

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Национальный исследовательский
Томский политехнический университет»



ISSN 1684-8519

ИЗВЕСТИЯ

ТОМСКОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Том 320, № 5, 2012

Управление,
вычислительная техника
и информатика

ИЗДАТЕЛЬСТВО  ТПУ

г. Томск

ИЗВЕСТИЯ ТОМСКОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Редакционный совет:

Чубик П.С. (председатель), д.т.н., ректор ТПУ (г. Томск)
Власов В.А. (зам. председателя), д.ф.-м.н.,
проректор ТПУ по НИИ (г. Томск)
Алексеенко С.В., д.ф.-м.н.,
чл.-корр. РАН (г. Новосибирск)
Болдырев В.В., д.х.н., академик РАН, (г. Новосибирск)
Боровиков Ю.С., к.т.н. (г. Томск)
Гвоздев Н.И., к.т.н. (г. Томск)
Гуляев Ю.В., д.ф.-м.н., академик РАН (г. Москва)
Дамамм Ж., д.н. (Франция)
Ершов Ю.Л., д.ф.-м.н., академик РАН (г. Новосибирск)
Клименов В.А., д.т.н. (г. Томск)
Конторович А.Э., д.г.-м.н., академик РАН (г. Новосибирск)
Крёнинг М., д.н. (Германия)
Кривобоков В.П., д.ф.-м.н. (г. Томск)
Летников Ф.А., д.г.-м.н., академик РАН (г. Иркутск)
Лопатин В.В., д.ф.-м.н. (г. Томск)
Мазуров А.К., д.г.-м.н. (г. Томск)
Месяц Г.А., д.ф.-м.н., академик РАН (г. Москва)
Михайленко Б.Г., д.ф.-м.н.,
академик РАН (г. Новосибирск)
Накоряков В.Е., д.т.н., академик РАН (г. Новосибирск)
Панин В.Е., д.ф.-м.н., академик РАН (г. Томск)
Сигов А.С., д.ф.-м.н., академик РАН (г. Москва)
Сигфуссон Т.И., д.н. (Исландия)
Сонькин М.А., д.т.н. (г. Томск)
Третьяков Ю.Д., д.х.н., академик РАН (г. Москва)
Турнаев В.И., д.и.н. (г. Томск)
Филиппов Г.А., д.т.н., академик РАН (г. Москва)
Шень Джаоли, д.н. (Китай)

Редакционная коллегия:

Власов В.А. (гл. редактор), д.ф.-м.н.
Коробейников А.Ф. (зам. гл. редактора), д.г.-м.н.
Могильницкий С.Б. (ученый секретарь), к.ф.-м.н.
Барышева Г.А., д.э.н.
Заворин А.С., д.т.н.
Григорьев В.П., д.ф.-м.н.
Корниенко А.А., д.ф.н.
Кривобоков В.П., д.ф.-м.н.
Лисицын В.М., д.ф.-м.н.
Погребной В.К., д.т.н.
Потылицын А.П., д.ф.-м.н.
Усов Ю.П., д.т.н.
Филимонов В.Д., д.х.н.

Журнал зарегистрирован Министерством
Российской Федерации по делам печати,
телерадиовещания и средств массовых коммуникаций.
Свидетельство ПИ № 77-16615 от 24 октября 2003 г.
Учредитель: Томский политехнический университет

Издается с 1903 г.

© ФГБОУ ВПО НИ ТПУ, 2012

BULLETIN OF THE TOMSK POLYTECHNIC UNIVERSITY

Editorial Board:

Chubik P.S. (Chairman), D.E., rector of TPU (Tomsk)
Vlasov V.A. (Deputy chairman), Phys. and Math. D. Sc.,
pro-rector of TPU for Research and Innovation (Tomsk)
Alekseenko S.V., Phys. and Math. D. Sc.,
corresponding member of RAS (Novosibirsk)
Boldyrev V.V., D. Chem., member of RAS (Novosibirsk)
Borovikov Yu.S., Candidate of Science (Tomsk)
Gvozdev N.I., Candidate of Science (Tomsk)
Gulyaev Yu.V., Phys. and Math. D. Sc., member of RAS (Moscow)
Damamme G., Phys. and Math. D. Sc. (France)
Ershov Yu.L., Phys. and Math. D. Sc., member of RAS (Novosibirsk)
Klimenov V.A., D.E. (Tomsk)
Kontorovich A.E., Geol. and Mineral. D. Sc., member of RAS (Novosibirsk)
Kröning M., Dr.h.c. (Germany)
Krivobokov V.P., Phys. and Math. D. Sc. (Tomsk)
Letnikov F.A., Geol. and Mineral. D. Sc., member of RAS (Irkutsk)
Lopatin V.V., Phys. and Math. D. Sc. (Tomsk)
Mazurov A.K., Geol. and Mineral. D. Sc. (Tomsk)
Mesyats G.A., Phys. and Math. D. Sc., member of RAS (Moscow)
Mikhailenko B.G., Phys. and Math. D. Sc.,
member of RAS (Novosibirsk)
Nakoryakov V.E., D. E., member of RAS (Novosibirsk)
Panin V.E., Phys. and Math. D. Sc., member of RAS (Tomsk)
Sigov A.S., Phys. and Math. D. Sc., member of RAS (Moscow)
Sigfusson T.I., Ph. D. (Island)
Sonkin M.A., D.E. (Tomsk)
Tretyakov Yu.D., D. Chem., member of RAS (Moscow)
Turnaev V.I., Ph. D. (Tomsk)
Filippov G.A., D.E., member of RAS (Moscow)
Shen Zhaoli, Ph. D. (China)

Editorial:

Vlasov V.A. (Editor in Chief), Phys. and Math. D. Sc.
Korobeinikov A.F. (Deputy Editor in Chief), Geol. and Mineral. D. Sc.
Mogilnitsky S.B. (Science Secretary), Candidate of Phys. and Math. Sc.
Barysheva G.A., Ec. D.
Zavorin A.S., D.E.
Grigoriev V.P., Phys. and Math. D. Sc.
Kornienko A.A., Ph. D.
Krivobokov V.P., Phys. and Math. D. Sc.
Lisitzyn V.M., Phys. and Math. D. Sc.
Pogrebnoy V.K., D.E.
Potylitzin A.P., Phys. and Math. D. Sc.
Usov Yu.P., D.E.
Filimonov V.D., D. Chem.

Подписной индекс по каталогу
Агентства «Роспечать» – 18054

Журнал рассылается в адреса 50-и библиотек РФ,
США, ФРГ, Великобритании, Франции
и 9-и стран ближнего зарубежья

Полнотекстовый доступ к электронной версии
журнала возможен на сайтах ТПУ: portal.tpu.ru/izvestiya/;
ООО «Научная электронная библиотека»:
www.elibrary.ru, www.e-library.ru,
а также поисковой системы scholar.google.com

Импакт-фактор РИНЦ 2011 г. 0,210

© Tomsk Polytechnic University, 2012

СОДЕРЖАНИЕ

УПРАВЛЕНИЕ, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И ИНФОРМАТИКА	
Разработка метода бинарной декомпозиции функциональных предикатов Рудометкина М.Н.	5
Асимптотический анализ систем массового обслуживания с неограниченным числом приборов и полумарковским входящим потоком Назаров А.А., Семенова И.А.	12
Построение и исследование алгоритма поиска первой крайней подсистемы для заданной совместной системы линейных неравенств Небаба С.Г., Вылегжанин О.Н.	17
Матричный алгоритм решения задачи разрезания двудольных графов Погребной Ан.В., Погребной Д.В.	21
Адресуемая ячейка однородной вычислительной структуры для решения дифференциальных уравнений в частных производных Хамухин А.А.	25
Архитектура распределенного вычислительного комплекса для двумерного анализа изображений дисков деревьев Ботыгин И.А., Попов В.Н., Тартаковский В.А.	30
Использование онтологии в электронных библиотеках Ле Хоай, Тузовский А.Ф.	36
Классификация текстов на основе оценки семантической близости терминов Нгуен Ба Нгок, Тузовский А.Ф.	43
Алгоритмическое и программное обеспечение для распознавания формы руки в реальном времени с использованием SURF-дескрипторов и нейронной сети Нгуен Тоан Тханг, Спицын В.Г.	48
Распознавание лиц на основе применения метода Виолы-Джонса, вейвлет-преобразования и метода главных компонент Буй Тхи Тху Чанг, Фан Нгок Хоанг, Спицын В.Г.	54
Алгоритмы для классификации отпечатков пальцев на основе применения фильтра Габора, вейвлет-преобразования и многослойной нейронной сети Фан Нгок Хоанг, Спицын В.Г.	60
Алгоритмы выделения лиц на статических RGB изображениях и в видеопотоке Друки А.А.	65
Простой алгоритм контроля битрейта кодера стандарта JPEG Сидоров Д.В., Осокин А.Н.	70
Построение многокомпонентных визуальных 3D-моделей с использованием разнородных источников информации, на примере создания геологических моделей Захарова А.А., Шкляр А.В.	73
Исследование разрешающей способности методов фазочастотного прослеживания сейсмических сигналов Иванченков В.П., Кочегуров А.И., Орлов О.В.	80
Параметрический синтез ПИ-регулятора линейной САУ на основе коэффициентных оценок степени устойчивости и заданной добротности Пушкарев М.И., Гайворонский С.А.	85
Идентификация объектов управления в форме дискретных передаточных функций на основе вещественного интерполяционного метода Рудницкий В.А., Алексеев А.С., Курганкин В.В.	89
Синтез математической модели подсистемы управления прогревом вакуумной камеры Казахстанского материаловедческого Токамака КТМ Коровиков А.Г., Павлов В.М., Ольховик Д.А.	94
Система управления промышленной установкой «КО-МКР» Пащенко О.В., Соловьев А.М., Юдаков С.В.	100
Использование имитационного комплекса для отладки системы управления процессом переработки натрия Кривцов П.Ю., Павлов В.М.	105

CONTENTS

CONTROL, COMPUTER SCIENCE AND TECHNOLOGY	
The development of binary decomposition method for function predicates Rudometkina M.N.	5
Asymptotic analysis of queuing systems with infinite number of servers and semi-Markovian input flow Nazarova A.A., Semenova I.A.	12
Construction and study of search algorithm of the first extreme subsystem for a given consistent system of linear inequalities Nebaba S.G., Vylegzhanin O.N.	17
Matrix algorithm for solving the problem of cutting bipartite graphs Pogrebnoy An.V., Pogrebnoy D.V.	21
Addressable cell of homogeneous computing structure for solving partial differential equations Khamukhin A.A.	25
Architecture of distributed computing system for two-dimensional analysis of tree disk image Botygin I.A., Popov V.N., Tartakovskiy V.A.	30
Using ontology in electronic libraries Le Hoai, Tuzovskiy A.F.	36
Text classification based on estimation of terms semantic similarity Nguyen Ba Ngoc, Tuzovskiy A.F.	43
Algorithms and software for hand shape on-line recognition using SURF descriptors and neural network Nguyen Toan Thang, Spitsyn V.G.	48
Face recognition based on application of Viola-Jones method, wavelet transform and principal component analysis Bui Thi Thu Trang, Phan Ngoc Hoang, Spitsyn V.G.	54
Algorithms for fingerprint classification based on application of Gabor filter, wavelet transform and multilayer neural network Phan Ngoc Hoang, Spitsyn V.G.	60
Algorithms of person selection on static RGB images and in video stream Druki A.A.	65
Simple algorithm for controlling bit rate of JPEG coder Sidorov D.V., Osokin A.N.	70
Construction of multicomponent visual 3D-models using heterogeneous data sources by the example of developing geological models Zakharova A.A., Shklyar A.V.	73
Studying the resolving ability in the techniques of seismic signal phase-frequency tracing Ivanchenkov V.P., Kochegurov A.I., Orlov O.V.	80
Parametric synthesis of PI-controller in linear automated control system based on coefficient estimations of stability degree and prescriber Q-factor Pushkarev M.I., Gayvoronskiy S.A.	85
Identification of controlled objects in the form of sampled-data transfer functions based on real interpolation method Rudnitskiy V.A., Alekseev A.S., Kurgankin V.V.	89
Synthesis of mathematical model of vacuum chamber heat control subsystem in Kazakhstan material testing tokamak KTM Korovikov A.G., Pavlov V.M., Olkhovik D.A.	94
The control system of industrial unit «KO-MKR» Pashchenko O.V., Solovyev A.M., Yudakov S.V.	100
Use of simulation complex for checking the control system of sodium process Krivtsov P.Yu., Pavlov V.M.	105

Передача данных в беспроводных сенсорных сетях с приоритетами на основе агрегирования предпочтений Муравьев С.В., Тараканов Е.В.	111	Data transmission in wireless sensor networks with priorities based on preference aggregation Muravyev S.V., Tarakanov E.V.
Цифровая обработка импульсных сигналов для определения частотных характеристик преобразователей тока Заревич А.И., Муравьев С.В., Бедарева Е.В., Величко О.Н.	116	Digital pulse signal processing for determining frequency response of current transducers Zarevich A.I., Muravyev S.V., Bedareva E.V., Velichko O.N.
Применение технологии CUDA для ускорения вычислений в нейронных сетях Парубец В.В., Берестнева О.Г., Девятых Д.В.	121	Application of CUDA technology for computational speedup in neural networks Parubets V.V., Berestneva O.G., Devyatykh D.V.
Применение методов визуализации при исследовании структуры экспериментальных многомерных данных Воловоденко В.А., Берестнева О.Г., Немеров Е.В., Осадчая И.А.	125	Imaging application in studying the structure of experimental multidimensional data Volovodenko V.A., Berestneva O.G., Nemerov E.V., Osadchaya I.A.
Оценивание состава изображений клеточных препаратов для медико-биологических исследований Тарков М.С., Тихонов Н.В., Половинкин В.Г.	130	Estimation of cell preparation image for medical-biological research Tarkov M.S., Tikhonov N.V., Polovinkin V.G.
Преимущества дифференциальных моделей в эколого-экономическом моделировании Затонский А.В.	134	The advantages of differential models in ecological-economic modeling Zatonskiy A.V.
НАШИ ЮБИЛЯРЫ		
Профессору В.К. Погребному – 70 лет	140	Professor V.K. Pogrebnoy is 70
Профессору А.М. Корикову – 70 лет	142	Professor A.M. Korikov is 70
Профессору А.А. Мицелю – 65 лет	144	Professor A.A. Mitsel is 65
РЕФЕРАТЫ СТАТЕЙ		
	147	SUMMARIES
СВЕДЕНИЯ ОБ АВТОРАХ		
	150	INFORMATION ABOUT AUTHOR

Управление, вычислительная техника и информатика

УДК 004.8

РАЗРАБОТКА МЕТОДА БИНАРНОЙ ДЕКОМПОЗИЦИИ ФУНКЦИОНАЛЬНЫХ ПРЕДИКАТОВ

М.Н. Рудометкина

Томский политехнический университет
Карагандинский государственный технический университет
E-mail: monika_7272@mail.ru

Проведен анализ единственного обоснованного метода бинарной декомпозиции конечных предикатов – метода декартовой декомпозиции. Впервые разработан и обоснован частный метод бинарной декомпозиции для функциональных предикатов, основанный на анализе зависимостей, при этом количество значений вспомогательной переменной сведено к минимуму. На основе предложенного метода создан алгоритм бинарной декомпозиции функциональных предикатов.

Ключевые слова:

Бинарная декомпозиция, предикат, функциональные предикаты, декартова декомпозиция.

Key words:

Binary decomposition, predicate, function predicate, Descartes decomposition.

Средства декомпозиции предикатов разрабатывались в работах Ю.Н. Павловского, М.Ф. Бондаренко, Ю.П. Шабанова-Кушнаренко, А.И. Мальцева [1–4]. Однако обзор работ этих авторов показал, что в области аппарата декомпозиции обнаруживается недостаток обоснованных, не требующих проверки эквивалентности полученного результата исходному предикату, и в то же время рациональных методов, позволяющих получать компактное бинарное представление. Следует отметить, что почти все имеющиеся формальные средства декомпозиции предикатов в указанных работах не учитывают зависимости между переменными. Учет таких особенностей в строении предиката позволяет выполнять его декомпозицию более рациональным путем. Между тем теория реляционных баз данных предлагает средства декомпозиции отношений в виде утверждений о зависимостях между атрибутами. Если взять во внимание наличие связи между реляционными отношениями и предикатами, то можно использовать средства декомпозиции отношений для декомпозиции предикатов. Как показал анализ литературы, такие средства в аппарате алгебры предикатов на данный момент отсутствуют.

Целью данной статьи является разработка бинарной конъюнктивной декомпозиции функциональных предикатов.

Функциональным (однозначным) назовем предикат $P \in \text{Pre}(S)$ с функциональной зависимостью $X \rightarrow Y$, для которого $X \cap Y = \emptyset$. Иначе говоря, в функциональном предикате должна содержаться функ-

циональная зависимость, которая охватывает все его переменные. Предикат $P \in \text{Pre}(S)$ может содержать функциональную зависимость $X \rightarrow Y$, но при этом не быть функциональным. Это возможно в случае, когда $X \cap Y \neq \emptyset$, т. е. когда функциональная зависимость охватывает не все переменные предиката P .

Для постановки задачи вводим обозначения, характеризующие предметное пространство. Пусть задан остов $T \subseteq V \times U$, а конечное множество V состоит из двух непересекающихся групп переменных: $X = \{x_1, x_2, \dots, x_m\}$ и $Z = \{z_1, z_2, \dots, z_n\}$, т. е. $V = \{x_1, x_2, \dots, x_m, z_1, z_2, \dots, z_n\}$. В качестве универсума предметов выбирается произвольное непустое не более чем счетное множество U , элементы которого называются предметами. Пусть предметное пространство S размерности n образуется с помощью остова T как декартово произведение областей изменения переменных из V .

Теперь задан произвольный конечный предикат $P \in \text{Pre}(S)$ с функциональной зависимостью $X \rightarrow Z$. Необходимо выполнить бинарную декомпозицию предиката $P(X, Z)$ с помощью одной вспомогательной переменной.

С точки зрения дальнейшей реализации модели важно, чтобы результат декомпозиции получился как можно более компактным (формулы полученных предикатов должны быть простыми, как и выражаемые ими отношения). Для этого надо, чтобы число значений вспомогательной переменной в пределах предикатов, полученных в результате декомпозиции, было минимальным.

Единственным обоснованным методом бинар-

ной декомпозиции предикатов, гарантирующим корректность результата, является, декартова декомпозиция [1], которая позволяет выполнять декомпозицию любых конечных предикатов, а не только функциональных. Хотя недостатки этого метода и обсуждались ранее, рассмотрим декартову декомпозицию как отправную точку для поиска более рационального метода в смысле экономии аппаратных ресурсов.

Для выполнения декартовой декомпозиции предиката $P \in \text{Pre}(S)$ необходимо в явном виде представить его область истинности. В качестве такого вида можно использовать совершенную нормальную дизъюнктивную форму, предиката P . Пусть P_T – область истинности предиката P , $P_T \subseteq S$. Тогда

$$P(X, Z) = \bigvee_{(a_1, a_2, \dots, a_m, c_1, c_2, \dots, c_n) \in P_T} x_1^{a_1} x_2^{a_2} \dots x_m^{a_m} z_1^{c_1} z_2^{c_2} \dots z_n^{c_n} = \bigvee_{i=1}^r x_1^{a_{i1}} x_2^{a_{i2}} \dots x_m^{a_{im}} z_1^{c_{i1}} z_2^{c_{i2}} \dots z_n^{c_{in}},$$

где r – количество кортежей отношения P_T (которое будет конечно в силу конечности предиката P).

Согласно методу декартовой декомпозиции вводится вспомогательная переменная y со значениями $B = \{b_1, b_2, \dots, b_r\}$, которые интерпретируются как уникальные имена кортежей отношения P_T . Каждому кортежу из P_T произвольным образом взаимно однозначно сопоставляется уникальное имя:

$$\begin{aligned} x_1^{a_{11}} x_2^{a_{12}} \dots x_m^{a_{1m}} z_1^{c_{11}} z_2^{c_{12}} \dots z_n^{c_{1n}} &= y^{b_1}, \\ x_1^{a_{21}} x_2^{a_{22}} \dots x_m^{a_{2m}} z_1^{c_{21}} z_2^{c_{22}} \dots z_n^{c_{2n}} &= y^{b_2}, \\ &\dots \\ x_1^{a_{r1}} x_2^{a_{r2}} \dots x_m^{a_{rm}} z_1^{c_{r1}} z_2^{c_{r2}} \dots z_n^{c_{rn}} &= y^{b_r}. \end{aligned} \quad (1)$$

На основе системы (1) определяется вспомогательный предикат P_y :

$$P_y(X, Z, y) = \bigvee_{i=1}^r y^{b_i} x_1^{a_{i1}} x_2^{a_{i2}} \dots x_m^{a_{im}} z_1^{c_{i1}} z_2^{c_{i2}} \dots z_n^{c_{in}}. \quad (2)$$

Так как предикат $P_y(X, Z, y)$ выражает взаимно однозначное соответствие между каждым кортежем (X, Z) и его именем y , то он удовлетворяет паре функциональных зависимостей $y \leftarrow \{x_1, x_2, \dots, x_m, z_1, z_2, \dots, z_n\}$. Тогда в силу следствия из теоремы Хеза предикат $P_y(X, Z, y)$ удовлетворяет зависимости конъюнкции $\wedge \{\{x_1, y\}, \{x_2, y\}, \dots, \{x_m, y\}, \{z_1, y\}, \{z_2, y\}, \dots, \{z_n, y\}\}$. Бинарная декомпозиция предиката P_y выполняется с помощью кванторов существования в соответствии с последней зависимостью.

Предикат P связан с исходным P равенством

$$P(X, Z) = \exists y P_y(X, Z, y). \quad (3)$$

Действительно, последовательно применяя к правой части (3) тождество (2) и равенство (1), имеем

$$\begin{aligned} \exists y P_y(X, Z, y) &= \bigvee_{b \in B} y / b(P_y(X, Z, y)) = \\ &= \bigvee_{i=1}^r (X, Z, b_i) = \\ &= \bigvee_{i=1}^r x_1^{a_{i1}} x_2^{a_{i2}} \dots x_m^{a_{im}} z_1^{c_{i1}} z_2^{c_{i2}} \dots z_n^{c_{in}} = P(X, Z). \end{aligned}$$

Таким образом, декартова декомпозиция $P(X, Z)$ выполняется путем построения вспомогательного предиката $P_y(X, Z, y)$ в соответствии с равенствами (1), (2) и последующей бинарной декомпозицией $P_y(X, Z, y)$ с помощью кванторов существования. Равенство (3) позволяет вернуться к исходному предикату $P(X, Z)$.

Недостатком такой декомпозиции является то, что переменная y принимает столько же значений в рамках предиката P_y и его проекций, сколько кортежей содержит область истинности предиката P . Будем считать, что это максимальное количество значений вспомогательной переменной, которое требуется для бинарной декомпозиции предиката, поскольку такого количества всегда достаточно. Кроме того, при декартовой декомпозиции не учитываются характерные особенности строения предиката для того, чтобы уменьшить число значений вспомогательной переменной.

Задача состоит в том, чтобы, используя функциональную зависимость (и другие зависимости при их наличии в предикате), выполнить бинарную декомпозицию либо непосредственно с помощью кванторов существования, либо с предварительным введением одной вспомогательной переменной с минимальным количеством значений.

Введем несколько обозначений. Через P_T обозначим область истинности предиката $P \in \text{Pre}(S)$, $P_T \subseteq S$. Предикат $P(X, Z)$ с зависимостью $X \rightarrow Z$ можно рассматривать как сюръективную функцию $p: A \rightarrow C$, где

$$A = P_T[X] \subseteq \text{Dom} X, C = P_T[Z] \subseteq \text{Dom} Z. \quad (4)$$

Пусть r – число кортежей отношения A , l – число кортежей отношения C . Т. к. для любого $\bar{c} \in C$ найдется $\bar{a} \in A$ такой, что $\bar{c} = p(\bar{a})$, то $p: A \rightarrow C$ – сюръекция. Для любого предиката $P(X, Z)$ с зависимостью $X \rightarrow Z$ справедливо неравенство $r \geq l$, поскольку в разных кортежах $(\bar{a}, \bar{c}) \in P_T$ значения c - могут повторяться, а значения \bar{a} – нет. Учитывая зависимость $X \rightarrow Z$, число кортежей в отношении P_T будет равно r , т. е. числу кортежей отношения A .

При декартовой декомпозиции предиката $P(X, Z)$ число значений вспомогательной переменной y в рамках полученных формул будет равно количеству всех кортежей отношения P_T , которое совпадает с количеством элементов области определения функции $p: A \rightarrow C$, т. е. $k=r$.

В силу принципа однозначности [3] все морфологические модели, построенные на его основе, являются функциями [2, 3]. При этом оказывается, что количество элементов области определения таких функций сильно превосходит количество элементов области значений.

Рассмотрим модель флексии как функцию $p: A \rightarrow C$. Область определения такой функции – это различные наборы признаков, характеризующие влияние контекста на флексию (окончание), область значений – множество всех флексий. Количество элементов множества A лежит в пределах от 10^3 до 10^4 , тогда как количество элементов множества C составляет от 10^1 до 10^2 .

Смысл предлагаемого ниже метода состоит в том, чтобы соотнести вспомогательную переменную, которая вводится с целью бинарной декомпозиции предиката $P(X,Z)$ не с областью определения функции $p:A \rightarrow C$ (как это происходит при декартовой декомпозиции), а с областью ее значений или со вспомогательной областью, которая строится на ее основе. Кроме того, предлагаемый метод учитывает функциональную зависимость в исходном предикате. Тем самым предполагается сократить количество значений вспомогательной переменной до минимума.

В предлагаемом методе будут повсеместно использоваться зависимости конъюнкции вида $X \rightarrow \gg x_1|x_2|\dots|x_m$, которые являются мощным инструментом декомпозиции предикатов. Непосредственными операторами декомпозиции в данном методе являются кванторы существования.

Перейдем к описанию метода бинарной декомпозиции предиката $P(X,Z)$ с функциональной зависимостью $X \rightarrow Z$, где $X = \{x_1, x_2, \dots, x_m\}$, $Z = \{z_1, z_2, \dots, z_n\}$. Прежде всего, рассмотрим тривиальную ситуацию – когда группа переменных состоит только из одной переменной, т. е. $m=1$. Тогда на основании следствия из теоремы Хеза функциональная зависимость $x \rightarrow \{z_1, z_2, \dots, z_n\}$ влечет зависимость конъюнкции $\wedge \{\{x, z_1\}, \{x, z_2\}, \dots, \{x, z_n\}\}$. Бинарная декомпозиция $P(x, Z)$ выполняется на основе последней зависимости с помощью кванторов существования. Вспомогательная переменная не требуется. Но это тривиальный случай, не характерный для задач реальной сложности, поэтому будем полагать, что $m \geq 2$.

С точки зрения строения исходного функционального предиката возможны три разные ситуации, которые определяют последовательность действий для достижения бинарной декомпозиции минимальной ценой. Рассмотрим каждый случай в порядке возрастания сложности.

При бинарной декомпозиции функционального предиката $P(X,Z)$ прежде всего надо проверить, являются ли аргументы $\{x_1, x_2, \dots, x_m\}$ описываемой функции $p: A \rightarrow C$ независимыми друг от друга. Взаимная независимость переменных x_1, x_2, \dots, x_m в предикате $P(X,Z)$ означает выполнение зависимости $Z \rightarrow \gg x_1|x_2|\dots|x_m$. Для строгой проверки предиката $P(X,Z)$ на наличие зависимости $Z \rightarrow \gg x_1|x_2|\dots|x_m$ необходимо выполнить дизъюнктивное разложение предиката $P(X,Z)$ по переменным Z . После этого с помощью утверждения этого разложения можно точно установить, выполняется зависимость $Z \rightarrow \gg x_1|x_2|\dots|x_m$ в предикате $P(X,Z)$ или нет. Если предикат $P(X,Z)$ удовлетворяет зависимости $Z \rightarrow \gg x_1|x_2|\dots|x_m$, то бинарная декомпозиция выполняется достаточно просто. При этом возможны две ситуации.

Первая ситуация имеет место, когда группа переменных Z состоит из одной переменной – z . Тогда вспомогательная переменная не требуется: бинарная декомпозиция $P(X,z)$ выполняется с помощью кванторов существования непосредственно на основе данной зависимости конъюнкции, которую можно представить в виде $\wedge \{\{x_1, z\}, \{x_2, z\}, \dots, \{x_m, z\}\}$. Это самый простой случай. Чуть более сложный слу-

чай – когда группа Z состоит из двух и более переменных; тогда имеет место вторая ситуация.

Во второй ситуации надо заменить группу переменных Z одной вспомогательной переменной y , после чего осуществить переход к двум вспомогательным предикатам – $P_1(X,y)$ и $P_2(y,Z)$, – которые являются результатом промежуточной декомпозиции предиката $P(X,Z)$. Бинарная декомпозиция предикатов $P_1(X,y)$ и $P_2(y,Z)$ осуществляется с помощью кванторов существования на основе зависимостей конъюнкции $\wedge \{\{y, x_1\}, \{y, x_2\}, \dots, \{y, x_m\}\}$ и $\wedge \{\{y, z_1\}, \{y, z_2\}, \dots, \{y, z_n\}\}$ соответственно, которые, как будет показано далее, выполняются в силу построения этих предикатов.

Если $P(X,Z)$ не удовлетворяет зависимости $Z \rightarrow \gg x_1|x_2|\dots|x_m$, то имеет место самая сложная ситуация из трех возможных. Здесь так же происходит переход к вспомогательным предикатам $P_1(X,y)$ и $P_2(y,Z)$ с последующей бинарной декомпозицией последних. Однако способ образования данных вспомогательных предикатов в этом случае гораздо сложнее.

В первой ситуации ясно как проводить бинарную декомпозицию функционального предиката $P(X,Z)$.

Ситуация, когда предикат $P(X,Z)$ удовлетворяет зависимости $Z \rightarrow \gg x_1|x_2|\dots|x_m$ и группа переменных Z включает более одной переменной, является довольно распространенной, поскольку аргументы функции, как правило, не зависят друг от друга. Однако если предикат $P(X,Z)$ выражает частичную, а не определенную всюду функцию из $\text{Dom}X$ в $\text{Dom}Z$, то указанная зависимость может не выполняться.

Выполним дизъюнктивное разложение предиката P по переменным Z :

$$P(X, Z) = \bigvee_{(c_1, c_2, \dots, c_n) \in C} z_1^{c_1} z_2^{c_2} \dots z_n^{c_n} P(x_1, x_2, \dots, x_m, c_1, c_2, \dots, c_n) = \bigvee_{i=1}^l z_1^{c_{i1}} z_2^{c_{i2}} \dots z_n^{c_{in}} P(x_1, x_2, \dots, x_m, c_1, c_2, \dots, c_n), \quad (5)$$

где множество C определяется выражением (4), а l – число кортежей отношения C . Тогда функцию $p:A \rightarrow C$, описываемую уравнением $P(X,Z)=1$, в явном виде можно выразить системой вида

$$\begin{aligned} P(x_1, x_2, \dots, x_m, c_{11}, c_{12}, \dots, c_{1n}) &= z_1^{c_{11}} z_2^{c_{12}} \dots z_n^{c_{1n}}, \\ P(x_1, x_2, \dots, x_m, c_{21}, c_{22}, \dots, c_{2n}) &= z_1^{c_{21}} z_2^{c_{22}} \dots z_n^{c_{2n}}, \\ &\dots \\ P(x_1, x_2, \dots, x_m, c_{l1}, c_{l2}, \dots, c_{ln}) &= z_1^{c_{l1}} z_2^{c_{l2}} \dots z_n^{c_{ln}}, \end{aligned} \quad (6)$$

где $(c_{i1}, c_{i2}, \dots, c_{in}) \in C$, $i=1, 2, \dots, l$.

Выполним декомпозицию этих уравнений за счет введения вспомогательной переменной y со значениями $B = \{b_1, b_2, \dots, b_l\}$, которые будем понимать как уникальные имена кортежей отношения C . Каждому кортежу $(c_{i1}, c_{i2}, \dots, c_{in}) \in C$ произвольным образом ставим во взаимно однозначное соответствие его уникальное имя $b \in B$ с помощью следующих уравнений алгебры предикатов:

$$\begin{aligned} z_1^{c_{11}} z_2^{c_{12}} \dots z_n^{c_{1n}} &= y^{b_1}, \\ z_1^{c_{21}} z_2^{c_{22}} \dots z_n^{c_{2n}} &= y^{b_2}, \\ &\dots \\ z_1^{c_{l1}} z_2^{c_{l2}} \dots z_n^{c_{ln}} &= y^{b_l}. \end{aligned} \quad (7)$$

На основании (7) образуем вспомогательный предикат P_2 :

$$P_2(y, Z) = \bigvee_{i=1}^l y^{b_i} z_1^{c_{i1}} z_2^{c_{i2}} \dots z_n^{c_{in}}. \quad (8)$$

В силу такого построения предикат $P_2(y, Z)$ содержит пару функциональных зависимостей $y \leftrightarrow \{z_1, z_2, \dots, z_n\}$.

Преобразуя систему (6) на основе уравнений (7), получим

$$\begin{aligned} P(X, c_{11}, c_{12}, \dots, c_{1n}) &= y^{b_1}, \\ P(X, c_{21}, c_{22}, \dots, c_{2n}) &= y^{b_2}, \\ &\dots \\ P(X, c_{l1}, c_{l2}, \dots, c_{ln}) &= y^{b_l}. \end{aligned} \quad (9)$$

С помощью выражения (9) образуем вспомогательный предикат P_1 :

$$P_1(X, y) = \bigvee_{i=1}^l y^{b_i} P(X, c_{i1}, c_{i2}, \dots, c_{in}). \quad (10)$$

Т. к. $P_1(X, y)$ был образован из $P(X, Z)$ с зависимостью $\{z_1, z_2, \dots, z_n\} \rightarrow x_1 | x_2 | \dots | x_m$ путем взаимно однозначной замены каждого значения переменной Z значением переменной y , то предикат $P_1(X, y)$ будет содержать зависимость $y \rightarrow x_1 | x_2 | \dots | x_m$.

Нетрудно показать, что предикат P выражается через P_1 и P_2 в виде

$$P(X, Z) = \exists (P_1(X, y) \wedge P_2(y, Z)). \quad (11)$$

Применив тождество (2) к правой части равенства (11) и пользуясь законами для операций подстановки, получим

$$\begin{aligned} &\exists y (P_1(X, y) \wedge P_2(y, Z)) = \\ &= \bigvee_{b \in B} y / b (P_1(X, y) \wedge P_2(y, Z)) = \\ &= \bigvee_{b \in B} (y / b (P_1(X, y)) \wedge y / b (P_2(y, Z))) = \\ &= \bigvee_{i=1}^l (P_1(X, b_i) \wedge P_2(b_i, Z)) = \\ &= \bigvee_{i=1}^l (P(X, c_{i1}, c_{i2}, \dots, c_{in}) \wedge z_1^{c_{i1}} z_2^{c_{i2}} \dots z_n^{c_{in}}). \end{aligned}$$

Последнее выражение представляет собой дизъюнктивное разложение предиката P в виде (5), из чего следует справедливость (11).

После построения предикатов P_1 и P_2 можно выполнить их бинарную декомпозицию. Бинарная декомпозиция предиката $P_1(X, y)$ выполняется с помощью кванторов существования на основе зависимости конъюнкции $\{x_1, y\}, \{x_2, y\}, \dots, \{x_m, y\}$, бинарная декомпозиция $P_2(y, Z)$ – на основе зависимости $\{y, z_1\}, \{y, z_2\}, \dots, \{y, z_n\}$, которая имеет место в силу функциональной зависимости

$y \rightarrow \{z_1, z_2, \dots, z_n\}$ и следствия из теоремы Хеза. Таким образом, бинарная декомпозиция предиката $P(X, Z)$ выполнена.

Определим границы, в пределах которых должно находиться количество значений вспомогательной переменной, введенной с целью бинарной декомпозиции функционального предиката.

Пусть k – число значений вспомогательной переменной y . Когда предикат $P(X, Z)$ с зависимостью $X \rightarrow Z$ так же удовлетворяет зависимости $Z \rightarrow x_1 | x_2 | \dots | x_m$, то при его бинарной декомпозиции описанным выше способом имеем $k=l$, где l – количество элементов множества $C = P_1[Z]$. Это минимальное количество значений вспомогательной переменной, так как оно равно количеству элементов области значений функции $p: A \rightarrow C$. Поэтому данная ситуация устанавливает нижний предел для параметра k .

Как показано выше, верхним пределом параметра k является r – количество элементов множества $A = P_1[X]$ (когда вспомогательная переменная y вводится на основе декартовой декомпозиции предиката $P(X, Z)$ и количество ее значений равно количеству элементов области определения функции $p: A \rightarrow C$. Следовательно, при выполнении бинарной декомпозиции функционального предиката с помощью одной вспомогательной переменной параметр k должен лежать в пределах $l \leq k \leq r$.

Последняя и самая сложная ситуация с точки зрения бинарной декомпозиции функционального предиката $P(X, Z)$ имеет место тогда, когда зависимость $Z \rightarrow x_1 | x_2 | \dots | x_m$ не выполняется. В данном случае так же требуется вспомогательная переменная. Она будет введена после того, как будет сформирована система специальных подмножеств множества A , определяемого выражением (4). Значения вспомогательной переменной в данном случае будут выступать в качестве имен этих подмножеств. При таком подходе, с одной стороны, во вспомогательных предикатах будут выстроены зависимости, которые позволяют выполнить их бинарную декомпозицию. С другой стороны, будет образовано минимально необходимое количество указанных подмножеств, что сведет количество значений вспомогательной переменной к минимуму.

Введем обозначения: $\bar{c}_i = (c_{i1}, c_{i2}, \dots, c_{in})$, $Z^{\bar{c}_i} = z_1^{c_{i1}} z_2^{c_{i2}} \dots z_n^{c_{in}}$, $i=1, 2, \dots, l$. Предикат $P(X, \bar{c}_i)$ описывает полный прообраз $\{X | (X, \bar{c}_i) = 1\}$ элемента $\bar{c}_i \in C$ при отображении $p: A \rightarrow C$. Обозначим множество $\{X | (X, \bar{c}_i) = 1\}$ через $s(\bar{c}_i)$. Множество всех полных прообразов $\{s(\bar{c}_1), s(\bar{c}_2), \dots, s(\bar{c}_l)\}$ является разбиением множества $A[M]$. Элементы разбиения называют слоями. Каждый элемент $\bar{c}_i = (c_{i1}, c_{i2}, \dots, c_{in}) \in C$ можно рассматривать как составное имя соответствующего ему слоя $s(\bar{c}_i)$, поскольку данное соответствие взаимно однозначно (это так называемое каноническое отображение [4]).

Чтобы выполнить бинарную декомпозицию предиката $P(X, Z)$, используя минимум значений вспомогательной переменной, предлагается следующий способ действий.

Для каждого слоя $s(\bar{c}_i) = \{X | P(X(\bar{c}_i) = 1)\}$, $i=1,2,\dots,l$, строим такое семейство его подмножеств $\{D_1^i, D_2^i, \dots, D_{k_i}^i\}$, которое удовлетворяет условию

$$s(\bar{c}_i) = D_1^i \cup D_2^i \cup \dots \cup D_{k_i}^i, \quad (12)$$

где

$$D_1^i = A_1^{i1} \times A_2^{i2} \times \dots \times A_m^{im},$$

$$D_2^i = A_1^{i2} \times A_2^{i2} \times \dots \times A_m^{im},$$

...

$$D_{k_i}^i = A_1^{ik_i} \times A_2^{ik_i} \times \dots \times A_m^{im}.$$

Каждое множество $A_1^q \subseteq \text{Dom}x_1, A_2^q \subseteq \text{Dom}x_2, \dots, A_m^q \subseteq \text{Dom}x_m$, $i=1,2,\dots,l$, $q=1,2,\dots,k_i$ выбирается произвольно в пределах своего домена. Множества $D_1^i, D_2^i, \dots, D_{k_i}^i$ назовем параллелепипедами слоя $s(\bar{c}_i)$ (по аналогии с термином из функционального анализа). В силу условия (12) семейство параллелепипедов $\{D_1^i, D_2^i, \dots, D_{k_i}^i\}$ является покрытием слоя $s(\bar{c}_i)$.

