петербург, а/я 29. Тел. 8(812) 294-25-32, e-mail: sales@atcsd.ru

ТОРМАСОВ Александр Геннадьевич – профессор кафедры теоретической и прикладной информатики Московского физико-технического института, доктор физико-математических наук.

141700, Московская обл., г. Долгопрудный, Институтский пер., д. 9. E-mail: tor@phystech.edu

ФИЛИППОВ Михаил Сергеевич – начальник 81 отдела Государственного научного центра Российской Федерации «Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики».

194064, Санкт-Петербург, Тихорецкий пр., д. 21. E-mail: mfill@npo-rtc.ru

ХОМЕНКО Татьяна Владимировна – заведующий кафедрой информатики Астраханского государственного технического университета, кандидат технических наук, доцент.

414025, г. Астрахань, ул. Татищева, д. 16. Тел. 8(8512)61-43-82, e-mail: t_v_khomenko@mail.ru

ЧЕРНОРУЦКИЙ Игорь Георгиевич – заведующий кафедрой информационных и управляющих систем, декан факультета технической кибернетики Санкт-Петербургского государственного политехнического университета, доктор технических наук, профессор.

195291, Санкт-Петербург, ул. Политехническая, д. 21. E-mail: dean@ftk.spbstu.ru

ЧУПРОВ Сергей Геннадьевич – аспирант кафедры мехатроники и роботостроения Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29. Тел. 8(812)552-41-98, e-mail: sergey3007@yandex.ru

ШЕБАЛОВ Алексей Андреевич – бакалавр кафедры прикладной математики Санкт-Петербургского государственного политехнического университета.

195251, Санкт-Петербург, ул. Политехническая, д. 29. E-mail: alex.shebalov@gmail.com

АННОТАЦИИ

КЛЮЧЕВЫЕ СЛОВА

Болдырев Ю.Я. РОЛЬ СУПЕРКОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ В ИНЖЕНЕРНОМ ОБРАЗОВАНИИ.

Рассмотрены ключевые вопросы о той роли, которую должны играть суперкомпьютерные технологии в процессе подготовки инженерных кадров в современной высшей технической школе.

СУПЕРКОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ. ИНЖЕНЕРНЫЕ КАДРЫ ОБРАЗОВАТЕЛЬНЫЙ ПРОЦЕСС. ВЫСШАЯ ТЕХНИЧЕСКАЯ ШКОЛА.

Сметанин А.Г., Тормасов А.Г. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ЭФФЕКТИВНЫХ АЛГОРИТМОВ ПОИСКА И РАЗМЕЩЕНИЯ ДАННЫХ В ДЕЦЕНТРАЛИЗОВАННОЙ РАСПРЕДЕЛЕННОЙ ФАЙЛОВОЙ СИСТЕМЕ.

Приведена математическая модель децентрализованной распределенной файловой системы. Доказаны теоретические оценки для длины поиска в данной системе и приведены результаты имитационного моделирования, подтверждающие теоретические оценки.

РАСПРЕДЕЛЕННЫЕ ФАЙЛОВЫЕ СИСТЕМЫ. АЛГОРИТМ ПОИСКА ДАННЫХ. ОТКАЗОУСТОЙЧИВОСТЬ. МАСШТАБИРУЕМОСТЬ. ДЕЦЕНТРАЛИЗАЦИЯ. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ.

Заборовский В.С., Кондратьев А.С., Мулюха В.А., Силиненко А.В., Ильяшенко А.С., Филиппов М.С. УДАЛЕННОЕ УПРАВЛЕНИЕ РОБОТОТЕХНИЧЕСКИМИ ОБЪЕКТАМИ В КОСМИЧЕСКИХ ЭКСПЕРИМЕНТАХ СЕРИИ «КОНТУР».

Рассмотрено использование различных режимов удаленного управления робототехническими объектами на примере космических экспериментов серии «Контур», проводимых специалистами ЦНИИ РТК, РКК «Энергия» и ЦНИИМаш.

УДАЛЕННОЕ УПРАВЛЕНИЕ РОБОТАМИ. СИЛОМОМЕНТНАЯ ОБРАТНАЯ СВЯЗЬ. КОСМИЧЕСКИЙ ЭКСПЕРИМЕНТ.

Гаазе М.В., Лычагин Н.И., Малаховский А.А. ИНТЕГРИРОВАННЫЕ ТЕЛЕКОММУНИКАЦИОННЫЕ СЛУЖБЫ ДЛЯ ВЕДОМСТВЕННЫХ СЕТЕЙ СВЯЗИ.

Рассмотрена методология создания ведомственных интегрированных телекоммуникационных служб и описан отечественный сертифицированный программный комплекс, созданный на основе методологии.

ТЕЛЕКОММУНИКАЦИОННЫЕ СИСТЕМЫ. ПРОГРАММНЫЙ КОМПЛЕКС. ВЕДОМСТВЕННЫЕ СЕТИ СВЯЗИ.

Алексеев В.Г. СИСТЕМА ЦИФРОВЫХ «ВОДЯНЫХ» ЗНАКОВ ДЛЯ АУДИОСИГНАЛОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ЯВЛЕНИЯ РЕВЕРБЕРАЦИИ.

Предложено использовать для вложения ЦВЗ в аудиосигналы явление умеренной реверберации, которое практически не ухудшает качество звучания музыкальных записей и устойчиво к сжатию с потерями в популярные форматы, такие, как mp3. Для извлечения ЦВЗ рекомендовано использовать кепстральное преобразование и корреляционный приемник, отличающийся от известного в классической теории связи.

СТЕГАНОГРАФИЯ. ЦИФРОВЫЕ ВОДЯНЫЕ ЗНАКИ. ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ. РЕВЕРБЕРАЦИЯ.

Воронин В.В. АВТОМАТИЗИРОВАННОЕ ДЕТЕКТИРОВАНИЕ ДЕФЕКТОВ ПРИ РЕКОНСТРУКЦИИ АРХИВНЫХ ФОТОДОКУМЕНТОВ.