Таким образом, предлагается для каждого слоя $s(\bar{c}_i)$ подобрать минимальное количество m -мерных параллелепипедов пространства $\text{Dom} X$, которые совместно покрывают этот слой.

Для выполнения декомпозиции вводим вспомогательную переменную y со значениями b_1, b_2, \dots, b_k , которые будем понимать как имена всех параллелепипедов, построенных в рамках данной модели; k – общее число построенных параллелепипедов, $k=k_1+k_2+\dots+k_l$, $\{b_1, b_2, \dots, b_k\} = B$. Мы стремимся сделать множество значений вспомогательной переменной y минимальным, поэтому задача состоит в том, чтобы количество элементов k_i в каждом семействе $\{D_1^i, D_2^i, \dots, D_{k_i}^i\}$, удовлетворяющем условию (12), было минимальным.

Определим границы, в пределах которых будет колебаться количество значений k вспомогательной переменной y при указанном способе ее задания.

Пусть r_i – количество всех кортежей слоя $s(\bar{c}_i)$ разбиения множества A , $i=1,2,\dots,l$. Тогда $r_1+r_2+\dots+r_l=r$, r – число кортежей множества A .

Так как отдельно взятый кортеж $(a_1, a_2, \dots, a_m) \in A$ можно рассмотреть как параллелепипед $\{a_1\} \times \{a_2\} \times \dots \times \{a_m\} \subset A$, то для каждого слоя всегда можно подобрать покрытие, состоящее из таких одноэлементных параллелепипедов. Число таких параллелепипедов для слоя $s(\bar{c}_i)$ совпадает с числом всех кортежей этого слоя и равно r_i , $i=1,2,\dots,l$. Поэтому число r_i возьмем в качестве максимального значения для параметра k_i .

Минимально возможное значение параметра k_i равно 1, когда слой $s(\bar{c}_i)$ сам по себе является параллелепипедом пространства $\text{Dom} X$. Но этот случай возможен не для всех полных прообразов $s(\bar{c}_i)$, иначе, выполнялась бы зависимость $Z \rightarrow \gg x_1|x_2|\dots|x_m$. Следовательно, каждое число k_i , $i=1,2,\dots,l$ удовлетворяет неравенству $1 \leq k_i \leq r_i$. Так как $k_1+k_2+\dots+k_l=k$, то, выполнив суммирование по всем индексам $i=1,2,\dots,l$, получим предельные значения параметра k : $l \leq k \leq r$.

В силу построения покрытий каждое значение k_i будет минимальным, $i=1,2,\dots,l$, поэтому и сумма $k_1+k_2+\dots+k_l=k$ будет минимальной.

Тогда в ситуации, когда зависимость $Z \rightarrow \gg x_1|x_2|\dots|x_m$ в функциональном предикате $P(X,Z)$ не выполняется, количество значений вспомогательной переменной, введенной для бинарной декомпозиции, будет находиться в допустимых пределах от l до r и при этом будет минимальным.

Перепишем выражение (13) в терминах алгебры предикатов:

$$P(X, \bar{c}_i) = D_1^i(X) \vee D_2^i(X) \vee \dots \vee D_{k_i}^i(X), \quad (13)$$

где

$$D_1^i(x_1, x_2, \dots, x_m) = A_1^{i1}(x_1) \wedge A_2^{i2}(x_2) \wedge \dots \wedge A_m^{im}(x_m),$$

$$D_2^i(x_1, x_2, \dots, x_m) = A_1^{i2}(x_1) \wedge A_2^{i2}(x_2) \wedge \dots \wedge A_m^{im}(x_m),$$

...

$$D_{k_i}^i(x_1, x_2, \dots, x_m) = A_1^{ik_i}(x_1) \wedge A_2^{ik_i}(x_2) \wedge \dots \wedge A_m^{im}(x_m),$$

$i=1,2,\dots,l$.

После того как для каждого слоя $P(X, \bar{c}_i)$, $i=1,2,\dots,l$, построено минимальное покрытие (13) можно перейти к построению вспомогательных предикатов.

Каждому параллелепипеду произвольным образом взаимно однозначно сопоставим уникальное имя $b \in B$ (уникальное для всей модели, а не только в пределах данного слоя). Последовательно рассматривая каждый слой $P(X, \bar{c}_i)$, $i=1,2,\dots,l$, положим, что

$$D_1^i(X) = y^{b_{i1}}, D_2^i(X) = y^{b_{i2}}, \dots, D_{k_i}^i(X) = y^{b_{ik_i}}. \quad (14)$$

На основании (14) образуем вспомогательный предикат P_1 :

$$P_1(X, y) = \bigvee_{i=1}^l (D_1^i(X) y^{b_{i1}} \vee D_2^i(X) y^{b_{i2}} \vee \dots \vee D_{k_i}^i(X) y^{b_{ik_i}}). \quad (15)$$

Так как $k_1+k_2+\dots+k_l=k$, то P_1 можно представить в более удобном виде:

$$P_1(X, y) = \bigvee_{j=1}^k D_j(X) y^{b_j}.$$

В соответствии с утверждением $X \rightarrow \gg Y$ и $Y \cup Z \rightarrow \gg W$, $X \cup Z \rightarrow \gg W \wedge (Y \cup Z)$, предикат $P_1(X, y)$ удовлетворяет зависимости $y \rightarrow \gg x_1|x_2|\dots|x_m$, т. к. для каждого $b_j \in B$

$$P_1(X, b_j) = D_j(X) = A_1^j(x_1) \wedge A_2^j(x_2) \wedge \dots \wedge A_m^j(x_m).$$

Благодаря тому, что покрытие каждого слоя состоит только из параллелепипедов, вспомогательный предикат $P_1(X, y)$ удовлетворяет зависимости $y \rightarrow \gg x_1|x_2|\dots|x_m$ или, что то же самое, $\hat{\wedge} \{x_1, y, \{x_2, y\}, \dots, \{x_m, y\}\}$. На основе этой зависимости можно выполнить бинарную декомпозицию P_1 .

Заметим, что система уравнений (14), $i=1,2,\dots,l$, равносильна одному уравнению $P_1(X, y) = 1$.

Строим второй вспомогательный предикат.

Как говорилось выше, функцию $p: A \rightarrow C$, описываемую уравнением $P(X, Z) = 1$, в явном виде

можно выразить системой равенств (6). С учетом (13) и (14) каждое уравнение из системы (6) можно представить в виде:

$$y^{b_{i1}} \vee y^{b_{i2}} \vee \dots \vee y^{b_{ik_i}} = Z^{\bar{c}_i}. \quad (16)$$

Уравнения (16), $i=1,2,\dots,l$, связывают имена параллелепипедов с именем слоя, покрытие которого они образуют.

На основании (16) образуем вспомогательный предикат P_2 :

$$P_2(y, Z) = \bigvee_{i=1}^l (y^{b_{i1}} y^{b_{i2}} \vee \dots \vee y^{b_{ik_i}}) Z^{\bar{c}_i}. \quad (17)$$

Система уравнений (16), $i=1,2,\dots,l$, равносильна одному уравнению $P_2(y, Z)=1$.

Заметим, что предикат $P_1(X, y)$ описывает всюду определенное многозначное отображение из множества A во множество B , т. к. параллелепипеды, покрывающие один и тот же слой, могут пересекаться, а значит, кортеж $\bar{a} \in A$ может входить в разные параллелепипеды. Но параллелепипеды, покрывающие разные слои разбиения пересекаться не могут, поэтому каждый параллелепипед является подмножеством только одного слоя. Следовательно, предикат $P_2(y, Z)$ описывает функцию из множества B во множество C , т. е. содержит функциональную зависимость $y \rightarrow \{z_1, z_2, \dots, z_m\}$. Тогда в силу следствия из теоремы Хеза предикат $P_2(y, Z)$ удовлетворяет зависимости конъюнкции $\bigwedge \{y, z_1, \{y, z_2\}, \dots, \{y, z_m\}\}$, которая позволяет выполнить его бинарную декомпозицию.

Предикаты P_1 и P_2 образованы так, что представляют собой результат декомпозиции исходного предиката P . Покажем, что P выражается через P_1 и P_2 в виде (11).

Действительно, пользуясь тождеством

$$\exists x_i(P) = \bigvee_{a \in \text{Dom} x_i} x_i / a(P)$$

и законами для операций подстановки, имеем

$$\begin{aligned} \exists y(P_1(X, y) \wedge P_2(y, Z)) &= \\ &= \bigvee_{b \in B} y / b(P_1(X, y) \wedge P_2(y, Z)) = \\ &= \bigvee_{b \in B} (y / b(P_1(X, y)) \wedge y / b(P_2(y, Z))) = \\ &= \bigvee_{i=1}^k (P_1(X, b_j) \wedge (P_2(b_j, Z))) = \\ &= \bigvee_{i=1}^l (P_1(X, b_{i1}) \wedge P_2(b_{i1}, Z) \vee P_1(X, b_{i2}) \wedge P_2(b_{i2}, Z) \vee \dots \vee P_1(X, b_{ik_i}) \wedge P_2(b_{ik_i}, Z)) = \\ &= \bigvee_{i=1}^l (D_{i1}(X) \wedge Z^{\bar{c}_i} \vee D_{i2}(X) \wedge Z^{\bar{c}_i} \vee \dots \vee \\ &\quad \vee D_{ik_i}(X) \wedge Z^{\bar{c}_i} = \\ &= \bigvee_{i=1}^l (D_{i1}(X) \vee D_{i2}(X) \vee \dots \vee D_{ik_i}(X) \wedge Z^{\bar{c}_i}). \end{aligned}$$

Далее, используя (13) и (5), получим

$$\exists y(P_1(X, y) \wedge P_2(y, Z)) = \bigvee_{i=1}^l (P(X, \bar{c}_i) Z^{\bar{c}_i} = P(X, Z),$$

что и требовалось.

После того, как вспомогательные предикаты построены, остается выполнить их бинарную декомпозицию. Бинарная декомпозиция предиката $P_1(X, y)$ выполняется с помощью кванторов существования на основе зависимости конъюнкции $\bigwedge \{x_1, y, \{x_2, y\}, \dots, \{x_m, y\}\}$, $P_2(y, Z)$ – на основе зависимости $\bigwedge \{y, z_1, \{y, z_2\}, \dots, \{y, z_m\}\}$. Тем самым бинарная декомпозиция исходного предиката $P(X, Z)$ выполнена; при этом k – число значений переменной y в полученных предикатах – является минимальным и лежит в пределах $l \leq k \leq r$.

На основе описанного метода разработан алгоритм выполнения бинарной декомпозиции функциональных предикатов.

Пусть задан конечный предикат $P(X, Z)$ с функциональной зависимостью $X \rightarrow Z$. Для выполнения его бинарной декомпозиции с помощью одной вспомогательной переменной, принимающей минимально возможное количество значений, надо выполнить следующие шаги.

1. Выполнить дизъюнктивное разложение предиката $P(X, Z)$ по переменным Z (5). С помощью утверждения: $X \rightarrow Y$ и $Y \cup Z \rightarrow W$, $X \cup Z \rightarrow W \wedge (Y \cup Z)$, по виду разложения проверить, выполняется зависимость $Z \rightarrow x_1 | x_2 | \dots | x_m$ в предикате $P(X, Z)$ или нет.

Если указанная зависимость выполняется, то дальнейшие шаги зависят от количества переменных $Z = \{z_1, z_2, \dots, z_m\}$: если $n=1$, то перейти к пункту 2; если $n>1$, то перейти к п. 3. Если зависимость $Z \rightarrow x_1 | x_2 | \dots | x_m$ не выполняется, то перейти к п. 5.

2. В данном случае предикат $P(X, z)$ удовлетворяет зависимости конъюнкции $\bigwedge \{z, x_1, \{z, x_2\}, \dots, \{z, x_m\}\}$. С помощью кванторов существования на основе данной зависимости выполнить бинарную декомпозицию $P(X, z)$. Бинарная декомпозиция предиката $P(X, Z)$ выполнена.

3. По формулам (7) ввести вспомогательную переменную.

4. По формулам (8) образовать вспомогательный предикат $P_2(y, Z)$. По формулам (9), (10) образовать вспомогательный предикат $P_1(X, y)$. Перейти к пункту 7.

5. Каждый предикат $P(X, \bar{c}_i)$, $i=1,2,\dots,l$ представить в виде выражения (13) так, чтобы число слагаемых было минимальным. Ввести вспомогательную переменную по формулам (14).

6. По формулам (15) образовать вспомогательный предикат $P_1(X, y)$. По формулам (16), (17) образовать вспомогательный предикат $P_2(y, Z)$.

7. Выполнить бинарную декомпозицию вспомогательных предикатов на основе содержащихся в них зависимостей конъюнкции с помощью кванторов существования: декомпозицию предиката $P_1(X, y)$ выполнить на основе зависимости $\bigwedge \{x_1, y, \{x_2, y\}, \dots, \{x_m, y\}\}$ декомпозицию предиката $P_2(y, Z)$ выполнить на основе зависимости $\bigwedge \{y, z_1, \{y, z_2\}, \dots, \{y, z_m\}\}$. Так как равенство

(11) выражает неэквивалентную декомпозицию предиката $P(X, Z)$ на предикаты $P_1(X, y)$ и $P_2(y, Z)$, то бинарная декомпозиция предиката $P(X, Z)$, выполнена.

Отметим некоторые достоинства и недостатки предложенного метода для декомпозиции функциональных предикатов.

Приведенный выше алгоритм является скорее инструкцией, чем алгоритмом в строгом смысле, поскольку пункт 5 является далеко не тривиальной задачей. Возможность полной алгоритмизации этой инструкции зависит от того, является ли задача, которая определяется пятым пунктом, алгоритмически разрешимой.

Достоинством данного метода является возможность обрабатывать формулу или отношение не целиком, а по частям: для каждого полного образа его минимальное покрытие параллелепипедами подбирается отдельно.

СПИСОК ЛИТЕРАТУРЫ

1. Павловский Ю.Н. Геометрическая теория декомпозиции и некоторые ее приложения. – М.: Вычислительный центр им. А.А. Дородницына РАН, 2011. – 93 с.
2. Бондаренко М.Ф. Модели языка // Бионика интеллекта. – 2004. – № 1. – С. 27–37.

В дальнейшем, на основе предложенного метода предполагается построить модель логической сети для морфологической модели путем бинарной декомпозиции последней. Полученные при этом результаты исследований будут опубликованы дополнительно.

Выводы

1. Проведен анализ единственного обоснованного метода бинарной декомпозиции конечных предикатов – метода декартовой декомпозиции.
2. Впервые разработан и обоснован частный метод бинарной декомпозиции для функциональных предикатов, который отличается от общего метода декартовой декомпозиции тем, что он основан на анализе зависимостей и количество значений вспомогательной переменной сведено к минимуму.
3. На основе предложенного метода разработан алгоритм бинарной декомпозиции функциональных предикатов.
3. Шабанов-Кушнаренко Ю.П. Теория интеллекта. Проблемы и перспективы. – Харьков: Высшая школа, 1987. – 159 с.
4. Мальцев А.И. Алгебраические системы. – М.: Наука, 1970. – 392 с.

Поступила 10.01.2012 г.

УДК 519.872

АСИМПТОТИЧЕСКИЙ АНАЛИЗ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С НЕОГРАНИЧЕННЫМ ЧИСЛОМ ПРИБОРОВ И ПОЛУМАРКОВСКИМ ВХОДЯЩИМ ПОТОКОМ

А.А. Назаров, И.А. Семенова

Томский государственный университет

E-mail: inna_ac@mail.ru

Рассматривается система массового обслуживания с входящим SM-поток и неограниченным числом обслуживающих приборов. Исследование проводится методом просеянного потока и методом асимптотического анализа в условии растущего времени обслуживания. Проведенный анализ позволяет получить асимптотические семиинварианты до третьего порядка.

Ключевые слова:

Метод просеянного потока, асимптотический анализ, входящий SM-поток, система Колмогорова.

Key words:

Method of the sifted flow, asymptotic analysis, input SM flow, Kolmogorov's system.

Математические модели систем массового обслуживания широко применяют при решении важных практических задач, возникающих в связи с бурным развитием систем коммуникаций, возникновением информационно-вычислительных систем, появлением и усложнением разнообразных технологических систем, созданием автоматизированных систем управления, для задач экономико-математического моделирования.

В качестве математических моделей страховых компаний, кредитно-депозитных организаций и многих других экономических и социально-экономических систем предлагается рассматривать системы с неограниченным числом приборов. Например, количество возможных договоров между клиентами и кредитно-депозитной организацией практически не ограничено. Сроки, на которые заключаются договоры, имеют весьма широкий спектр продолжительностей, поэтому достаточно адекватно могут моделироваться некоторой случайной величиной с заданной функцией распределения их значений. Поток клиентов, обращающихся в кредитно-депозитную организацию, может быть как пуассоновским, так и коррелированным.

Наиболее общим ординарным потоком однородных событий является полумарковский или SM-поток (Semi-Markovian process). Идея введения такого потока была выдвинута П. Леви (1954) и В. Смитом (1955). Анализ распределения числа событий, наступивших в SM-потоке за некоторое время, можно найти в работах [1, 2]. Системы массового обслуживания с таким входящим потоком интенсивно изучаются в настоящее время [3, 4].

В данной работе проводится исследование системы массового обслуживания с неограниченным числом обслуживающих приборов, на вход которой поступает SM-поток, функция распределения времени обслуживания произвольная. Исследование проводится методом просеянного потока и методом асимптотического анализа в условии растущего времени обслуживания.

Математическая модель

Рассмотрим систему массового обслуживания с неограниченным числом приборов, на вход которой поступает SM-поток событий, заданный полумарковской матрицей $A(x)$. Продолжительности обслуживания заявок стохастически независимы, одинаково распределены и имеют произвольную (но не экспоненциальную) функцию распределения $B(x)$. Поступающая заявка занимает любой из свободных приборов. Завершив обслуживание, заявка покидает систему (рис. 1).

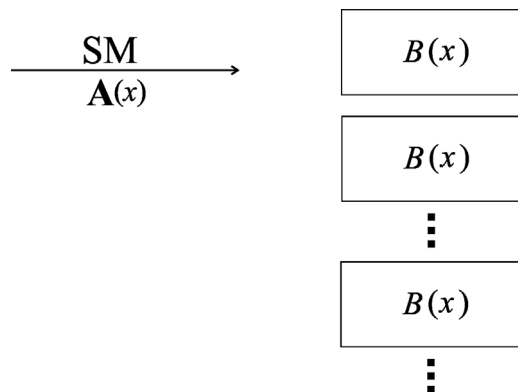


Рис. 1. Математическая модель системы массового обслуживания с неограниченным числом обслуживающих приборов

Рассматривается трёхмерный случайный процесс $\{s(t), z(t), i(t)\}$, который является марковским с непрерывным временем, где $z(t)$ — длина интервала от момента времени t до момента наступления очередного события в SM-потоке, а дискретный процесс $\sigma(t)$ определяется как $s(t) = \xi(n+1)$, если $t_n < t \leq t_{n+1}$, где моменты восстановления t_n определяются равенством

$$t_n = \sum_{i=1}^n \tau(i),$$

то есть процесс $s(t)$ на интервале $t_n < t \leq t_{n+1}$ принимает и сохраняет значение $\xi(n+1)$, здесь $\xi(n)$ – эргодическая цепь Маркова с дискретным временем и матрицей $\mathbf{P}=[p_{jk}]$ вероятностей перехода за один шаг, процесс $\tau(t)$ принимает неотрицательные значения из непрерывного множества и определяет длины интервалов в SM-потоке. Обозначим $i(t)$ – число занятых приборов в момент времени t . Матрицу $\mathbf{A}(x)$ с элементами $A_{jk}(x)$ будем называть полумарковской. Матрица переходных вероятностей цепи Маркова $\xi(n)$ определяется равенством

$$\mathbf{P} = \mathbf{A}(\infty).$$

Для элементов полумарковской матрицы имеет место равенство

$$A_{jk}(x) = G_{jk}(x)P_{jk},$$

где $G_{jk}(x)$ – условная функция распределения длины интервала полумарковского потока при условии, что в начале этого интервала вложенная цепь Маркова приняла значение v , а в конце его примет значение k . Матрицу $\mathbf{A}(x)$ можно записать в виде произведения Адамара двух матриц $\mathbf{G}(x)$ и \mathbf{P}

$$\mathbf{A}(x) = \mathbf{G}(x) * \mathbf{P},$$

и можно полагать, что полумарковский поток задан двумя матрицами $\mathbf{G}(x)$ и \mathbf{P} .

Для исследования такой системы воспользуемся методом просеянного потока.

Метод просеянного потока

Пусть на вход системы с неограниченным числом приборов поступает SM-поток заявок. Для реализации метода просеянного [5] потока рассмотрим две оси времени (рис. 2), на первой из которых обозначим моменты наступления событий входящего потока (верхняя ось, рис. 2).

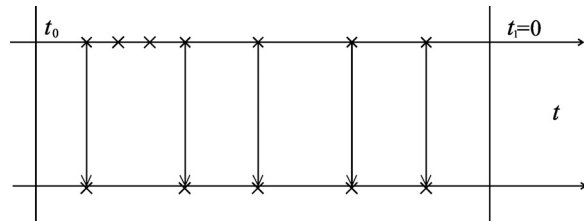


Рис. 2. Схематическая модель применения метода просеянного потока

Выделим некоторый момент времени t_1 . Не нарушая общности, можно считать, что $t_1=0$. Обозначим вероятность

$$S(t) = 1 - B(t_1 - t), \quad (1)$$

которая имеет смысл вероятности того, что заявка входящего потока, поступившая в систему в момент времени $t < t_1$, в момент времени t_1 будет находиться в системе, занимая для своего обслуживания один из приборов системы.

Просеянный поток на второй оси времени, рис. 2, будем формировать следующим образом. Выделим некоторый момент времени $t_0 < t_1$. Каждая заявка входящего потока, поступившая в систему в момент времени $t_0 < t < t_1$, с вероятностью $S(t)$ формиру-

ет событие просеянного потока в тот же момент времени t .

Очевидно, что заявки, не попавшие в просеянный поток, завершат обслуживание и покинут систему до момента t_1 , в то время как все заявки просеянного потока в момент t_1 будут находиться в системе, занимая её приборы.

Для просеянного потока обозначим $n(t)$ – число событий этого потока, наступивших до момента времени t , то есть на интервале $[t_0, t)$. В силу способа формирования просеянного потока, полагая, что в момент времени t_0 система обслуживания свободна, т. е. в ней нет обслуживаемых заявок $i(t_0)=0$, имеет место равенство числа $n(t_1)$ событий, наступивших в просеянном потоке к моменту времени t_1 , и числа $i(t_1)$ приборов, занятых в рассматриваемой системе, в момент времени t_1 , т. е.

$$i(t_1) = n(t_1). \quad (2)$$

Равенство (2) является основным для дальнейших исследований, т. к. проблему исследования наиболее сложной системы обслуживания с неограниченным числом приборов сводит к задаче анализа просеянного нестационарного потока, определяемого процессом $n(t)$. Найдя характеристики этого случайного процесса в произвольный момент времени t , где $t_0 \leq t \leq t_1$, положим $t=t_1$, тогда, в силу равенства (2), его характеристики совпадают с характеристиками величины $i(t_1)$.

Метод просеянного потока исследования системы $SM|GI|\infty$

Для исследования рассматриваемой системы $SM|GI|\infty$ применим метод просеянного потока. Для распределения вероятностей $P(s, z, n, t)$ трёхмерного марковского процесса $\{s(t), z(t), n(t)\}$,

$$P(s, z, n, t) = P\{s(t) = s, z(t) < z, n(t) = n\}$$

запишем систему дифференциальных уравнений Колмогорова

$$\frac{\partial P(s, z, n, t)}{\partial t} = \frac{\partial P(s, z, n, t)}{\partial z} - \frac{\partial P(s, 0, n, t)}{\partial z} + \sum_v \left\{ \frac{\partial P(v, 0, n-1, t)}{\partial z} S(t) + \frac{\partial P(v, 0, n, t)}{\partial z} (1 - S(t)) \right\} A_{vs}(z). \quad (3)$$

Начальные условия для решения этой системы в момент времени t_0 определим равенством

$$P(s, z, n, t_0) = \begin{cases} R(s, z), & \text{если } n = 0, \\ 0, & \text{если } n > 0. \end{cases}$$

где компоненты $R(s, z)$ вектора $\mathbf{R}(z)$ по определению равны

$$R(s, z) = P\{s(t) = s, z(t) < z\}.$$

Обозначив

$$H(s, z, u, t) = \sum_{n=0}^{\infty} e^{jun} P(s, z, n, t),$$

из (3) получим следующую задачу

$$\frac{\partial H(s, z, u, t)}{\partial t} = \frac{\partial H(s, z, u, t)}{\partial z} - \frac{\partial H(s, 0, u, t)}{\partial z} + \{1 + S(t)(e^{ju} - 1)\} \sum_v \frac{\partial H(v, 0, u, t)}{\partial z} A_{v,s}(z),$$

$$H(s, z, u, t_0) = R(s, z).$$

Эту систему запишем в виде матричного уравнения, обозначив вектор-строку

$$\mathbf{H}(z, u, t) = \{H(1, z, u, t), H(2, z, u, t), \dots\},$$

получим

$$\frac{\partial \mathbf{H}(z, u, t)}{\partial t} = \frac{\partial \mathbf{H}(z, u, t)}{\partial z} + \frac{\partial \mathbf{H}(0, u, t)}{\partial z} \times \{\mathbf{A}(z) - \mathbf{I} + (e^{ju} - 1)S(t)\mathbf{A}(z)\},$$

$$\mathbf{H}(z, u, t_0) = \mathbf{R}(z),$$

где распределение $\mathbf{R}(z)$ имеет вид [6]

$$\mathbf{R}(z) = \kappa_1 \mathbf{r} \int_0^z (\mathbf{P} - \mathbf{A}(x)) dx,$$

здесь $\mathbf{A}(x)$ – полумарковская матрица; \mathbf{P} – стохастическая матрица вероятностей переходов вложенной цепи Маркова; \mathbf{r} – стационарное распределение вероятностей значений вложенной цепи Маркова, а величина κ_1 определяется равенством

$$\kappa_1 = \frac{1}{\mathbf{r}\mathbf{A}\mathbf{E}},$$

где матрица \mathbf{A} имеет вид

$$\mathbf{A} = \int_0^\infty (\mathbf{P} - \mathbf{A}(x)) dx.$$

Уравнение (4), определяющее характеристики системы $SM|GI|_\infty$, будем решать в асимптотическом условии растущего времени, полагая, что среднее значение времени обслуживания $b \rightarrow \infty$.

Метод асимптотических семиинвариантов исследования системы $SM|GI|_\infty$

Методом асимптотического анализа в теории массового обслуживания (в теории потоков) называется исследование уравнений, определяющих какие-либо характеристики системы (потока) при выполнении некоторого асимптотического (предельного) условия, вид которого конкретизируется для различных моделей и поставленных задач исследования [4].

В данной работе предлагается модификация метода асимптотического анализа – метод асимптотических семиинвариантов. Этот метод реализуется в нахождении последовательности асимптотик возрастающего порядка, в которой асимптотика первого порядка, аналогично закону больших чисел, определяет асимптотическое среднее значение числа занятых приборов. Асимптотика второго порядка, аналогично центральной предельной теореме, позволяет построить гауссовскую аппроксимацию распределения вероятностей состояний системы. Асимптотики более высокого порядка опреде-

ляют соответствующие семиинварианты и аппроксимации распределения вероятностей, выполняющие более детальное исследование рассматриваемой характеристики [7].

Асимптотика первого порядка

Обозначим $\varepsilon=1/b$ и в уравнении (4) выполним замены

$$t\varepsilon = \tau, \quad t_0 = \tau_0, \quad S(t) = S_1(\tau), \quad u = \varepsilon w,$$

$$\mathbf{H}(z, u, t) = \mathbf{F}_1(z, w, \tau, \varepsilon),$$

получим задачу

$$\varepsilon \frac{\partial \mathbf{F}_1(z, w, \tau, \varepsilon)}{\partial \tau} = \frac{\partial \mathbf{F}_1(z, w, \tau, \varepsilon)}{\partial z} + \frac{\partial \mathbf{F}_1(0, w, \tau, \varepsilon)}{\partial z} \times \{\mathbf{A}(z) - \mathbf{I} + (e^{j\varepsilon w} - 1)S_1(\tau)\mathbf{A}(z)\},$$

$$\mathbf{F}_1(z, w, \tau_0, \varepsilon) = \mathbf{R}(z).$$

Теорема 1. *Предельное, при $\varepsilon \rightarrow 0$, значение $\mathbf{F}_1(z, w, \tau)$ решения $\mathbf{F}_1(z, w, \tau, \varepsilon)$ уравнения (6) имеет вид*

$$\mathbf{F}_1(z, w, \tau) = \mathbf{R}(z) \exp \left\{ jw\kappa_1 \int_{\tau_0}^{\tau} S_1(z) dz \right\},$$

где вектор-функция $\mathbf{R}(z)$ и параметр κ_1 определены выше.

Доказательство. В задаче (6) выполним предельный переход при $\varepsilon \rightarrow 0$

$$\frac{\partial \mathbf{F}_1(z, w, \tau)}{\partial z} + \frac{\partial \mathbf{F}_1(0, w, \tau)}{\partial z} \{\mathbf{A}(z) - \mathbf{I}\} = 0,$$

отсюда получим, что $\mathbf{F}_1(z, w, \tau)$ является решением однородного уравнения, поэтому имеет вид

$$\mathbf{F}_1(z, w, \tau) = \mathbf{R}(z)\Phi_1(w, \tau),$$

здесь вектор-функция $\mathbf{R}(z)$ определена выше, а скалярную функцию $\Phi_1(w, \tau)$ определим следующим образом. В задаче (6) выполним предельный переход при $z \rightarrow 0$ и покомпонентное суммирование векторов, получим

$$\varepsilon \frac{\partial \mathbf{F}_1(\infty, w, \tau, \varepsilon)}{\partial \tau} \mathbf{E} = \frac{\partial \mathbf{F}_1(0, w, \tau, \varepsilon)}{\partial z} \mathbf{E} (e^{j\varepsilon w} - 1) S_1(\tau),$$

$$\mathbf{F}_1(\infty, w, \tau_0, \varepsilon) \mathbf{E} = 1.$$

Выполнив здесь предельный переход при $\varepsilon \rightarrow 0$ и замену (7), получим равенства

$$\frac{\partial \Phi_1(w, \tau)}{\partial \tau} \mathbf{R}(\infty) \mathbf{E} = \Phi_1(w, \tau) \frac{\partial \mathbf{R}(0)}{\partial z} \mathbf{E} jw S_1(\tau),$$

$$\Phi_1(w, \tau_0) \mathbf{R}(\infty) \mathbf{E} = 1,$$

которые перепишем в следующем виде

$$\frac{\partial \Phi_1(w, \tau)}{\partial \tau} = \Phi_1(w, \tau) \kappa_1 jw S_1(\tau),$$

$$\Phi_1(w, \tau_0) = 1.$$

Решение задачи (8) очевидно имеет вид

$$\Phi_1(w, \tau) = \exp \left\{ jw\kappa_1 \int_{\tau_0}^{\tau} S_1(z) dz \right\}.$$

Теорема доказана.

В силу замен (5), а также равенства (7), можно записать приближённое (асимптотическое) равенство

$$\begin{aligned} \mathbf{H}(z, u, t) &= \mathbf{F}_1(z, w, \tau, \varepsilon) \approx \mathbf{F}_1(z, w, \tau) = \\ &= \mathbf{R}(z) \exp \left\{ j u \kappa_1 \int_{t_0}^t S(z) dz \right\}, \end{aligned}$$

поэтому для характеристической функции величины $n(t)$ запишем

$$Me^{ju n(t)} = \mathbf{H}(\infty, u, t) \mathbf{E} \approx \exp \left\{ j u \kappa_1 \int_{t_0}^t S(z) dz \right\}.$$

При $t=t_0=0$ для характеристической функции процесса $i(t)$ в стационарном режиме получим

$$\begin{aligned} h_1(u) &= Me^{ju i(t)} = Me^{ju i(0)} \approx \\ &\approx \exp \left\{ j u \kappa_1 \int_{-\infty}^0 S(z) dz \right\} = \exp \{ j u \kappa_1 \beta_1 \}. \end{aligned}$$

Определение. Функцию

$$h_1(u) = \exp \{ j u \kappa_1 \beta_1 \}$$

будем называть асимптотикой первого порядка характеристической функции числа занятых приборов в системе, а величину $\kappa_1 \beta_1$ – асимптотическим семиинвариантом первого порядка.

Асимптотика второго порядка

Для более детального исследования рассматриваемой системы обслуживания, получим асимптотику второго порядка.

В уравнении (4) выполним замену

$$\mathbf{H}(z, u, t) = \mathbf{H}_2(z, u, t) \exp \left\{ j u \kappa_1 \int_{t_0}^t S(z) dz \right\}, \quad (9)$$

получим

$$\begin{aligned} \frac{\partial \mathbf{H}_2(z, u, t)}{\partial t} + \mathbf{H}_2(z, u, t) j u \kappa_1 S(t) &= \\ = \frac{\partial \mathbf{H}_2(z, u, t)}{\partial z} + \frac{\partial \mathbf{H}_2(0, u, t)}{\partial z} \times \\ \times \{ \mathbf{A}(z) - \mathbf{I} + (e^{ju} - 1) S(t) \mathbf{A}(z) \}, \\ \mathbf{H}_2(z, u, t_0) &= \mathbf{R}(z). \end{aligned} \quad (10)$$

Обозначив $\varepsilon^2 = 1/b$, в задаче (10) выполним замены

$$\begin{aligned} \varepsilon^2 t = \tau, \quad \varepsilon^2 t_0 = \tau_0, \quad S(t) = S_1(\tau), \quad u = \varepsilon w, \\ \mathbf{H}_2(z, u, t) = \mathbf{F}_2(z, w, \tau, \varepsilon), \end{aligned} \quad (11)$$

получим

$$\begin{aligned} \varepsilon^2 \frac{\partial \mathbf{F}_2(z, w, \tau, \varepsilon)}{\partial \tau} + \mathbf{F}_2(z, w, \tau, \varepsilon) j \varepsilon w \kappa_1 S_1(\tau) &= \\ = \frac{\partial \mathbf{F}_2(z, w, \tau, \varepsilon)}{\partial z} + \frac{\partial \mathbf{F}_2(0, w, \tau, \varepsilon)}{\partial z} \times \\ \times \{ \mathbf{A}(z) - \mathbf{I} + (e^{j \varepsilon w} - 1) S_1(\tau) \mathbf{A}(z) \}, \\ \mathbf{F}_2(z, w, \tau_0, \varepsilon) &= \mathbf{R}(z). \end{aligned} \quad (12)$$

Теорема 2. Предельное, при $\varepsilon \rightarrow 0$, значение $\mathbf{F}_1(z, w, \tau)$ решения $\mathbf{F}_1(z, w, \tau, \varepsilon)$ уравнения (12) имеет вид

$$\begin{aligned} \mathbf{F}_2(z, w, \tau) &= \\ = \mathbf{R}(z) \exp \left\{ \frac{(jw)^2}{2} \left[\kappa_1 \int_{\tau_0}^{\tau} S_1(z) dz + 2\kappa_2 \int_{\tau_0}^{\tau} S_1^2(z) dz \right] \right\}, \end{aligned}$$

где величина κ_2 определяется равенством

$$\kappa_2 = \frac{\partial f_2(0)}{\partial z} \mathbf{E},$$

а вектор-функция $f_2(z)$ удовлетворяет условию $f_2(\infty) \mathbf{E} = 0$ и является решением уравнения

$$\frac{\partial f_2(z)}{\partial z} + \frac{\partial f_2(0)}{\partial z} (\mathbf{A}(z) - \mathbf{I}) + \frac{\partial \mathbf{R}(0)}{\partial z} \mathbf{A}(z) - \kappa_1 \mathbf{R}(z) = 0.$$

Доказательство аналогично доказательству теоремы 1.

Определение. Функцию

$$h_2(u) = \exp \left\{ j u \kappa_1 \beta_1 + \frac{(ju)^2}{2} [\kappa_1 \beta_1 + 2\kappa_2 \beta_2] \right\}, \quad (13)$$

будем называть асимптотикой второго порядка характеристической функции числа занятых приборов в системе, а величину $[\kappa_1 \beta_1 + 2\kappa_2 \beta_2]$ – асимптотическим семиинвариантом второго порядка,

где $\beta_2 = \int_0^{\infty} (1 - B(z))^2 dz$.

Асимптотика третьего порядка

Для нахождения асимптотики третьего порядка формулируем следующее утверждение:

Теорема 3. Предельное, при $\varepsilon \rightarrow 0$, значение $\mathbf{F}_1(z, w, \tau)$ решения $\mathbf{F}_1(z, w, \tau, \varepsilon)$ уравнения

$$\begin{aligned} \varepsilon^3 \frac{\partial \mathbf{F}_3(z, w, \tau, \varepsilon)}{\partial \tau} + \\ + \mathbf{F}_3(z, w, \tau, \varepsilon) \left\{ \left[j \varepsilon w + \frac{(j \varepsilon w)^2}{2} \right] \kappa_1 S_1(\tau) + \right. \\ \left. + (j \varepsilon w)^2 \kappa_2 S_1^2(\tau) \right\} = \\ = \frac{\partial \mathbf{F}_3(z, w, \tau, \varepsilon)}{\partial z} + \frac{\partial \mathbf{F}_3(0, w, \tau, \varepsilon)}{\partial z} \times \\ \times \{ \mathbf{A}(z) - \mathbf{I} + (e^{j \varepsilon w} - 1) S_1(\tau) \mathbf{A}(z) \}, \end{aligned}$$

имеет вид

$$\mathbf{F}_3(z, w, \tau) = \mathbf{R}(z) \exp \left\{ \frac{(jw)^3}{6} \left[\begin{aligned} &\kappa_1 \int_{\tau_0}^{\tau} S_1(z) dz + \\ &+ 6\kappa_2 \int_{\tau_0}^{\tau} S_1^2(z) dz + \\ &+ 6\kappa_3 \int_{\tau_0}^{\tau} S_1^3(z) dz \end{aligned} \right] \right\},$$

где величина κ_3 определяется равенством

$$\kappa_3 = \frac{\partial f_3(0)}{\partial z} \mathbf{E},$$

а вектор-функция $f_3(z)$ удовлетворяет условию $f_3(\infty)\mathbf{E}=0$ и является решением уравнения

$$\frac{\partial f_3(z)}{\partial z} + \frac{\partial f_3(0)}{\partial z}(\mathbf{A}(z) - \mathbf{I}) + \frac{\partial f_2(0)}{\partial z} \mathbf{A}(z) - \kappa_1 f_2(z) - \kappa_2 \mathbf{R}(z) = 0.$$

Определение. Функцию

$$h_3(u) = \exp \left\{ \begin{aligned} &juk_1\beta_1 + \frac{(ju)^2}{2}[\kappa_1\beta_1 + 2\kappa_2\beta_2] + \\ &+ \frac{(ju)^3}{6}[\kappa_1\beta_1 + 6\kappa_2\beta_2 + 6\kappa_3\beta_3] \end{aligned} \right\} \quad (14)$$

будем называть асимптотикой третьего порядка характеристической функции числа занятых приборов в системе, а величину $[\kappa_1\beta_1 + 6\kappa_2\beta_2 + 6\kappa_3\beta_3]$ – асимптотическим семиинвариантом третьего порядка, где $\beta_3 = \int_0^\infty (1 - B(z))^3 dz$.