Разработан метод детектирования искажений на архивных фотодокументах с помощью текстурного анализа и выделения локальных дескрипторов на изображениях. Алгоритм разработанного модифицированного метода SURF позволяет выделять различные типы локальных особенностей с помощью предварительной обработки изображения энергетическими характеристиками Лавса. Предлагаемый подход позволяет определять особые точки на различных типах текстур и анализировать дефекты на поверхности фотографий как один из типов текстуры, таких, как края, пятна, рябь, волны и т. д.

РЕКОНСТРУКЦИЯ ИЗОБРАЖЕНИЯ. ДЕФЕКТЫ. АРХИВНЫЕ ФОТОДОКУМЕНТЫ. ВОССТАНОВЛЕНИЕ.

Мухин В.Е., Корнага Я.И. МНОГОУРОВНЕВАЯ МОДЕЛЬ С ДИФФЕРЕНЦИАЛЬНЫМ УРОВНЕМ ДОВЕРИЯ К СУБЪЕКТАМ ДЛЯ ПОВЫШЕНИЯ ЗАЩИЩЕННОСТИ РАСПРЕДЕЛЕННЫХ БАЗ ДАННЫХ.

Рассмотрены средства поддержки безопасной обработки данных в распределенных базах данных (РБД). Предложен подход к реализации безопасной среды РБД с использованием многоуровневой модели безопасности. Для повышения защищенности обработки данных в РБД предложены специальные механизмы формирования уровня доверия к субъектам. Реализация средств защиты на основе предложенных механизмов обеспечивает поддержку высокого уровня защищенности в условиях динамического изменения числа субъектов РБД.

БАЗЫ ДАННЫХ. БЕЗОПАСНОСТЬ. МНОГОУРОВНЕВАЯ МОДЕЛЬ. УРОВЕНЬ ДОВЕРИЯ.

Заборовский В.С., Кондратьев А.С., Мулюха В.А., Силиненко А.В., Ильяшенко А.С. УПРАВЛЕНИЕ ГРУППИРОВКАМИ НАПЛАНЕТНЫХ РОБОТОВ С БОРТА ПИЛОТИРУЕМОЙ ОРБИТАЛЬНОЙ СТАНЦИИ В РАМКАХ ПРОЕКТА «METERON».

Рассмотрено использование сетецентрического подхода к задачам управления группировками напланетных робототехнических объектов, управляемых с борта пилотируемой орбитальной станции.

УПРАВЛЕНИЕ ГРУППИРОВКАМИ РОБОТОВ, СЕТЕЦЕНТРИЧНОСТЬ, КОМБИНИРОВАННЫЕ СПУТНИКОВЫЕ КАНАЛЫ СВЯЗИ.

Кислицын А.Б., Рашич А.В. ПЕРЕДАЧА И ПРИЕМ СИГНАЛОВ С DOQPSK И CDOQPSK С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА ВИТЕРБИ.

Изучена задача повышения помехоустойчивости приема сигналов с методами манипуляции DOQPSK и CDOQPSK посредством перехода от поэлементной демодуляции к приему «в целом» с использованием алгоритма Витерби. Показано, что энергетический выигрыш с учетом пик-фактора, получаемый в результате применения демодулятора Витерби, составляет 2,1 дБ для сигналов DOQPSK и 3,4 дБ для сигналов CDOQPSK по сравнению с сигналами $\pi/4$ -DQPSK при вероятности ошибки $P_{\text{ош}} = 10^{-5}$.

ПИК-ФАКТОР. DOQPSK. CDOQPSK. АЛГОРИТМ ВИТЕРБИ.

Голубев А.Г., Смирнов А.С. О РАСЧЕТЕ ВЕРХНЕЙ ГРАНИЦЫ ЧУВСТВИТЕЛЬНОСТИ ИНТЕРФЕРЕНЦИОННЫХ ВОЛОКОННО-ОПТИЧЕСКИХ ГИДРОФОНОВ ПРИЕМНОЙ ГИДРОАКУСТИЧЕСКОЙ АНТЕННЫ.

Решена задача расчета верхней границы чувствительности гидрофонов, являющейся оптимальной по критерию достижения требуемого динамического диапазона тракта при заданном минимальном уровне акустических шумов.

ВОЛОКОННО-ОПТИЧЕСКИЙ ГИДРОФОН. МЕТОДИКА. ЧУВСТВИТЕЛЬНОСТЬ. ДИНАМИЧЕСКИЙ ДИАПАЗОН.

Чупров С.Г., Иванов А.А. МОДЕЛИРОВАНИЕ ДИНАМИКИ ДВУХДВИГАТЕЛЬНОГО СЕРВОПРИВОДА С ЭФФЕКТОМ САМОТОРМОЖЕНИЯ.

Рассмотрены проблемы управления двухдвигательным сервоприводом с эффектом самоторможения. Построена математическая модель динамики системы, позволяющая сформулировать подход к управлению такого привода.

УПРАВЛЕНИЕ. МНОГОДВИГАТЕЛЬНЫЙ СЕРВОПРИВОД. САМОТОРМОЖЕНИЕ. ДИНАМИКА. МОДЕЛИРОВАНИЕ.

Петросян Г.С. АНАЛИЗ ОБЪЕКТНО-ОРИЕНТИРОВАННОЙ ПАРАДИГМЫ И ПОИСК ЛУЧШЕЙ АЛЬТЕРНАТИВЫ.

Изучены основные проблемы промышленного программирования и проведен анализ объектно-ориентированной парадигмы с точки зрения их решения.

ЯЗЫКИ ПРОГРАММИРОВАНИЯ. ОБЪЕКТНО-ОРИЕНТИРОВАННОЕ ПРОГРАММИРОВАНИЕ. ИНКАПСУЛЯЦИЯ. ПОВТОРНОЕ ИСПОЛЬЗОВАНИЕ. ПОЛИМОРФИЗМ.

Пышкин Е.В. ПРОБЛЕМЫ АВТОМАТИЗАЦИИ ПРИЕМОЧНОГО ТЕСТИРОВАНИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ПРИ ИСПОЛЬЗОВАНИИ ПОДХОДА «РАЗРАБОТКА, УПРАВЛЯЕМАЯ ОПИСАНИЕМ ПОВЕДЕНИЯ».

Изучены проблемы, возникающие при внедрении в проектную практику подхода «разработка, управляемая описанием поведения» (behavior driven development – BDD). Представлены основные идеи подхода, проиллюстрирован процесс разработки и тестирования с использованием BDD. Особое внимание уделено трудностям, возникающим при переходе от описанных в текстовой форме сценариев приемочных тестов, понятных заказчику, к модульным тестам, конструируемым разработчиками. Проанализирован процесс разметки тестового сценария инженером по контролю качества для его последующего преобразования к тестовым методам и классам, сформулированы задачи по разработке средств автоматизации этого процесса, определены требования к улучшению языка пользовательских сценариев.

ПРОГРАММНАЯ ИНЖЕНЕРИЯ. ТЕСТИРОВАНИЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ. ГИБКИЕ МЕТОДОЛОГИИ. КОНТРОЛЬ КАЧЕСТВА.

Лабошин Л.Ю., Лукашин А.А., Семеновский В.Б. ПОДДЕРЖКА ВЫЧИСЛЕНИЙ НА ГРАФИЧЕСКИХ УСКОРИТЕЛЯХ В ВИРТУАЛЬНОЙ СРЕДЕ ХЕН ДЛЯ ОБРАБОТКИ ПОТОКОВ ДАННЫХ РОБОТОТЕХНИЧЕСКИХ СИСТЕМ.

Освещены проблемы использования графического процессора в виртуальной среде. Описан метод передачи графического контроллера в виртуальную машину. Произведена оценка эффективности использования графического процессора в качестве акселератора потоковых вычислений в виртуальной машине. Предложено использование разработанного метода для обработки данных робототехнических систем.

ВИРТУАЛИЗАЦИЯ. ВИРТУАЛЬНАЯ МАШИНА. ГИПЕРВИЗОР. ГРАФИЧЕСКИЙ ПРОЦЕССОР. ПРОБРОС. CUDA. PCI. IOMMU.

Валынец Е.В., Орженковский М.В., Иванков А.А. СТОХАСТИЧЕСКАЯ МОДЕЛЬ СТРУКТУРЫ ПУЛА ЗАЯВОК ПОСТАВЩИКОВ НА ЗАКРЫТОМ АУКЦИОНЕ СПОТОВЫХ ЦЕН.

Изложено построение стохастической модели в форме смеси распределений для структуры пула заявок поставщиков на закрытом аукционе спотового рынка электроэнергии. Структура пула идентифицирована по косвенной информации – агрегированным данным о предложениях по каждому из типов сырья, которое используется для генерации энергии.

СПОТОВЫЕ ЦЕНЫ НА ЭЛЕКТРОЭНЕРГИЮ. ЗАКРЫТЫЙ АУКЦИОН. ПОСТАВЩИКИ ЭЛЕКТРОЭНЕРГИИ. СТОХАСТИЧЕСКАЯ МОДЕЛЬ. СТРУКТУРА ПУЛА ЗАЯВОК.

Иванов А.А. МЕТОД ПЛАНИРОВАНИЯ ФОРМЫ ГИПЕРИЗБЫТОЧНОГО МАНИПУЛЯТОРА С ОГРАНИЧЕННЫМ ДИАПАЗОНОМ ИЗМЕНЕНИЯ ШАРНИРНЫХ КООРДИНАТ.

Предложены постановка и метод решения обратной задачи кинематики для манипулятора с произвольным числом звеньев и ограниченным диапазоном изменения шарнирных координат. Приведены результаты численного моделирования, демонстрирующие работоспособность и эффективность метода

МАНИПУЛЯТОР. КИНЕМАТИЧЕСКАЯ ИЗБЫТОЧНОСТЬ. ОБРАТНАЯ ЗАДАЧА КИНЕМАТИКИ. МОДЕЛИРОВАНИЕ. ОПТИМИЗАЦИЯ.

Кадырова Н.О., Павлова Л.В. ЭФФЕКТИВНАЯ МЕТОДИКА ОБРАБОТКИ МНОГОРАЗМЕРНЫХ ДАННЫХ БОЛЬШОГО ОБЪЕМА.

На основании современных эффективных подходов предложена методика обработки данных большого объема и большой размерности при решении задачи бинарной классификации. Приведены результаты ее применения к реальным данным.

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ. ДАННЫЕ БОЛЬШОГО ОБЪЕМА И БОЛЬШОЙ РАЗМЕРНОСТИ. НАУЧНАЯ МЕТОДОЛОГИЯ РАБОТЫ С ДАННЫМИ. МАШИННОЕ ОБУЧЕНИЕ.

Петрова И.Ю., Хоменко Т.В. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ВЫБОРА И МЕТОД ОЦЕНКИ ВАРИАНТОВ ФИЗИЧЕСКОГО ПРИНЦИПА ДЕЙСТВИЯ ЧУВСТВИТЕЛЬНОГО ЭЛЕМЕНТА.

Решение задачи выбора физического принципа действия (ФПД) чувствительного элемента (ЧЭ), встречающееся на начальных этапах процесса проектирования датчиковой аппаратуры и сводящееся к выбору одного или нескольких вариантов из полученного/синтезированного набора ФПД ЧЭ, существенно зависит от информации о критериях выбора. Предложена математическая модель возможных вариаций условий выбора и информационный метод оценки выбранного решения.

ПРОЕКТИРОВАНИЕ. ВЫБОР РЕШЕНИЙ. КОЛИЧЕСТВО ИНФОРМАЦИИ. МЕТОД ОЦЕНКИ.

Черноруцкий И.Г. НЕКОТОРЫЕ СТАНДАРТНЫЕ СХЕМЫ ПАРАМЕТРИЧЕСКОЙ ОПТИМИЗАЦИИ.

Изложена техника построения гладких критериев оптимальности в минимаксных задачах аппроксимации и в задачах решения систем неравенств, часто используемых для формализации технических требований к оптимизируемым объектам. Обсуждаемые в статье гладкие аппроксимации критериев позволяют для решения широкого класса задач параметрической оптимизации ограничиваться простейшими реализациями стандартных алгоритмов нелинейного программирования.