Область применимости асимптотических результатов в допредельной ситуации

С помощью полученных асимптотик (13), (14) и обратного преобразования Фурье запишем асимптотическое распределение вероятностей числа занятых приборов в системе

$$P_2(i) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ju i} h_2(u) du, \quad (15)$$

$$P_3(i) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ju i} h_3(u) du. \quad (16)$$

Полученные распределения будем называть асимптотической аппроксимацией второго (15) и третьего (16) порядков допредельного распределения вероятностей.

Для системы с неограниченным числом приборов и полумарковским входящим потоком при детерминированном времени обслуживания, продолжительности b , стационарное распределение вероятностей $P(i) = P\{i(t)=i\}$ числа $i(t)$ приборов, занятых в момент времени t , имеет вид

$$P(i) = \frac{\kappa_1 r}{2\pi} \int_{y^2}^{\infty} \frac{1}{y^2} (1 - e^{-jyb}) (\mathbf{I} - \mathbf{A}^*(y))^2 \mathbf{A}^{*n-1}(y) \mathbf{E} dy, \quad (17)$$

где $\mathbf{A}^*(\alpha) = \int_0^\infty e^{j\alpha z} d\mathbf{A}(z)$.

Теперь сравним распределения вероятностей числа занятых приборов, полученные методом асимптотического анализа и допредельным способом. Для этого найдем расстояние Колмогорова между этими распределениями

$$\Delta_n = \max_{0 \leq m < \infty} \left| \sum_{i=0}^m P_n(i) - \sum_{i=0}^m P(i) \right|, \quad n = 2, 3,$$

где $P_n(i)$ – функция распределения, полученная с помощью асимптотического анализа, а $P(i)$ – функция распределения для допредельной ситуации (17).

Например, полумарковская матрица $\mathbf{A}(x) = \mathbf{G}(x) \mathbf{P}$, где $\mathbf{G}(x)$ – гамма-функция распределения $F(x) = \Gamma(\alpha, \beta)$ с параметрами $\alpha_1=5, \beta_1=5, \alpha_2=10, \beta_2=20, \alpha_3=15, \beta_3=45$,

$$\mathbf{G}(x) = \begin{bmatrix} \Gamma(5,5) & \Gamma(10,20) & \Gamma(10,20) \\ \Gamma(15,45) & \Gamma(5,5) & \Gamma(10,20) \\ \Gamma(15,45) & \Gamma(15,45) & \Gamma(5,5) \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} 0,3 & 0,3 & 0,4 \\ 0,4 & 0,5 & 0,1 \\ 0,3 & 0,2 & 0,5 \end{bmatrix}.$$

Для различных значений b значения Δ_n составили (таблица).

Таблица. Область применимости асимптотических результатов в допредельной ситуации

$N \backslash b$	5	10	50	100
Δ_2	0,0475	0,0284	0,0119	0,0090
Δ_3	0,0343	0,0192	0,0058	0,0042

На рис. 3 показана одна из графических реализаций распределения вероятностей состояний системы, полученных асимптотически $P_2(i), P_3(i)$ и в допредельной ситуации $P(i)$ при заданных значениях параметров.

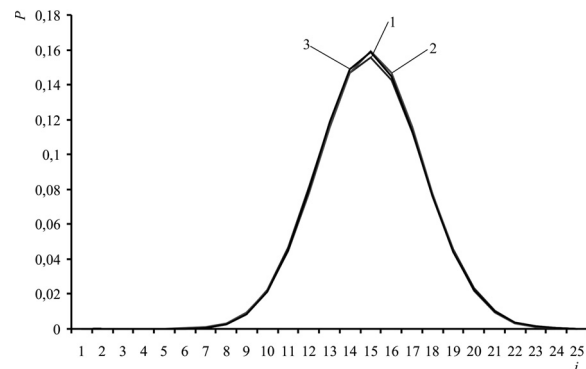


Рис. 3. Графическая реализация полученных результатов при $b=10$: 1) допредельное распределение вероятностей $P(i)$; 2) асимптотическая аппроксимация второго порядка $P_2(i)$ допредельного распределения вероятностей; 3) асимптотическая аппроксимация третьего порядка $P_3(i)$ допредельного распределения вероятностей

Полагая приемлемой погрешность аппроксимации, равную значению 0,02 расстояния Колмогорова, можно сделать вывод о том, что применение метода асимптотического анализа к исследованию системы с $SM|G|_\infty$ целесообразно при $b \geq 50$, применяя асимптотическую аппроксимацию второго порядка, и уже при $b \geq 10$ для асимптотики третьего порядка.

Выводы

Проведено исследование системы с неограниченным числом обслуживающих приборов и вхо-

длящим SM-поток методом асимптотических семиинвариантов. Предложен метод просеянного потока для исследования таких систем массового обслуживания. Получены асимптотические семиинварианты первого, второго и третьего порядков,

а также показана область применимости асимптотических результатов в допредельной ситуации.

Работа выполнена при поддержке АВЦП «Развитие научного потенциала высшей школы (2009–2011 гг.)», проект № 11803.

СПИСОК ЛИТЕРАТУРЫ

1. Лопухова С.В. Исследование полумарковского потока асимптотическим методом третьего порядка // Информационные технологии и математическое моделирование: Матер. VI Междунар. научно.-практ. конф. – Томск, 2007. – Ч. 2. – С. 30–34.
2. Назаров А.А., Лопухова С.В., Гарайшина И.Р. Исследование полумарковского потока событий // Вычислительные технологии. – 2008. – Т. 13. – Спецвыпуск 5. – С. 56–62.
3. Дудин А.Н., Клименок В.И. Системы массового обслуживания с коррелированными потоками. – Минск: БГУ, 2000. – 175 с.
4. Назаров А.А., Моисеева С.П. Метод асимптотического анализа в теории массового обслуживания. – Томск: Изд-во НТЛ, 2006. – 112 с.
5. Назаров А.А., Семенова И.А. Исследование системы $MMP|G|_{\infty}$ методом просеянного потока // Вестник Томского государственного университета. Сер. Управление, вычислительная техника и информатика. – 2011. – № 4 (17). – С. 74–84.
6. Лопухова С.В. Асимптотические и численные методы исследования специальных потоков однородных событий: дис. ... канд. физ.-мат. наук. – Томск, 2008. – 167 с.
7. Назаров А.А., Семенова И.А. Исследование RQ-систем методом асимптотических семиинвариантов // Вестник Томского государственного университета. Сер. Управление, вычислительная техника и информатика. – 2010. – № 3 (12). – С. 85–96.

Поступила 20.01.2012 г.

УДК 519.9

ПОСТРОЕНИЕ И ИССЛЕДОВАНИЕ АЛГОРИТМА ПОИСКА ПЕРВОЙ КРАЙНЕЙ ПОДСИСТЕМЫ ДЛЯ ЗАДАННОЙ СОВМЕСТНОЙ СИСТЕМЫ ЛИНЕЙНЫХ НЕРАВЕНСТВ

С.Г. Небаба, О.Н. Вылегжанин

Томский политехнический университет
E-mail: onv@am.tpu.ru

Разработан алгоритм поиска крайней подсистемы для заданной системы линейных неравенств, определена область допустимых значений для этой подсистемы и предложен метод нахождения зависимых неравенств. Алгоритм включает учет возможных ограничений-равенств, а также приведение матрицы системы неравенств к матрице полного столбцового ранга. Работа алгоритма демонстрируется на тестовом примере.

Ключевые слова:

Система линейных неравенств, крайняя подсистема, учет ограничений-равенств, область допустимых значений, зависимые неравенства.

Key words:

System of linear inequalities, extreme subsystem, accounting of equality constraints, range of permitted values, dependent inequalities.

Подавляющее большинство задач оценки и прогноза реальных производственно-экономических ресурсов и результатов их использования сводится к нахождению решения (области допустимых решений) для некоторой системы неравенств, выражающих ограничения по тем или иным ресурсам. Эти неравенства представляются в виде линейных функций, либо легко могут быть приведены к ним с использованием ряда допущений.

Нахождение решения системы линейных неравенств при наличии некоторой целевой функции обычно относят к задачам математического программирования. Эти задачи относятся к задачам исследования конкретных производственно-хозяйственных ситуаций, которые в том или ином виде интерпретируются как задачи об оптимальном использовании ограниченных ресурсов.

На практике нередко отсутствует целевая функция, на нахождение оптимального значения которой направлено большинство моделей математического программирования, и на основании учета ограничений в виде системы линейных неравенств определяется множество точек, выпуклая оболочка которых является областью допустимых значений задачи.

В связи с этим возникает задача построения алгоритма, позволяющего найти область допустимых значений для любой заданной системы линейных неравенств при наличии ограничений-равенств.

Отдельные свойства систем линейных неравенств рассматривались еще в первой половине XIX в. в связи с некоторыми задачами аналитической механики, решение которых сводилось к системам такого рода. Их систематическое изучение

началось в самом конце XIX в., однако о теории линейных неравенств стало возможным говорить лишь в конце двадцатых годов XX в., когда уже накопилось достаточное количество связанных с ними результатов.

В тридцатых годах XX в. появился эскиз Г. Вейля теории конечных систем линейных неравенств, построенной финитными методами без привлечения топологических свойств поля действительных чисел. Дальнейшее развитие теории линейных неравенств показало, что теория конечных систем линейных неравенств может рассматриваться как ветвь линейной алгебры, исходящая из нее при дополнительном требовании упорядоченности основного поля (поля коэффициентов). Излагаемая в [1] алгебраическая теория линейных неравенств охватывает все основные результаты, относящиеся к конечным системам линейных неравенств с действительными коэффициентами и свободными членами, давая им алгебраическую трактовку. Центральным предложением этой теории был установлен автором [1] принцип граничных решений, утверждающий, что в каждой совместной конечной системе линейных неравенств над полем действительных чисел, имеющей любой отличный от нуля ранг, можно выделить хотя бы одну такую подсистему того же ранга и с равным ему числом неравенств, каждое решение которой, обращающее все ее неравенства в равенства, удовлетворяет исходной системе. За последнее время появились новые математические средства, позволяющие выделять крайние подсистемы системы линейных неравенств более просто и эффективно.

Рассмотрим задачу нахождения области допустимых значений для системы линейных неравенств и равенств. В этом случае математическая постановка задачи может быть сформулирована следующим образом:

$$\hat{A}x \leq \hat{b}, \quad x \in R^n, \quad \hat{b} \in R^m, \quad (1)$$

$$\check{A}x = \check{b}, \quad x \in R^n, \quad \check{b} \in R^m. \quad (2)$$

Будем полагать, что системы линейных неравенств (1) и ограничений-равенств (2) определены или переопределены, и первые k столбцов \hat{A} линейно независимы.

Необходимо найти множество крайних подсистем, узловых решения которых образуют множество точек, выпуклая оболочка которых совпадает с областью допустимых значений решаемой системы.

В работе [2] предложен метод нахождения крайней подсистемы, сводящийся к решению задачи квадратичного программирования, которую решали методом субоптимизации на многообразиях. Однако этот метод не учитывает в явном виде наличие ограничений-равенств. В данной работе предложен метод решения задачи отыскания крайней подсистемы путем последовательного перемещения рабочей точки вдоль области пересечения удерживающих неравенств.

Для обоснования алгоритма получения крайней подсистемы, т. е. получения точки, являющейся вершиной многогранника области допустимых значений M , сформулируем несколько утверждений:

Утверждение 1. Пусть $\hat{A}x \leq \hat{b}$ – крайняя подсистема системы линейных неравенств, тогда областью допустимых значений этой крайней подсистемы будет множество точек, удовлетворяющих условию

$$(x - x^*) \leq 0,$$

где $x^* = \hat{A}^{-1} \cdot \hat{b}$.

Доказательство. Пусть D – конус допустимых значений, определенный условием $\hat{A}x \leq \hat{b}$, с вершиной в точке $x^* = \hat{A}^{-1} \cdot \hat{b}$. Пусть x – произвольная точка, принадлежащая этому конусу ($x \in D$), тогда, вычитая из $\hat{A}x \leq \hat{b}$ почленно $\hat{A}x^* = \hat{b}$ получим $\hat{A}(x - x^*) \leq 0$ и, умножив обе части слева на \hat{A}^{-1} , получим доказываемое.

Утверждение 2. Если x_0 – начальная точка, а ближайшая к ней плоскость многогранника M имеет направляющим вектором a_i , то x_1 – проекция начальной точки на эту плоскость может быть найдена как

$$x_1 = x_0 + \frac{b_i - a_i^T x_0}{a_i^T a_i} a_i.$$

Утверждение 3. Если точка x_k принадлежит пересечению k плоскостей, заданных направляющими векторами $a_j, j=1 \dots k$, то ее проекция на пересечение плоскости, заданной строкой a_i , с пересечением плоскостей, заданных строками $a_j, j=1 \dots k$, составляющими матрицу A_k , есть:

$$x_{k+1} = x_k + \frac{b_j - a_i^T x_k}{a_i^T \Omega a_i} \Omega a_i, \quad (3)$$

где $\Omega = (I - A_k \cdot A_k^+)$ – проектор на пространство, перпендикулярное пространству, натянутому на строки матрицы A_k , а A_k^+ – матрица, псевдообратная к матрице A_k , I – единичная матрица.

Доказательства утверждений 2 и 3 приведены в [3]. Повторив процедуру, определенную утверждением 3, не более $k-1$ раз, получим подсистему, образованную k линейно независимыми строками исходной системы.

Определим вектор разностей \bar{y}

$$\bar{y} = \bar{b} - Ax,$$

показывающий, каким неравенствам системы, определенной матрицей A и вектором \bar{b} , удовлетворяет точка x . Решение полученной подсистемы может нарушать некоторые неравенства исходной системы, т. е. при подстановке его в исходную систему вектор разностей может содержать отрицательные элементы.

Для вектора \bar{y} , в зависимости от того, где находится точка x , возможны следующие варианты:

- если $y_i < 0$ для $\forall i | i \in 1 \dots n$, то точка x является внутренней точкой многогранника M ;
- если $y_i = 0$ только для некоторой части из $i | i \in 1 \dots n$, а для остальных элементов i $y_i < 0$, то

точка x принадлежит поверхности многогранника M ;

- если $\exists i | i \in 1...n$, для которых $y_i > 0$, то точка x находится вне области многогранника M .

Рассмотрим общий случай, когда начальная точка находится вне многогранника допустимых значений задачи. Для отыскания крайней подсистемы необходимо найти точку, соответствующую вершине многогранника M . Для этого, исключив из матрицы A_k те строки, которые соответствуют положительным значениям вектора \bar{y} , будем повторять операцию (3), пока все элементы вектора \bar{y} не станут неположительными.

Когда такая точка найдена, то все неравенства исходной системы могут быть разделены на два множества: множество неравенств, которые не изменяют область допустимых значений D , и множество неравенств, которые ее изменяют. Назовем первые неравенства зависимыми от данной крайней подсистемы. Рассмотрим задачу выделения зависимых неравенств.

Утверждение 4. Пусть в линейном пространстве L размерности r задана система линейных неравенств $Ax \leq \bar{b}$ ранга r . Тогда для того, чтобы линейное неравенство $\alpha x \leq \beta$ было зависимым от этой системы необходимо и достаточно, чтобы для вектора $\gamma^T = \alpha A^+$ выполнялось условие $\gamma \geq 0$.

Доказательство.

Достаточность. При умножении любого неравенства системы $a_i^T x \leq b_i$ на неотрицательный множитель $\gamma \geq 0$ неравенство не нарушится. Отсюда справедливо $\gamma^T A x \leq \gamma^T b$, где γ – вектор неотрицательных множителей. Обозначив $\alpha = \gamma^T A$ и $\beta = \gamma^T b$, получим доказываемое.

Необходимость. Так как ранг системы равен размерности пространства векторов L ($\text{rk} A = r$), то любой вектор из L может быть представлен в виде линейной комбинации строк A , т. е. $\alpha = \gamma A$ или, умножив обе части на A^+ , получим $\gamma^T = \alpha^T A^+$. Если среди элементов γ есть хотя бы один отрицательный, то для него справедливо $\gamma A_i^T x > \gamma b_i$, что противоречит условиям теоремы.

Для учета ограничений-равенств, проверки матрицы системы на полноту столбцового ранга, поиска первой крайней подсистемы и учета зависимых неравенств был разработан следующий алгоритм:

1. Задаются начальные данные: система линейных неравенств (матрица A , вектор \bar{b}), система ограничений-равенств (матрица A , вектор \bar{b}).

2. Выполняется учет ограничений-равенств по методу, предложенному в [4].

3. Проверяется полученная матрица системы на полноту столбцового ранга и, если необходимо, она приводится к матрице полного столбцового ранга методом, предложенным в [3].

4. В результате описанных выше шагов исходная система ограничений равенств и неравенств приводится к системе линейных неравенств полного столбцового ранга: с матрицей A и вектором \bar{b} .

Выбирается начальная точка x_0 и далее:

5. Вычисляется вектор

$$\bar{y} = \bar{b} - Ax_0.$$

6. Выбирается ближайшая плоскость (минимальный отрицательный элемент $\bar{y} \neq 0$) с номером i .

7. Вычисляется проекция начальной точки на выбранную плоскость [3]:

$$x_1 = x_0 + \frac{b_i - a_i^T x_0}{a_i^T a_i} a_i,$$

где x_0 – начальная точка; a_i – направляющий вектор ближайшей к ней плоскости многогранника M ; b_i – i -й элемент вектора \bar{b} ; x_1 – проекция начальной точки на эту плоскость.

9. Начало рекуррентного вычисления, начиная с l , где $l=2, \dots, n$; n – число строк матрицы A .

10. Вычисляется вектор

$$\bar{y} = \bar{b} - Ax_l.$$

11. Если среди элементов вектора \bar{y} нет положительных и k элементов равны нулю, где k – размерность задачи (число столбцов матрицы A), то множество строк матрицы A , для которых элементы вектора \bar{y} равны нулю, составляют крайнюю подсистему, а точка x_l – искомая вершина многогранника M . Иначе переход к шагу 12.

12. Выбирается ближайшая к точке x_l плоскость j (по минимальному отрицательному элементу вектора \bar{y}).

13. Вычисляется вектор направления от точки x_l к плоскости j [3]:

$$\Omega a_j = (I - A_1^+ A_1) a_j,$$

где I – единичная матрица, A_1^+ – матрица, псевдообратная к матрице A_1 .

14. Вычисляется проекция точки x_l на плоскость j [2]:

$$x_{l+1} = x_l + \frac{b_j - a_j^T x_l}{a_j^T \Omega a_j} \Omega a_j.$$

15. Переход к следующему шагу:

$$l=l+1,$$

начиная с шага 10.

16. Вывод найденной точки x_n .

17. Выполняется поиск зависимых неравенств: умножив матрицу системы на матрицу, псевдообратную найденной подсистеме, получим матрицу, которая показывает зависимость всех неравенств от неравенств, составляющих крайнюю подсистему. В этом случае строки матрицы, содержащие только отрицательные элементы, будут соответствовать зависимым неравенствам.

В качестве тестового примера была взята система из [4].

Исходные ограничения-равенства:

$$\begin{cases} 1, 2x_1 - 3x_2 + x_3 - x_5 = -0, 5 \\ 1, 4x_1 - 1, 5x_2 + x_3 - x_4 - x_5 = 0, 5 \\ -0, 7x_1 + 0, 5x_2 + x_3 - x_4 + x_5 = 4, 5 \\ 0, 26x_1 - 0, 45x_2 + 0, 2x_3 - 0, 1x_4 - 0, 2x_5 = 0 \end{cases}$$

Ограничения-неравенства:

$$\begin{cases} 0,4x_1 - 2x_2 + 0,1x_3 - 0,501x_4 + 0,35x_5 \leq 2,225 \\ -x_1 - 0,1x_2 - 2x_3 + 9,994x_4 - 3,575x_5 \leq 11,037 \\ -0,5x_1 + 0,1x_2 + 0,5x_3 - 7,481x_4 + 5,425x_5 \leq 6,537 \\ 4,5x_1 - 3x_2 - 10,1x_3 - 16,551x_4 - 30,05x_5 \leq 110,225 \\ 4,5x_1 - 3x_2 - 2,398x_4 - 5x_5 \leq -12 \\ -x_1 + 2x_2 + 0,5x_3 - 1,2x_4 - 0,35x_5 \leq -1,125 \\ 2,4x_1 - x_2 - 3,5x_3 + 4x_4 - 5,2x_5 \leq 0,2 \\ -x_1 + x_2 - 3x_3 + 4,687x_4 - 1,25x_5 \leq -2,625 \end{cases}$$

Или в обозначениях (1) и (2):

$$\hat{A} = \begin{pmatrix} 1,2 & -3 & 1 & 0 & -1 \\ 1,4 & -1,5 & 1 & -1 & -1 \\ -0,7 & 0,5 & 1 & -1 & 1 \\ 0,26 & -0,45 & 0,2 & -0,1 & -0,2 \end{pmatrix}, \hat{b} = \begin{pmatrix} -0,5 \\ 0,5 \\ 4,5 \\ 0 \end{pmatrix},$$

$$\hat{A} = \begin{pmatrix} 0,4 & -2 & 0,1 & 0,501 & 0,35 \\ -1 & 0,1 & -2 & 9,994 & -3,575 \\ -0,5 & 0,1 & 0,5 & -7,481 & 5,425 \\ 4,5 & -3 & -10,1 & -16,551 & -30,05 \\ 4,5 & -3 & 0 & -2,398 & -5 \\ -1 & 2 & 0,5 & -1,2 & -0,35 \\ 2,4 & -1 & -3,5 & 4 & -5,2 \\ -1 & 1 & -3 & 4,687 & -1,25 \end{pmatrix}, \hat{b} = \begin{pmatrix} 2,225 \\ 11,037 \\ 6,537 \\ 110,225 \\ -12 \\ -1,125 \\ 0,2 \\ -2,625 \end{pmatrix}.$$

Система линейных неравенств после учета ограничений-равенств и проверки полноты ранга матрицы \hat{A} :

$$A = \begin{pmatrix} -0,998 & -1 \\ 8 & -3 \\ -7 & 5 \\ -26,998 & -25 \\ 2,998 & -5 \\ -0,162 & 0,4 \\ 0,455 & -3 \\ 2 & 0,5 \end{pmatrix}, \bar{b} = \begin{pmatrix} 2 \\ 16 \\ 4,999 \\ 135 \\ 15 \\ 1 \\ 9,5 \\ 7 \end{pmatrix}.$$

В качестве начальной точки выбрана:

$$x_0 = (2 \quad -8).$$

Применяя предложенный алгоритм, получим следующую последовательность векторов \bar{y} и точек x_i :

1 шаг. Вектор $\bar{y} = \bar{b} - Ax_0$:

$$\bar{y} = (-4,004 \quad -24 \quad 58,999 \quad -11,004 \quad -30,996 \quad 4,524 \quad -15,41 \quad 7).$$

Ближайшая плоскость соответствует первой строке матрицы A $(-0,998 \quad -1)$, проекция начальной точки на соответствующую этой строке плоскость (точка x_1) есть:

$$x_1 = (4,002 \quad -5,994).$$

2 шаг. Вектор $\bar{y} = \bar{b} - Ax_1$

$$\bar{y} = (0 \quad -33,998 \quad 62,983 \quad 93,196 \quad -26,968 \quad 4,046 \quad -10,303 \quad 1,9 \quad 93).$$

Выбирается седьмая строка матрицы A $(0,455 \quad -3)$, вычисляется точка x_2 :

$$x_2 = (1,015 \quad -3,013).$$

3 шаг. Вычисляется вектор $\bar{y} = \bar{b} - Ax_2$:

$$\bar{y} = (0 \quad -1,157 \quad 27,166 \quad 87,078 \quad -3,106 \quad 2,369 \quad 0 \quad 6,477).$$

Выбирается вторая строка матрицы A $(8 \quad -3)$, вычисляется точка x_3 :

$$x_3 = (0,91 \quad -2,908).$$

4 шаг. Вычисляется вектор $\bar{y} = \bar{b} - Ax_3$:

$$\bar{y} = (0 \quad 0 \quad 25,905 \quad 86,863 \quad -2,266 \quad 2,31 \quad 0,363 \quad 6,635).$$

Выбирается пятая строка матрицы A $(2,998 \quad -5)$, вычисляется точка x_4 :

$$x_4 = (0,626 \quad -2,625).$$

5 шаг. Вычисляется вектор $\bar{y} = \bar{b} - Ax_4$:

$$\bar{y} = (0 \quad 3,118 \quad 22,504 \quad 86,282 \quad 1,776 \cdot 10^{-15} \quad 2,151 \quad 1,341 \quad 7,06).$$

Вектор \bar{y} содержит только неотрицательные элементы, из которых два равны нулю, следовательно, найдена первая крайняя подсистема, состоящая из первого и пятого неравенств системы.

Для нахождения неравенств, зависимых от этой крайней подсистемы, найдем матрицу, псевдообратную матрице крайней подсистемы:

$$A_n^+ = \begin{pmatrix} -0,626 & 0,125 \\ -0,375 & -0,125 \end{pmatrix}.$$

Умножив ее справа на исходную матрицу ограничений-неравенств получим:

$$A \cdot A_n^+ = \begin{pmatrix} 1 & 0 \\ -3,882 & 1,376 \\ 2,505 & -1,501 \\ 26,282 & -0,256 \\ 0 & 1 \\ -0,049 & -0,07 \\ 0,841 & 0,432 \\ -1,44 & 0,188 \end{pmatrix}.$$

Следовательно, зависимым является седьмое неравенство $(0,455 \quad -3)$.

Выводы

Разработан алгоритм расчета крайней подсистемы для заданной системы линейных неравенств, включающий учет ограничений равенств и приведение полученной матрицы системы неравенств к полному столбцовому рангу. Предложен способ определения неравенств системы, зависимых от неравенств, входящих в найденную крайнюю подсистему, и доказаны соответствующие теоремы. Результаты расчета тестового примера подтверждают работоспособность предложенного алгоритма.

СПИСОК ЛИТЕРАТУРЫ

1. Черников С.Н. Линейные неравенства. – М.: Наука, 1968. – 488 с.
2. Вен В.Л., Литвинцев П.И., Фролов К.К. Решение системы линейных неравенств по методу наименьших квадратов // Журнал вычислительной математики и математической физики. – 1984. – Т. 24. – № 1. – С. 144–148.
3. Вылегжанин О.Н., Шкатова Г.И. Решение задачи линейного программирования с использованием оператора-проектора //

- Известия Томского политехнического университета. – 2009. – Т. 314. – № 5. – С. 37–40.
4. Вылегжанин О.Н., Шкатова Г.И. Учет ограничений равенств при решении оптимизационных задач с линейными ограничениями // Известия Томского политехнического университета. – 2008. – Т. 312. – № 5. – С. 76–78.

Поступила 13.03.2012 г.

УДК 004.94

МАТРИЧНЫЙ АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ РАЗРЕЗАНИЯ ДВУДОЛЬНЫХ ГРАФОВ

Ан.В. Погребной, Д.В. Погребной

Томский политехнический университет
E-mail: avpogrebnoy@gmail.com

Предложен алгоритм решения задачи разрезания двудольного графа на заданную совокупность минимально связанных подграфов. Алгоритм использует матричное представление двудольного графа и учитывает его специфику. Это позволило свести задачу разрезания к задаче линейного математического программирования. Алгоритм может работать как с произвольно сформированным, так и улучшенным исходным вариантом разрезания двудольного графа.

Ключевые слова:

Двудольный граф, исходное разрезание, матрица связности вершин графа, задача линейного математического программирования.

Key words:

Bipartite graph, the original cut, connectivity matrix of the vertices, problem of linear mathematical programming.

Введение

Двудольные графы [1] используют при построении графовых моделей, описывающих отношения между объектами разбитыми на два множества. В таких моделях отношения (рёбра) могут связывать объекты (вершины) только из разных множеств. Отношения между объектами внутри множеств отсутствуют или не учитываются. Примером графовых моделей, имеющих двудольную природу, являются сети Петри [2] с непересекающимися множествами позиций и переходов. Дуги в сети Петри могут связывать только переходы и позиции. Другим примером являются графы потока данных [3] с множествами модулей и данных, в которых дуги отражают потребление или формирование данных модулями.

Формально двудольный граф G описывается четвёркой (N, M, V, R) , где $N = \{e_i\}$, $i = 1, 2, \dots, n$; $M = \{e_j\}$, $j = 1, 2, \dots, m$ – множества вершин двудольного графа; $V = \{v_{ij}\}$ – множество рёбер v_{ij} , связывающих неупорядоченные пары вершин (e_i, e_j) , $e_i \in N$, $e_j \in M$; $R = \{r_{ij}\}$ – множество весов r_{ij} рёбер v_{ij} .

Если для всех $v_{ij} \in V$, $r_{ij} = 1$, то имеет место не взвешенный двудольный граф. Заметим также, что граф G рассматривается как неориентированный, т. к. пара вершин (e_i, e_j) , связанная ребром v_{ij} , принимается неупорядоченной. Ориентация дуг, например в графе потока данных, означает направление передачи данных. Вместе с тем, загрузка се-

ти при передаче данных по разрезанной дуге зависит от веса дуги и не зависит от её ориентации. Как в этом, так и во многих других примерах, ориентация дуг не учитывается и при решении задачи разрезания используется неориентированный взвешенный двудольный граф. Такой граф полностью описывается матрицей связности вершин $R = \|r_{ij}\|_{N \times M}$. Матрица R несимметрична и все её элементы (веса рёбер) $r_{ij} \geq 0$. Особенность матрицы R заключается в том, что она допускает перестановки строк R_i и столбцов R_j независимо друг от друга.

Задача разрезания возникает всякий раз, когда графовую модель, описывающую отношения между объектами распределённой системы, необходимо разбить на совокупность минимально связанных между собой подсистем. Если графовая модель представлена двудольным графом, то это, как правило, связано с наличием в системе объектов двух типов, которые нельзя смешивать в ходе решения задачи разрезания. В частности поэтому, для её решения нельзя использовать матричный алгоритм разрезания обыкновенного графа [4], в котором вершины не различаются по какому-либо признаку и образуют единое множество. Кроме того, двудольные графы в сравнении с обыкновенными обладают спецификой, которая даёт возможность разработать алгоритм разрезания, эффективно работающий для данного класса графов.

Предложенный в статье алгоритм можно отнести к алгоритмам последовательного типа. На первом этапе формируется исходный вариант разрезания графа G на подграфы $G_q, q=1,2,\dots,k$. Здесь фиксируется число k подграфов G_q , на которые разрезается граф G и для каждого q -го подграфа устанавливается число и состав вершин из множеств N и M . На последующих этапах выполняется улучшение разрезания за счёт изменения состава вершин в подграфах исходного разрезания $\{G_q\}$ так, чтобы сумма весов разрезанных рёбер была минимальной.

Алгоритм назван матричным, т. к. основан на представлении двудольного графа матрицей связности вершин R и при реализации поисковых процедур также использует матричное представление.

Формирование исходного варианта разрезания

Алгоритм разрезания будем рассматривать на примере взвешенного двудольного графа, представленного матрицей R на рис. 1, а, в котором множество N содержит 8, а M – 12 вершин. Граф G требуется разрезать на 3 подграфа $G_q, q=1, 2, 3$ с заданным числом вершин из множеств N и M .

На матрице R (рис. 1, а) вариант разрезания по аналогии с алгоритмом [4] можно представить совокупностью блоков, каждый из которых соответствует подграфу G_q с множеством вершин $N(G_q)$ и $M(G_q)$. В данном примере все блоки содержат по 4 вершины из множества M , а 8 вершин множества N распределены по блокам в количестве 3, 3, 2. Число блоков и их размерность, т. е. соотношение числа вершин из множеств N и M в соответствующих подграфах устанавливается пользователем. Последовательность расположения блоков разной размерности в матрице R формируется произвольно.

Элементы r_{ij} матрицы R , расположенные за пределами блоков, соответствуют межблочным связям, т. е. разрезанным рёбрам, связывающим разные подграфы. Эти элементы составляют часть матрицы, названную областью сечения. Таким образом, матрица R разбивается на область блоков и область сечения. Задача разрезания сводится к перестановкам строк R_i и столбцов R_j матрицы R

так, чтобы сумма элементов r_{ij} в области блоков была максимальной, а в области сечения – минимальной. Качество разрезания оценивается соотношением суммы элементов в блоках к сумме в области сечения. Чем больше величина этого соотношения, тем лучше вариант разрезания.

Приведённый на рис. 1, а, исходный вариант разрезания получен произвольным образом. Вершины $e_i \in N$ и $e_j \in M$ включены в очередную блок в порядке их нумерации. К тому же блок 3 в последовательности блоков мог быть поставлен первым или вторым. Соотношение сумм элементов в блоках и в области сечения для данного варианта разрезания равно 26/56. При визуальном анализе области блоков и области сечения легко обнаружить, что данный вариант разрезания можно улучшить. Так, например, если столбцы R_{11} и R_{12} поменять местами, то соотношение станет равным 35/47, т. е. улучшится на 9 единиц. Взаимная перестановка строк R_{11} и R_{12} также улучшает исходный вариант на 4 единицы.

Операции по поиску в множествах строк и столбцов перестановок, улучшающих качество разрезания, могут быть формализованы аналогично тому как это было сделано в [4]. Однако в данном случае более эффективным оказалось применение аппарата математического программирования. Такая возможность была найдена благодаря учёту отмеченной выше особенности матрицы R .

Метод улучшения качества разрезания

Рассмотрим подробнее операцию перестановки строк R_{11} и R_{12} . Каждая строка относительно блоков делится на подмножества элементов, принадлежащих столбцам соответствующих блоков $M(G_1), M(G_2), M(G_3)$. При перестановке строк элементы этих подмножеств всегда совместно попадают в некоторый блок или исключаются из него. В нашем случае при перемещении строки R_{11} на место R_{12} одно подмножество (0,1,2,0) исключается из 1-го блока, а другое (3,0,2,0) включается в 3-й блок. Аналогично для строки R_{12} при перемещении её на место R_{11} подмножество (0,0,3,1) включается в 1-й блок, а подмножество (0,1,0,1) исключается из 3-го блока.

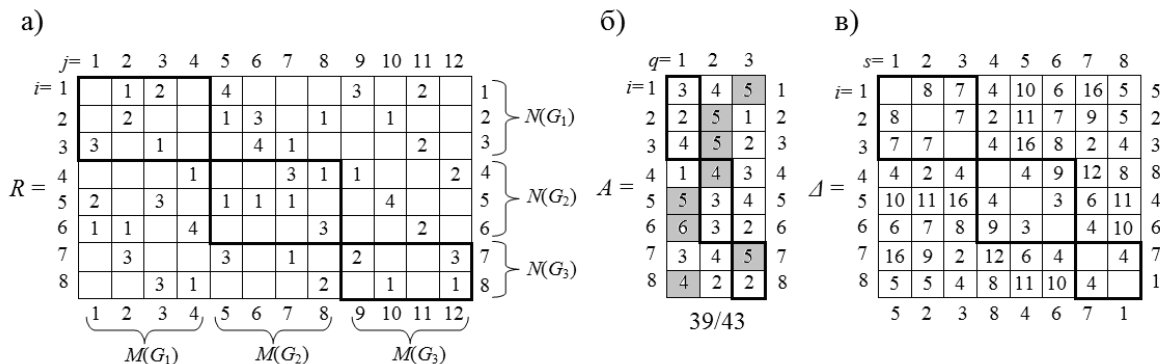


Рис. 1. Исходный вариант разрезания

нумерацией снизу матрицы. Попытка улучшить для этой матрицы качество разрезания путём перестановки строк на основе решения задачи *TN* не привела к успеху. На рис. 2, *г*, видно, что в матрице *A* все строки сохранили свои места в блоках. Визуальный анализ области сечения матрицы на рис. 2, *в* также показывает, что путём перестановки какой-либо строки или столбца достигнутое качество разрезания нельзя улучшить. Ниже предлагается попытка достигнуть разрезание с более высоким качеством за счёт использования улучшенного исходного разрезания.

Методика поиска варианта исходного распределения строк по блокам.

Вариант разрезания с оценкой 56/26 получен на основе применения метода улучшения качества исходного разрезания, сформированного произвольным образом. В связи с этим возникает вопрос, имеется ли возможность улучшить качество разрезания, если исходное разрезание формировать не произвольным образом, а на основе некоторой процедуры рационального распределения строк или столбцов по блокам.

Рассматривая вопрос о предпочтениях в распределении строк по блокам, легко обнаружить, что в один блок должны попадать те строки, вершины которых соответствуют максимальным суммам весов связей с общими вершинами столбцов. В этом случае при попадании в общий блок данных столбцов и строк суммарный вес блока возрастает. Связность двух строк R_i и R_s , $i, s=1,2,\dots,n$, обозначим величиной δ_{is} , $i \neq s$. Значение величины δ_{is} определяется согласно выражениям:

$$\forall e_i, e_s \in N \ \& \ \forall e_j \in M [r_{ij} > 0 \ \& \ r_{sj} > 0] \Rightarrow (\delta_{isj} = r_{ij} + r_{sj}), \ \text{иначе} \ (\delta_{isj} = 0); \tag{9}$$

$$\delta_{is} = \sum_{j=1}^m \delta_{isj}. \tag{10}$$

На рис. 3 приведены пояснения к вычислению δ_{is} на примере величины $\delta_{1,7}$. Вершины e_1 и e_7 , соответствующие строке R_{i_1} и R_{i_7} , связаны через общие вершины e_2, e_3, e_9 столбцов $R_{j_2}, R_{j_3}, R_{j_9}$, которые помечены штриховкой. Сумма весов r_{ij} и r_{sj} для общих вершин e_j , связывающих вершины e_1 и e_7 составила 16 единиц, т. е. $\delta_{1,7}=16$.

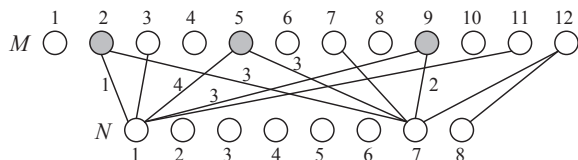


Рис. 3. Иллюстрация к определению $\delta_{1,7}$

Величины δ_{is} , вычисленные по выражениям (9), (10), формируются в матрицу связности строк Δ . Матрица Δ , полученная для исходной матрицы R (рис. 1, *а*), приведена на рис. 1, *в*. Для матрицы Δ с помощью матричного алгоритма (программное средство *CutGraf*) [6] решается задача разрезания, т. е. определяется распределение строк по блокам. При этом последовательность блоков и их размерность должна совпадать с этими параметрами в рассматриваемом примере, что и отражено на матрице Δ . Справа (и снизу) от матрицы указано полученное распределение строк по блокам.

Используя полученное распределение строк на матрице R формируется исходное разрезание и для него решается задача распределения столбцов, т. е. задача *TM*. Для нашего примера решение задачи *TM* привело к распределению столбцов $M(G_1)=(1,3,6,10)$, $M(G_2)=(4,7,8,11)$, $M(G_3)=(2,5,9,12)$. Оценка разрезания при данном распределении составила 58/24. Аналогичная методика формирования улучшенного исходного разрезания может быть применена и в отношении строк с последующим решением задачи *TN*. Достигнутое улучшение оценки на 2 единицы вовсе не означает, что с улучшенным исходным разрезанием всегда будет получено разрезание с более высокой оценкой.

Следует отметить, что после получения решения на основе произвольно сформированного исходного разрезания попытки поиска разрезания с более высокой оценкой являются вполне оправданными, т. к. не требуют больших вычислительных затрат. К таким попыткам, наряду с изложенной выше методикой улучшения исходного разрезания как относительно строк так и столбцов, относятся изменения последовательности решения задач *TN* и *TM*, а также изменения последовательности расположения блоков, имеющих разные размерности.