ПАРАМЕТРИЧЕСКАЯ ОПТИМИЗАЦИЯ. МИНИМАКСНЫЕ ЗАДАЧИ. СИСТЕМЫ НЕРАВЕНСТВ. ГЛАДКИЕ КРИТЕРИИ ОПТИМАЛЬНОСТИ.

Марценюк М.А., Поляков В.Б., Селетков И.П. МАТРИЧНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМОВ НЕЧЕТКОГО ВЫВОДА.

Рассмотрена матричная реализация алгоритмов нечеткого вывода, естественным образом обобщающая «четкие» логические правила, такие, как *modus ponens*, *modus tollens* и др. Показано, что в этом представлении логический вывод сводится к решению линейных алгебраических уравнений с обычными условиями существования решения. Приведены примеры различных вариантов нечеткого вывода, детально демонстрирующие эффективность предлагаемого метода. Показано, что в матричном представлении упрощается программирование приложений нечеткой логики на контроллерах.

НЕЧЕТКАЯ ЛОГИКА. МАТРИЧНАЯ РЕАЛИЗАЦИЯ НЕЧЕТКОЙ ПЕРЕМЕННОЙ. НЕЧЕТКИЙ ВЫВОД. НЕЧЕТКИЕ ПРАВИЛА ВЫВОДА. ПРОГРАММИРОВАНИЕ НЕЧЕТКОЙ ЛОГИКИ НА КОНТРОЛЛЕРЕ. НЕЧЕТКИЕ ПРЕДИКАТЫ.

Черноруцкий И.Г. АЛГОРИТМИЧЕСКИЕ ПРОБЛЕМЫ ЖЕСТКОЙ ОПТИМИЗАЦИИ.

Введено понятие жесткого целевого функционала в задачах параметрической оптимизации. Указаны источники жесткости в задачах компьютерного моделирования.

ПАРАМЕТРИЧЕСКАЯ ОПТИМИЗАЦИЯ. ЖЕСТКИЕ ФУНКЦИОНАЛЫ. ЖЕСТКИЕ ДИФФЕРЕНЦИАЛЬНЫЕ СИСТЕМЫ.

Шебалов А.А., Баженов А.Н. ИССЛЕДОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ МЕТОДОВ ВЫЧИСЛЕНИЯ ОПТИЧЕСКОГО ПОТОКА.

Описаны методы вычисления оптического потока и оценка их производительности. Проведен вычислительный эксперимент с целью сравнения классических и современных алгоритмов, результаты которого проанализированы с точки зрения точности и скорости работы.

ОБРАБОТКА ИЗОБРАЖЕНИЙ. ОПТИЧЕСКИЙ ПОТОК. АЛГОРИТМЫ. МЕТОДЫ. ТЕХНИКИ. СРАВНЕНИЕ. АНАЛИЗ.

Аникеева М.С., Софронов А.Н., Дремов С.С., Тер-Мартirosян А.Л. ЦИФРОВАЯ СИСТЕМА ВИЗУАЛИЗАЦИИ ДЛЯ ФОТОДИНАМИЧЕСКОЙ ТЕРАПИИ.

Представлена система визуализации пространственного распределения флюоресценции при диагностике и лечении методами фотодинамической терапии. Приведена объектно-ориентированная архитектура системы визуализации. Произведен выбор оптимального варианта конкретной реализации с активным использованием сторонних библиотек с открытым исходным кодом. Представлено описание программного обеспечения системы. Изготовлен опытный образец системы визуализации, включающий аппаратную часть и программное обеспечение, который успешно прошел клинические испытания.

ВИЗУАЛИЗАЦИЯ. ФОТОДИНАМИЧЕСКАЯ ТЕРАПИЯ. ОБЪЕКТНО-ОРИЕНТИРОВАННАЯ АРХИТЕКТУРА.

Латыпова Э.А., Бондарева И.О. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ КАК СРЕДСТВО УПРАВЛЕНИЯ ЛОГИСТИЧЕСКИМ ПРЕДПРИЯТИЕМ НА ПРИМЕРЕ ГРУЗОВОГО ПОРТА.

Рассмотрена возможность повышения эффективности управления бизнес-процессами грузового порта. Описана технология создания сбалансированной системы показателей. Предложено использовать методы имитационного моделирования с целью оценки, прогнозирования и корректировки деятельности порта.

ВСЕОБЩЕЕ УПРАВЛЕНИЕ КАЧЕСТВОМ. СБАЛАНСИРОВАННАЯ СИСТЕМА ПОКАЗАТЕЛЕЙ. ГРУЗОВОЙ ПОРТ. БИЗНЕС-ПРОЦЕССЫ. СТРАТЕГИЯ ПОРТА. ДЕРЕВО ЦЕЛЕЙ. СТРАТЕГИЧЕСКАЯ КАРТА. УПРАВЛЕНЧЕСКИЕ РЕШЕНИЯ. ПЕРСПЕКТИВЫ. ЭФФЕКТИВНОСТЬ. КОНКУРЕНТОСПОСОБНОСТЬ. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ.

Аксенов К.А., Аксенова О.П., Ван Кай ПЛАНИРОВАНИЕ ПОРТФЕЛЯ ПРОЕКТОВ В СТРОИТЕЛЬСТВЕ НА ОСНОВЕ МУЛЬТИАГЕНТНОГО ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ.

Рассмотрена разработка гибридного метода планирования и анализа узких мест портфеля объектов строительства, основанная на мультиагентном имитационном моделировании, методе критического пути и операционном анализе вероятностных сетей.

ПОРТФЕЛЬ ПРОЕКТОВ. ПЛАНИРОВАНИЕ. ЖИЛИЩНОЕ СТРОИТЕЛЬСТВО. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ. МУЛЬТИАГЕНТНЫЕ СИСТЕМЫ. АНАЛИЗ УЗКИХ МЕСТ.

Рыбак В.А. Сулейман Х.М. АНАЛИЗ И ОПТИМИЗАЦИЯ ЭКОЛОГО-ЭКОНОМИЧЕСКИХ ПАРАМЕТРОВ ИННОВАЦИОННЫХ ПРОЕКТОВ.

Описана новая технология анализа и оптимизации эколого-экономических параметров инновационных проектов с позиций рационального природопользования и информационного управления. Рассмотренные методы и программные средства позволяют осуществлять поддержку принятия управленческих решений при выборе наиболее эффективных проектов.

РАЦИОНАЛЬНОЕ ПРИРОДОПОЛЬЗОВАНИЕ. УПРАВЛЕНИЕ. ТЕОРИЯ НЕЧЕТКИХ МНОЖЕСТВ. ЭКОЛОГО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ. ИННОВАЦИОННЫЕ ПРОЕКТЫ.

Маслов В.И., Минаков В.Ф. КРИТЕРИЙ ЭФФЕКТИВНОСТИ В СИСТЕМЕ УПРАВЛЕНИЯ КАЧЕСТВОМ ПРОДУКЦИИ ПРЕДПРИЯТИЯ.

Принципы менеджмента качества развиваются количественными показателями (эластичностью качества по цене и по затратам), характеризующими прирост качества на единицу роста цены и затрат соответственно, например, в процентном выражении. Соотношение эластичностей качества разных предприятий-производителей продукции характеризует их конкурентоспособность. Соотношение ценовой эластичности качества (спроса на качество) с эластичностью качества по затратам (предложения качества) является критерием экономической эффективности, позволяющим принимать решение о целесообразности мероприятий по обеспечению качества, осуществлять их сопоставление и выбор.

ЦЕНОВАЯ ЭЛАСТИЧНОСТЬ КАЧЕСТВА. ЭЛАСТИЧНОСТЬ КАЧЕСТВА ПО ЗАТРАТАМ.

Калмыкова С.В., Соколицин А.С. ОПТИМИЗАЦИЯ СОДЕРЖАНИЯ И СТРУКТУРЫ УЧЕБНОГО ПРОЦЕССА В ВУЗЕ С ИСПОЛЬЗОВАНИЕМ МЕТАМОДЕЛИ.

Сформирован алгоритм оптимизации учебного процесса на основе предложенных критериев оценки эффективности усвоения учебного материала.

МЕТОДЫ ОПТИМИЗАЦИИ. СОДЕРЖАНИЕ И СТРУКТУРА УЧЕБНОГО ПРОЦЕССА. ВУЗ. МЕТАМОДЕЛЬ. ПРОГРАММИРОВАННОЕ ОБУЧЕНИЕ. АЛГОРИТМ.

ABSTRACTS

KEYWORDS

Boldyrev Yu.Ya. THE ROLE OF THE SUPERCOMPUTER TECHNOLOGY IN THE ENGINEERING EDUCATION.

This article contains of common questions of the supercomputer technology and its function in the education process in the modern higher technical education.

SUPERCOMPUTER TECHNOLOGY. ENGINEERING SKILLS. EDUCATIONAL PROCESS. HIGHER TECHNICAL EDUCATION.

Smetanin A.G., Tormasov A.G. MATHEMATICAL MODEL AND EFFICIENT ALGORITHM TO FIND AND DISTRIBUTE DATA IN THE DECENTRALIZED DISTRIBUTED FILE SYSTEM.

The article describes the mathematical model of a decentralized, distributed file system. Theoretical estimates for the length of the search in this system are proven. The results of simulation are relevant to theoretical predictions.

DISTRIBUTED FILE SYSTEMS. DATA SEARCH ALGORITHM. FAULT TOLERANCE. SCALABILITY. DECENTRALIZATION. SIMULATION.

Zaborovsky V.S., Kondratiev A.S., Muliukha V.A., Silinenko A.V., Ilyashenko A.S., Filippov M.S. REMOTE ROBOTIC OBJECTS CONTROL WITHIN SPACE EXPERIMENTS «KONTUR» SERIES.

The article concerns the usage of different remote control modes of robotic objects on an example of Space Experiments «Kontur» Series, held by CNII RTC, RSC «Energia» and CNIIMash.

REMOTE ROBOTIC CONTROL. FORCE FEEDBACK. SPACE EXPERIMENT.

Gaaze M.V., Lychagin N.I., Malakhovskiy A.A. INTEGRATED TELECOMMUNICATION SERVICES FOR DEPARTMENTAL NETWORK COMMUNICATIONS.

It discusses the methodology of departmental integrated telecommunications services and describes the domestic certified software systems, based on this methodology.

TELECOMMUNICATION SYSTEMS. SOFTWARE PACKAGE. DEPARTMENTAL NETWORKS

Alekseyev V.G. DIGITAL WATERMARKS SYSTEM FOR AUDIO SIGNALS BASED ON THE REVERBERATION EFFECT.

Paper introduces an approach for embedding watermarks into audio signals by using the moderate reverberation effect which doesn't affect the music recordings quality significantly. The process is tolerant to using lossy bit-rate reduction based formats such as mp3 as well. To extract digital watermarks, we propose to use cepstral transformation and correlation receiver that differs from canonical models known from the communication theory. We use modeling techniques to calculate errors frequency and to optimize watermarks embedding parameters.

STEGANOGRAPHY. DIGITAL WATERMARKS. INFORMATION SECURITY. REVERBERATION.

Voronin V.V. AUTOMATED DEFECT DETECTION IN CASE OF RECONSTRUCTION OF ARCHIVE DOCUMENTARY PHOTOGRAPHS.

The paper developed a method for detecting defects in the archival documentary photographs by texture analysis and allocation of local descriptors in images. The algorithm developed by the modified method SURF allows selecting different types of local features with image pre-processing power characteristics Low's. The proposed approach allows us to find points of interests on the different types of textures and analyze defects on the surface of the images as a type of texture, such as edges, spots, ripples, waves, etc.

IMAGE RECONSTRUCTION. DEFECTS. ARCHIVAL PHOTOGRAPHS. RECOVERY.



Mukhin V.E., Kornaga Ya.I. MULTI-LEVEL MODEL WITH DIFFERENTIAL TRUST LEVEL TO SUBJECTS FOR THE DISTRIBUTED DATABASES DATA SECURITY IMPROVING.

In this paper are described the mechanisms for the data security ensuring in the distributed databases (DDB). There is suggested the approach to realization of the secure environment in DDB based on multi-level security model. To improve the data security in the DDB also is suggested the special trust-forming mechanisms. The implementation of the suggested security mechanisms allows to provide the high level of data security in terms of the dynamic changes of subjects of DDB.