Заключение

1. Расширены возможности матричного алгоритма разрезания графов в направлении разработки более эффективного алгоритма решения этой задачи для двудольных графов. Учёт специфики двудольных графов сделал возможным применение в алгоритме улучшения разрезания графов этого типа методы линейного математического программирования.
2. Введение оценки связности вершин в одном из подмножеств двудольного графа через общие вершины другого подмножества позволило формировать исходное разрезание с более высокой оценкой. Для решения задачи поиска варианта распределения строк или столбцов по блокам исходного разрезания применен матричный алгоритм разрезания графов.

Работа выполнена в рамках государственного задания «Наука».

СПИСОК ЛИТЕРАТУРЫ

1. Зыков А.А. Основы теории графов. – М.: Вузовская книга, 2004. – 644 с.
2. Питерсон Дж. Теория сетей Петри и моделирование систем. Пер. с англ. – М.: Мир, 1984. – 264 с.
3. Погребной В.К. Визуальный уровень представления алгоритмов функционирования распределённой системы реального времени на языке структурного моделирования // Известия Томского политехнического университета. – 2009. – Т. 314. – № 5. – С. 140–146.
4. Погребной В.К. Матричный алгоритм решения задачи разрезания графов // Известия Томского политехнического университета. – 2007. – Т. 310. – № 5. – С. 91–96.
5. Галкина В.Н. Дискретная математика: комбинаторная оптимизация на графах. – М.: Гелиос АРВ, 2003. – 232 с.
6. Погребной А.В., Погребной Д.В. Исследование матричного алгоритма решения задачи разрезания графов // Молодёжь и современные информационные технологии: Труды VIII Всерос. научно-практ. конф. молодых ученых. – Томск, 2010. – С. 140–141.

Поступила 27.02.2012 г.

УДК 519.673

АДРЕСУЕМАЯ ЯЧЕЙКА ОДНОРОДНОЙ ВЫЧИСЛИТЕЛЬНОЙ СТРУКТУРЫ ДЛЯ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ

А.А. Хамухин

Томский политехнический университет

E-mail: aaxtru@tpu.ru

Предложена функциональная схема адресуемой ячейки однородной вычислительной структуры для решения дифференциальных уравнений в частных производных. Показана возможность ускорения решения за счет снижения длительности цикла загрузки данных в регистры ячеек. Представлены сравнительные результаты моделирования линейки адресуемых и не адресуемых ячеек на примере решения тестовой краевой задачи с постоянными и переменными коэффициентами.

Ключевые слова:

Адресуемая ячейка, однородная вычислительная структура, параллельные вычисления, дифференциальные уравнения в частных производных, постоянные и переменные коэффициенты.

Key words:

Addressable cell, homogeneous computing structure, parallel computation, partial differential equations, independent and dependent coefficients.

Введение

В статьях [1–3] описаны примеры реализации вычислительных устройств для решения дифференциальных уравнений в частных производных на основе одинаковых ячеек однородной вычислительной структуры (ОВС), показаны возможности реконfigurирования ОВС в соответствии с заданными граничными условиями без аппаратных добавлений в схему ячейки, приведены дополнения, необходимые для решения задач с переменными коэффициентами. Основу такой ячейки составляют сумматоры и масштабные множители, а информация о коэффициентах дифференциальных уравнений хранится в нескольких регистрах каждой ячейки. Разработан ряд схем реализации таких ячеек, на которые получены патенты [4–6].

Главным преимуществом однородной вычислительной структуры является синхронное аппаратное распараллеливание решения дифференциального уравнения, которое позволяет добиться наибольшей производительности на данном классе задач при прочих равных условиях по сравнению с другими типами вычислительных устройств. Вторым немаловажным преимуществом ОВС являются относительно низкие аппаратные затраты в пе-

ресчете на одну операцию по сравнению с универсальным многоядерным микропроцессором, в котором большая часть транзисторов просто простаивает при решении данного класса задач.

В этих статьях также сделано предположение, что реализация ОВС возможна на современных микросхемах типа БМК (известных за рубежом под аббревиатурой *ASIC*) или ПЛИС (известных за рубежом под аббревиатурой *FPGA*).

Наиболее известные реализованные проекты в этой области – это платформа *RC100* фирмы *SGI* и суперкомпьютер фирмы *Cray*. Платформа *RC100* содержит 35 модулей, каждый из которых имеет по две ПЛИС *Altera Stratix III* и 10 блоков локальной памяти. По данным фирмы *SGI* платформа решает задачи биоинформатики в 900 раз быстрее, чем кластер из 68 узлов на базе универсальных микропроцессоров *AMD Opteron* [7].

Суперкомпьютер *Cray XT5h* объединяет в единую систему скалярные процессоры (на основе *AMD Opteron*) и векторные процессоры на основе ПЛИС. Он комплектуется из серверов двух типов: *Cray X2* и *Cray XR1*. Последний состоит из двух узлов, в каждый из которых входит микропроцессор *AMD Opteron*, связанный с помощью высокоско-

ростной шины *Hyper Transport* с двумя ПЛИС *Xilinx Virtex-4*. Масштабируемость системы позволяет объединять до 30 тыс. ПЛИС-процессоров в одной системе [8].

В России также ряд научных коллективов выполняет подобные разработки. Так, например, в НИИ многопроцессорных вычислительных систем Южного федерального университета (НИИ МВС ЮФУ, г. Таганрог) выполняется проект по созданию модульно-наращиваемой вычислительной системы с производительностью 20 Тфлопс, содержащей 1536 ПЛИС *Virtex-5* в 19-дюймовой стойке [7]. Основу системы составляют базовые модули *16V5-1250P* с производительностью 250 Гфлопс каждый, содержащие 16 ПЛИС *Virtex-5* и соединяемые между собой в единый вычислительный ресурс с помощью скоростных *LVDS*-каналов. Четыре базовых модуля *16V5-1250P* объединяются в конструктиве *1U* вычислительного блока «Орион».

Хотя подобные вычислительные блоки, состоящие из реконфигурируемых полей ПЛИС, предназначены для решения более широкого спектра задач, чем решение дифференциальных уравнений в частных производных, их архитектура практически совпадает с архитектурой ОВС. Поэтому необходимо учитывать проблемы, возникающие с увеличением количества параллельно работающих блоков такой архитектуры.

В работе [9], представленной также НИИ МВС ЮФУ, эта проблема сформулирована как «кризис архитектуры реконфигурируемых вычислительных систем (РВС) на базе ортогональной коммутации». В работе отмечается, что с ростом степени интеграции ПЛИС растет потребность в увеличении количества каналов распределенной памяти. Однако конструктивные особенности базовых модулей не позволяют разместить достаточное количество каналов, поэтому реально наблюдается даже снижение производительности при увеличении количества вычислительных модулей при неизменном количестве каналов [9].

Объект и цель исследования

Принципиальное отличие предлагаемых ячеек однородной вычислительной структуры от рассмотренных выше вычислительных блоков обусловлено ограниченным классом задач – решением дифференциальных уравнений в частных производных. Относительно небольшое количество коэффициентов дифференциальных уравнений и особенность матрицы, получаемой при аппроксимации дифференциального уравнения, позволяют использовать для хранения коэффициентов внутренние быстрые регистры. Это избавляет от необходимости обращения к относительно медленным внешним блокам памяти и ускоряет весь процесс решения.

Но при этом необходимо выполнить запись данных в регистры ячеек ОВС перед началом цикла решения и считывание результатов из всех яче-

ек ОВС после окончания цикла решения. Поэтому при существенном увеличении количества ячеек в структуре можно ожидать эффект снижения общей производительности из-за влияния длительности загрузки данных в регистры ячеек. При частом повторении этой процедуры загрузки реальная производительность ОВС может существенно снизиться.

Объектом исследования является процесс загрузки данных в регистры однородной вычислительной структуры перед началом каждого цикла решения заданного дифференциального уравнения в частных производных.

Целью исследования является модификация схемы ячейки таким образом, чтобы снизить негативный эффект влияния длительности загрузки данных в регистры ячеек на общее время решения задачи.

Модификация схемы ячейки

Традиционно все внутренние регистры ячеек ОВС для хранения начальной и текущей информации выполняются в виде сдвиговых регистров параллельно-последовательного действия. Параллельные выходы регистров используются масштабными множителями в цикле решения задачи. Через последовательные входы и выходы все регистры соединены в цепь, первый вход и последний выход которой соединены с выводами ПЛИС. Через эти выводы и каналы связи внешнее устройство управления записывает коэффициенты дифференциального уравнения и считывает результат его решения [2]. Это известное применение сдвиговых регистров для преобразования последовательного кода в параллельный код и наоборот широко используется при передаче информации на большие расстояния, при записи информации на внешние носители, при работе с мониторами, видеокамерами и пр.

В одной микросхеме ПЛИС можно разместить несколько ячеек ОВС. Но для решения реальных задач требуются структуры с десятками тысяч таких ячеек. Поэтому необходимо комплексирование таких микросхем на платах, в блоках, в стойках. Количество выводов каждой микросхемы ограничено. Поэтому, с одной стороны, упомянутое выше экономное аппаратное объединение регистров в цепь позволяет существенно уменьшить количество выводов микросхемы, занятых на подготовительных операциях. Это позволяет увеличить соответственно количество выводов микросхемы, предназначенных для обмена данными между ячейками в цикле решения задачи и уменьшить время его выполнения. Но, с другой стороны, при переходе от параллельного кода к последовательному коду увеличивается время передачи данных. При этом циклы обмена данными чередуются с циклами решения уравнения [3]. Поэтому требуется дополнительное исследование влияния времени загрузки на общее время решения задачи для разных схем реализации ячейки.

Сочетание параллельного и последовательного кода можно применить ко всей ОВС либо к ее части, используя те или иные схемы коммутации. Все варианты соединения также требуют дополнительного исследования их влияния на общее время решения задачи. Кроме того, в некоторых случаях (при слишком большой длине цепи регистров) возрастает вероятность ошибок загрузки данных, для ликвидации которых может потребоваться многократная перезагрузка данных в регистры ОВС. В случае однородной вычислительной структуры, состоящей из сотен тысяч ячеек, это приведет к заметному увеличению общего времени решения.

Для снижения влияния последнего фактора предлагается применить техническое решение, хо-

рошо себя зарекомендовавшее в течение последних десятков лет при передаче данных в локальных компьютерных сетях и известное под названием *Ethernet*.

Это решение было предложено Р. Меткалфом и Д. Боггсом в 1975 г. [10]. В данном случае предлагается использовать не сам протокол *Ethernet* в полном объеме, а только принцип разбиения всей передаваемой информации на пакеты с обязательным заголовком в каждом пакете. В заголовке должен быть (кроме прочего) адрес ячейки-получателя и контрольная сумма. Таким образом, можно почти полностью исключить ошибки передачи данных, связанных с неправильной адресацией или сбоем при передаче данных.

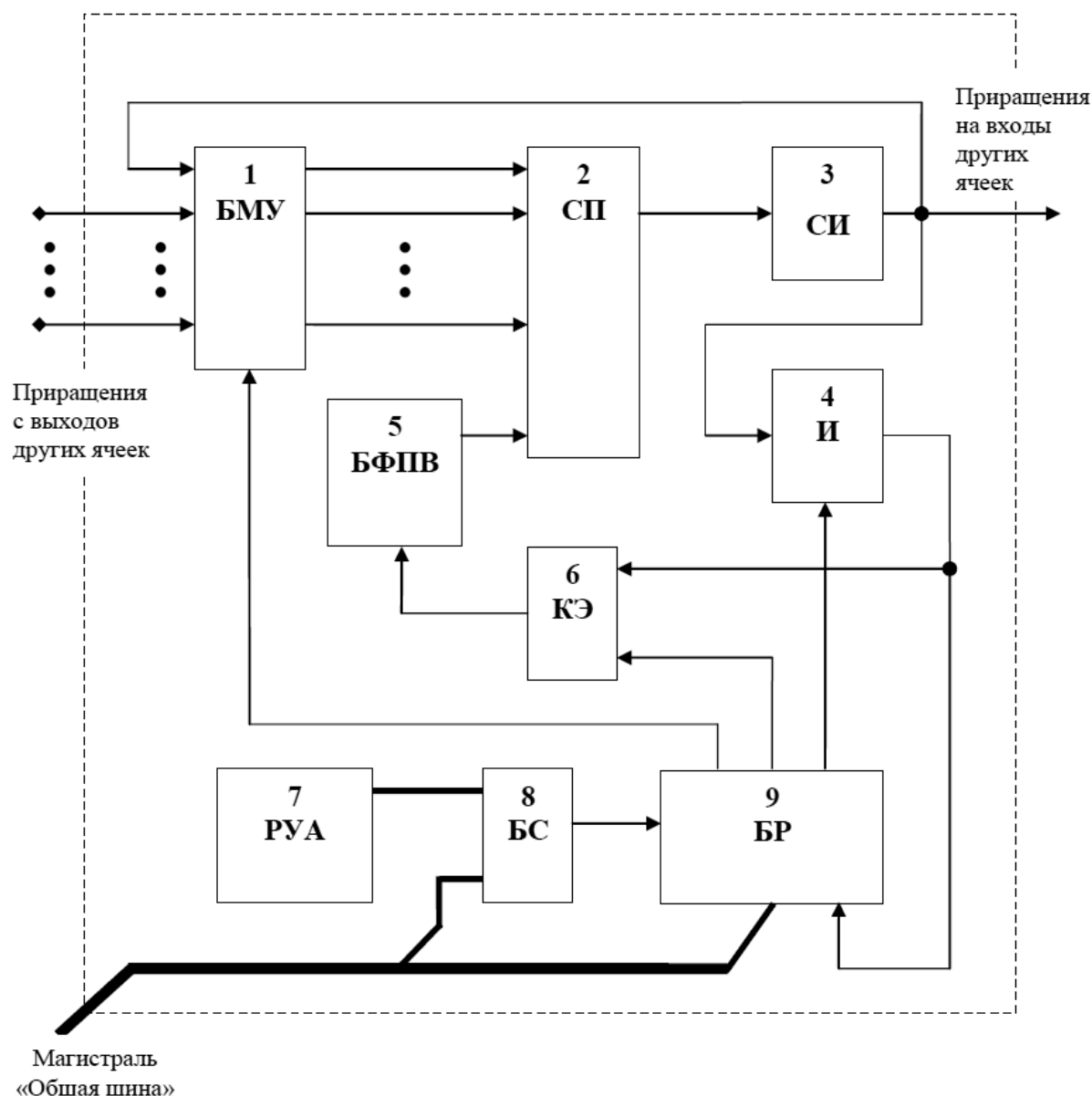


Рисунок. Функциональная схема адресуемой ячейки однородной структуры для решения дифференциальных уравнений в частных производных: 1) блок масштабных умножителей – БМУ; 2) сумматор приращений – СП; 3) следящий интегратор – СИ; 4) интегратор – И; 5) блок формирования приращений с предыдущих шагов по времени – БФПВ; 6) ключевые элементы – КЭ; 7) регистр уникального адреса – РУА; 8) блок совпадения – БС; 9) блок распределения – БР

Реализация этого принципа потребует модификации схемы ячейки, которую в отличие от известной схемы можно считать адресуемой ячейкой. Эта модификация связана с введением дополнительных аппаратных элементов в схему ячейки, что приводит к увеличению ее стоимости и размеров. Но это увеличение может быть оправдано при создании крупных однородных вычислительных структур, в которых время загрузки данных в регистры становится значительным и снижает реальную производительность всей ОВС. На схему адресуемой ячейки, представленной на рисунке, получен патент РФ [11].

Первые шесть элементов адресуемой ячейки полностью совпадают с известной схемой [3]. Новыми элементами ячейки являются: регистр уникального адреса РУА, блок совпадения БС и блок распределения БР.

Адресуемая ячейка, также как и не адресуемая, может находиться в двух состояниях, задаваемых внешним устройством управления. Первое состояние – это рабочий цикл вычислений (цикл решения уравнения), во время которого сигналы с выходов других ячеек (приращения) умножаются на коэффициенты в БМУ, суммируются в сумматоре СП, передаются в интегратор и при этом формируются выходные сигналы (приращения) для соседних ячеек. По сигналу окончания шага по времени от внешнего устройства управления ячейка переходит во второе состояние – цикл чтения/записи (цикл обмена данными). Во время этого цикла результат интегрирования передается в блок формирования приращений с предыдущих шагов по времени БФПВ и используется уже в следующем рабочем цикле вычислений ячейки, а также передается во внешнее устройство управления. Если необходимо (из постановки решаемой задачи), то во время этого цикла чтения/записи в БМУ записываются новые значения коэффициентов, полученные от внешнего устройства управления.

Все сдвиговые регистры, входящие в состав БМУ и других блоков, объединены в 1 или n цепей по последовательным входам-выходам. Это позволяет занимать под обмен данными с внешним устройством управления 2 или $2n$ выводов микросхемы. При сбое или сдвиге хотя бы на 1 бит последовательного кода все коэффициенты этой цепи будут неправильными, и возникнет ошибка в решении задачи, которую можно выявить только после следующего цикла обмена данными. Эту ошибку необходимо исправлять и загружать данные в регистры заново. Чтобы исключить эту потерю времени в каждую ячейку добавлен регистр уникального адреса РУА, который может быть закодирован при изготовлении микросхемы, и блок совпадения БС, который выявляет, совпадает ли значение, хранящееся в РУА, с адресом заголовка очередного пакета данных, поступающего в ячейку через магистраль «Общая шина». Если адреса совпадают, то блок БС дает разрешающий сигнал блоку распределения БР, который также подключен к ма-

гистрала. Блок БР распределяет данные, находящиеся в идентифицированном пакете данных по блокам ячейки: на ключевой элемент КЭ, в сдвиговые регистры блока масштабных умножителей БМУ и интегратора И. Временная диаграмма работы ячейки, описанная в работе [3], при этом не изменяется.

Снижение времени загрузки данных в ячейку достигается за счет того, что из магистрали «Общая шина» данные в параллельном коде поступают в блок распределения БР, который затем передает их в последовательном коде в цепь (цепи) сдвиговых регистров ячейки. Во время последней передачи данных внутри ячейки магистраль начинает обслуживать уже следующие ячейки.

Также обеспечивается и снижение потерь времени из-за необходимости повторять циклы чтения/записи в результате ошибок, связанных с отсутствием адресации при передаче данных от внешнего устройства в ячейку однородной вычислительной структуры.

Исследование влияния длительности загрузки на общее время решение задачи

При решении краевых задач с постоянными коэффициентами загрузка данных в ячейку производится один раз в начале решения и общее время загрузки не зависит от количества итераций. Для задач с переменными коэффициентами загрузку новых коэффициентов необходимо делать после каждой итерации (группы итераций). В этом случае снижение времени загрузки данных в ячейку особенно важно. Модификация схемы ячейки, описанная выше, позволяет уменьшить длительность цикла обмена данными за счет того, что коэффициенты в каждую ячейку передаются по магистрали «Общая шина» параллельно. Для получения количественной оценки влияния цикла загрузки данных на общее время решения задачи применялось математическое моделирование решения краевых задач на ОВС.

Исследования проводились на математических моделях, представленных в работах [1–3]. Результаты моделирования в работе [1] приведены в виде таблицы, в которой показано количество итераций, полученных при решении задачи на модели ОВС, состоящей из 9 внутренних и двух граничных ячеек, при различной разрядности ячеек. При расчете общего времени решения на ОВС считалось, что на каждый такт цикла вычислений приходится один такт цикла обмена данными с внешним устройством, как это принято для не адресуемых ячеек.

Для сравнения времени решения задач на адресуемых и не адресуемых ячейках эта таблица была дополнена новыми результатами моделирования. При моделировании использовались уравнения, полностью идентичные описанным в работе [1] (задача с постоянными коэффициентами) и описанным в работе [3] (задача с переменными коэффициентами) при прочих равных условиях, указанных

в работе [1]. В таблице приведены результаты моделирования решения тестовой краевой задачи на однородной вычислительной структуре, состоящей из двух граничных ячеек и 9 внутренних (адресуемых и не адресуемых), при различной разрядности ячеек.

Таблица. Сравнение времени решения тестовой краевой задачи с постоянными и переменными коэффициентами на модели однородной вычислительной структуры

Разрядность ячейки	Число тактов в одной итерации	Количество итераций	Время решения задачи с постоянными/переменными коэффициентами, мс	
			Не адресуемые ячейки	Адресуемые ячейки
Заданная точность 0,01				
8	28	*	–	–
10	210	*	–	–
12	212	148	0,61/18,37	0,31/15,11
14	214	112	1,84/17,52	0,93/12,13
16	216	108	7,08/24,36	3,55/14,35
18	218	106	27,79/46,87	13,90/24,50
20	220	106	111,15/132,35	55,58/66,18
Заданная точность 0,1				
8	28	*	–	–
10	210	70	0,07/7,07	0,05/7,05
12	212	57	0,23/6,07	0,13/5,83
14	214	56	0,92/8,76	0,47/6,07
16	216	60	3,93/13,53	1,98/7,98
18	218	60	15,73/26,53	7,87/13,87
20	220	60	62,91/74,91	31,47/37,47

*Заданная точность решения недостижима.

Как видно из таблицы, при увеличении разрядности ячейки количество итераций решения краевой задачи стабилизируется. Время решения задачи с постоянными коэффициентами на адресуемых и не адресуемых ячейках растет пропорционально разрядности ячейки. Время решения задачи с переменными коэффициентами при небольшой разрядности ячеек существенно возрастает по

сравнению со временем решения задачи с постоянными коэффициентами, но при увеличении разрядности они становятся сопоставимыми. Это можно объяснить тем, что при увеличении разрядности ячеек длительность цикла вычислений существенно увеличивается, а длительность цикла обмена данными с ячейкой остается прежней.

Во всех сопоставимых примерах расчетов, приведенных в таблице, модель ОВС на адресуемых ячейках показала меньшее время решения задачи, чем модель однородной вычислительной структуры на не адресуемых ячейках.

Есть основания предположить, что при увеличении количества ячеек в ОВС влияние длительности загрузки на общее время решения задачи будет снижаться, поскольку длительность цикла вычисления увеличивается более значительно, чем длительность цикла обмена данными.

Выводы

1. Предложена модифицированная функциональная схема адресуемой ячейки однородной вычислительной структуры для решения дифференциальных уравнений в частных производных.
2. Модификация заключается в том, что данные из магистрали «Общая шина» в ячейку поступают в параллельном коде и при их распределении по сдвиговым регистрам внутри ячейки магистраль освобождается для обслуживания других ячеек.
3. Применение схемы адресуемой ячейки ускоряет решение дифференциальных уравнений в частных производных за счет снижения длительности цикла загрузки данных, особенно в задачах с переменными коэффициентами.
4. Результаты моделирования на примере решения тестовой краевой задачи подтвердили работоспособность и эффективность модифицированной ячейки.

Представленная работа выполнена в рамках государственного задания «Наука».

СПИСОК ЛИТЕРАТУРЫ

1. Хамухин А.А. Ячеечная модель устройства для решения дифференциальных уравнений в частных производных // Известия Томского политехнического университета. – 2010. – Т. 316. – № 5. – С. 62–67.
2. Хамухин А.А. Реконфигурирование однородной вычислительной структуры с непрограммируемыми ячейками для решения дифференциальных уравнений в частных производных // Известия Томского политехнического университета. – 2010. – Т. 316. – № 5. – С. 68–72.
3. Хамухин А.А. Модификация ячейки однородной структуры для решения дифференциальных уравнений в частных производных с переменными коэффициентами // Известия Томского политехнического университета. – 2011. – Т. 318. – № 5. – С. 37–41.
4. Ячейка однородной структуры для решения дифференциальных уравнений в частных производных: пат. 2359322 Рос. Федерация. № 2007141832/09; заявл. 12.11.07; опубл. 20.06.09, Бюл. № 17. – 6 с.
5. Ячейка однородной структуры для решения дифференциальных уравнений в частных производных с переменными коэффициентами: пат. 2419141 Рос. Федерация. № 2009128437/08; заявл. 22.07.09; опубл. 27.01.11, Бюл. № 14. – 7 с.
6. Устройство для вычисления дискретизированного непрерывного вейвлет-преобразования: пат. 2437147 Рос. Федерация. № 2010127068/08; заявл. 01.07.2010; опубл. 20.12.2011, Бюл. № 35. – 9 с.
7. Каляев И.А., Левин И.И. Реконфигурируемые мультиконвейерные вычислительные системы для решения потоковых задач обработки информации и управления // Параллельные вычисления и задачи управления (РАСО'2010): Матер. V Международ. конф. – М., 26–28 октября 2010. – М., 2010. – С. 5–19.
8. Суперкомпьютер Cray XT5h // Лаборатория Параллельных информационных технологий НИВЦ МГУ. 2009. URL: <http://parallel.ru/computers/reviews/CrayXT5h.html> (дата обращения: 26.04.2011).
9. Левин И.И. Реконфигурируемые вычислительные системы с открытой масштабируемой архитектурой // Параллельные

- вычисления и задачи управления (РАСО'2010): Матер. V Междунар. конф. – М., 26–28 октября 2010. – М., 2010. – С. 83–95.
10. Metcalfe R.M., Boggs D.R. Ethernet: Distributed Packet Switching for Local Computer Networks. 1975. URL: <http://ethernethistory.typepad.com/papers/EthernetPaper.pdf> (дата обращения: 11.01.2011).
11. Адресуемая ячейка однородной структуры для решения дифференциальных уравнений в частных производных: пат. 2427033 Рос. Федерация. № 2010107933/08; заявл. 03.03.10; опубл. 20.08.11, Бюл. № 23. – 7 с.

Поступила 02.05.2012 г.

УДК 004.42

АРХИТЕКТУРА РАСПРЕДЕЛЕННОГО ВЫЧИСЛИТЕЛЬНОГО КОМПЛЕКСА ДЛЯ ДВУМЕРНОГО АНАЛИЗА ИЗОБРАЖЕНИЙ ДИСКОВ ДЕРЕВЬЕВ

И.А. Ботыгин, В.Н. Попов, В.А. Тартаковский*

Томский политехнический университет

*Институт мониторинга климатических и экологических систем СО РАН, г. Томск

E-mail: botygin@ad.cctpu.edu.ru

Разработана архитектура и алгоритмы реализации распределенного вычислительного комплекса обработки дендрологических данных. Представлена GPSS-модель комплекса для оценки эффективности его функционирования. Практическая реализация комплекса осуществлялась с использованием стека свободно распространяемых программных продуктов. Проиллюстрирована работа комплекса на таких задачах двумерного анализа изображений дисков деревьев, как вычисление азимута и среднеквадратичной ширины области максимального прироста, а также вычисление значений индексов прироста ширины годичных колец деревьев.

Ключевые слова:

Дендрология, дендрохронология, древесные спилы, годовые кольца деревьев, математическая модель, серверное приложение.

Key words:

Dendroecology, dendrochronology, tree stem disk, tree-rings, mathematical model, server application.

Программное обеспечение в области дендрохронологии

Проведенный обзор и сравнительный анализ существующих аппаратно-программных комплексов и систем для анализа и обработки данных в области дендрохронологии показал, что они являются ограниченными для использования и ориентированы на решение новых задач, связанных с неоднородностью окружающего пространства. Особенностью таких задач является большой объем данных, которые необходимо хранить, обрабатывать и сопоставлять между собой, пространственная распределенность мест сбора образцов.

В таких условиях оптимальным системным решением, обеспечивающим повышение вычислительной мощности, увеличение объема хранимых данных, является использование технологии распределенной их обработки. Эта технология подразумевает физическое распределение хранения и обработки данных в пространстве на нескольких вычислительных машинах, которые связаны между собой каналами передачи данных, координацию их вычислительных мощностей, использование стандартных протоколов и служб сетевого взаимодействия. В данное время существует множество инструментальных средств технологий распределенных вычислений, а также проектов, реализованных с их использованием, но для мониторинга

климатических и экологических изменений на основе биоиндикации такие технологии не применялись. Таким образом, реализация технологии распределенных вычислений и разработка алгоритмического обеспечения для дендрохронологических исследований, связанных с двумерным анализом изображений спилов деревьев, даст возможность получать новые результаты и качественные оценки параметров окружающей среды.

В настоящей работе описывается реализация технологии распределенных вычислений и разработка алгоритмического обеспечения для дендрохронологических исследований. Безусловно, дендрохронологические исследования – это только часть мониторинга, моделирования и прогнозирования климатических и экосистемных изменений под воздействием природных и антропогенных факторов. Но особенность задач дендрологического анализа, заключающаяся в необходимости математической обработки очень большого объема данных (временные ряды наблюдений могут достигать сотен гигабайт), широком спектре решаемых задач, коллективной работе многих сотрудников на всех этапах дендрологических исследований, а также в необходимости хранения и систематизации больших объемов неоднородной структурированной информации (собственно хронологические ряды наблюдений, результаты обработки,

сопутствующие метеорологические, геологические, геофизические, аэрокосмические и т. п. ряды наблюдений), однозначно подразумевает в качестве одного из системных решений использование технологии распределенной обработки (*grid*-технологии), обеспечивающей динамическое изменение основных компонентов инфраструктуры системы обработки (от структур хранимых данных — до схем и алгоритмов решаемых задач) и повышение вычислительной мощности.

Сравнительный обзор и анализ аппаратно-программных средств обработки и анализа годичных колец деревьев

Анализ научных исследований в области дендрологии, дендрохронологии и дендроклиматологии показывает, что в настоящее время для моделирования и анализа дендрологических данных используется достаточно широкий спектр аппаратно-программных средств. Использование программных средств зависит от задач исследования. Это может быть статистическая обработка первичных данных замеров характеристик прироста и получение надёжных обобщённых хронологий, сопоставление характеристик прироста с факторами внешней среды, моделирование процессов роста, изучение пространственного распределения характеристик прироста и визуализация результатов анализа. Для решения этих задач одни исследователи предпочитают использовать широко распространённые статистические пакеты (SAS, MatLab, SyStat, STATISTICA и др.), электронные таблицы (QUATTRO, LOTUS, Excel и др.), универсальные системы математической обработки результатов, предназначенные для численного и символьного решения математических задач различной сложности (MathCAD, Mathematica и др.) и интегрированные программные решения, такие как DPL, PRECON 5.17C, TREERING 3.0, LignoVision 1.32, TSAP-Win Professional 0.30, DendroClim 2002, WinDENDRO, DendroLab 470, PAST 32, OSM 3.10, MeasureJ2X, или решения на базе географических информационных систем (ESRI ARC/INFO, ArcView, Mapinfo, AutoCAD Map и др.) с форматами различных баз данных (Microsoft Access, Oracle, dBASE, FoxPro и др.) для связи обрабатываемых данных с географическим местоположением их сбора и отображением на электронных картах. В настоящее время широко распространено, в том числе и в России, оборудование и программное обеспечение для диагностики, контроля и исследования внутреннего состояния деревьев и древесины, разработанное компанией RINNTech [1].

Рассмотренные программные комплексы и системы достаточно универсальны, ориентированы на широкий класс исследований и выполнение часто используемых операций обработки и анализа дендрологических данных (например, измерение ширины годичных колец и др.). Современные

инструментальные средства автоматизируют многие этапы получения, сбора и обработки дендрологических данных, но и одновременно создают большие потоки информации, требующие не только обширных баз для их хранения, но и серьезных математических методов их обработки.

GPSS-модель для оценки эффективности функционирования распределенного вычислительного комплекса обработки дендрологических данных

Исследование задач обработки дендрологических данных и инструментально-программных средств их решения выявили трудоёмкость и сложность создания соответствующего программного обеспечения. Для успешного проектирования и реализации программного обеспечения обработки дендрологических данных должны быть построены его полные и непротиворечивые как функциональные, так и информационные модели, структура основных компонентов программного комплекса и алгоритмы их функционирования.

Для оценки эффективности функционирования распределенного вычислительного комплекса произвольной структуры была разработана его имитационная модель на языке GPSS World [2].

На рис. 1 отражена укрупненная модель распределенного вычислительного комплекса обработки дендрологических данных.

Основными объектами моделируемой системы являются: потоки входных заданий, вычислительные сегменты, вычислительные серверы (ВС) сегментов, коммуникационный сервер (КС), серверы баз данных (БД), менеджер БД. Вычисление моментов появления заданий (случайной величины, связанной с промежутком времени между появлениями двух соседних заданий) осуществляется по нормальному закону распределения. Интервалы обслуживания также являются случайной величиной.

Изменяемыми параметрами модели являются (всего 18): количество ВС, максимальное время обработки задания, количество типов заданий, средний интервал между заданиями, доля параллельных заданий, доля отклонённых заданий, среднее время обработки заданий КС, отклонение от среднего времени обработки заданий КС, вероятность сбоя ВС, вероятность восстановления ВС, среднее время восстановления ВС, отклонение от среднего времени восстановления ВС, количество КС, максимальное время обработки задания администратора, количество типов заданий, средний интервал между заданиями, среднее время восстановления ВС, отклонение от среднего времени восстановления ВС. Типовыми заданиями в настоящей модели являлись задачи вычисления азимута и среднеквадратичной ширины области максимального прироста годичных колец деревьев, вычисления значений индексов прироста ширины годичных колец деревьев и т. д.

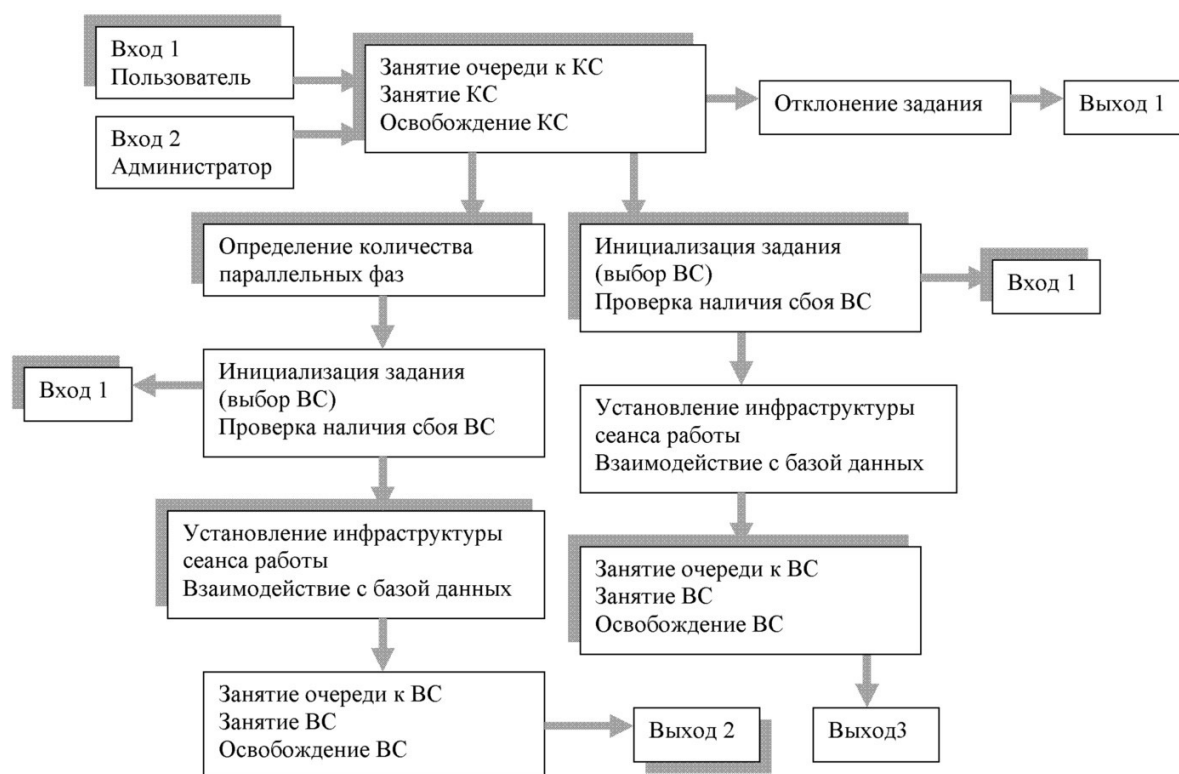


Рис. 1. GPSS-модель для оценки эффективности функционирования распределенного вычислительного комплекса обработки дендрэкологических данных

В работе проведено исследование зависимостей таких критериев как (всего 9): количество заданий общее, отклонённых, без параллелизма, с параллелизмом, аварийных ситуаций с восстановлением, без восстановления, с восстановлением (для заданий с параллелизмом), без восстановления (для заданий с параллелизмом), заданий в очереди, от следующих параметров (всего 5): количество ВС, максимальное время обработки задания, количество типов заданий, интервал между заданиями, количество КС (рис. 2).

По оси Y приведены полученные значения исследуемых критериев – абсолютные численные значения количества, а по оси X – абсолютные численные значения соответствующих изменяемых параметров.

Во всех экспериментах моделирование осуществлялось в течении 24 ч (модельное время).

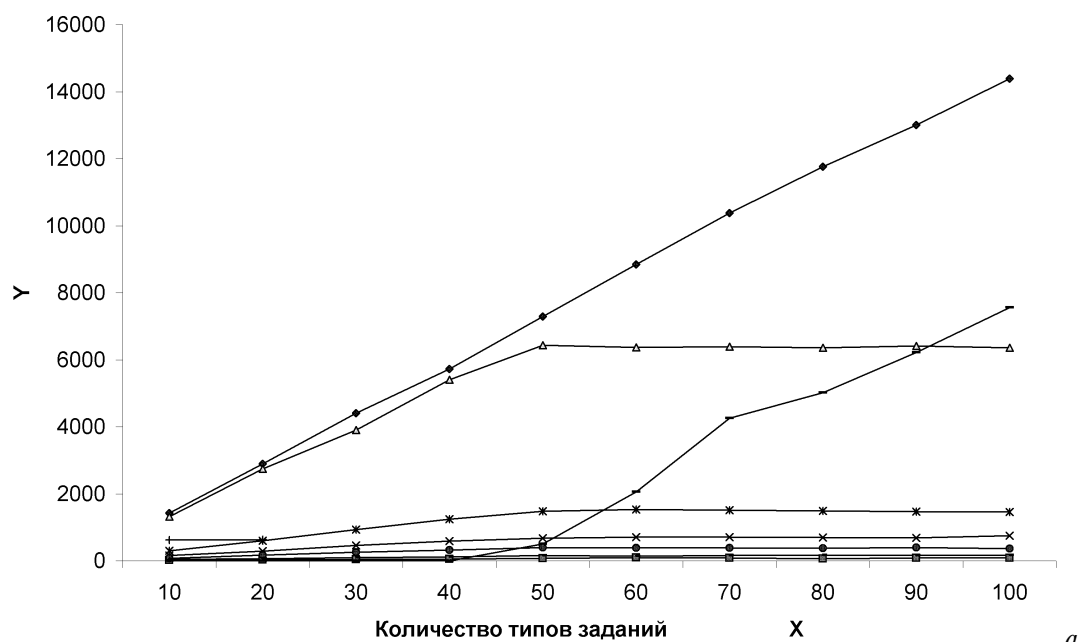
При выбранных в модели параметрах быстрого действия ВС: их количество (свыше 100), максимального времени обработки задания (до 100 единиц), количество КС (до 10) практически не влияет на исследуемые критерии. Количество типов выполняемых заданий значительно влияет на число выполненных заданий, а также на очередь к КС. Изменение интервала поступления заданий значительно влияет на исследуемые критерии, особенно на начальных этапах его увеличения.