DATABASE. SECURITY. MULTI-LEVEL MODEL. TRUST LEVEL.

Zaborovsky V.S., Kondratiev A.S., Muliukha V.A., Silinenko A.V., Ilyashenko A.S. SURFACE ROBOTIC GROUPS CONTROL FROM MANNED SPACE STATION WITHIN «METERON» PROJECT.

The article concerns the usage of the network-centric approach to the tasks of surface robotic groups control from manned space station.

ROBOTIC GROUPS CONTROL. NETWORK-CENTRIC. COMBINATED SATELLITE CHANNELS.

Kislitsin A.B., Rashich A.V. VITERBI DEMODULATOR FOR DOQPSK AND CDOQPSK SIGNALS.

The aim of this paper is to propose a method for reducing BER for signals with DOQPSK and CDOQPSK. In this method Viterbi algorithm is used to provide the reception of signals in general. It is shown that for BER 10^{-5} Viterbi algorithm gives 2,1–3,4 dB energy gain (including peak-to-average power ratio) compared to $\pi/4$ -DQPSK signals.

PEAK-TO-AVERAGE POWER RATIO. DOQPSK. CDOQPSK. VITERBI ALGORITHM.

Golubev A.G., Smirnov A.S. CALCULATION OF UPPER LIMIT OF SENSITIVITY FOR INTERFERENCE FIBER-OPTIC HYDROPHONE ARRAY.

The paper is devoted to calculation of upper limit of optimal hydrophone sensitivity to achieve the required dynamic range under minimum specified acoustic noise.

FIBER-OPTIC HYDROPHONE. PROCEDURE. SENSITIVITY. DYNAMIC RANGE.

Chuprov S.G., Ivanov A.A. MODELLING OF DYNAMICS OF TWO-MOTOR SERVO DRIVE WITH SELF-LOCKING EFFECT.

In clause problems of management of two-motor servo drive with self-locking effect are considered. We construct a mathematical model of the drive. This model allows to formulate an approach to the management of such a drive.

MANAGEMENT. MULTIMOTOR SERVO DRIVE. SELF-LOCKING. DYNAMICS. MODELLING.

Petrosyan G.S. ANALYSIS OF THE OBJECT-ORIENTED PARADIGM AND SEARCH FOR A BETTER ALTERNATIVE.

Major problems of industrial programming are outlined and object-oriented paradigm of solving them is evaluated. Drawbacks of the solutions are considered, and more simple and agile approach is proposed.

PROGRAMMING LANGUAGES. OBJECT-ORIENTED PROGRAMMING. ENCAPSULATION. CODE REUSE. POLYMORPHISM.

Pyshkin E.V. PROBLEMS OF SOFTWARE ACCEPTANCE TESTING AUTOMATION IN BEHAVIOR DRIVEN DEVELOPMENT.

Paper is focused on problems of acceptance tests automation in behavior driven development (BDD). We illustrate basic ideas of BDD by some development and testing examples. Special attention is being paid to difficulties of converting customer-side test stories into executable unit tests. We analyze process of marking up of test stories by a QA engineer for their subsequent automated conversion into the corresponding test classes and test methods. We define tasks for developing further tools aimed to support acceptance tests automation and to improve BDD based test story language.

SOFTWARE ENGINEERING. SOFTWARE TESTING. AGILE PRACTICES. QUALITY ANALYSIS.

Laboshin L.Yu., Lukashin A.A., Semenovskiy V.B. METHOD TO PROVIDE GPU SUPPORT FOR VIRTUAL MACHINES IN XEN HYPERVISOR FOR DATA STREAM PROCESSING IN ROBOTIC SYSTEMS.

Introduced problems of the graphic processor passthrough in a virtual environment. Developed method for transferring the graphics controller to the virtual machine. Evaluated performance of the calculations using GPU for scientific tasks in virtual environment. Proposed appliance of developed method to support data processing in robotic systems.

VIRTUALIZATION. VIRTUAL MACHINE. HYPERVISOR. GPU. PCI. IOMMU. PASSTHROUGH.

Valynets E.V., Orzhenovsky M.V., Ivankov A.A. STOCHASTIC MODEL OF SUPPLIER BIDS POOL AT THE ENERGY SPOT PRICE AUCTION.

Stochastic model was developed in order to identify the structure of supplier bids pool at the sealed-bid auction of energy spot price. The data on power generation committed by spot market participants to be in use for estimation.

ENERGY SPOT-PRICE. SEALED-BID AUCTION. ENERGY SUPPLIERS. STOCHASTIC MODEL. STRUCTURE OF THE POOL OF BIDS.

Ivanov A.A. METHOD OF A FORM'S PLANNING OF THE HYPER REDUNDANT MANIPULATOR'S WITH A LIMITED VARIATION RANGE OF JOINT COORDINATES.

Formulation and solution method of the inverse kinematics problem for manipulators with arbitrary number of connections and with limited range of joint coordinates variation are proposed. The results of numerical simulations are demonstrated the efficiency and effectiveness of the method.

MANIPULATOR. KINEMATIC REDUNDANCY. INVERSE PROBLEM OF KINEMATICS. SIMULATION. OPTIMIZATION.

Kadyrova N.O., Pavlova L.V. METHODS FOR EFFICIENT PROCESSING AND MINING OF BIG DATA.

A new methodology has been developed for processing and mining of very large data sets of high dimensionality based on the modern efficient approaches for binary classification. Several examples are given of the real life applications.

BIG DATA (ANALYTICS). DATA MINING. DATA SCIENCE. MACHINE LEARNING.

Petrova I.Yu., Khomenko T.V. MATHEMATICAL MODEL OF THE SELECTION AND METHOD OF EVALUATION OF OPTIONS FOR THE PHYSICAL PRINCIPLE OF THE ACTION OF THE SENSING ELEMENT.

The solution of the problem of choosing physical principle of the action (PPA) of the sensing element (SE), which can be found in the initial stages of the design process sensor equipment and comes to the selection of one or more variants of the received/synthesized set of PPA SE, essentially depend on the information on the selection criteria. We propose a mathematical model of the possible variations of conditions for the selection and informative method of evaluation of the selected solution.

DESIGN. SOLUTION CHOICE. AMOUNT OF INFORMATION. EVALUATION METHOD.

Chernorytsky I.G. STANDARD SCHEMES OF PARAMETRIC OPTIMIZATION.

The methods of smooth criteria constructing for minimax optimization problems and the problems of systems of inequalities solutions are presented. Discussed in the article smooth approximations of criteria allow to use the simplest implementations of the standard algorithms for nonlinear programming.

PARAMETRIC OPTIMIZATION. MINIMAX PROBLEMS. THE SYSTEM OF INEQUALITIES. SMOOTH OPTIMALITY CRITERIA.

Martsenyuk M.A., Polyakov V.B., Seletkov I.P. MATRIX IMPLEMENTATION OF FUZZY INFERENCE ALGORITHMS.

Matrix implementation of fuzzy inference algorithms that generalize by natural way of crisp logic derivation rules as modus ponens, modus tollens and like that are considered. It is shown that in suggested representation fuzzy inference procedure is equivalent to solving some system of linear algebraic equation with the usual existence condition for the solution. Examples that demonstrate effectiveness of method are considered in details. It is also performed that matrix representation simplifies programming fuzzy logic applications on controllers.

FUZZY LOGIC. MATRIX IMPLEMENTATION OF FUZZY VARIABLE. FUZZY INFERENCE PROCEDURE. FUZZY LOGIC DERIVATION PROCEDURE. PROGRAMMING OF FUZZY LOGIC ON CONTROLLERS. FUZZY PREDICATE.

Chernorytsky I.G. ALGORITHMIC PROBLEMS OF STIFF OPTIMIZATION.

Introduced the definition of stiff functionals in problems of parametric optimization. Specify the sources of stiffness in the problems of computer modelling.

PARAMETRIC OPTIMIZATION. STIFF FUNCTIONALS. STIFF DIFFERENTIAL SYSTEMS.

Shebalov A.A., Bazhenov A.N. PERFORMANCE ANALYSIS OF OPTICAL FLOW TECHNIQUES.

Paper is devoted to methods of optical flow calculation and evaluation of their performance. Experimental comparison of classical and modern algorithms performed, results analyzed in terms of precision and speed.

IMAGE PROCESSING. OPTICAL FLOW. ALGORITHMS. METHODS. TECHNIQUES. COMPARISON. ANALYSIS.

Anikeeva M.S., Sofronov A.N., Dremov S.S., Ter-Martirosyan A.L. DIGITAL IMAGING SYSTEM FOR PHOTODYNAMIC THERAPY.

System for visualization of spatial distribution of fluorescence during photodynamic therapy procedures is presented. Object-oriented analysis and design of the system is performed, visualization software is presented. Optimal realization is created using external open source components. Developed visualisation system including software and hardware was successfully finished clinical trials.

VISUALISATION. PHOTODYNAMIC THERAPY. OBJECT-ORIENTED ARCHITECTURE.

Latypova E.A., Bondareva I.O. SIMULATION AS A MEANS OF LOGISTICS ENTERPRISES BY THE EXAMPLE OF THE CARGO PORT.

We consider the possibility of increasing the efficiency of cargo port business process management. We describe the technology of balanced scorecard creating. Suggested using simulation modelling to assess, predict and adjust the operation of the port.

TOTAL QUALITY MANAGEMENT. BALANCED SCORECARD. CARGO PORT. BUSINESS PROCESS. PORT STRATEGY. THE TREE OF GOALS. STRATEGIC MAP. MANAGEMENT DECISIONS. THE EFFICIENCY. COMPETITIVENESS. SIMULATION.

Aksyonov K.A., Aksyonova O.P., Wang Kai HOUSE-BUILDING PROJECTS PORTFOLIO SCHEDULING BASED ON MULTI AGENT SIMULATION.

In this work is considered the development of hybrid scheduling and bottlenecks analysis method of projects portfolio in house-building based on multi agent simulation, critical path method and operational analysis of probabilistic networks.

PROJECTS PORTFOLIO. SCHEDULING. HOUSE-BUILDING. SIMULATION MODELLING. MULTI AGENT SYSTEMS. BOTTLENECKS ANALYSIS.

Rybak V.A., Suleyman H.M. ANALYSIS AND OPTIMIZATION OF ENVIRONMENTAL AND ECONOMIC PARAMETERS OF INNOVATIVE PROJECTS.

In article a new technology for analysis and optimization of environmental and economic parameters of innovative projects from the standpoint of environmental management and information management is described. Methods and tools allow for the support of decision-making choosing the most effective projects.

ENVIRONMENTAL MANAGEMENT. MANAGEMENT. FUZZY SETS. ENVIRONMENTAL AND ECONOMIC INDICATORS. INNOVATIVE PROJECTS.

Maslov V.I., Minakov V.F. AN EFFICIENCY CRITERION IN THE ENTERPRISE PRODUCTION QUALITY CONTROL SYSTEM.

Quality management principles quantitative indices develop: elasticity of quality at the price and on the expenses, characterizing a quality gain on unit of growth of the price and expenses respectively, for example, in percentage expression. The ratio of elasticity of quality of the different manufacturing enterprises of production characterizes their competitiveness. The ratio of price elasticity of quality (demand for quality) with elasticity of quality on expenses (the quality offer) is criterion of the economic efficiency, allowing to make the decision on expediency of actions on ensuring quality, to carry out their comparison and a choice.

PRICE ELASTICITY OF QUALITY. ELASTICITY OF QUALITY ON EXPENSES.

Kalmykova S.V., Sokoliscin A.S. OPTIMIZING THE CONTENT AND STRUCTURE OF THE EDUCATIONAL PROCESS AT THE UNIVERSITY WITH USING METAMODELS.

Optimization algorithm of the learning process based on the proposed criteria for evaluation of the learning formed in the article.

METHODS OF OPTIMIZATION. CONTENT AND STRUCTURE OF THE EDUCATIONAL PROCESS. HIGH SCHOOL. METAMODEL. PROGRAM EDUCATION. ALGORITHM.