Распределенный вычислительный комплекс обработки дендрэкологических данных

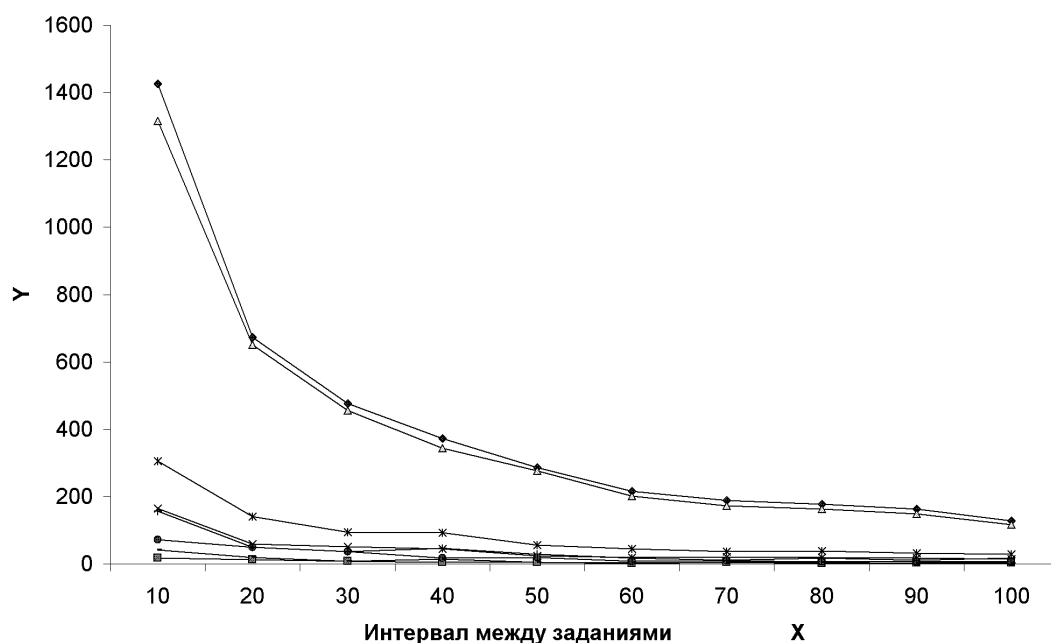
На основе результатов моделирования был спроектирован распределенный вычислительный комплекс обработки дендрэкологических данных (рис. 3). Архитектура комплекса базируется на компонентах, используемых при построении современных инструментальных средств распределенных вычислений (*grid*-систем). Но внутренняя структура комплекса, содержание функциональных компонентов (*middleware*) и протоколы их взаимодействия являются оригинальными. Архитектура разработанного комплекса виртуализирует три основных технических ресурса, из которых строится высокопроизводительный центр обработки данных (вычислительные системы, системы хранения данных и глобальные коммуникации), а затем собирает их в единый виртуальный компьютер, чтобы предоставлять его ресурсы в виде сервисов пользователям центра.

В структуре распределенного комплекса обработки дендрэкологических данных выделены вычислительные серверы, коммуникационный сервер и сервер базы данных.

Планирование и диспетчирование процессами обработки дендрэкологических данных возложено на специальный коммуникационный сервер. Взаимодействие всех пользователей с комплексом



a



б

- ◆— Количество заданий вообще
- Количество отклоненных заданий
- △— Количество заданий без параллелизма
- ×— Количество заданий с параллелизмом
- *— Количество аварийных ситуаций с восстановлением
- Количество аварийных ситуаций без восстановления
- +— Количество аварийных ситуаций с восстановлением (для заданий с параллелизмом)
- |— Количество аварийных ситуаций без восстановления (для заданий с параллелизмом)
- |— Количество заданий в очереди

Рис. 2. Результаты моделирования распределенного вычислительного комплекса обработки дендрозкологических данных зависимости критериев от: а) количества типов заданий; б) интервала между заданиями

обработки осуществляется только через коммуникационный сервер. Основной задачей коммуникационного сервера является обеспечение оптимальной загрузки имеющихся в его распоряжении вычислительных серверов и обеспечение режима работы в реальном времени (*online*) пользователей. В такой ситуации пользователю уже не важно, на каком конкретном узле сети исполняется его задача; он просто потребляет определенное количество виртуальной процессорной мощности, имеющейся в сети.

На вычислительных серверах комплекса производится математическая обработка данных.

Для полноценного функционирования комплекса анализа дендрозоологических данных организован автоматизированный сбор, систематизация и хранение научной информации в области дендрозоологического мониторинга, а также формирование и ведение базы дендрозоологических данных. Выполнение этих задач возложено на менеджера данных, управляющий работой распределенных серверов баз данных.

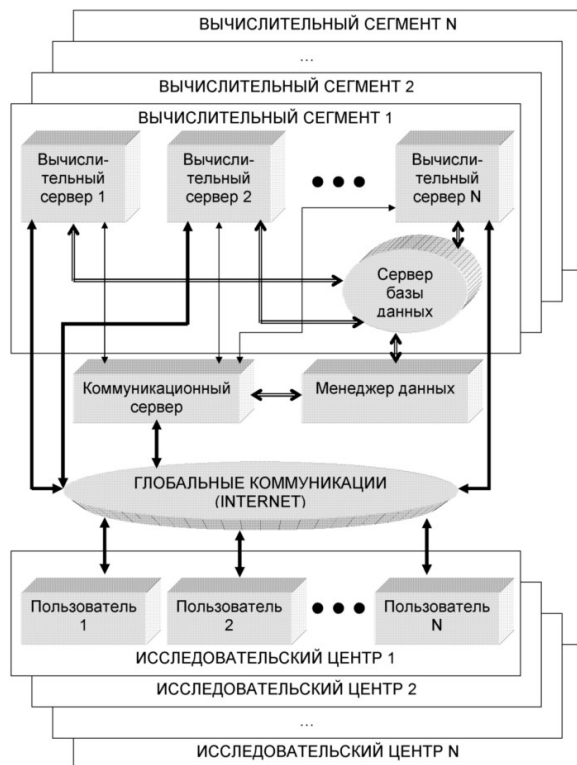


Рис. 3. Распределенный вычислительный комплекс обработки дендрозоологических данных

На рис. 4 представлена схема разработанного алгоритма выбора вычислительного сервера коммуникационным сервером.

Алгоритм использует информацию, которая хранится в профилях (метаописаниях) вычислительных серверов (таблица процессоров, таблица соответствия задач и таблица задач).

Основные шаги алгоритма:

Шаг 1. Формирование списка вычислительных серверов для решения задачи.

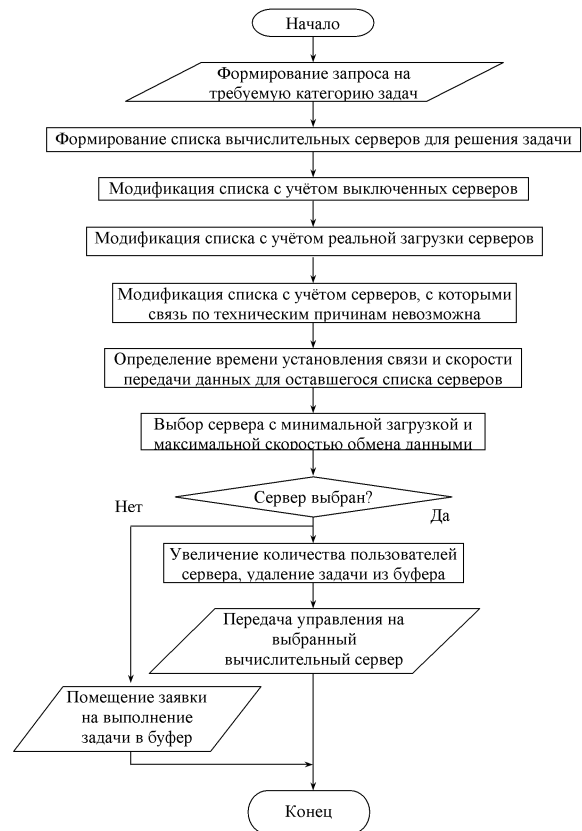


Рис. 4. Алгоритм функционирования коммуникационного сервера

Шаг 2. Исключение из списка выключенных серверов.

Шаг 3. Исключение из списка занятых серверов.

Шаг 4. Исключение из списка серверов, с которыми связь по техническим причинам невозможна.

Шаг 5. Определение времени установления связи и скорость передачи данных с серверами.

Шаг 6. Определение сервера с максимальной свободной вычислительной мощностью.

Шаг 7. При наличии нескольких серверов с одинаковой вычислительной мощностью выбирается тот сервер, с которым скорость обмена данными выше.

Шаг 8. Увеличение на единицу значения поля Количество пользователей.

Шаг 9. При отсутствии серверов со свободными вычислительными мощностями заявка на выполнение задачи помещается в буфер коммуникационного сервера.

Шаг 10. Передача управления на выбранный вычислительный сервер.

Критерий выбора заключается в поиске такого сервера, нагрузка на который минимальна, а скорость обмена данными – максимальна.

На рис. 5 представлена схема разработанного алгоритма менеджера данных (СУБД – система управления базой данных).



Рис. 5. Алгоритм функционирования менеджера данных

На менеджер данных (рис. 6) возложена работа по взаимодействию с серверами баз данных. Именно менеджер данных координирует распределение и использование информации, находящейся в локальных базах данных и тем самым виртуализирует накопители данных, объединяя их в единый логический информационный ресурс. Функционирование менеджера данных основывается на информации, формируемой в системной базе данных в таблицах серверов баз данных, таблицах соответствия классов информации и описания классов информации и СУБД.

Для практической реализации комплекса в настоящей работе использовался стек программных продуктов LAMP – набор свободно распространяемого инструментария: HTTP сервер Apache 1.3.14, SQL СУБД MySQL 3.22.21, язык сценариев PHP 4.2, а также сервер баз данных Oracle 8i для постоянного хранения информации. Отметим, что *grid*-системой, реализованной с использованием стека LAMP, является, например, компонент Grid-Premis французского *grid*-проекта Grid'5000.

Таким образом, с использованием разработанного алгоритма двумерного анализа изображений дисков деревьев созданный распределенный вычислительный комплекс обработки дендрологических данных способен решать такие задачи, как вычисление азимута и среднеквадратичной ширины области максимального прироста, а также вычисление значений индексов прироста ширины годичных колец деревьев. Также, с помощью разработанных программных средств математического анализа годичных колец деревьев появилась возможность выявления изменений параметров окружающей среды, отраженных в приросте дерева,

и решения задач дендрологической диагностики с использованием дополнительной картографической и метеорологической информации. Более подробная информация о решенных с помощью комплекса задач двумерного анализа изображений дисков деревьев представлена в работах [3, 4].

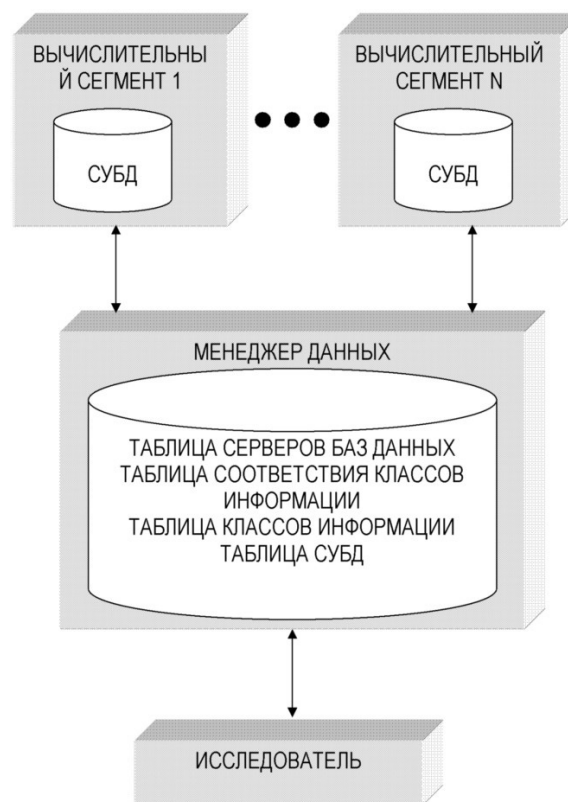


Рис. 6. Менеджер данных

Выводы

1. Разработаны алгоритмы и программные средства обработки дендрологических данных, которые могут быть применены в мониторинге экологического состояния окружающей среды.
2. Разработанная и постоянно пополняемая база данных изображений годичных колец деревьев обеспечивает использование полученной невозобновимой информации в будущих работах на основе новых методов и разработок.
3. Предложенная функциональная структура распределенного вычислительного комплекса может быть использована для решения широкого круга задач, требующих значительных вычислительных мощностей.
4. Использование разработанного распределенного вычислительного комплекса для решения задач дендрологического анализа позволяет упростить их технологические аспекты решения и значительно увеличить мощность вычислительного комплекса.
5. Созданные программные средства могут быть использованы в качестве лабораторной базы

в различных учебных дисциплинах, связанных с охраной окружающей среды и экологией человека и рекомендованы к внедрению в лесохозяйственных организациях и при подготовке специалистов в области лесоведения в университетах России.

зайственных организациях и при подготовке специалистов в области лесоведения в университетах России.

СПИСОК ЛИТЕРАТУРЫ

1. RINNTech // Products and services for tree and wood analysis. 2010. URL: <http://www.rinntech.de/> (дата обращения: 27.08.2010).
2. GPSS // Имитационное моделирование систем. 2010. URL: <http://www.gpss.ru/> (дата обращения: 27.08.2010).
3. Ботыгин И.А., Попов В.Н., Тартаковский В.А. Математические модели в задачах обработки дендроэкологических данных. Ч. I // Известия Томского политехнического университета. – 2011. – Т. 319. – № 5. – С. 118–122.
4. Ботыгин И.А., Попов В.Н., Тартаковский В.А. Математические модели в задачах обработки дендроэкологических данных. Ч. II // Известия Томского политехнического университета. – 2011. – Т. 319. – № 5. – С. 123–125.

Поступила 29.06.2011 г.

УДК 004.4:004.89

ИСПОЛЬЗОВАНИЕ ОНТОЛОГИИ В ЭЛЕКТРОННЫХ БИБЛИОТЕКАХ

Ле Хоай, А.Ф. Тузовский

Томский политехнический университет
E-mail: lehotomsk@yahoo.com

Рассматривается использование онтологий в семантических электронных библиотеках, дается их определение и назначение. Анализируются виды онтологий таких библиотек, в том числе системы организации знаний и структурная таксономия. Обосновывается вариант набора онтологий для разработки семантических электронных библиотек.

Ключевые слова:

Онтология, контрольный словарь электронных ресурсов, семантические технологии, электронная библиотека, семантическая электронная библиотека.

Key words:

Ontology, glossary of electronic resources, semantic technology, electronic library, semantic digital library.

Под электронными библиотеками (ЭБ) понимаются информационные системы, позволяющие автоматизировать работу пользователей с электронными ресурсами (ЭР), такими, как документы, изображения, аудио-, и видеофайлы и т. д. С появлением семантических технологий (СТ), предоставляющих средства работы с семантикой документов, возникла возможность разработки подходов к автоматизации работы с этими ресурсами на новом уровне. Разработка семантических электронных библиотек (СЭБ) представляет собой решение комплекса задач, целью которых являются повышение возможностей взаимодействия с пользователями и расширение функциональности ЭБ, особенно в поиске данных. Многие ЭР содержат метаданные, в том числе документы включают данные об авторе, издании, дате создания и т. д. Такие метаданные часто хранятся в виде XML файлов, позволяющих выполнять универсальное описание ЭР. Язык XML позволяет описывать только структуру объектов, а не их семантику. В свою очередь, СТ представляют возможность описывать семантику ЭР (аннотировать их) и выполнять программную обработку таких метаданных.

Для описания семантики метаданных необходимо использовать онтологические модели (онто-

логии), определяющие наборы понятий предметной области и их взаимосвязи. Для работы с онтологиями в СТ имеются специальные языки для описания семантики (OWL, RDFS, RDF) [1–3] и запросов к семантическим данным (SPARQL) [4], а также набор инструментов редактирования и работы с семантическими данными (Protégé, Sesame, Jena и т. д.) [5–7].

При разработке СЭБ с использованием СТ одной из наиболее важных и сложных задач является разработка онтологий, описывающих области знаний, с которыми связано функционирование ЭБ и содержание имеющихся в них ресурсов. В данной статье рассматриваются особенности и проблемы, связанные с задачей построения онтологий для СЭБ, и обосновывается базовый набор таких онтологий.

1. Использование онтологии в СЭБ

В области искусственного интеллекта под онтологией понимается специальная система понятий и взаимосвязей между ними, описывающая определенную предметную область. Содержание понятий определяется с помощью концептов. Формально в онтологии концепт отождествляется с объектом (классом), имеющим связи с другими класса-

ми. Класс определяется как множество экземпляров с общими свойствами и содержит описания собственно экземпляров и их свойств [8].

Для обеспечения возможности программной работы с семантикой был разработан язык RDF для описания высказываний, а также язык OWL для описания схемы отношений между концептами (классами) и их атрибутами (свойствами) [9]. С помощью этих языков и стандартов программа понимает данные и может выполнять логические выводы.

На основе различных метаданных СЭБ интегрирует разнородную информацию, например, ресурсы, пользовательские профили, закладки и таксономии. В связи с этим в СЭБ онтологии играют основную роль для решения проблем, вызванных структурными различиями существующих систем и семантическими различиями стандартов метаданных.

2. Типы онтологий СЭБ

Из анализа ранее выполненных исследовательских проектов по разработке СЭБ (например, JeromeDL [10], BRICKS [11]) можно определить три типа онтологий, используемых в библиотеках: 1) библиографические; 2) структурные; 3) пользовательские.

2.1. Библиографические онтологии

Для хранения и использования информации о ресурсах библиотеки создаются библиографические форматы (стандарты). Примером набора таких стандартов является СИБИБД (система стандартов по информации, библиотечному и издательскому делу), которая основана на аналогических международных стандартах, устанавливающих термины и определения в области информационной деятельности, библиотечного дела и библиографии.

Существует ряд форматов описания библиографической информации, например, MARC (MARC21, RUSMARC, UNIMARC), ONIX XML, BibTEX, Dublin Core, каждый из которых имеет свою область применения. Библиотеки обычно используют форматы семейства MARC; исследователи хранят информацию о публикациях в виде BibTEX; а пользователи Интернет в виде Dublin Core. Данные стандарты определяют набор резервированных полей и позволяют хранить данные в виде обычного текста, но компьютерные программы не могут осуществлять интеллектуальную обработку этих значений. В связи с этим возникает необходимость использования специальных полей с семантическими описаниями. Набор таких полей формирует основу онтологии, позволяющей описывать понятия и отношения предметной области.

Онтология дополняет существующие форматы семантикой и позволяет осуществлять интеграцию данных для них [10].

Онтология MarcOnt была разработана в качестве однородного языка описания библиографиче-

ских данных путем объединения таких существующих онтологий, как BibTeX-RDF, Dublin Core и MarcRDF. Одной из причин разработки MarcOnt [12] явилась сложность языка MARC. Онтология MarcOnt использовалась в качестве базовой в соответствующем проекте MarcOnt, разрабатывающем подход к работе с описаниями метаданных. В проекте был также разработан набор программных средств Mediation Services, предоставляющих возможности выполнения преобразования между разнородными форматами описаний ЭР.

Онтология SWRC (Semantic Web for Research Communities) [13] была разработана институтом AIFB для описания простых библиографических записей и дополнительной информации об исследовательских сообществах. В терминах библиографической выразительности онтология похожа на BibTeX и Dublin Core. В исследовательском проекте Bibster онтология SWRC объединена с онтологическим представлением ACM иерархии тем.

Библиографическая онтология ViVo предоставляет главные концепты и атрибуты для описания цитат и библиографических ссылок. Онтология ViVo использует такие словари (онтологии), как Dublin Core и FOAF. Она позволяет описать ресурсы (документы, книги, статьи), их коллекции и другие объекты. Например, эта онтология включает такое понятие, как Event для описания события, связанного с ресурсами.

Из рассмотрения данной структуры можно отметить, что онтология ViVo является сочетанием структурной и библиографической онтологий, из которых структурная онтология может быть выделена отдельно.

2.2. Структурная онтология

Современные электронные библиотеки хранят не только библиографические метаданные ЭР, но и представление их содержания. В зависимости от их типов, содержание ресурса обычно следует некоторой структуре, которую можно занести в метаданные описания. Например, книги, хранимые в библиотеке, обычно делятся на главы, и для каждой главы можно предоставить индивидуальное описание и хранить информацию об отношениях данной главы с другими главами.

Одним из примеров такой онтологии является структурная онтология библиотеки JeromeDL. Кроме структурного описания ресурсов, в этой онтологии можно подробнее описать статус каждого ресурса, данные о дате поступления в библиотеку и другую информацию, необходимую для управления работой системы с данным ресурсом [14].

2.3. Пользовательская онтология

Данный тип онтологий предназначен для описания профилей пользователей библиотеки, включая их интересы в определенных ресурсах и другую информацию. Существуют онтологии FOAF [15] и FOAFRealm [16], которые позволяют описать профили пользователей. Онтология FOAFRealm

является расширенным вариантом FOAF, с ее помощью можно определить насколько пользователи хорошо знают друг друга, а также она позволяет хранить зашифрованные данные. Такая онтология использовалась в исследовательском проекте JermDL, и для работы с ней был разработан набор программных средств. В этих онтологиях можно добавлять и другие атрибуты для целей решения собственных задач.

Следует отметить, что пользователи библиотеки могут иметь заинтересованность в определенных ресурсах или темах. Подробнее вопрос будет рассмотрен при описании задач разработки СЭБ.

2.4. Другие онтологии в СЭБ

Кроме основных рассмотренных онтологий, в СЭБ используются и другие онтологии, с помощью которых создаются структурные контрольные словари, важные для категоризации библиотечных ресурсов. Структурные контрольные словари представляют собой тезаурусы, схемы классификации, таксономии, выражаемые в виде схемы концептов (SKOS/RDF). Существуют несколько видов контрольных словарей [17]. Чаще всего используемыми словарями для классификации ресурсов являются таксономическая классификация ACM (для компьютерных наук), DDC (десятичная классификация Дьюи) и UDC (универсальная десятичная классификация).

Применение семантических технологий для этих словарей предоставляет возможность аннотировать ресурсы и решить задачи категоризации и категоризации. Существуют варианты кодирования этих словарей в формат SKOS/RDF для применения в СЭБ [18]. Так как эти варианты применяются только для аннотирования ресурсов на английском языке, то при использовании русского языка необходимо изменить их кодирование в формате SKOS/RDF. Кроме этого, требуется внести в них и ряд дополнений. Например, для диссертационных работ необходимо иметь возможность указывать код специальностей ВАК [19]. Для этого необходимо разработать и включить в SKOS/RDF контрольный словарь, в котором каждый код является экземпляром класса концепта и имеет название – этикетку (на русском или другом языке).

В СЭБ базовыми являются онтологии, на основе которых решаются основные задачи их функционирования.

3. Основные задачи функционирования семантических ЭБ

- **Ведение онтологической базы знаний:** задача может быть решена только после создания всех онтологий СЭБ. Для решения задачи требуется разработать программные инструменты для создания базы знаний.
- **Ведение иерархии каталогов:** задача представляет собой создание иерархической структуры

электронных документов с использованием какой-либо категоризации. Для каждой категоризации должна быть разработана возможность представления ЭР в виде иерархии каталогов.

- **Аннотирования ресурсов:** задача аннотирования ЭР должна быть решена путем представления подходящих пользовательских интерфейсов для внесения требуемых данных в описания ЭР.
- **Категоризация:** задача тесно связана с задачей ведения иерархии каталогов, суть задачи заключается в создании специальных словарей для распределения ресурсов по типам, видам и особенностям.
- **Просмотр:** предоставляется возможность удобного перехода между данными электронных документов (автор, издание, ключевые слова и т. д.).
- **Поиск:** это типичная задача каждой системы управления данными. Новые технологии и СТ должны способствовать более точному и полному поиску документов.
- **Формирование профилей пользователей:** задача включает заполнение профилей пользователей данными с использованием удобного пользовательского интерфейса.
- **Рекомендация:** задача решается с использованием учетных данных профилей пользователей. Интересы пользователей являются ключевыми для рекомендации им соответствующих документов.

Выше перечисленные задачи должны учитываться при создании онтологий для СЭБ, а также необходимо использовать некоторые методологии [20–22].

4. Предложенный вариант базовых онтологий СЭБ

4.1. Пользовательская онтология

В качестве онтологии пользователей рекомендуется модифицированная онтология FOAF для описания пользовательских профилей. Схема предложенной онтологии, построенная с помощью редактора Protégé, приведена на рис. 1. Базовым классом онтологии является класс «Агент», одним из подклассов которого – класс «Человек». В связи с этим все связи класса «Агент» с другими классами можно рассматривать как связи класса «Человек».

Полное описание онтологии FOAF приведено в [23], а далее описаны только новые добавленные, выделенные подчеркиванием, классы (концепты) и атрибуты:

- **Домен интереса** «URI: <http://www.w3.org/2004/02/skos/core#Concept>»: включение класса объясняется тем, что в СЭБ должны храниться интересы пользователя для рекомендации ему новых документов. Так как каждый документ в СЭБ относится к какой-либо области и если она интересует пользователя, то все новые поступления документов будут рекомендованы ему, и он будет проинформирован об этих документах.

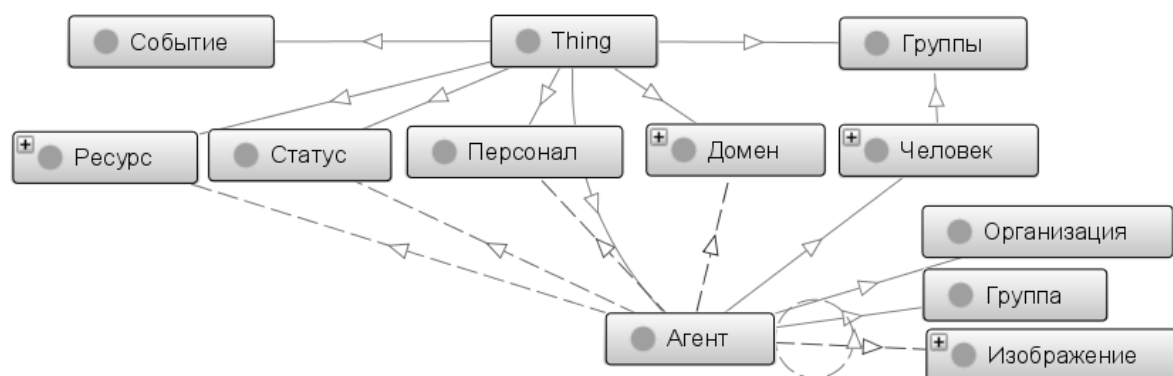


Рис. 1. Схема онтологии пользовательских профилей

- **Персонал библиотеки** «URI: <http://semdl.tpu.ru/ontology/structure/1.0/LibraryStaff>»: класс имеет только три экземпляра (**Admin**, **Editor**, **User**). Каждый зарегистрированный пользователь будет являться одним из них и иметь соответствующие права.
- **Группы** «URI: <http://semdl.tpu.ru/ontology/structure/1.0/Groups>»: Пользователь библиотеки может находиться в дружеских отношениях с другими пользователями, которые группируются по пользовательским целям для обмена информацией об электронных документах.
- **Статус** «URI: <http://semdl.tpu.ru/ontology/structure/1.0/StatusType>»: класс обладает тремя индивидуумами, такими как **Online**, **Busy**, **Offline** для обозначения статуса пользователя в системе.
- **Событие** «URI: <http://purl.org/NET/c4dm/event.owl#Event>»: класс представляет событие, которое может произойти, и о котором система будет информировать пользователей.
- **Ресурсы**: «URI: <http://www.marcont.org/ontology/2.0#Resource>»: класс включен в онтологию, т. к. в онтологии ресурсов он является основным и суперклассом других классов (книги, статьи, доклады и т. д.). В данной онтологии все атрибуты связей остаются прежними, а основное внимание было уделено таким атрибутам, как «**интересуется-Темой**» (**topic_interest**) и «**интересует**» (**interest**):
 - **ИнтересуетсяТемой** «URI: http://xmlns.com/foaf/0.1/topic_interest, Domains: **Агент**, Ranges: **Событие**, **Домен** **Интереса**»: заполнение данной связи пользователем позволит системе рекомендовать и информировать его о новом поступлении интересующих его документов.
 - **Интересует** «URI: <http://xmlns.com/foaf/0.1/interest>, Domains: **Агент**, Ranges: **Thing**»: для избавления от путаницы при выполнении логического вывода были добавлены два атрибута (**interestAgent**, **interestResource**), которые позволяют пользователям библиотеки описывать свои интересы (издания, книги, автора и т. п.). Данная связь также позволяет рекомендовать и информировать пользователя о новом поступлении документов конкретного автора или новых комментариев к ним.

В онтологии пользователей центральным является класс «Агент–Человек», то в онтологии ресурсов или онтологии предметных областей класс «Ресурс».

4.2. Библиографическая онтология

Библиографическая онтология (или онтология ресурсов) основывается на библиотечной онтологии Marcont [24], и при решении задач СЭБ будет использоваться эта онтология с некоторыми изменениями. В данной онтологии содержатся необходимые классы и атрибуты для описания электронных документов, таких, как книга, статья, диссертация и т. д. Схема онтологии приведена на рис. 2.

Класс «Ресурс» представляет все типы электронных документов в СЭБ и имеет связь с классом «Агент–Человек», в котором описываются издание, автор, редактор, соучастник. Этот класс также связывается с классом «Домен», представляющим собой концепт системы организации знаний (SKO); концептом могут быть ключевые слова, домен области документов.

Каждый документ в СЭБ может быть одним из типов ресурсов (книга, статья, диссертация и другие), а это значит, что задачу каталогизации документов можно решить с использованием таких типов. Задача категоризации заключается в аннотировании документов по своим доменам (класс «Домен»).

4.3. Структурная онтология (онтология структуры)

Онтология позволяет описывать каждый документ метаданными о его структуре, типе, оценке пользователя, а также коллекции. Кроме этого введены новые классы, такие, как класс «Сообщение» для представления разных оповещений от системы к пользователю и класс «Тип Статуса Документа» для задания статуса документа (**IPublished**, **ISubmitted**, **IToBeSubmitted**, **IAccepted**) при его редактировании библиотечным персоналом.

Схема онтологии структуры СЭБ показана на рис. 3

В отличие от других данная онтология характеризует работу конкретной системы и повторно не может быть использована. Вариант онтологии подобного типа можно найти в [25].

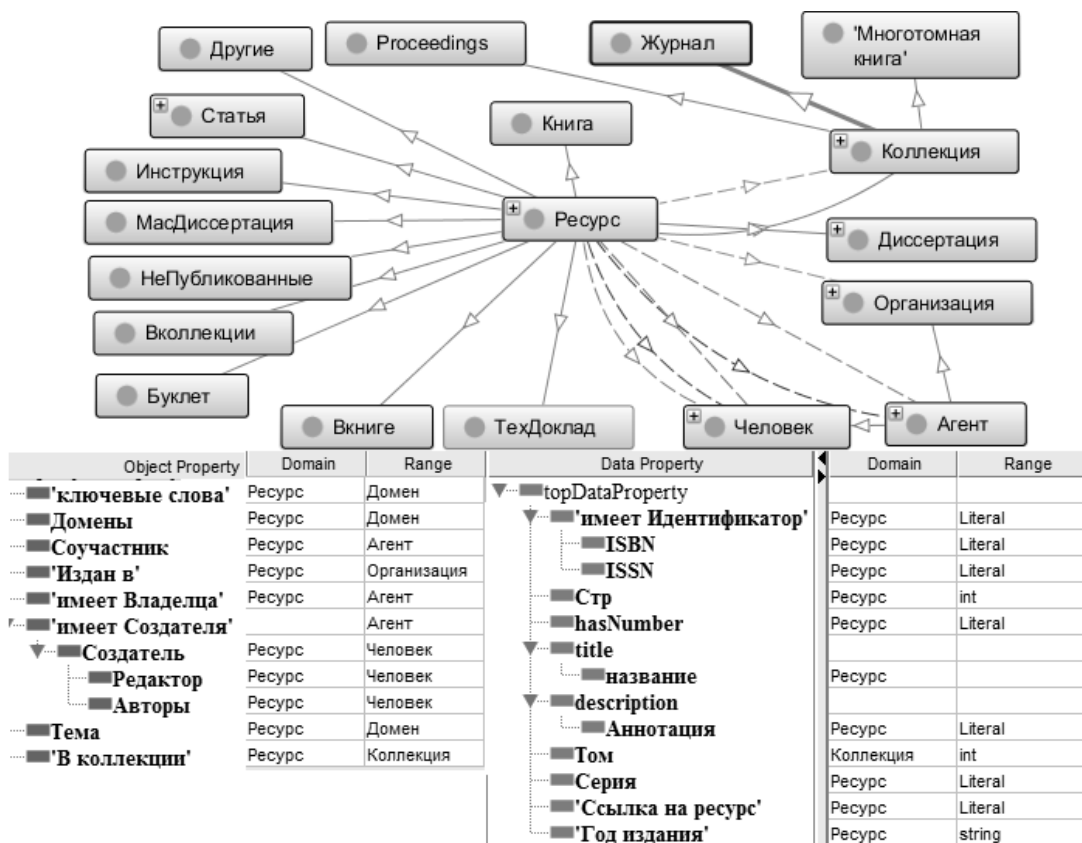


Рис. 2. Схема онтологии ресурсов (электронных документов) и их атрибуты

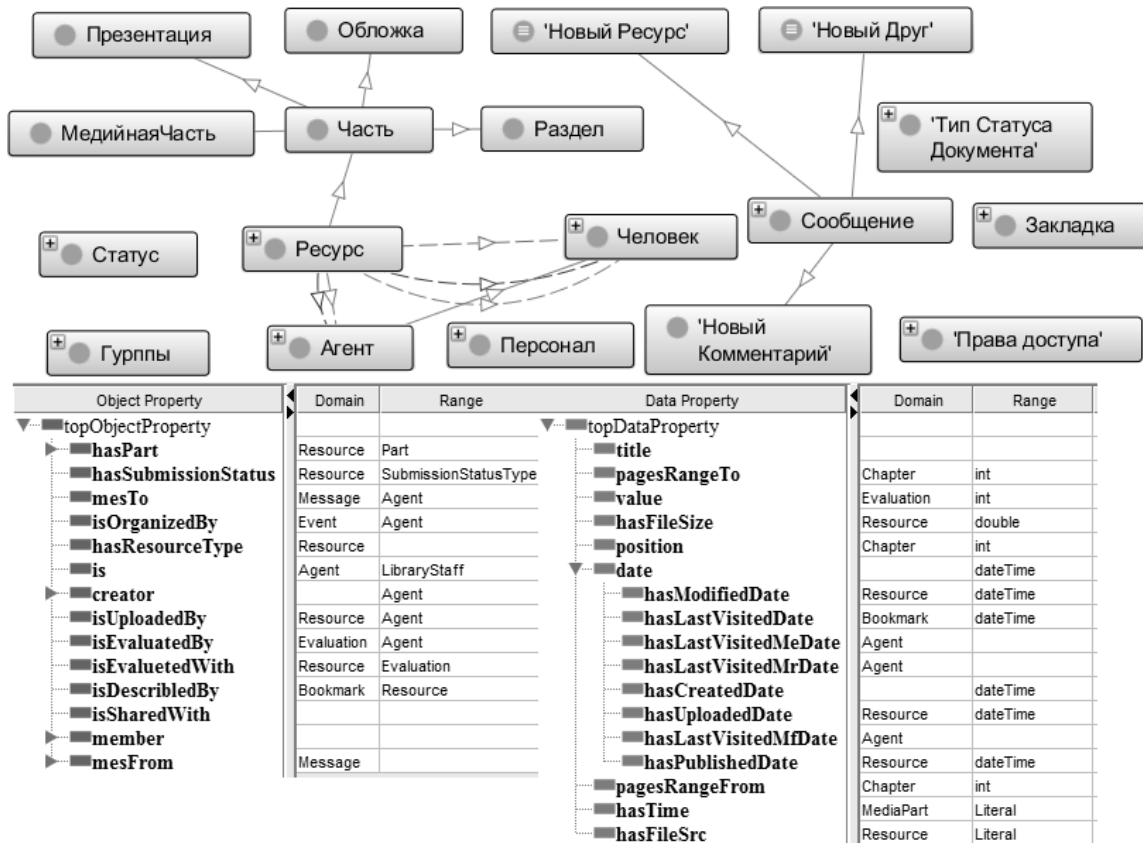


Рис. 3. Схема онтологии структуры электронных ресурсов и их атрибуты

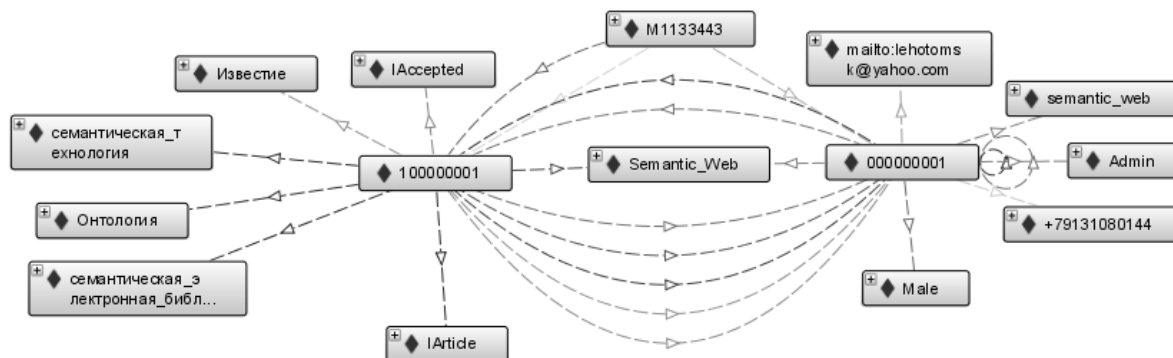


Рис. 4. Схема примера использования онтологий

4.4. Пример использования

Рассмотрим пример использования представленных выше онтологий для описания пользователя, документа и его системных данных. На рис. 4 показана схема отношений между экземплярами классов этих онтологий, где **000000001** – **Человек**, **100000001** – **Статья**.