НАУЧНОЕ ИЗДАНИЕ
«НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ
САНКТ-ПЕТЕРБУРГСКОГО
ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА.
ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ»
«ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL.
COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS»

№ 6 (162) 2012

Учредитель – Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Санкт-Петербургский государственный политехнический университет»

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор).
Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Редакция журнала

канд. техн. наук, д-р экон. наук, профессор *А.В. Бабкин* – научный редактор
Е.А. Калинина – литературный редактор, корректор
Г.А. Пышкина – ответственный секретарь, выпускающий редактор

Телефон редакции (812)552-62-16, 297-18-21

E-mail: infocom@spbstu.ru

Компьютерная верстка *А.Н. Смирнов*

Директор Издательства Политехнического университета *А.В. Иванов*

Лицензия ЛР № 020593 от 07.08.97

Подписано в печать 25.12.2012. Формат 60×84 1/8. Бум. тип. № 1.
Печать офсетная. Усл. печ. л. 24,18. Уч.-изд. л. 24,18. Тираж 1000. Заказ

Санкт-Петербургский государственный политехнический университет
Издательство Политехнического университета
член Издательско-полиграфической ассоциации университетов России
Адрес университета и издательства: 195251, Санкт-Петербург, ул. Политехническая, д. 29.

УСЛОВИЯ ПУБЛИКАЦИИ СТАТЕЙ

в журнале «Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление»

1. ОБЩИЕ ПОЛОЖЕНИЯ

Журнал «Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление» является периодическим печатным научным рецензируемым изданием. Зарегистрировано Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19 октября 2012 г. С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» (ISSN 1994-2354).

Издание с 2002 года входит в Перечень ведущих научных рецензируемых журналов и изданий (перечень ВАК) и принимает для печати материалы научных исследований, а также статьи для опубликования основных результатов диссертаций на соискание ученой степени доктора наук и кандидата наук по следующим основным научным направлениям: ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА, РАДИОТЕХНИКА И СВЯЗЬ, ЭЛЕКТРОНИКА, ИЗМЕРИТЕЛЬНАЯ ТЕХНИКА, УПРАВЛЕНИЕ В СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ СИСТЕМАХ. Научные направления журнала учитываются ВАК Минобрнауки РФ при защите докторских и кандидатских диссертаций в соответствии с Номенклатурой специальностей научных работников.

Сведения о публикации представлены в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich's Periodical Directory».

Периодичность выхода журнала – 6 номеров в год.

2. ТРЕБОВАНИЯ К ПРЕДОСТАВЛЯЕМЫМ МАТЕРИАЛАМ

2.1. Оформление материалов

1. Объем статей докторов наук, профессоров, докторантов, соискателей ученой степени доктора наук, как правило, 12–20 страниц формата А-4. Количество рисунков не должно превышать 5, таблиц – 4, литературных источников – 15.

2. Объем статей преподавателей, сотрудников, аспирантов, соискателей ученой степени кандидата наук, как правило, 8–15 страниц формата А-4, объем статей аспирантов – 8 страниц формата А-4. Количество рисунков не должно превышать 4, таблиц – 3, литературных источников – 10.

3. Авторы должны придерживаться следующей обобщенной структуры статьи: вводная часть (0,5–1 стр., актуальность, существующие проблемы); основная часть (постановка и описание задачи, изложение и суть основных результатов); заключительная часть (0,5–1 стр., предложения, выводы), список литературы (оформление по ГОСТ 7.1-2003).

4. Число авторов статьи не должно превышать трех человек.

5. Набор текста осуществляется в редакторе **MS Word**, формул – в редакторе **MS Equation 3.0** или **MythType**. Таблицы набираются в том же формате, что и основной текст.

6. Шрифт – TNR, размер шрифта основного текста – 14, интервал – 1,5; таблицы большого размера могут быть набраны 12 кеглем. Параметры страницы: поля слева – 3 см, сверху, снизу – 2,5 см, справа – 2 см, текст размещается без переносов. Абзацный отступ – 1 см.

2.2. Предоставление материалов

Вместе с материалами статьи должны быть обязательно предоставлены:

- номер УДК в соответствии с классификатором (в заголовке статьи);
- аннотация (2–3 предложения) на русском и английском языках;
- ключевые слова (5–7) на русском и английском языках;

- сведения об авторах на русском и английском языках: ФИО, место работы, должность, ученое звание, ученая степень, контактные телефоны, e-mail;
- аспиранты представляют документ отдела аспирантуры, заверенный печатью;
- рекомендация на имя зам. главного редактора журнала, подписанная научным руководителем автора и/или руководителем подразделения. Рекомендация должна быть **ОБЯЗАТЕЛЬНО** заверена в отделе кадров;
- рецензия на имя зам. главного редактора, подписанная специалистом, имеющим ученую степень доктора наук и/или ученое звание профессора. Рецензия должна быть **ОБЯЗАТЕЛЬНО** заверена в отделе кадров. Рецензент несет ответственность за содержание статьи, достоверность представленных материалов.

При необходимости редколлегия может потребовать представления акта экспертизы, о чем она сообщает автору (авторам).

С авторами статей заключается издательский лицензионный договор.

Предоставление всех материалов осуществляется по электронной почте на адрес редакции: **infocom@spbstu.ru**

2.3. Рассмотрение материалов

Предоставленные материалы (п. 2.2) первоначально рассматриваются редакционной коллегией и передаются для рецензирования. После одобрения материалов, согласования различных вопросов с автором (при необходимости) редакционная коллегия сообщает автору решение об опубликовании статьи. В случае отказа в публикации статьи редакция направляет автору мотивированный отказ.

При отклонении материалов из-за нарушения сроков подачи, требований по оформлению или как не отвечающих тематике журнала материалы не публикуются и не возвращаются.

Редакционная коллегия не вступает в дискуссию с авторами отклоненных материалов.

Публикация материалов аспирантов очной бюджетной формы обучения осуществляется бесплатно в соответствии с очередностью.

При поступлении в редакцию значительного количества статей их прием в очередной номер может закончиться **ДОСРОЧНО**.

Более подробную информацию можно получить по телефону/факсу редакции:

8(812) 552-62-16 с 10⁰⁰ до 18⁰⁰ Галина Александровна

или по e-mail: infocom@spbstu.ru