На схеме не показаны свойства, но их можно посмотреть в виде OWL файла:

Выводы

Для создания семантических электронных библиотек требуется разрабатывать набор специальных онтологий. В качестве таковых могут использоваться библиографическая, структурная и пользовательская онтологии. Их применение позволяет решать основные задачи разработки семантических электронных библиотек. Разработанный вариант онтологий будет применен для создания базы знаний семантических электронных библиотек.

```
<Person rdf:about="&users;000000001">
<rdf:type rdf:resource="&owl;NamedIndividual"/>
<phone rdf:resource="&FoafUser;+79131080144"/>
< users:interestAgent rdf:resource="& users;000000001"/>
< users:interestResource rdf:resource="&Resource;100000001"/>
< users:gender rdf:resource="& users;Male"/>
<mbox rdf:resource="& users;mailto:lehotomsk@yahoo.com"/>
<topic_interest rdf:resource="&resource;Semantic_Web"/>
<structure:is rdf:resource="&structure;Admin"/>
</Person>
<owl:NamedIndividual rdf:about="&resource;100000001">
<rdf:type rdf:resource="http://www.marcont.org/ontology/2.0#Article"/>
<hasPages rdf:datatype="&xsd:int">8</hasPages>
<hasAbstract rdf:datatype="&xsd:string">В статье рассматривается вариант онтологий
для разработки семантических электронных библиотек (СЭМ), их спецификации для решения
основных задач в системе управления электронными документами, а также проводится изучение
методологии создания онтологий.</hasAbstract>
<resource:hasTitle rdf:datatype="&xsd:string">Разработка онтологий для семантических
электронных библиотек</Resource:hasTitle>
<hasYear>2011</hasYear>
<hasAuthor rdf:resource="&users;000000001"/>
<hasDomain rdf:resource="&resource;Semantic_Web"/>
<hasPublisher rdf:resource="&resource;Известие"/>
<hasKeyword rdf:resource="&resource;Онтология"/>
<hasKeyword rdf:resource="&resource;семантическая_технология"/>
<hasKeyword rdf:resource="&resource;семантическая_электронная_библиотека"/>
<rdf:type rdf:resource="&ontology;Resource"/>
<isUploadedBy rdf:resource="&users;000000001"/>
<hasSubmissionStatus df:resource="http://semDl.tpu.ru/ontology/structure/1.0/IAccepted"/>
<hasResourceType rdf:resource="http://semDl.tpu.ru/ontology/structure/1.0/IArticle"/>
</owl:NamedIndividual>
```

СПИСОК ЛИТЕРАТУРЫ

1. OWL Web Ontology Language Overview. 2011. URL: <http://www.w3.org/TR/owl-features/> (дата обращения: 20.04.2011).
2. RDF Vocabulary Description Language 1.0: RDF Schema. 2011. URL: <http://www.w3.org/TR/rdf-schema/> (дата обращения: 20.04.2011).
3. Resource Description Framework (RDF). 2011. URL: <http://www.w3.org/RDF/> (дата обращения: 20.04.2011).
4. SPARQL Query Language for RDF. 2011. URL: <http://www.w3.org/TR/rdf-sparql-query/> (дата обращения: 20.04.2011).
5. Welcome to the Protege wiki. 2011. URL: http://protegewiki.stanford.edu/wiki/Main_Page (дата обращения: 20.04.2011).
6. Java RDF Framework and Database System. 2011. URL: <http://sourceforge.net/projects/sesame/> (дата обращения: 20.04.2011).
7. Jena – A Semantic Web Framework for Java. 2011. URL: <http://jena.sourceforge.net/> (дата обращения: 20.04.011).
8. Кузнецов О.П., Суховеров В.С., Шипилина Л.Б. Онтология как систематизация научных знаний: структура, семантика, задачи // Технические и программные средства систем управления, контроля и измерения: Труды II Росс. конф. с международным участием. – М., 2010. – С. 762–773.
9. RIF RDF and OWL Compatibility. 2011. URL: <http://www.w3.org/TR/rif-rdf-owl/> (дата обращения: 20.04.2011).
10. Kruk S.R., Decker S., Zieborak L. JeromeDL – Adding Semantic Web Technologies to Digital Libraries // Lecture Notes in Computer Science. – 2005. – V. 3558/2005. – P. 716–725.
11. BRICKS Project: Building Resources for Integrated Cultural Knowledge Services (IST 507457). 2011. URL: <http://www.brickcommunity.org> (дата обращения: 20.04.2011).
12. Kruk S.R., Synak M., Kerstin Z. MarcOnt – Integration Ontology for Bibliographic Description Formats // International Conference on Dublin Core and Metadata Applications: Proc. – Madrid, Spain, 2005. – P. 231–234.
13. York S., Stephan B., Peter H., Jens H., Daniel O. The SWRC ontology – semantic web for research communities // 12th Portuguese Conference on Artificial Intelligence – Progress in Artificial Intelligence: Proc. – 2005. – V. 3803 of LNCS. – Covilha, Portugal, DEC Springer 2005. – P. 218–231.
14. The Role of Ontologies in Semantic Digital Libraries. 2011. URL: <http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2006/papers/paper-13.pdf> (дата обращения: 23.04.2011).
15. OWL Web Ontology Language. 2011. URL: <http://www.w3.org/TR/owl-guide/> (дата обращения: 23.04.2011).
16. FOAF-Realm – control your friends' access to resources. 2011. URL: http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/foaf_realm/ (дата обращения: 23.04.2011).
17. Linking Public Vocabularies. 2011. URL: <http://www.openvocabulary.info/vocabularies/> (дата обращения: 23.04.2011).
18. JOnto project. 2011. URL: <http://jonto.sourceforge.net/> (дата обращения: 25.04.2011).
19. Паспорта специальностей ВАК. 2011. URL: <http://teacode.com/online/vak/> (дата обращения: 25.04.2011).
20. Dean A., James H. Semantic web for the working ontologist modeling in RDF, RDFS and OWL. – Morgan Kaufmann, 2008. – 352 p.
21. Yu Liyang. A Developer's Guide to the Semantic Web. – Springer, 2011. – 628 p.
22. Ontology editor. 2011. URL: http://en.wikipedia.org/wiki/Ontology_editor (дата обращения: 25.04.2011).
23. FOAF Vocabulary Specification 0.98. 2011. URL: <http://xmlns.com/foaf/spec/> (дата обращения: 26.04.2011).
24. Marcont Vocabulary Specification 2.0. 2011. URL: <http://www.marcont.org/ontology/2.0> (дата обращения: 26.04.2011).
25. JeromeDL Ontology Specification. 2011. URL: <http://www.jeromedl.org/structure/2.1> (дата обращения: 27.04.2011).

Поступила 27.01.2012 г.

КЛАССИФИКАЦИЯ ТЕКСТОВ НА ОСНОВЕ ОЦЕНКИ СЕМАНТИЧЕСКОЙ БЛИЗОСТИ ТЕРМИНОВ

Нгуен Ба Нгок, А.Ф. Тузовский

Томский политехнический университет

E-mail: nguyen_bn@hotmail.com

Рассматривается способ увеличения точности классификации текстов по алгоритму kNN путем применения оценки семантической близости на основе матрицы совместной встречаемости терминов. Предлагается метод уменьшения размера матрицы совместной встречаемости путем фильтрации терминов по частям речи. Выполнена проверка влияния метода фильтрации на точность классификации.

Ключевые слова:

Классификация текстов, kNN , семантическая близость, матрица совместной встречаемости, аннотация по частям речи.

Key words:

Text classification, kNN , semantic similarity, co-occurrence matrix, parts of speech annotation.

Введение

Классификация текстов по тематическим категориям является актуальной задачей, которая требуется в различных информационных системах. Общая задача классификации данных рассматривается в книге [1]. В данной статье под термином *классификатор* понимается алгоритм решения задачи классификации текстов по тематическим категориям.

Среди существующих классификаторов, таких, как *Naïve Bayes* [1, 2], *Rocchio* [1, 3], *SVM* [1, 4], kNN [1] и т. д., самым быстрым и интуитивно понятным является классификатор kNN (*k-Nearest Neighbor* – ближайший из k соседей).

По алгоритму kNN результирующим тематическим классом классификатора является класс, имеющий максимальную оценку близости

$$kNN(D) = K_{\max},$$

при этом K_{\max} определяется по условию:

$$sim(D, K_{\max}) = \max_{i=1..n} (sim(D, K_i)),$$

где $sim(D, K_i)$ – оценка близости документа D корпусу K_i .

Целью данной работы является повышение точности работы алгоритма kNN с помощью применения оценки семантической близости на основе матрицы совместной встречаемости терминов в тексте. Предложенный алгоритм сравнивается по точности с классическим текстовым классификатором kNN .

Постановка задачи

Заданы k тематических классов, при этом, каждый i -й тематический класс определяется коллекцией (корпусом) документов K_i , которая состоит из множества текстовых документов по i -й теме. Каждая коллекция K_i также называется обучающей выборкой i -го тематического класса:

$$K_i = \{D_1, D_2, \dots, D_n\},$$

где каждый документ D_j состоит из множества терминов:

$$D_j = \{t_1, t_2, \dots, t_m\}.$$

Для любого текстового документа D , который не входит в обучающую выборку K_i , требуется определить наиболее подходящий тематический класс.

1. Классификатор kNN в векторном пространстве

Классический алгоритм kNN для классификации текстов базируется на оценки близости в векторном пространстве. Для этого выполняется представление документа в виде вектора терминов.

1.1. Представление документов в векторном пространстве

Пусть T – множество всех уникальных терминов системы:

$$T = \{t_1, t_2, \dots, t_n\},$$

которое определяет n -мерное пространство, где каждый термин t_i соответствует одной размерности пространства.

В данном n -мерном пространстве каждый документ D коллекции K соответствует точке, которая определяется вектором v_D – представление документа D :

$$v_D = \langle (t_1, \omega_1), (t_2, \omega_2), \dots, (t_n, \omega_n) \rangle,$$

где ω_i – весовой коэффициент термина t_i .

Одним из наиболее эффективных методов вычисления весовых коэффициентов является метод *tf.idf*, по которому значение ω_i определяется следующим образом:

$$\omega_i = \begin{cases} tf.idf_{D,K}(t_i), & \text{если } t_i \in D; \\ 0, & \text{в противном случае.} \end{cases}$$

1.2. Схема вычисления весовых коэффициентов *tf.idf*

Весовой коэффициент термина по схеме *tf.idf* вычисляется как произведение нормализованной частоты встречаемости термина *tf* (*term frequency*) и обратного значения нормализованной частоты документов *idf* (*inverted document frequency*).

Коэффициент tf термина t документа D определяется по следующей формуле:

$$tf_D(t) = \frac{freq_D(t)}{|D|},$$

где $freq_D(t)$ – частота встречаемости термина t в документе D ; $|D|$ – количество всех терминов документа D (размер документа D).

Нормализованная частота документов df (*document frequency*) термина t в корпусе K определяется по формуле:

$$df_K(t) = \frac{|\{D_i | D_i \in K, t \in D_i\}|}{|K|},$$

где $|K|$ – количество документов коллекции K ; обратное значение нормализованной частоты документов вычисляется следующим образом:

$$idf_K(t) = \log\left(\frac{1}{df_K(t)}\right).$$

Весовой коэффициент ω_i термина t_i документа D в корпусе K вычисляется по следующей формуле:

$$tf \cdot idf_{D,K}(t) = \frac{freq_D(t)}{|D|} \cdot \log\left(\frac{|K|}{|\{D_i | t \in D_i, D_i \in K\}|}\right).$$

1.3. Обучение классификатора

Обучение классификатора kNN заключается в вычислении центральных векторов обучающих выборок. Под центральным вектором C_K обучающей выборки K понимается следующий вектор:

$$C_K = \frac{\sum_{D \in K} v_D}{|K|} = \langle (t_1, c_1), (t_2, c_2), \dots, (t_n, c_n) \rangle,$$

где v_D – вектор представления документа D ; t_i – термины; c_i – весовой коэффициент, который определяется следующим образом:

$$c_i = \frac{\sum_{D \in K} \omega_{i,D}}{|K|},$$

где $\omega_{i,D}$ – весовой коэффициент термина t_i документа D .

1.4. Классический алгоритм kNN для классификации текстов

Оценка близости классифицируемого документа D и обучающей выборки K определяется как близость их векторов представления в векторном пространстве

$$sim(D, K) = sim_{vector}(v_D, C_K).$$

На основе данной оценки близости результат классификатора kNN определяется следующим образом:

$$kNN(D) = K_{max},$$

при этом K_{max} определяется по условию:

$$sim_{vector}(v_D, C_{K_{max}}) = \max_{K \in K_i} (sim_{vector}(v_D, C_K)),$$

где K_i – множество обучающих выборок;

1.5. Оценки близости векторов

Оценка близости векторов a и b в векторном пространстве обозначается $sim_{vector}(a, b)$. Методы вычисления близости векторов представляются в таблице.

2. Классификатор kNN с использованием оценки семантической близости

По предлагаемой схеме для вычисления оценки семантической близости, сначала требуется определение матрицы совместной встречаемости терминов. При этом отличаются два типа оценки семантической близости: 1) терминов; 2) коллекций.

2.1. Матрица совместной встречаемости терминов

Матрица совместной встречаемости M_K коллекции K определяется в множестве уникальных терминов T следующим образом:

$$M_K(i, j) = M_K(t_i, t_j) = freq_K(t_i, t_j),$$

где $t_i, t_j \in T$; $freq_K(t_i, t_j)$ – частота совместной встречаемости, т. е. частота, по которой термин t_i встречается вместе с термином t_j в коллекции K .

Считается, что термин t_i встречается вместе с термином t_j , если расстояние между ними в тексте не превышает предела z : $|i-j| \leq z$.

Таблица. Методы вычисления оценки близости между векторами

Название метода	Формула
<i>Inv. Sq. City-block</i> – обратное значения расстояния по Сити-Блок метрики (city-block distance) в квадрате	$sim_{inv.sq.city-block}(a, b) = \frac{1}{(\sum_i a_i - b_i)^2 + 1}$
<i>Inv. Sq. Euclidean</i> – обратное значение евклидова расстояния в квадрате	$sim_{inv.sq.euclidean}(a, b) = \frac{1}{\sum_i (a_i - b_i)^2 + 1}$
<i>Cosine</i> – косинус угла между векторами	$sim_{cosine}(a, b) = \frac{\sum_i a_i b_i}{\sqrt{\sum_i a_i^2} \sqrt{\sum_i b_i^2}}$
<i>Correlation</i> – корреляция между векторами	$sim_{correlation}(a, b) = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2} \sqrt{\sum_i (b_i - \bar{b})^2}}$

Имеется следующий псевдокод алгоритма определения матрицы совместной встречаемости M_K коллекции документов K :

```

 $M_K = [0]$  // в начале алгоритма, все элементы матрицы равны нулю
for  $D \in K$  do
  for  $t_i \in D$  do
    for  $j = i - z$  to  $i + z$  do
      if  $i \neq j$  then
         $M_K(t_i, t_j)++$  // увеличение на единицу
    
```

Главной проблемой создания матрицы совместной встречаемости является большой требуемый объем оперативной памяти для её сохранения. В следующем разделе представляется решение данной проблемы путем фильтрации терминов.

2.2. Метод фильтрации терминов

Целью применения метода фильтрация терминов является уменьшение размера матрицы совместной встречаемости путем снижения количества терминов коллекции, при этом поддерживается минимальная потеря точности классификации.

В данной статье предлагается метод фильтрации терминов фильтрации терминов по частям речи, согласно которому из коллекции документов удаляются все термины, относящиеся к части речи, которая является исключением. Для чего, сначала требуется выполнение аннотирования текстов по частям речи (*parts of speech tagging* [5]).

Одним из эффективных инструментов для выполнения аннотирования текстов на английском языке является библиотека *Stanford CoreNLP* (<http://nlp.stanford.edu/software/corenlp.shtml>), аналогом которой для текстов на русском языке является библиотека *MyStem* (<http://company.yandex.ru/technologies/mystem>).

2.3. Оценка семантической близости терминов

На основе матрицы совместной встречаемости терминов M_K , значение термина t_i определяется i -й строкой матрицы M_K следующим образом:

$$mean_K(t_i) = \langle (t_1, \omega_1), (t_2, \omega_2), \dots, (t_n, \omega_n) \rangle,$$

где $\omega_j = M_K(t_i, t_j)$ – частота совместной встречаемости терминов t_i и t_j .

Семантическая близость терминов определяется как близость между их значениями, представленными в виде векторов. Семантическая близость термина t_i коллекции K_1 и термина t_j коллекции K_2 определяется следующим образом:

$$sim_{sem}(t_{i_{K_1}}, t_{j_{K_2}}) = sim_{vector}(mean_{K_1}(t_i), mean_{K_2}(t_j)).$$

Другие методы определения семантической близости терминов на основе матрицы совместной встречаемости представляются в работах [6, 7, 8].

2.4. Оценка семантической близости коллекции документов

Оценка семантической близости коллекций документов K_1 и K_2 определяется как сумма оценки

семантической близости терминов следующим образом:

$$sim_{sem}(K_1, K_2) = \sum_{t \in T} sim_{vector}(mean_{K_1}(t), mean_{K_2}(t)).$$

Идентично, оценка семантической близости документа D и коллекции K вычисляется по следующей формуле:

$$sim_{sem}(D, K) = \sum_{t \in T} sim_{vector}(mean_D(t), mean_K(t)).$$

2.5. Алгоритм kNN с использованием оценки семантической близости

С использованием оценки семантической близости предлагается следующая комбинация (*combination*) оценок близости документа D и коллекции K :

$$sim_{comb}(D, K) = sim_{vector}(v_D, C_K) + sim_{sem}(D, K),$$

результат классификатора kNN определяется следующим образом:

$$kNN(D) = K_{max},$$

при этом, K_{max} определяется по условию:

$$sim_{comb}(D, K_{max}) = \max_{K \in K_i} (sim_{comb}(D, K)).$$

3. Эксперимент

В качестве тестовых данных используется коллекция *20Newsgroups* (<http://people.csail.mit.edu/jrennie/20Newsgroups>), которая предназначена для тестирования метода классификации. Коллекция *20Newsgroups* содержит 20 тысяч текстовых сообщений, которые разделяются на 20 тематических групп новостей. Каждая группа имеет приблизительно 1000 сообщений содержащихся 400 документов для тестирования и 600 документов для обучения классификатора.

С использованием данной коллекции были выполнены три эксперимента для проверки точности предлагаемого метода классификации и размера матрицы совместной встречаемости.

Оценка точности классификации в одной группе новостей определяется как отношение между количеством правильных результатов и общим количеством документов группы. Итоговая оценка точности классификации вычисляется как среднее значение оценок точности классификации в отдельных группах.

В данной статье предлагается определение размера матрицы совместной встречаемости M как количество её ненулевых элементов:

$$size(M) = |\{(t_i, t_j) | t_i, t_j \in T, M(t_i, t_j) > 0\}|.$$

3.1. Методы вычисления близости векторов

Значения точности предлагаемого алгоритма и классического алгоритма в зависимости от используемого метода вычисления близости векторов (см. таблица) и коэффициента z (см. раздел 2.1) представлены на рис. 1.

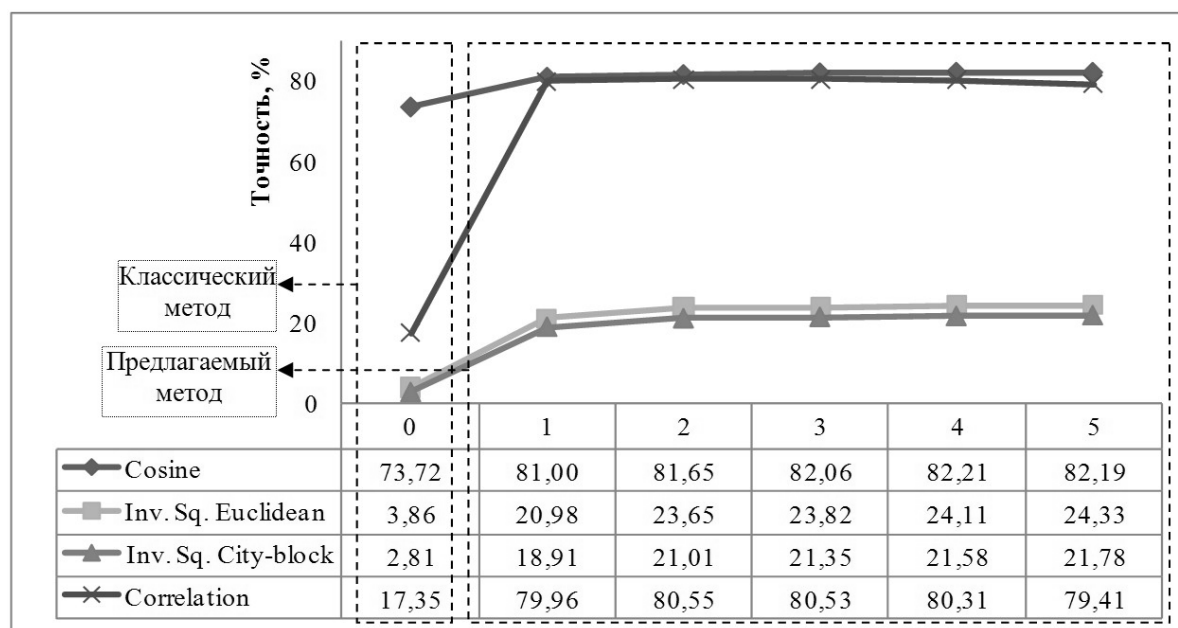


Рис. 1. Оценки точности классификации

Горизонтальная ось представляет собой значения коэффициента z , а вертикальная – значения точности классификации. Четыре серии данных соответствуют значениям точности классификации при использовании разных оценок близости векторов sim_{vector} : *cosine* – косинус; *correlation* – значение корреляции; *inv.sq.euclidean* – обратное значение евклидова расстояния в квадрате; *inv.sq.city-block* – обратное значение расстояния по сити-блок метрики в квадрате.

Исходя из полученных данных, для всех методов вычисления близости векторов можно отметить существенное увеличение точности классификации в результате использования оценки семантической близости по сравнению с классическим методом, т. е. точность классификации в случаях $z > 0$ значительно выше, чем точность классификации в случае $z = 0$ (разница $> 7\%$).

Наивысшая точность для классического метода классификации ($\approx 74\%$) достигается при использовании метода вычисления близости векторов по косинусу (*cosine*), $z = 0$.

По сравнению с классическим методом наивысшая точность для предлагаемого метода классификации ($\approx 82\%$) достигается при использовании метода вычисления близости векторов по косинусу и $z = 4$. Кроме того, если $z > 0$, то идентичные результаты также получаются с применением значения корреляции между векторами.

Самая низкая точность получается при использовании метода вычисления близости на основе сити-блок-метрики и метода вычисления близости на основе евклидова расстояния ($< 30\%$).

3.2. Размер матрицы совместной встречаемости

Зависимость размера матрицы совместной встречаемости от метода фильтрации терминов и значения расстояния z представлены на рис. 2.

Горизонтальная ось представляет собой значения коэффициента z , а вертикальная ось – значения размера матрицы совместной встречаемости. Три серии данных соответствуют значениям размера матрицы, полученным при использовании разных методов фильтрации терминов, в каждом сохранены: *All* – все термины; *Noun & Verb* – только существительные и глаголы; *Noun* – только существительные.

Исходя из полученных данных, видно, что в результате фильтрации терминов значительно уменьшается размер матрицы совместной встречаемости по сравнению с случаем использования всех терминов: приблизительно 30% в случае использования только существительных и глаголов и приблизительно 50% в случае использования только существительных.

Кроме того, также видно, что с условием фиксированного метода фильтрации, если увеличивается расстояние z , то резко увеличивается размер матрицы совместной встречаемости (рис. 2), однако точность классификации увеличивается незначительно (рис. 3).

3.3. Точность классификатора с использованием метода фильтрации терминов

Зависимость точности классификатора от используемого метода фильтрации терминов и значения расстояния z представлено на рис. 3. В каче-

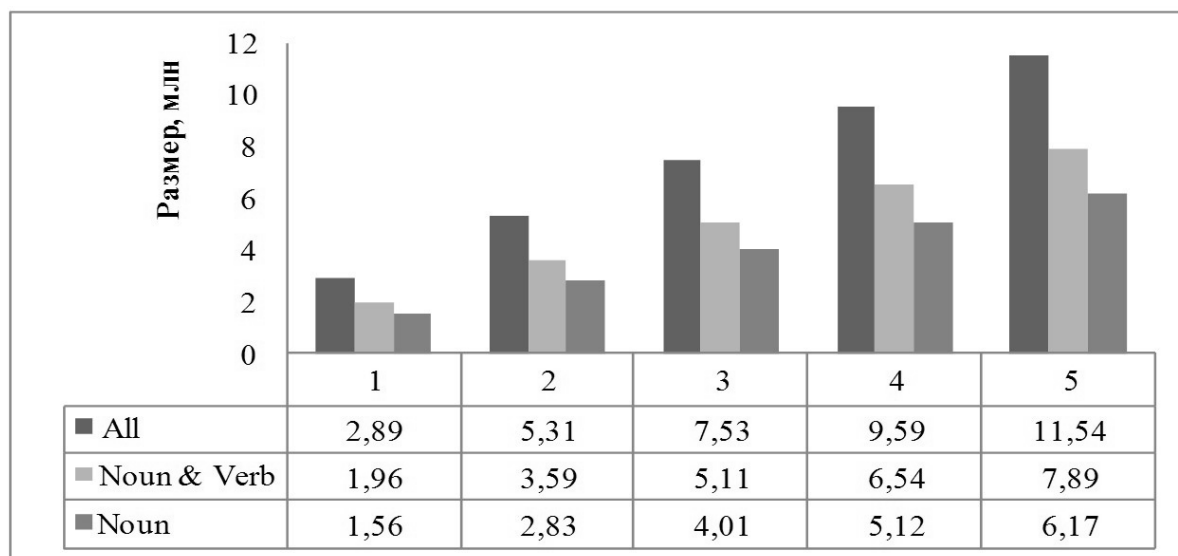


Рис. 2. Размер матрицы совместной встречаемости

стве оценки близости векторов используется оценка близости по косинуса, так как, исходя из первого эксперимента, она является лучшим методом определения близости векторов.

Горизонтальная ось представляет собой значения коэффициента z , а вертикальная ось – значения точности классификации. Три серии данных соответствуют значениям точности, полученным при использовании разных методов фильтрации терминов: *All* – используются все терминов; *Noun & Verb* – используются только существительные и глаголы; *Noun* – используются только существительные.

Исходя из полученных данных видно, что в случае $z=0$ (классический алгоритм *kNN*), максимальная точность (75,56 %) получается при сохранении в коллекции документов только существительных.

Однако в остальных случаях $z>0$ (используется оценка семантической близости), наивысшая точность ($\approx 82\%$) получается при использовании всех терминов, и если применяется метод фильтрация, то снижается точность классификации. Кроме того, в случае использования метода фильтрации терминов при увеличении z значение точности увеличивается незначительно в пределах 3...4 %, по сравнению с 7 % в случае без использования метода фильтрации (*All*).

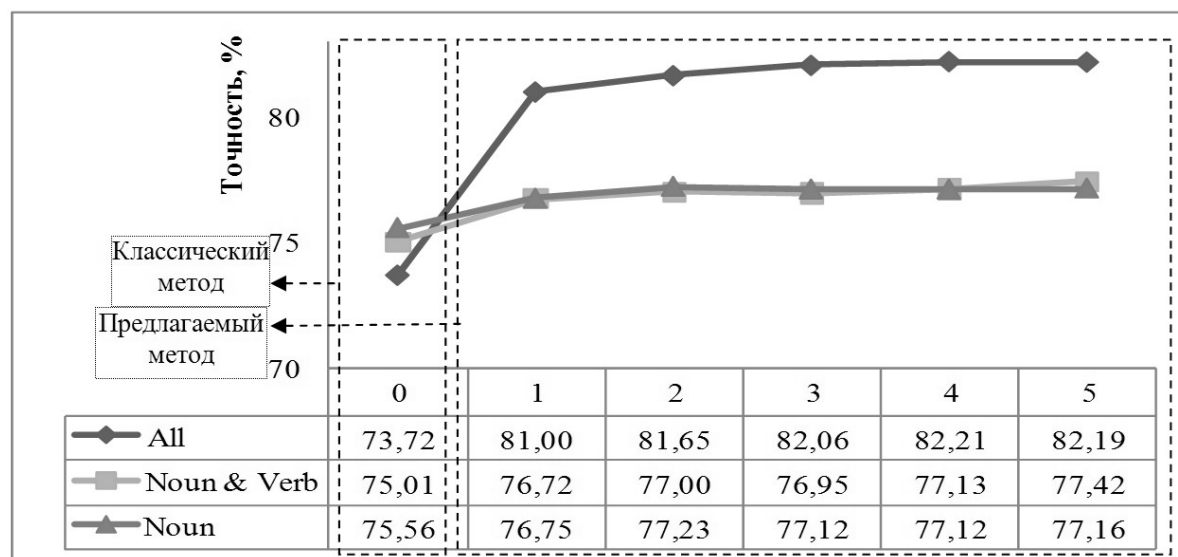


Рис. 3. Влияние метода фильтрации на точность классификатора

Выводы

Показано, что: 1) лучшим методом оценки векторной близости (среди рассмотренных) является оценка близости по косинусу; 2) метод фильтрации существительных имеет большое практическое значение для классического алгоритма *kNN* в векторном пространстве, так как, при сохранении только существительных из коллекции документов, точность классификации является максимальной,

одновременно количество используемых терминов для классификации значительно сокращается; 3) с использованием оценки семантической близости точность классификатора *kNN* значительно увеличивается по сравнению с классической реализацией (>7%), однако требуется дополнительные затраты для вычисления матрицы совместной встречаемости и оценки семантической близости.

СПИСОК ЛИТЕРАТУРЫ

1. Nisbet R., Elder J., Miner G. Handbook of statistical analysis and data mining applications. – Amsterdam: Academic press, 2009. – 824 p.
2. McCallum A., Nigam K. A comparison of event models for naïve bayes text classification // multinomial-aaaiws98. 1998. URL: <http://www.cs.cmu.edu/~knigam/papers/multinomial-aaaiws98.pdf> (дата обращения: 10.01.2012).
3. Joachims T. A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization // Joachims_97a. 1997. URL: http://www.cs.cornell.edu/People/tj/publications/joachims_97a.pdf (дата обращения: 10.01.2012).
4. Joachims T. Text Categorization with Support Vector Machines: Learning with Many Relevant Features // Joachims_98a. 1998. URL: http://www.cs.cornell.edu/People/tj/publications/joachims_98a.pdf (дата обращения: 10.01.2012).
5. Handbook of natural language processing / Ed. by N. Indurkha, F.J. Damerau. – London: Chapman & Hall/CRC, 2010. – 692 p.
6. Douglas L.T.R., Laura M.G., David C.P. An improved method for deriving word meaning from lexical co-occurrence // Cognitive Science. – 2009. – V. 7. – № 2. – P. 573–605.
7. Foltz P.W., Kintsch W., Landauer T.K. The measurement of textual coherence with Latent Semantic Analysis // Discourse Processes. – 1998. – V. 25. – № 2. – P. 285–307.
8. Lund K., Burgess C. Producing high-dimensional semantic spaces from lexical co-occurrence // Behavior Research Methods, Instruments and Computers. – 1996. – V. 28. – № 2. – P. 203–208.

Поступила 26.01.2012 г.

УДК 004.931

АЛГОРИТМИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ РАСПОЗНАВАНИЯ ФОРМЫ РУКИ В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ SURF-ДЕСКРИПТОРОВ И НЕЙРОННОЙ СЕТИ

Нгуен Тоан Тханг, В.Г. Спицын

Томский политехнический университет
E-mail: thangngt.cntt@gmail.com; spvg@tpu.ru

Разработан оригинальный алгоритм распознавания формы руки в реальном времени на основе SURF-дескрипторов и нейронной сети. Предложен новый метод генерации дескрипторов для нейронной сети. Создано программное обеспечение для распознавания формы руки в режиме реального времени на основе предложенного алгоритма. Численные эксперименты по распознаванию формы руки на видеопоследовательности в реальном времени показали, что средняя точность распознавания составляет 92 %.

Ключевые слова:

Распознавание руки, SURF-дескриптор, обнаружение признаков, выделение признаков, описание признаков, кластеризация, многослойная нейронная сеть.

Key words:

Hand recognition, SURF-descriptor, feature detection, feature extraction, feature description, clustering, multilayer neural network.

Введение

Распознавание позы и жестов является одной из центральных задач в человеко-машинном взаимодействии и привлекает внимание многих исследователей. Для решения этой проблемы были предложены различные методы и алгоритмы. Они могут быть сгруппированы в две категории: методы на основе внешнего вида руки (*Vision-based approach*), и методы на основе 3D модели руки (*3D hand*

model based approach) [1]. Методы на основе внешности руки используют двумерные признаки изображения для моделирования визуальной внешности руки и сравнивают эти параметры с теми же признаками, выделенными из входного изображения. В методах на основе 3D модели применяют 3D кинематические модели руки, чтобы оценить параметры руки, сравнивая эти параметры с двумерными проекциями 3D моделей. В первую

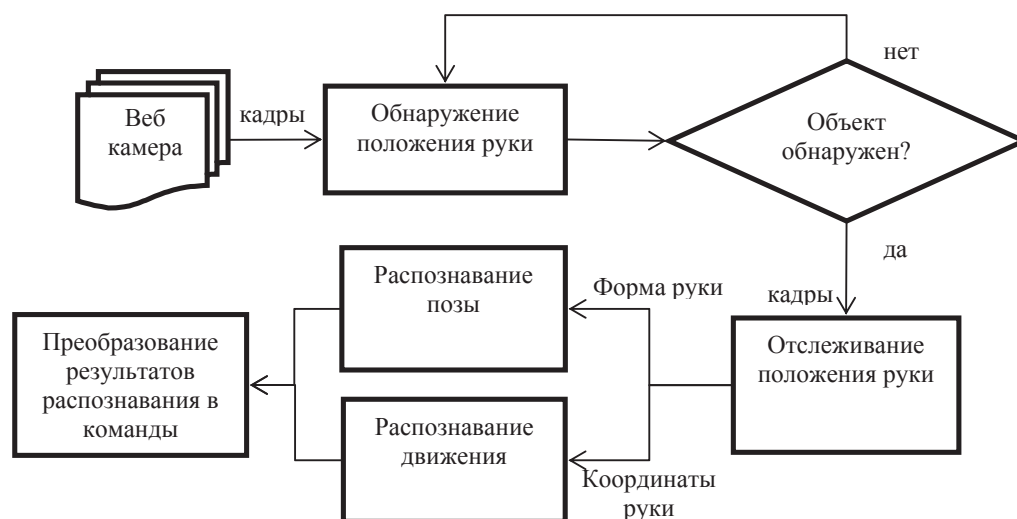


Рис. 1. Алгоритм распознавания жестов на видеопоследовательности

категорию входят методы, в которых используются такие признаки как образы руки [2, 3], локальные признаки [4, 5].

В данной работе предлагается новый алгоритм распознавания жестов на видеопоследовательностях в реальном времени на основе использования SURF-дескрипторов и многослойной нейронной сети. Алгоритм включает следующие этапы (рис. 1): обнаружение объекта (руки) в видеопотоке с помощью детектора Джонса–Виолы, отслеживание положения руки с использованием метода *CAMShift*; выделение характеристических признаков методом *SURF (Speeded Up Robust Features)*; генерация «словаря» с помощью метода кластеризации *K-means (K-means clustering)*; создание дескрипторов с совместным использованием *SURF* и *K*-ближайших соседей (*K-nearest neighbors*) на основе созданного «словаря»; обучение нейронной сети с вновь созданными дескрипторами. На основе предложенного алгоритма создан пакет программ, который позволяет распознавать жесты на видеопоследовательностях в реальном времени.

Обнаружение руки в видеопотоке

Обнаружение руки на изображении является первоначальной задачей всех систем распознавания жестов. В настоящее время метод Джонса–Виолы с использованием признаков Хаара и каскадного *AdaBoost* классификатора [6] для обнаружения человеческих лиц считается одним

из лучших алгоритмов для таких задач. Метод Джонса–Виолы был разработан для задачи обнаружения человеческих лиц на изображении. Главное преимущество этого метода состоит в его высокой скорости выполнения и точности обнаружения. В данной работе мы модифицируем детектор Джонса–Виолы с использованием нового набора признаков Хаара для обнаружения руки в видеопотоке в режиме реального времени. В традиционном методе Джонса–Виолы (и в модифицированном *Lienhard* и *Maydt* детекторе) применяется набор признаков, указанный на рис. 2, а. В данной работе мы применяем новый набор признаков, предложенный в работе [7] (рис. 2, б).

Человеческая рука не имеет постоянных контуров, что усложняет задачу обнаружения. Чтобы избежать этой трудности, мы предложили создать два детектора: один – для обнаружения кулака, а второй – для ладони. Кадр из видеокamеры передается сначала в детектор ладони. Если интересующий объект обнаружен, то детектор кулака отключится. В противном случае, детектор кулака включится для обнаружения присутствия руки на кадре.

Для обучения первого детектора (каскадного классификатора *AdaBoost*) мы использовали 1300 позитивных изображений (изображения ладони человеческой руки) размером 20×20 пикселей и 1000 негативных изображений (без объекта). Для второго детектора – 1000 позитивных изображений и 1000 негативных изображений размером 24×24 пикселя.

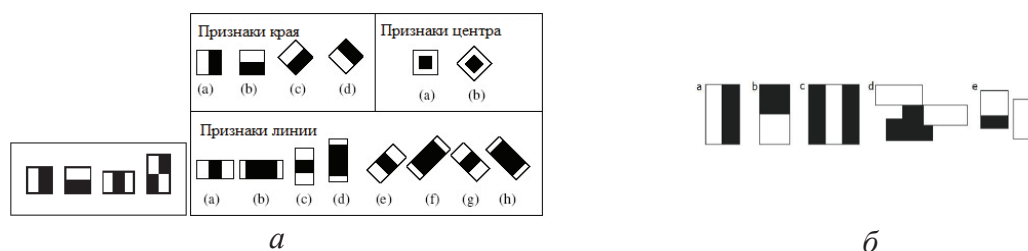


Рис. 2. Набор признаков Хаара, используемых: а) в традиционной модели; б) авторами статьи

Отслеживание положения руки в видеопотоке

Методы отслеживания объекта на видеопоследовательностях могут быть разделены на три большие группы: слежение на основе точек (*point tracking*), слежение с использованием ядра (*kernel tracking*) и слежение за контуром (*silhouette tracking*) [8]. Из-за того, что форма руки может резко изменяться на кадрах, слежение на основе точек и контура не подходят для решения данной задачи. В данной работе мы применили метод трекинга на основе метода *CAMShift* с использованием цветowych признаков.

Как только произошел захват руки на кадре, ее положение и размер используются для инициализации поискового окна, содержащего ладонь. Вычисляются две гистограммы: *HM* – гистограмма внутри найденного окна, *HI* – гистограмма целого изображения. Г. Брадски [9] предложил использовать *H*-компонент цветового пространства *HSV* для построения гистограммы. В данной работе цветовые пространства *RGB* и *HSL* используются для сравнения эффективности алгоритма трекинга. Из двух выше построенных гистограмм вычисляется «гистограмма вероятности цветов» (*color probability histogram*) по формуле:

$$P_i = \min\left(\frac{HM_i}{HI_i}, 1\right). \quad (1)$$

Здесь P_i – i -й компонент гистограммы вероятности цветов; HM_i – компонент гистограммы модели; HI_i – компонент гистограммы целого кадра.

Если заменить каждый пиксель входного изображения соответствующим значением P_i , получим «изображение вероятности» (*probability image*). Для каждого последовательного входного кадра алгоритм повторно вычисляет «центр масс» в обработанной области вокруг предыдущего центра масс в изображении вероятности по формуле (1) и перемещает поисковое окно в это новое положение. Для того, чтобы избежать «вечного» вычисления центра масс, алгоритм прекращает работу после определенной итерации, либо когда изменение центра масс окажется меньше, чем определенный порог:

$$(x_c, y_c) = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}\right),$$

$$M_{00} = \sum_x \sum_y I(x, y),$$

$$M_{10} = \sum_x \sum_y xI(x, y),$$

$$M_{01} = \sum_x \sum_y yI(x, y).$$

Здесь (x_c, y_c) – координаты центра масс; $I(x, y)$ – интенсивность пикселя (x, y) изображения вероятности; M_{00} , M_{10} , M_{01} – нулевой, а также первый момент по направлениям x и y соответственно.

Алгоритм распознавания формы руки

Основная идея предложенного алгоритма состоит в том, что изображение рассматривается как текстовый документ, в котором каждый визуальный признак учитывается как слово, образующее данный документ. Класс документа определяется путем вычисления частоты появления некоторых «ключевых слов». Для применения этой идеи в распознавании объекта вначале из набора изображений (базы данных для обучения) методом *SURF* выделяются характеристические признаки [10]. Затем все выделенные признаки разделяются на группы, используя *K-means* кластеризацию. Каждая группа служит «ключевым словом», из которых строится так называемый «словарь ключевых слов». Дескрипторы для нейронной сети создаются путем сопоставления выделенных признаков входного изображения с ключевыми словами из словаря признаков с помощью поиска *K*-ближайших соседей. Набор дескрипторов и соответствующий класс в дальнейшем используются для обучения нейронной сети. Общая схема алгоритма указана на рис. 3.

Детектор *SURF* базируется на вычислении матрицы Гессена с простой аппроксимацией. Для каждой точки $P=(x, y)$ матрица Гессена $H(P, \sigma)$ на P с масштабом σ определяется по следующей формуле:

$$H(P, \sigma) = \begin{bmatrix} L_{xx}(P, \sigma) & L_{xy}(P, \sigma) \\ L_{xy}(P, \sigma) & L_{yy}(P, \sigma) \end{bmatrix}.$$

Здесь $L_{xx}(P, \sigma)$, $L_{yy}(P, \sigma)$, $L_{xy}(P, \sigma)$ – свертка второй производной Гаусса $\frac{\partial^2}{\partial x^2} g(\sigma)$, $\frac{\partial^2}{\partial y^2} g(\sigma)$,

$\frac{\partial^2}{\partial x \partial y} g(\sigma)$ с изображением в точке P , соответственно.

С целью сокращения времени обработки, вторая производная Гаусса приближенно вычисляется с использованием «*box filter*» и «интегрального изображения». Интегральное изображение также упрощает процесс анализа в пространстве масштабов. Детерминант матрицы Гессена затем используется для выбора координат точек интереса и масштаба. Точки интереса локализируются с помощью «подавления не-максимумов» (*non-maximum suppression*). Это делает *SURF* признаки устойчивыми к масштабированию. *SURF* также строит окружность вокруг выделенной точки интереса, чтобы определить уникальную ориентацию, и тем самым дает *SURF* признакам инвариантность к вращению. *SURF* признаки описываются векторами, длина которых фиксирована (64 в оригинальном *SURF* и 128 в расширенной версии *SURF*).

В связи с тем, что *SURF* выделяет точки интереса на основе локальных признаков, количество обнаруженных точек не фиксировано, а меняется

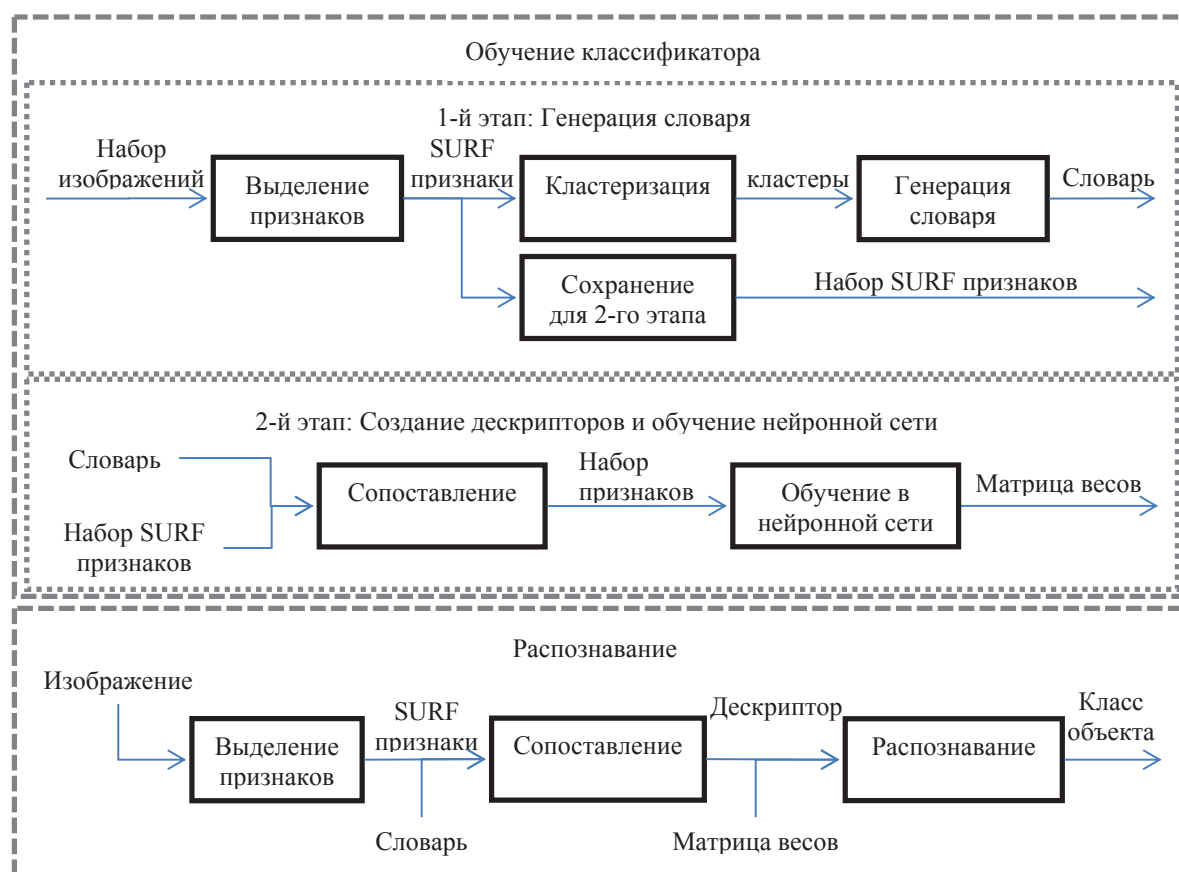


Рис. 3. Общая схема алгоритма распознавания формы руки

в зависимости от изображения, даже для одного и того же объекта. Поэтому нельзя напрямую подавать данные дескрипторы в нейронную сеть для распознавания. Для решения такой проблемы предлагается разделить все обнаруженные признаки из набора изображений на группы. Это обусловлено тем, что визуальные признаки, выделенные из одного класса объекта должны иметь сходство. Признаки, полученные при анализе одного изображения, также могут быть похожи друг на друга. Разделение признаков на группы позволяет «соединить» подобные признаки вместе и выбрать одного «представителя» для всей группы. Описание объекта тогда можно осуществлять с помощью только этих «представителей». В данной работе используется популярный метод кластеризации *K-means*.

Метод *K-means* разбивает множество элементов векторного пространства на заранее известное число кластеров k . Основная идея заключается в том, что на каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения кластеров. Это происходит за конечное число итераций, так как количество возможных разбиений конечного множества конечно, причем на каждом шаге суммарное

квадратичное отклонение уменьшается, поэтому заикливание невозможно. *K-means* стремится минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров:

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2.$$

Здесь k – число кластеров; S_i – полученные кластеры; μ_i – центры масс векторов $x_j \in S_i$; x – набор входных векторов.

При применении *K-means* кластеризации нужно вначале задать количество кластеров. Если кластером считается множество однозначных слов (как в текстовом документе), то количество кластеров – это общее число «ключевых слов». Центром масс каждого кластера является само ключевое слово, а совокупность всех полученных ключевых слов составляет словарь. Если считается, что выделенные *SURF* методом признаки из объектов определенного класса уникальны и характерны только для данного класса объекта, то количество слов в словаре (число кластеров) приблизительно равняется сумме средних чисел признаков всех классов объекта:

$$K = \sum_{i=1}^N \frac{\sum_{j=1}^{M_i} k_j}{M_i}.$$

Здесь K – число кластеров; N – количество классов объектов; M_i – количество образцов i -го класса; k_j – число выделенных признаков из j -го образца i -го класса.

При завершении процесса кластеризации получается набор векторов центра масс всех групп *SURF*-дескрипторов. Сами векторы центра масс являются *SURF*-дескрипторами и могут служить «представителем» для всех векторов в своей группе. Набор векторов центра масс (словарь) используется для генерации новых дескрипторов для классификатора.

При поступлении нового изображения, *SURF* выделяет характеристические признаки и описывает эти признаки набором векторов, размерность которых фиксирована. Для каждого *SURF* вектора (дескриптора) вычисляется квадрат расстояния Евклида до каждого из центров масс по формуле (2) и выбирается самое короткое расстояние.

$$d(p, q_k) = \sum_{i=1}^N (q_{ki} - p_i)^2. \quad (2)$$

Здесь $d(p, q_k)$ – расстояние от входного *SURF* вектора до k -го центра масс; N – размерность *SURF*-вектора (64 или 128 в зависимости от выбранного способа описания); p – входной *SURF* вектор; q_k – k -й центр масс.

Если значение этого расстояния меньше чем определенный порог, то соответствующий центр выбирается в качестве представителя для поступающего *SURF*-вектора. Другими словами, данный *SURF*-вектор заменяется соответствующим центром масс. При создании базы данных для генерации словаря используются изображения без фона, чтобы выделенные признаки были характерны только для объектов. Однако на изображениях в базе данных для обучения нейронной сети могут присутствовать фон или другие объекты. Поэтому нужно установить пороговое значение расстояния, чтобы исключить «чужие признаки», которые выделяются, к примеру, из фона, а не из объекта. Этот порог определяется как среднее значение всех расстояний между центрами масс:

$$\partial = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N d(q_i, q_j), i \neq j.$$

Здесь ∂ – пороговое значение; $d(q_i, q_j)$ – расстояние между центрами q_i и q_j ; N – число центров.

Когда все входные *SURF*-векторы данного изображения заменяются и обозначаются соответствующими «словами», проводится построение гистограммы, в которой указывается частота появления каждого ключевого слова из словаря. Эту гистограмму будем называть «*BOW*-дескриптором». Для каждого объекта *BOW*-дескриптор имеет фиксированную длину, равную числу слов (кластеров) в словаре признаков. Следует отметить, что *BOW*-дескриптор принимает значения, которые характерны для каждого класса объекта, и, таким образом, указанные значения идеально подходят для нейронной сети.

Поскольку *SURF* выделяет локальные признаки, не учитывая относительные связи между ними, и в гистограмме вычисляется частота появления определенных *SURF* признаков, пространственные отношения между признаками объекта также не учитываются. В распознавании руки, данное свойство *BOW*-дескрипторов очень полезно, так как нужно обучать нейронную сеть только один раз, скажем, для распознавания левой руки. Существенным является тот факт, что при тестировании нейронная сеть оказывается способной распознавать еще и правую руку. Таким образом, мы получаем возможность распознавать обе руки одновременно в одном классификаторе.

При разработке алгоритма для распознавания формы руки в режиме реального времени скорость работы классификатора является приоритетным показателем. Многослойная нейронная сеть удовлетворяет этому требованию. Данный тип нейронной сети также обеспечивает высокую точность распознавания. Кроме того, *BOW*-дескрипторы с низкой размерностью и фиксированной длиной идеально подходят для нейронной сети. Поэтому в данной работе в качестве классификатора применяется многослойная нейронная сеть с обратным распространением ошибки.

Используемая нейронная сеть в предложенном алгоритме состоит из N входных нейронов, где N – количество слов в словаре, K выходных нейронов, где K – число классов. Количество нейронов в скрытом слое определяется эмпирическим способом. В качестве функции активизации применяется симметричная биполярная сигмоидальная функция:

$$f(x) = \frac{\beta(1 - e^{-\alpha x})}{(1 + e^{-\alpha x})}.$$

Здесь α и β – свободные параметры.

Для повышения скорости сходимости сети применяются модификации: *Nguyen-Widrow* инициализация, моментум, и групповое обновление.

Реализация и результаты экспериментов

В работе реализована система распознавания четырех классов позы руки (рис. 4) с помощью предложенного алгоритма.

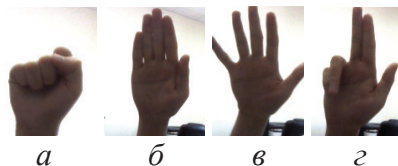


Рис. 4. Классы используемых объектов в программе: а) камень; б) бумага; в) открытая бумага; г) ножницы

На этапе генерации «словаря» для каждого класса отбираются по 25 «чистых» изображений (без фона) с объектом, снятым в различных ракурсах и масштабах. Количество выделенных признаков зависит от значения порога Гессена. Если установить маленькое значение порога, то количество признаков

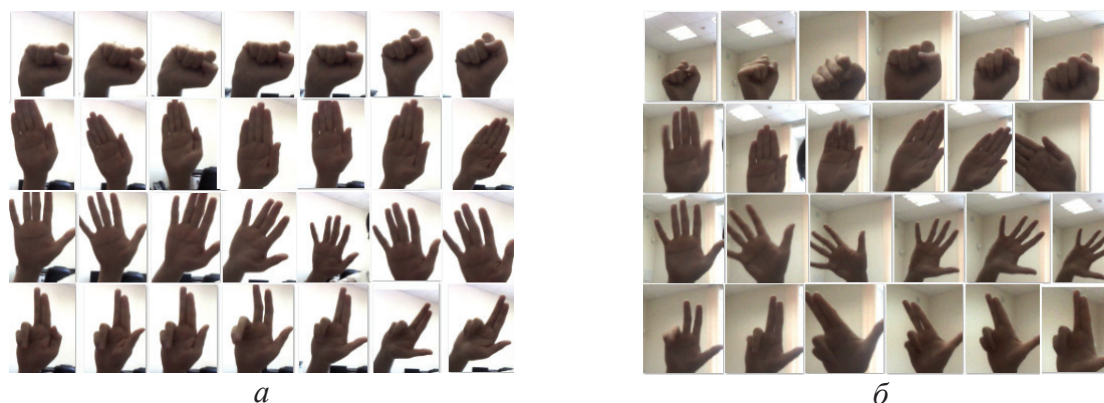


Рис. 5. Изображения: а) часть базы данных для обучения; б) тестовые

увеличивается, но при этом требуется длительное время обработки. Если установить большой порог, то число выделенных признаков может быть не выгодным для распознавания. Для изображения заданного размера порог Гессена был выбран таким образом, чтобы для изображений класса (а) SURF выделял в среднем 30 признаков; для класса (б) – 70; для класса (в) – 90; для класса (г) – 80. В этом случае число слов в словаре равно 270.

Для обучения нейронной сети в базу данных добавляются еще 75 изображений для каждого класса. Изображения для обучения были сняты на простом фоне для повышения робастности классификатора (рис. 5, а). Для тестирования классификатора веб-камерой были сняты 1000 изображений для каждого класса (рис. 5, б). Размер каждого изображения составляет 100×100 пикселей.

Программа для апробации работы алгоритма была реализована на языке С#. Программа предоставляет возможность генерации словаря признаков, формирования базы данных для обучения, создания и обучения нейронной сети, а также име-

ет отдельный интерфейс для проверки и тестирования работоспособности. Тестирования были выполнены на ноутбуке ASUS UL VT80 с операционной системой Windows 7 64 bit и встроенной веб-камерой. Результаты тестирования выборки из 1000 различных изображений при времени обработки одного кадра порядка 60 мс показывают: камень – правильных ответов 874; бумага – 912; открытая бумага – 945; ножницы – 932.

Таким образом, средняя точность распознавания составляет ~92 %. Самая высокая точность получается при распознавании «открытой бумаги». Это можно объяснить тем, что для «открытой бумаги» число выделенных признаков больше всех. Для «камня» количество признаков намного меньше остальных. Это и объясняет низкую точность распознавания класса «камней».

Время обработки для всех изображений составляет ~60 мс, что позволяет работать в реальном времени со скоростью около 15 кадров в секунду. На рис. 6 представлены результаты работы программы в реальном времени.



Рис. 6. Распознавание руки в реальном времени

Выводы

Предложен и реализован новый алгоритм для распознавания руки на основе *SURF*-дескрипторов и нейронной сети. Изложен новый метод генерации дескрипторов для нейронной сети, которые устойчивы к поворотам и масштабированию, частично к изменению освещения, и робастны на сложном фоне. Созданные дескрипторы позволяют распознавать левую и правую руки одновременно, используя только один классификатор, при этом обучение требуется только один раз для одной руки. Алгоритм

позволяет распознавать форму руки с высокой точностью. На основе предложенного алгоритма реализовано программное обеспечение для распознавания формы руки в видеопотоке в режиме реального времени. Программа поддерживает генерацию «словаря» признаков, формирование базы данных для обучения, создание и обучение нейронной сети. Проведенные численные эксперименты по распознаванию формы руки на видеопоследовательности в реальном времени показали, что средняя точность распознавания составляет 92 %.

СПИСОК ЛИТЕРАТУРЫ

1. Zhou H., Huang T.S. Tracking articulated hand motion with Eigen dynamics analysis // Proc. of Intern. Conf. on Computer Vision. – Nice (France), 2003. – V. 2. – P. 1102–1109.
2. Chen F., Fu C., Huang C. Hand gesture recognition using a real-time tracking method and Hidden Markov Models // Image and Vision Computing. – Oxford: Elsevier, 2003. – V. 21. – № 8. – P. 745–758.
3. Ng C.W., Ranganath S. Gesture recognition via pose classification // Proc. 15th Intern. Conf. on Pattern Recognition. – Barcelona (Spain), 2000. – V. 3. – P. 699–704.
4. Oka K., Sato Y., Koike H. Real-time fingertip tracking and gesture recognition // Proc. IEEE Computer Graphics and Applications. – 2002. – V. 22. – № 6. – P. 64–71.
5. Huang C., Jeng S. A model-based hand gesture recognition system // Machine Vision and Application. – 2001. – V. 12. – № 5. – P. 243–258.
6. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // IEEE CVPR01. – 2001. – P. 511–518.
7. Kolsch M., Turk M. Robust hand detection // Proc. of the 6th IEEE Intern. Conf. on Automatic Face and Gesture Recognition. – Seoul (Korea), May 2004. – Seoul, 2004. – P. 614–619.
8. Alper Y., Omar J., Mubarak S. Object tracking: A Survey // ACM Computing Surveys (CSUR). – 2006. – V. 38. – Iss. 4. – Art. № 13.
9. Bradski G.R. Real time face and object tracking as a component of a perceptual user interface // Proc. of the 4th IEEE Workshop on Applications of Computer Vision. – New Jersey (USA), October 1998. – New Jersey, 1998. – P. 214–219.
10. Bay H., Ess A., Tuytelaars T., Van Gool L. SURF: Speeded Up Robust Features // Computer Vision and Image Understanding. – Oxford: Elsevier, 2008. – V. 110. – № 3. – P. 346–359.

Поступила 11.03.2012 г.

УДК 004.932

РАСПОЗНАВАНИЕ ЛИЦ НА ОСНОВЕ ПРИМЕНЕНИЯ МЕТОДА ВИОЛЫ–ДЖОНСА, ВЕЙВЛЕТ–ПРЕОБРАЗОВАНИЯ И МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

Буй Тхи Тху Чанг, Фан Нгок Хоанг, В.Г. Спицын

Томский политехнический университет
E-mail: trangbt.084@gmail.com

На основе совместного применения метода Виолы–Джонса, вейвлет-преобразования и метода главных компонент предложен новый алгоритм распознавания лиц на цифровых изображениях и видеопоследовательностях в режиме реального времени. Описан алгоритм и разработано программное обеспечение для распознавания лиц. Приведен пример работы и представлены результаты тестирования программы. Показано, что использование предложенного оригинального алгоритма дает возможность эффективного распознавания лиц на цифровых изображениях и видеопоследовательностях.

Ключевые слова:

Распознавание лиц, метод Виолы–Джонса, метод главных компонент, вейвлет-преобразование.

Key words:

Face recognition, method Viola-Jones, principal component analysis, wavelet transform.

Введение

Распознавание лиц является одной из самых изученных задач в таких областях как цифровая обработка изображений, компьютерное зрение, биометрия, организация видеоконференций, создание интеллектуальных систем безопасности и контроля доступа и т. п. Процесс распознавания лиц

обычно состоит из двух этапов: поиск области лица на изображении, и сравнение найденного лица с лицами, находящимися в базе данных. В настоящее время метод Виолы–Джонса [1] является самым популярным методом для поиска области лица на изображении из-за его высокой скорости и эффективности. Детектор лица Виолы–Джонса

основан на основных идеях: интегральном представлении изображения, методе построения классификатора на основе алгоритма адаптивного бустинга (*AdaBoost*), и методе комбинирования классификаторов в каскадную структуру. Эти идеи позволяют построить детектор лица, способный работать в режиме реального времени. Метод главных компонент и вейвлет-преобразование используются для получения характеристик изображения. В задаче распознавания лиц они успешно применяются для сравнения компонент, характеризующих цветные изображения, с компонентами, описывающими неизвестные изображения.

Целью данной работы является создание нового алгоритма, основанного на комбинации метода Виолы–Джонса, вейвлет-преобразования и метода главных компонент (МГК) для распознавания лиц на цифровых изображениях и видеопоследовательностях в режиме реального времени.

1. Метод Виолы–Джонса

Метод был разработан и представлен в 2001 г. Полом Виолой и Майклом Джонсом. Он до сих пор является эффективным методом для поиска объектов на изображениях и видеопоследовательностях в режиме реального времени [1, 2]. Следует отметить, что этот детектор обладает крайне низкой вероятностью ложного обнаружения лица. Метод хорошо работает и обнаруживает черты лица даже при наблюдении объекта под небольшим углом, примерно до 30° . При угле наклона больше 30° вероятность обнаружения лица резко падает. Указанная особенность метода не позволяет в стандартной реализации детектировать лицо человека, повернутое под произвольным углом, что в значительной мере затрудняет или делает невозможным использование алгоритма в современных производственных системах с учетом их растущих потребностей.

1.1. Интегральное представление изображений

Для того, чтобы рассчитать яркость прямоугольного участка изображения, используют интегральное представление [3]. Такое представление используется часто и в других методах, например, в вейвлет-преобразованиях, *Speeded up robust feature* (SURF), фильтрах Хаара и многих разработанных алгоритмах. Интегральное представление позволяет быстро рассчитывать суммарную яркость произвольного прямоугольника на данном изображении, причем время расчета не зависит от площади прямоугольника.

Интегральное представление изображения представляет собой матрицу, совпадающую по размерам с исходным изображением. В каждом ее элементе хранится сумма интенсивностей всех пикселей, находящихся левее и выше данного элемента. Элементы матрицы рассчитываются по следующей формуле:

$$I(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'),$$

где $I(x, y)$ – значение точки (x, y) интегрального изображения; $i(x, y)$ – значение интенсивности исходного изображения. На основе применения интегрального представления изображения вычисление признаков одинакового вида, но с разными геометрическими параметрами, происходит за одинаковое время.

Каждый элемент матрицы $I(x, y)$ представляет собой сумму пикселей в прямоугольнике от $i(0, 0)$ до $i(x, y)$, т. е. значение каждого элемента $I(x, y)$ равно сумме значений всех пикселей левее и выше данного пикселя $i(x, y)$. Расчет матрицы занимает линейное время, пропорциональное числу пикселей в изображении и его можно производить по следующей формуле:

$$I(x, y) = i(x, y) - I(x-1, y-1) + \\ + I(x, y-1) + I(x-1, y).$$

Интегральное представление имеет интересную особенность. По интегральной матрице можно очень быстро вычислить сумму пикселей произвольного прямоугольника.

1.2. Хаар-подобные характеристики

С точки зрения необходимости использования достаточно простых алгоритмов получения признаков, перспективным является использование Хаар-подобных характеристик, представляющих собой результат сравнения яркостей в двух прямоугольных областях изображения.

Предположим, что задано множество объектов A и множество допустимых ответов B . Пусть $g: A \rightarrow B$ называется решающей функцией. Решающая функция g должна допускать эффективную компьютерную реализацию, по этой причине её также называют алгоритмом. Признак (*feature*) f объекта a – отображение $f: A \rightarrow D_f$, где D_f – множество допустимых значений признака. В частности, любой алгоритм $g: A \rightarrow B$ также можно рассматривать как признак. Если задан набор признаков f_1, \dots, f_n , то вектор $x = (f_1(a), \dots, f_n(a))$ называется признаковым описанием объекта $a \in A$. Признаковые описания допустимо отождествлять с самими объектами. При этом множество $A = D_{f_1} \times \dots \times D_{f_n}$ называют признаковым пространством [4].

В стандартном методе Виолы–Джонса используются прямоугольные признаки, рис. 1. Эти признаки называются примитивами Хаара.

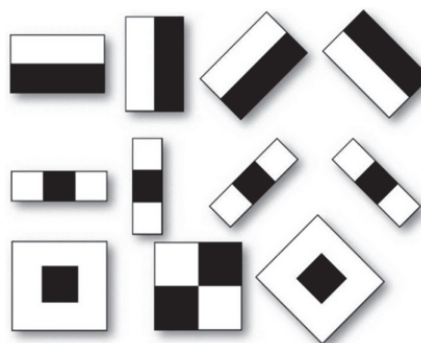


Рис. 1. Примитивы признаков Хаара

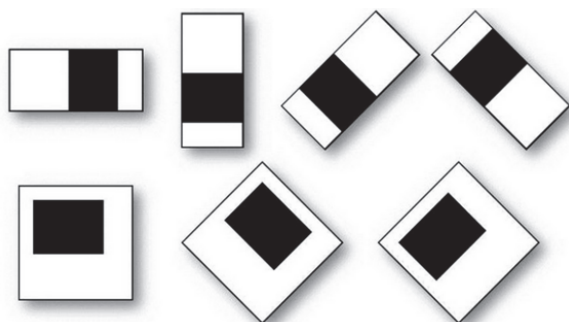


Рис. 2. Дополнительные признаки Хаара

В расширенном методе Виолы–Джонса, представленном в библиотеке OpenCV, используются дополнительные признаки (рис. 2).

Вычисляемым значением такого признака будет:

$$F = U - V,$$

где U – сумма значений яркостей точек, закрываемых светлой частью признака, а V – сумма значений яркостей точек, закрываемых темной частью признака. Для их вычисления используется понятие интегрального изображения. Хаар-подобные признаки описывают значение перепада яркости по оси X и Y изображения соответственно.

1.3. Метод построения классификатора на основе алгоритма бустинга

Бустинг – комплекс методов, способствующих повышению точности аналитических моделей. Бустинг (*boosting*) означает дословно «усиление» «слабых» моделей [5] – это процедура последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих алгоритмов. Идея бустинга была предложена Робертом Шапиро (*Schapire*) в конце 90-х гг. прошлого века [6], когда надо было найти решение вопроса о том, каким образом имея множество плохих (незначительно отличающихся от случайных) алгоритмов обучения, получить один хороший.

В результате работы алгоритма бустинга на каждой итерации формируется простой классификатор вида:

$$h_j(z) = \begin{cases} 1, & \text{если } p_j f_j(z) < p_j \theta_j, \\ 0, & \text{иначе} \end{cases},$$

где p_j – показывает направление знака неравенства; θ_j – значение порога; $f_j(z)$ – вычисленное значение признака; z – окно изображения размером 24×24 пикселей. Полученный классификатор имеет минимальную ошибку по отношению к текущим значениям весов, задействованным в процедуре обучения для определения ошибки.

1.4. Метод комбинирования классификаторов в каскадную структуру

Каскадная структура повышает скорость обнаружения, фокусируя свою работу на наиболее информативных областях изображения.

Структура каскадного детектора приведена на рис. 3. Каскад состоит из слоев, которые представляют собой классификаторы, обученные с помощью процедуры бустинга.

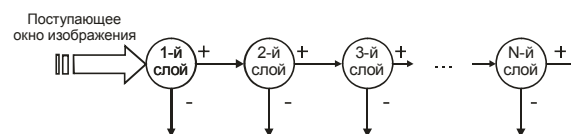


Рис. 3. Структура каскадного детектора

2. Вейвлет-преобразование

Вейвлет-преобразование широко используется для анализа нестационарных процессов. Оно показало свою эффективность для решения широкого класса задач, связанных с обработкой изображения. Коэффициенты вейвлет-преобразования содержат информацию об анализируемом процессе и используемом вейвлете. Поэтому выбор анализируемого вейвлета определяется тем, какую информацию необходимо извлечь из процесса. Каждый вейвлет имеет характерные особенности во временной и частотной областях, поэтому иногда с помощью разных вейвлетов можно полнее выявить и подчеркнуть те или иные свойства анализируемого процесса.

В работах [7, 8] представлены разложение изображения и извлечение его признаков для классификации изображений самолетов на основе применения вейвлет-преобразования Хаара и многослойной нейронной сети. В данной работе используются вейвлет-преобразования Хаара и Добеши для извлечения признаков изображения лиц. Примеры применения вейвлет-преобразования Хаара и вейвлет-преобразования Добеши для извлечения признаков изображения лица представлены на рис. 4.

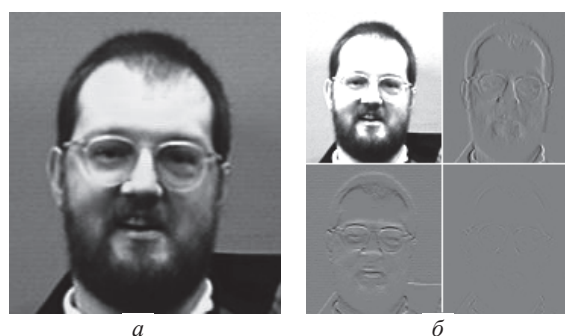


Рис. 4. Пример извлечения признаков лица: а) исходное изображение лица; б) результат после применения вейвлет-преобразования Хаара

3. Метод главных компонент

Метод главных компонент (*Principal Component Analysis, PCA*) – один из наиболее распространенных методов для уменьшения размерности данных, потери наименьшего количества информации. Он заключается в линейном ортогональном преобразовании входного вектора P размерности N в вы-

ходной вектор Q размерности M , $M < N$. Компоненты вектора Q являются некоррелированными, и общая дисперсия после преобразования остаётся неизменной.

Главная идея МГК состоит в представлении изображений лиц людей в виде набора главных компонент изображений, называемых «собственные лица» (*Eigenfaces*). Собственные лица имеют полезное свойство, заключающееся в том, что изображение, соответствующее каждому такому вектору имеет лицеподобную форму, рис. 5. Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы, которая рассчитывается из изображения. Сумма главных компонент, умноженных на соответствующие собственные вектора, является реконструкцией изображения, рис. 6.



Рис. 5. Пример изображений собственных векторов (собственные лица)

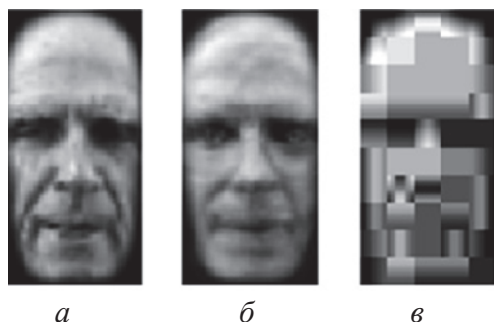


Рис. 6. Пример изображений лица: а) выровненное изображение; б) реконструкция по 85-и главным компонентам; в) JPEG-реконструкция (530 байт)

Для каждого изображения лица вычисляются его главные компоненты. Обычно берётся от 5 до 200 главных компонент. Остальные компоненты кодируют мелкие различия между лицами и шумами. Процесс распознавания заключается в сравнении главных компонент неизвестного изображения с компонентами всех известных изображений. При этом предполагается, что изображения лиц, соответствующих одному человеку, сгруппированы в кластеры в собственном пространстве. Из базы данных выбираются изображения-кандидаты, имеющие наименьшее расстояние от входного (неизвестного) изображения [9].

4. Алгоритм распознавания лиц

Целью данной работы является распознавание лиц с высокой скоростью. Для решения задачи распознавания лиц предложен алгоритм на основе применения метода Виолы–Джонса, вейвлет-преобразования и метода главных компонент. Предло-

женный алгоритм состоит из двух процессов: сохранения признаков известных лиц в базе данных и распознавания лиц.

Процесс сохранения признаков известных лиц происходит следующим образом:

- Шаг 1.** Преобразование изображения видео фрейма в полутоновое изображение.
- Шаг 2.** Применение к полутоновому изображению метода Виолы–Джонса для поиска области лица.
- Шаг 3.** Уменьшение размера области лица до 64×64 пикселей.
- Шаг 4.** Применение к полученному на шаге 3 изображению вейвлет-преобразования для извлечения признаков лица (вейвлет-коэффициентов).
- Шаг 5.** Сохранение извлеченных признаков в базе данных.

В процессе распознавания неизвестного лица осуществляются шаги 1–4, затем на основе применения метода главных компонент происходит сокращение числа признаков и их сравнение с признаками, хранящимися в базе данных. Функциональная схема предложенного алгоритма представлена на рис. 7.

5. Эксперименты

Для распознавания лиц на изображениях и видеопоследовательностях в реальном времени на языке объектно-ориентированного программирования C# (*Visual studio 2010*) с использованием библиотеки OpenCV разработано программное обеспечение.

5.1. Распознавание изображений лиц

Целью исследования является поиск изображений лиц, которые соответствуют заданным тестовым образам. Численные эксперименты выполнены на основе базы изображений лиц [10]. Примеры изображений представлены на рис. 8.

База содержит 366 изображений лиц различных людей, по 20 изображений каждого лица. При формировании базы данных размер изображений и условия съемки были одинаковы. Применялся 24 битный формат JPEG. В базе [10] содержатся снимки людей мужского и женского пола, разных национальностей и возрастов. В ней отражены изменения внешности человека: различные причёски, присутствие бороды и очков.

В процессе подготовки к экспериментам были созданы две обучающие выборки. Первая из них содержит 5 изображений каждого человека (всего $5 \times 366 = 1830$ изображений). Вторая – 10 изображений лиц каждого человека для обучения (всего $10 \times 366 = 3660$ изображений). Тестовая выборка состоит из 20-ти изображений лиц каждого человека (всего $20 \times 366 = 7320$ изображений). Используются два типа вейвлет-преобразования: Хаара и Добеши. Результаты экспериментов показаны в таблице.

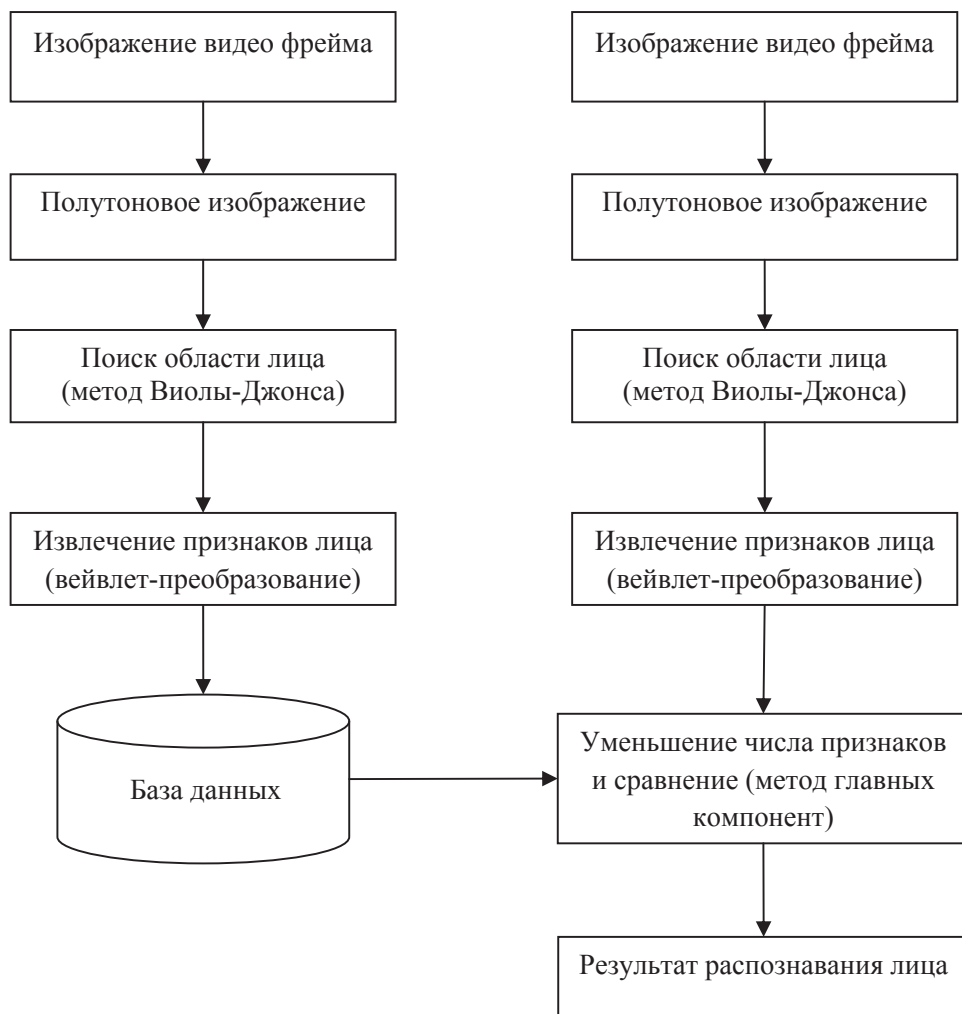


Рис. 7. Функциональная схема предложенного алгоритма распознавания лиц в видеопоследовательностях в режиме реального времени



Рис. 8. Примеры изображений лиц

Таблица. Количество распознаваний лиц

Тип вейвлет-преобразования	Хаар		Добеши	
	5 изоб.	10 изоб.	5 изоб.	10 изоб.
Достоверность распознаваний лиц, %	94,57	98,40	94,69	98,40

5.2. Распознавание лиц в видеопоследовательностях в режиме реального времени

На этапе обучения каждый человек находится перед камерой, совершая поступательные и вращательные движения головой. Программа записывает 10 различных изображений человека и сохраняет

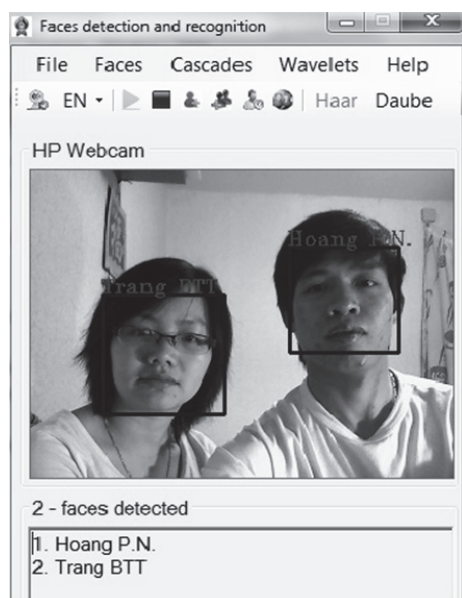


Рис. 9. Пример распознавания лиц в видеопоследовательностях в режиме реального времени

СПИСОК ЛИТЕРАТУРЫ

1. Viola P., Jones M.J. Rapid object detection using a boosted cascade of simple features // IEEE Conf. on Computer Vision and Pattern Recognition. – Kauai, Hawaii, USA, 2001. – V. 1. – P. 511–518.
2. Viola P., Jones M.J. Robust real-time face detection // International Journal of Computer Vision. – 2004. – V. 57. – № 2. – P. 137–154.
3. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005. – 1072 с.
4. Местецкий Л.М. Математические методы распознавания образов. – М.: МГУ, ВМиК, 2002–2004. – С. 42–44.
5. Sochman J., Matas J. AdaBoost with totally corrective updates for fast face detection // Center for Machine Perception, Czech Technical University, Prague. 2004. URL: <http://cmp.felk.cvut.cz/~sochmj1/> (дата обращения: 10.02.2012).
6. Freund Y., Schapire R.E. A Short Introduction to Boosting // Journal of Japanese Society for Artificial Intelligence – September 1999. – V. 14. – № 5. – P. 771–780.
7. Буй Тхи Тху Чанг, Спицын В.Г. Разложение цифровых изображений с помощью двумерного дискретного вейвлет-преобра-

в базе данных с именем этого человека, рис. 9. Для проверки правильности работы программы этот процесс повторяется, и программа в реальном времени определяет – кто в данный момент находится перед видеокамерой. Скорость работы программы составляет 7–14 кадров/с.

Выводы

1. Предложен и описан новый алгоритм на основе совместного применения метода Виолы–Джонса, вейвлет-преобразования и метода главных компонент для распознавания лиц на изображениях и видеопоследовательностях в режиме реального времени.
2. Разработана программа, реализующая предложенный алгоритм распознавания лиц на изображениях и видеопоследовательностях в режиме реального времени на языке объектно-ориентированного программирования C#.
3. Анализ результатов компьютерных экспериментов позволяет говорить об успешной работе созданного алгоритма и программы при распознавании лиц.

зования и быстрого преобразования // Известия Томского политехнического университета. – 2011. – Т. 318. – № 5. – С. 73–76.

8. Буй Тхи Тху Чанг, Фан Нгок Хоанг, Спицын В.Г. Алгоритмическое и программное обеспечение для классификации цифровых изображений с помощью вейвлет-преобразования Хаара и нейронных сетей // Известия Томского политехнического университета. – 2011. – Т. 319. – № 5. – С. 103–106.
9. Брилюк Д., Старовойтов В. Распознавание человека по изображению лица и нейросетевые методы. 2002. URL: <http://daily.sec.ru/publication.cfm?pid=4425> (дата обращения: 10.02.2012).
10. Spacek L. Description of the Collection of Facial Images. 2008. URL: <http://cswww.essex.ac.uk/mv/allfaces/index.html> (дата обращения: 10.02.2012).

Поступила 24.02.2012 г.

УДК 004.932

АЛГОРИТМЫ ДЛЯ КЛАССИФИКАЦИИ ОТПЕЧАТКОВ ПАЛЬЦЕВ НА ОСНОВЕ ПРИМЕНЕНИЯ ФИЛЬТРА ГАБОРА, ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ И МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ

Фан Нгок Хоанг, В.Г. Спицын

Томский политехнический университет

E-mail: hoangpn285@gmail.com

Предложены алгоритмы для классификации изображений отпечатков пальцев по типам папиллярных узоров на основе применения фильтра Габора, вейвлет-преобразования Хаара, Добеши и многослойной нейронной сети. Проведены численные эксперименты и представлены результаты работы предложенных алгоритмов. Показано, что использование алгоритма, основанного на совместном применении фильтра Габора, пятиуровневого вейвлет-преобразования Добеши и многослойной нейронной сети, дает возможность эффективной классификации отпечатков пальцев.

Ключевые слова:

Классификация отпечатков пальцев, фильтр Габора, вейвлет-преобразование, нейронные сети.

Key words:

Fingerprint classification, Gabor filter, wavelet transform, neural networks.

Введение

Отпечатки пальцев являются одним из надежных биометрических признаков, успешно применяемыми для идентификации личности [1]. Преимуществами отпечатка пальца при идентификации личности являются:

- Уникальность отпечатков пальцев, отличающихся друг от друга и от других отпечатков пальцев любого другого человека. Даже у близнецов отпечатки пальцев отличаются.
- В отличие от паролей, ПИН кодов отпечатки пальцев невозможно потерять или забыть.
- Отпечатки пальцев не изменяются со временем.
- Отпечатки пальцев использовались в течение многих лет для идентификации личности, поэтому есть возможность апробации разрабатываемых алгоритмов с использованием существующих баз данных.

В каждом отпечатке пальца можно определить два типа признаков – глобальные и локальные. Глобальные признаки – характеристики отпечатка пальца, которые можно увидеть невооруженным глазом. Глобальные признаки включают в себя область образа, ядро, пункт «дельта», счетчик линий, папиллярный узор. Локальные признаки, называемые минуциями, являются маленькими уникальными точками для каждого отпечатка пальца, которые успешно используются для идентификации личности. У отпечатка пальца могут быть одинаковые глобальные признаки, но локальные признаки всегда являются уникальными.

Целью данной работы является создание алгоритма для классификации отпечатков пальцев по типам папиллярных узоров на основе совместного применения фильтра Габора, вейвлет-преобразования и нейронной сети. Решение этой задачи позволит ускорить поиск отпечатков пальцев в больших базах данных.

Глобальные признаки отпечатка пальца

Область образа является фрагментом отпечатка пальца, в котором расположены все глобальные признаки [1]. Отпечатки пальцев могут быть прочитаны и классифицированы на основе информации области образа. Минуции, которые используются для идентификации личности, могут находиться вне области образа, поэтому при идентификации личности лучше использовать информацию целого отпечатка пальца.

Ядро является точкой, которая находится в приблизительном центре отпечатка пальца, и используется в качестве ориентира для чтения и классификации.

Пункт «дельта» является начальной точкой, где происходит разделение или соединение бороздок папиллярных линий, может иметь вид очень короткой бороздки, в предельном случае – точки.

Счетчик линий – число папиллярных линий на области образа, либо между ядром и пунктом «дельта».

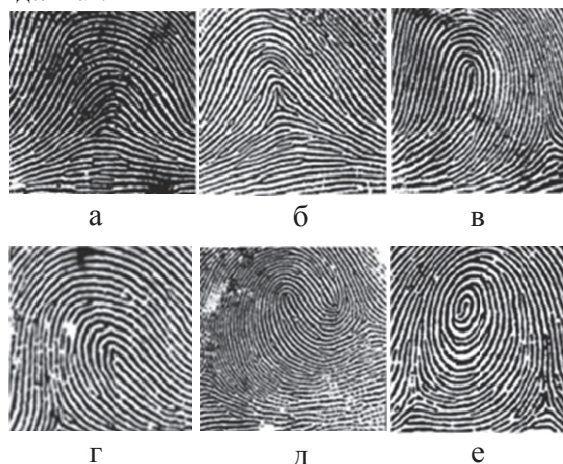


Рис. 1. Типы папиллярных узоров: а) простая арка; б) треугольная арка; в) левая петля; г) правая петля; д) двойная петля; е) завиток

Папиллярные узоры разделены на три типа: арки (дуги), петли и завитки [2]. Арки редко встречаются и занимают 5...10 % всех отпечатков пальцев. Петли обнаруживаются у большинства людей (60...65 %). Завитки проявляются значительно реже – 30 % из всех отпечатков пальцев. Примеры типов папиллярных узоров арки, петли и завитка представлены на рис. 1.

Локальные признаки отпечатка пальца

Линии отпечатков пальцев не являются прямыми. Они часто сломаны, разветвлены, меняют направление и имеют разрывы. Точки, в которых линии кончаются, разветвляются или изменяют направление, называются точками минуции. Эти точки минуции обеспечивают уникальную информацию об отпечатке пальца при идентификации личности. Примеры типов этих точек представлены на рис. 2.



Рис. 2. Типы точек минуции: 1) фрагмент папиллярной линии; 2) начало папиллярной линии; 3) глазок; 4) бифуркация; 5) крючок; 6) мостик; 7) островок; 8) точка; 9) окончание папиллярной линии; 10) бифуркация-слияние; 11) включение

Практика показывает, что отпечатки пальцев разных людей могут иметь одинаковые глобальные признаки, но совершенно невозможно наличие одинаковых локальных признаков, т. е. точек минуции. Поэтому процесс идентификации личности обычно состоит из двух этапов. Первым этапом является классификация отпечатков пальцев по глобальным признакам, использующая базы данных для разделения на классы. Вторым этапом заключается в распознавании отпечатка пальца на основе

сравнения структуры и коэффициента совпадения точек минуции.

Фильтр Габора

Фильтр Габора – линейный фильтр, импульсная переходная характеристика которого представляется в виде произведения функции Гаусса на гармоническую функцию [3]:

$$g(x, y) = \text{Gauss}(x', y') \cos\left(\frac{2\pi x'}{\lambda} + \varphi\right),$$

$$\text{Gauss}(x', y') = e^{-\left(\frac{x'^2}{2\sigma_x^2} + \frac{y'^2}{2\sigma_y^2}\right)},$$

$$x' = x \cos \theta + y \sin \theta,$$

$$y' = -x \sin \theta + y \cos \theta,$$

где λ – длина волны; φ – фаза; угол θ указывает ориентацию нормали к параллельным полосам функции Габора; γ – коэффициент сжатия. Изменение ориентации θ дает возможность менять направление обнаружения края. Примеры фильтра Габора с ориентациями в диапазоне $0... \pi$ с шагом $\pi/16$ представлены на рис. 3.

В обработке изображений фильтр Габора обычно используется для выделения краев, обнаружения контура объекта, извлечения признаков текстуры, выделения области образа отпечатка пальца, извлечения локального направления и других целей [4, 5]. В данной работе для улучшения изображения отпечатка пальца используются различные ориентации фильтра Габора, рис. 3. Пример результата применения фильтра Габора к изображению отпечатка пальца приведен на рис. 4.



Рис. 4. Пример результата применения фильтра Габора: а) исходное изображение отпечатка пальца; б) изображение, обработанное фильтром Габора

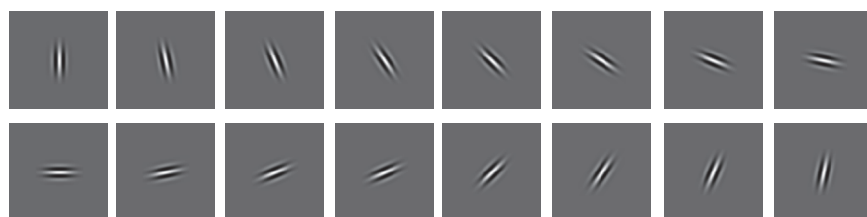


Рис. 3. Примеры фильтра Габора с различными ориентациями

Вейвлет-преобразование

Главной задачей в каждом виде обработки изображения является нахождение эффективного представления, позволяющего отобразить его в компактной форме. В современной теории и практике сигналов, в частности при спектральном анализе, используются сигналы специального вида – вейвлеты. В работах [6, 7] представлены разложение изображения и извлечение его признаков для классификации изображений самолетов на основе применения вейвлет-преобразования Хаара и многослойной нейронной сети. В данной работе используются вейвлет-преобразования Хаара и Добеши для извлечения признаков изображения отпечатка пальца. Пример применения вейвлет-преобразования Добеши 5-ого уровня для извлечения признаков изображения отпечатка пальца представлен на рис. 5.

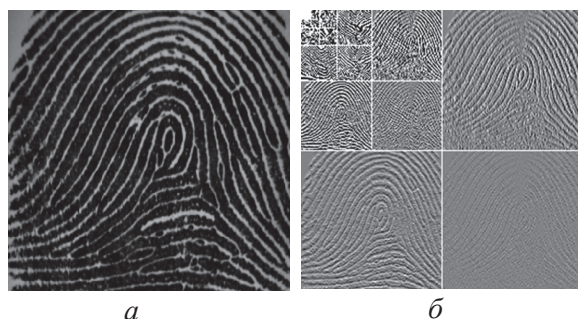


Рис. 5. Извлечение признаков отпечатка пальца: а) исходное изображение; б) полученные признаки на основе применения вейвлет-преобразования Добеши

Способ классификации отпечатков пальцев

В данной работе предложен способ классификации изображений отпечатков пальцев по типам папиллярных узоров на основе применения фильтра Габора, вейвлет-преобразования и нейронной сети. Функциональная схема предложенного способа представлена на рис. 6.

В предложенном способе используются вейвлет-преобразования Хаара и Добеши для извлечения признаков изображения отпечатка пальца. Для того чтобы оценивать эффективность извлечения признаков, используются вейвлет-преобразования 5-го и 6-го уровня. Создана нейронная сеть с 192 входами для вейвлет-преобразования 5-го уровня. Число скрытых нейронов для этой сети изменяется от 200 до 250. Для вейвлет-преобразования 6-го уровня создана нейронная сеть с 48 входами, для которой число скрытых нейронов изменяется от 80 до 120. Обе сети имеют 7 выходов в соответствии с числом категорий классификации отпечатков пальцев.

Эксперименты

Для тестирования работы алгоритмов используется часть базы данных отпечатков пальцев FVC2006 [8], содержащая 9 изображений каждой

категории для обучения (всего $9 \times 7 = 63$ изображений), и 14 изображений каждой категории для тестирования (всего $14 \times 7 = 98$ изображений). Результаты сравнения эффективности извлечения признаков представлены на рис. 7 и 8. Анализ приведенных результатов показывает, что алгоритм, использующий вейвлет-преобразование Добеши, дает лучшие результаты, чем алгоритм, основанный на вейвлете Хаара.

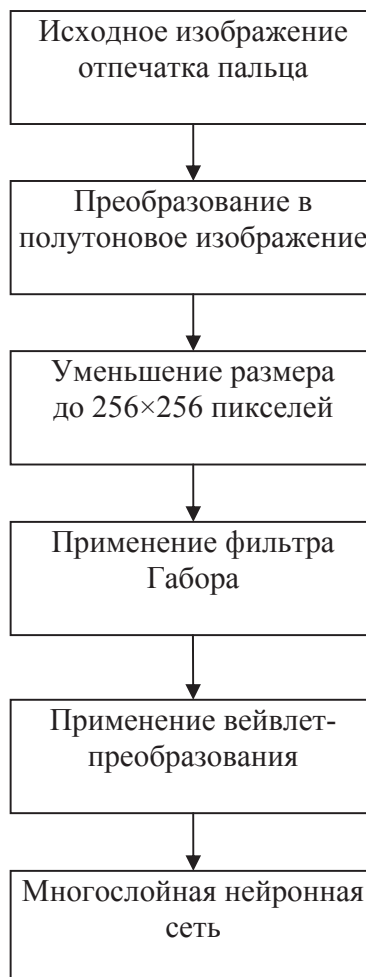


Рис. 6. Функциональная схема предложенного способа классификации отпечатков пальцев

Также проведено сравнение работоспособности алгоритма, использующего фильтр Габора, с алгоритмом без применения этого фильтра, рис. 9 и 10. Приведенные результаты показывают, что алгоритм, использующий фильтр Габора, работает эффективнее, чем алгоритм без применения этого фильтра.

Выводы

1. Предложены и описаны алгоритмы классификации изображений отпечатков пальцев по типам папиллярных узоров, основанные на применении фильтра Габора, вейвлет-преобразования Хаара, Добеши и многослойной нейронной сети.

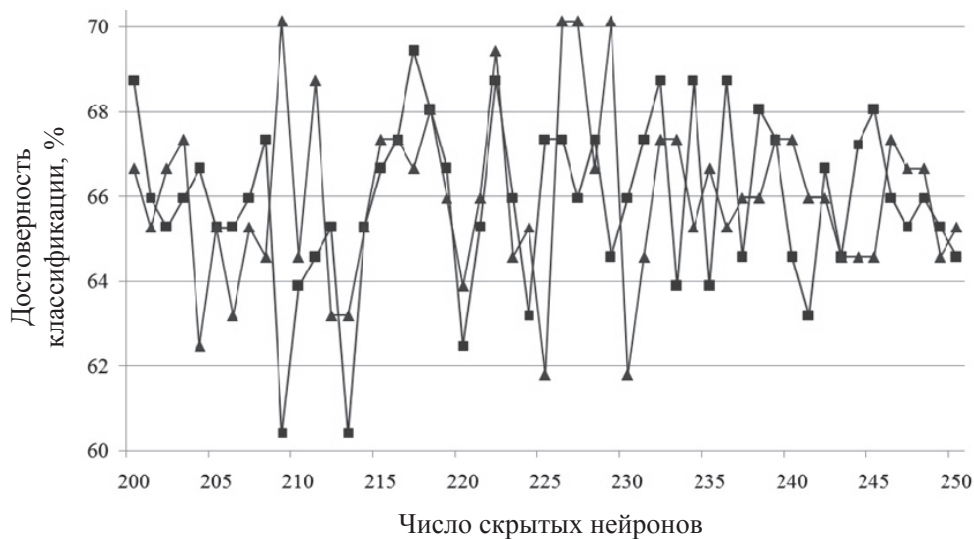


Рис. 7. Результат сравнения эффективности алгоритма на основе применения пятиуровневого вейвлет-преобразования: ■ – Добеши; ▲ – Хаара

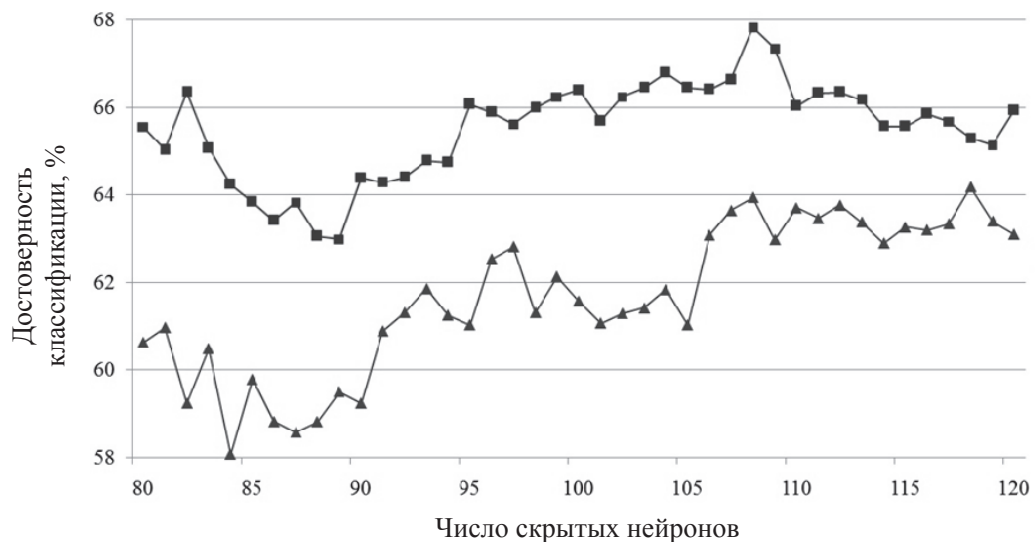


Рис. 8. Результат сравнения эффективности алгоритма на основе применения шестиуровневого вейвлет-преобразования: ■ – Добеши; ▲ – Хаара

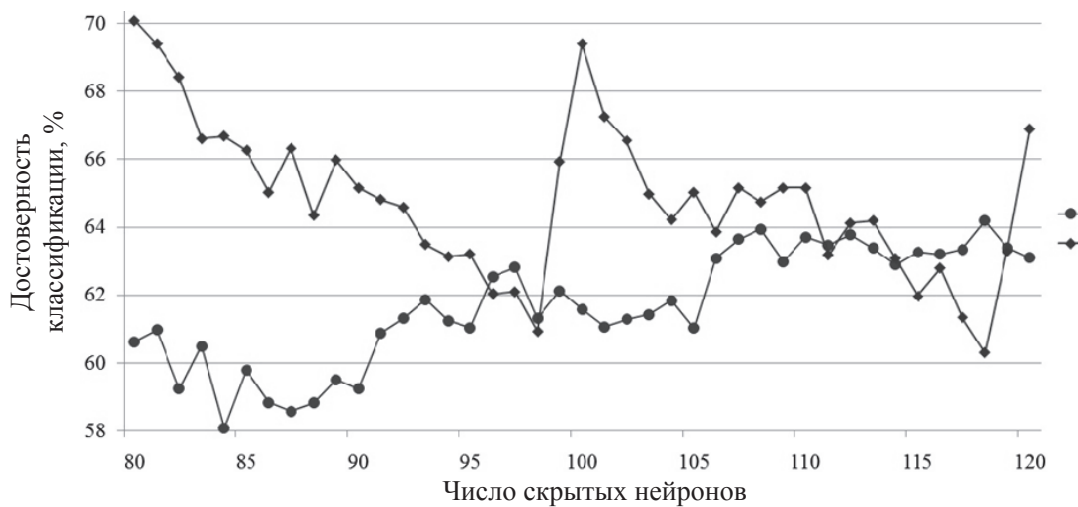


Рис. 9. Результат сравнения эффективности алгоритма на основе применения шестиуровневого вейвлет-преобразования Хаара и: ◆ – с фильтром Габора; ● – без фильтра Габора

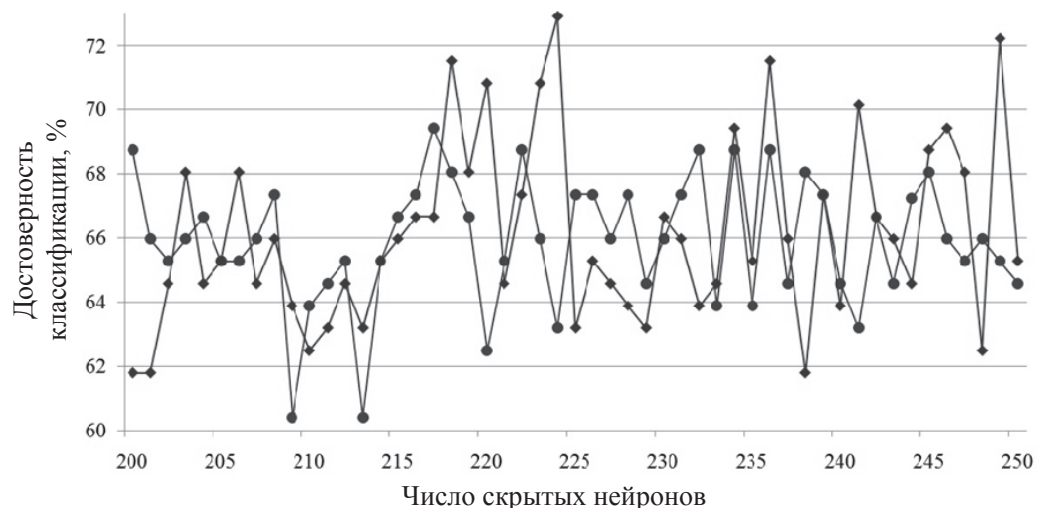


Рис. 10. Результат сравнения эффективности алгоритма на основе применения пятиуровневого вейвлет-преобразования Добеши и: ◆ – с фильтром Габора; ● – без фильтра Габора

2. На основе анализа результатов численных экспериментов установлено, что наилучшей достоверностью классификации отпечатков пальцев обладает алгоритм, основанный на совместном

применении фильтра Габора, пятиуровневого вейвлет-преобразования Добеши и многослойной нейронной сети.

Список литературы

1. Maltoni D., Maio D., Jain A.K., Prabhakar S. Handbook of fingerprint recognition. – N.Y.: Springer, 2003. – 348 p.
2. Ashbaugh D.R. Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology. – CRC Press, 1999. – 234 p.
3. Movellan J.R. Tutorial on Gabor filters. 2008. URL: <http://mplab.ucsd.edu/tutorials/gabor.pdf> (дата обращения: 01.02.2012).
4. Dolezel M., Hejtmanikova D., Busch C., Drahansky M. Segmentation procedure for fingerprint area detection in image based on enhanced Gabor filter // Intern. Conf. of Bio-Science and Bio-Technology. – Daedock, Korea, 2010. – № 3. – P. 39–50.
5. Bernard S., Boujemaa N., Vitale D., Bricot C. Fingerprint segmentation using the phase of multiscale Gabor wavelets // The 5th Asian Conf. on Computer Vision. – Melbourne, Australia, 2002. – P. 27–32.
6. Буй Тхи Тху Чанг, Спицын В.Г. Разложение цифровых изображений с помощью двумерного дискретного вейвлет-преобразования и быстрого преобразования // Известия Томского политехнического университета. – 2011. – Т. 318. – № 5. – С. 73–76.
7. Буй Тхи Тху Чанг, Фан Нгок Хоанг, Спицын В.Г. Алгоритмическое и программное обеспечение для классификации цифровых изображений с помощью вейвлет-преобразования Хаара и нейронных сетей // Известия Томского политехнического университета. – 2011. – Т. 319. – № 5. – С. 103–106.
8. Fierrez J., Ortega-Garcia J., Torre-Toledano D., Gonzalez-Rodriguez J. BioSec baseline corpus: A multimodal biometric database // Pattern Recognition. – 2007. – V. 40. – № 4. – P. 1389–1392.

Поступила 01.03.2012 г.

УДК 004.931

АЛГОРИТМЫ ВЫДЕЛЕНИЯ ЛИЦ НА СТАТИЧЕСКИХ RGB ИЗОБРАЖЕНИЯХ И В ВИДЕОПОТОКЕ

А.А. Друки

Томский политехнический университет

E-mail: druki2008@yandex.ru

Разработаны алгоритмы выделения лиц на статических изображениях и в видеопотоке: алгоритм обнаружения лица с помощью цветовой сегментации, алгоритм выделения лиц на статических RGB изображениях с помощью деформируемых эллиптических моделей, метод статических моментов для выделения лиц в видеопотоке.

Ключевые слова:

Обработка изображений, выделение лиц, метод деформируемых эталонов, метод статических моментов.

Key words:

Image recognition, selection of person, method of deformable standards, method of static moments.

Введение

Многие направления науки, техники и производства в значительной степени ориентируются на развитие систем, в которых информация носит характер поля (изображения). При обработке такой информации возникает ряд сложных научных, технических и технологических проблем. Одной из самых сложных задач является обработка и распознавание изображений. О важности этой проблемы говорит тот факт, что исследования по распознаванию образов, анализу изображений и речи включены в перечень приоритетных направлений развития науки и техники и критических технологий федерального уровня [1].

Распознавание изображений находит широкое применение в различных приложениях – это может быть контроль топологии печатных плат, текстуры ткани, контроль доступа к информации по идентификации личности (биометрическая идентификация), доступ к объектам ограниченного доступа, оперативный поиск в картотеке изображений, дактилоскопия и др. [2].

Широкое распространение получают биометрические системы идентификации человека. Традиционные системы идентификации требуют знания пароля, наличия ключа, идентификационной карточки, либо иного идентифицирующего предмета, который можно забыть или потерять. В отличие от них биометрические системы основываются на уникальных биологических характеристиках человека, которые трудно подделать и которые однозначно определяют конкретного человека. К таким характеристикам относятся отпечатки пальцев, форма ладони, узор радужной оболочки, изображение сетчатки глаза, индивидуальные характеристики лица [3].

Лицо человека является важным источником информации при общении между людьми. Выражение лица, мимика, артикуляция при разговоре, движения головой являются удобным, естественным и, что важно, необременительным способом передачи информации. Неспособность компьютера воспринять и распознать столь естественные для человека способы общения затрудняет передачу и восприятие информации при работе с ПЭВМ.

Для того, чтобы научить компьютер распознавать объекты и реагировать на их движения, требуются устойчивые алгоритмы анализа и распознавания изображения лица человека [4].

Цель работы является разработка алгоритмов обнаружения и выделения лица на статических и динамических изображениях низкого разрешения (порядка 320×240 пикселей) при наличии высокого уровня шума, характерного для бытовых фото- и видеокамер.

Разработка алгоритма обнаружения лица с помощью цветовой сегментации

Существующие алгоритмы обнаружения лица можно разбить на две категории: методы эмпирического распознавания и методы моделирования изображения лица.

Среди эмпирических методов обнаружения лица получили широкое распространение методы, использующие цветовую сегментацию областей кожи [5]. Причина этой популярности состоит в том, что использование цвета кожи как признака для обнаружения лица сочетает в себе несколько важных преимуществ:

- малая вычислительная сложность и, как следствие, высокая скорость обработки;
 - устойчивость к изменению ориентации и масштаба лица;
 - устойчивость к изменению освещения;
 - устойчивость к изменению выражения лица и частичного перекрытия лица другим объектом сцены.
- К недостаткам этого метода относятся:
- потенциальная возможность ложного обнаружения предметов с цветом, близким к цвету кожи;
 - зависимость результата от цветового баланса камеры и цвета освещения [6, 7].

Вышеперечисленные достоинства обнаружения лиц по цвету делают использование характерного цвета кожи как признака для распознавания весьма привлекательным, даже несмотря на имеющиеся недостатки.

Обнаружение лиц по цвету кожи производится в два этапа:

1. Выделение пикселей, близких по цвету к коже.
2. Поиск лиц на изображении с использованием априорного знания о структуре человеческого лица.

Для выделения пикселей с цветом кожи используется модель распределения оттенка кожи в цветовом пространстве. На основании сравнения существующих методов цветовой сегментации кожи оптимальным представляется метод моделирования распределения с помощью Байесовской карты вероятностей.

Этот метод заключается в вычислении отношения

$$\frac{P(\text{skin} / c)}{P(-\text{skin} / c)}$$

(соотношение вероятностей принадлежности цвета с координатами c к оттенку «кожи» или «не-кожи») для каждой точки дискретизированного цветового пространства по формуле

$$P_{out} = 2 \sum_{r \in \Omega_2} \frac{P_{skin}(x)}{S},$$

и последующего сравнения $\frac{P(\text{skin} / c)}{P(-\text{skin} / c)}$ с некоторым

выбранным порогом θ . $P(c/\text{skin})$ и $P(c/-\text{skin})$ напрямую высчитываются из набора тренировочных изображений, с областями кожи, сегментированными вручную. Априорные вероятности $P(\text{skin})$ и $P(-\text{skin})$ также можно вычислить, исходя из тренировочного набора. Однако, как легко показать, их можно принять равными 0,5, поскольку изменение априорных вероятностей влияет лишь на величину порога, используемого для обнаружения кожи, не изменяя соотношения процента ложных обнаружений и ложных промахов. Данная модель обладает высокими показателями вероятности успешного распознавания, малой вероятностью ложного обнаружения, и является инвариантной по отношению к выбору цветового пространства, в котором проводится сегментация.

Результатом цветовой сегментации кожи является информация о степени близости цвета каждого пикселя к цвету кожи. На основе этой информации формируется полутоновое изображение, где интенсивности пикселей установлены равными характеристике близости пикселя к цвету кожи.

Разработка алгоритма выделения лиц с помощью деформируемых эллиптических моделей

К сожалению, цветовая информация не может дать достаточно данных для устойчивого обнаружения лица. Неизбежные ошибки цветовой сегментации и области, близкие по оттенку к коже, не являющиеся лицами, создают помехи для распознавания [8]. Поэтому для повышения устойчивости обнаружения лиц был разработан дополнительный метод. Метод заключается в выделении эллиптических областей максимального размера, внутри которых плотность количества пикселей цвета кожи максимальна, в то время как на границе количество пикселей минимально.

На рис. 1 представлен пример желаемого выделения эллиптической области из неидеальных исходных данных: 1) выделенная эллиптическая область; 2) область цвета кожи

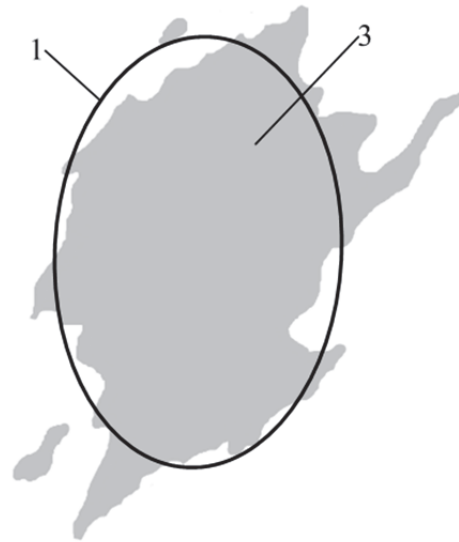


Рис. 1. Пример выделения эллиптической области из неидеальных исходных данных: 1) выделенная эллиптическая область; 2) область цвета кожи

Представляется естественным сформулировать эту проблему, как задачу оптимизации в пространстве параметров эллипса (x, y, θ, a, b) . Чтобы избежать трудностей, связанных с поиском явного задания целевой функции и разработки метода оптимизации, не подверженного проблеме локальных экстремумов, было разработано два алгоритма, использующих последовательные приближения с помощью деформируемой эллиптической модели лица, жестко фиксирующие возможные формы областей лиц.

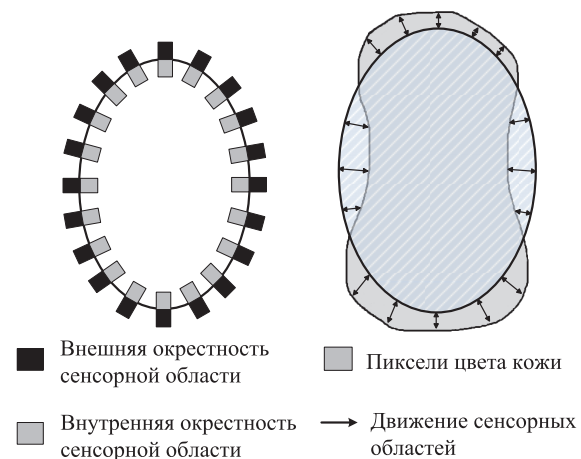


Рис. 2. Деформируемая эллиптическая модель для выделения области лица

Вдоль границы модели расположены сканируемые области, в которых происходит оценка плотности пикселей, близких по цвету к коже. Модель инициализируется вблизи ожидаемого положения лица на изображении (например, в центре масс связанной компоненты пикселей кожи). Деформация

модели производится по шагам. Для каждой сканируемой области производится расчёт плотности пикселей во внутренней и внешней окрестности границы модели. В зависимости от полученных плотностей сканируемая область перемещается перпендикулярно границе модели.

$$P_{in} = 2 \sum_{r \in \Omega 1} \frac{P_{skin}(x)}{S},$$

$$P_{out} = 2 \sum_{r \in \Omega 2} \frac{P_{skin}(x)}{S},$$

где $\Omega 1$ и $\Omega 2$ – внутренняя и внешняя части сканирующей области; S – площадь сканирующей области; $P_{skin}(x)$ – вероятность принадлежности цвета пикселя x к цвету кожи. Исходя из рассчитанных величин, вычисляется вектор перемещения сканирующей области:

$$v_i = \begin{cases} -k_{in} \vec{n}_i, & \text{если } P_{in}^i < T_1, \\ k_{out} \vec{n}_i, & \text{если } P_{in}^i \geq T_1 \text{ и } P_{out} < T_2, \\ 0, & \text{иначе.} \end{cases}$$

где i – номер сканирующей области; \vec{n}_i – единичная нормаль к границе модели в точке центра сканирующей области, направленная от центра модели; P_{in}^i и P_{out}^i – плотности пикселей кожи во внутренней и внешней частях сканирующей области; T_1 и T_2 – пороговые величины плотности пикселей. Коэффициенты K_{in} и K_{out} (≥ 0) управляют скоростью и характером перемещения сканирующих областей.

Полный цикл обнаружения потенциального региона лица модели задается следующим алгоритмом:

- 1) инициализация модели;
- 2) количество итераций $k=0$;
- 3) пока $k < n$:
- 4) для каждой сканирующей области $i=1, n$:
- 5) подсчитать P_{in}^i, P_{out}^i ;
- 6) если $P_{in}^i < T_{in}$ перемещение сканирующей области:

$$\mu_{ij} = \frac{\sum_{(x,y) \in S} (x - \mu_x)^i (y - \mu_y)^j I(x,y)}{\sum_{(x,y) \in SI} I(x,y)}, \text{ иначе:}$$

- 7) если $P_{out}^i \geq T_{out}$ перемещение сканирующей области: $\vec{v}_i = -\vec{n}_i v_{out}$, иначе:
- 8) $\vec{v}_i = \theta$;
- 9) пересчитать координаты центров сканирующих областей $x_i^{k+1} = x_i^k + v_i$;
- 10) аппроксимировать центры сканирующих областей $x_i^{k+1}, i=1, n$ эллипсом;
- 11) новые параметры формы модели = параметры полученного эллипса;
- 12) $k=k+1$, перейти на 2.

Для приближения центров сканирующих областей эллипсом был использован метод наименьших квадратов, обладающий высоким быстродействием и точностью.

Поведение модели можно варьировать и настраивать, модифицируя следующие параметры:

- количество, расположение и размер сканируемых областей;
- пороговые величины T_{in} и T_{out} ;
- коэффициенты длины векторов перемещения сканирующих областей k_{in} и k_{out} ;
- дополнительные ограничения на форму, размер, ориентацию и положение эллипса.

Путем модификации перечисленных параметров можно добиться тонкой настройки функционирования модели.

Модификация алгоритма

Для увеличения устойчивости метода в случае присутствия на изображении регионов кожных пикселей, значительно отличающихся из-за шума по форме от эллиптических, введены несколько усовершенствований базового метода:

- малое случайное перемещение сканирующих областей вдоль границы модели на каждом шаге;
- предобработка изображения кожи – заполнение дефектов в связанных компонентах кожи заданного меньшего размера, чем ожидаемый размер лица.

Для того, чтобы увеличить влияние сканирующих областей, находящихся на границе области, введена третья модификация метода. Пересчет параметров эллипса производится методом наименьших квадратов по набору координат центров сканирующих областей $x_i^{k+1}, i=1, n$. Особенность метода наименьших квадратов заключается в том, что ошибочное перемещение одной сканирующей области на большое расстояние может значительно сказаться на форме модели и привести к смещению модели с границы области кожи. Чтобы избежать такой ситуации и увеличить влияние стационарных сканирующих областей (в предположении, что они остановились на границе интересующей нас области), их центры участвуют несколько раз при пересчете параметров эллипса. Координаты центров стационарных сканирующих областей x_i дублируются l раз в массиве, передаваемом процедуре аппроксимации центром сканирующих областей эллипсом. Эксперименты показали, что $l=5$ дает хороший компромисс между устойчивостью к несвязностям внутри областей и точностью выделения границ регионов.

Метод статических моментов

для отслеживания перемещения лица

Задача отслеживания перемещения лица в видеопотоке обладает определенными особенностями по сравнению с обнаружением лица на статическом изображении. Во-первых, мы можем отталкиваться от положения лица на предыдущем кадре, во-вторых мы можем рассчитывать на то, что изменения в размере и, главное, пропорциях лица между кадрами будут невелики. С другой стороны, требуется значительно более высокое быстродействие и сходимость метода. Также, частым отличием при обработке видеопотока является заметно меньшее разрешение и худшее качество изображения по сравнению со статикой [9, 10].

Учитывая перечисленные особенности, был разработан метод отслеживания перемещения лица, также основанный на методе последовательных приближений и использующий жесткое ограничение на возможную форму выделяемых областей, но использующий всю информацию внутри области, а не только на ее границе.

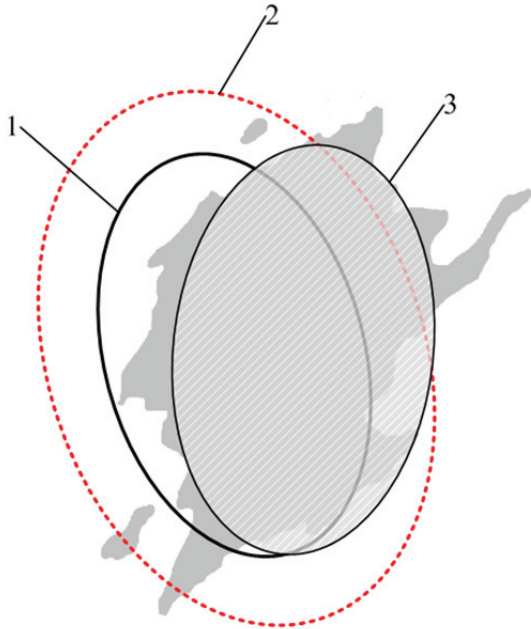


Рис. 3. Метод статистических моментов для выделения области лица: 1) текущее положение модели; 2) новое положение модели; 3) область для вычисления моментов

Метод основан на вычислении статистических моментов множества пикселей цвета кожи, попадающих в текущую обрабатываемую область:

$$\mu_x = \frac{\sum_{(x,y) \in SyI(x,y)} (x,y)}{\sum_{(x,y) \in SI(x,y)},}$$

$$\mu_y = \frac{\sum_{(x,y) \in SxI(x,y)} (x,y)}{\sum_{(x,y) \in SI(x,y)},}$$

$$\mu_{ij} = \frac{\sum_{(x,y) \in S(x-\mu_x)^i (y-\mu_y)^j I(x,y)} (x,y)}{\sum_{(x,y) \in SI(x,y)},}$$

где $I(x,y)$ – вероятность принадлежности пикселя (x,y) к цвету кожи.

Инициализация модели производится либо положением лица на предыдущем кадре (в случае обработки видео), либо в центре масс связной компоненты пикселей кожи (как в случае с предыдущим методом). На каждом шаге деформации производится пересчет новых параметров эллипса. Пусть \vec{v}_k – вектор параметров эллипса на n шаге (x,y,θ,a,b) . Здесь θ – угол наклона главной оси эллипса; a, b – длины осей эллипса.

Полный цикл обнаружения области лица указанным методом описан в следующем алгоритме:

0. Инициализация модели.
1. Количество итераций $k=0$.

2. Пока $k < n$.
3. Рассмотреть эллипс увеличенного размера $(x_k, y_k, \theta_k, \alpha a_k, \alpha b_k)$, $\alpha > 1$.

4. Подсчитать центр масс (μ_x, μ_y) и центральные моменты второго порядка $\mu_{20}, \mu_{02}, \mu_{11}$ и множества пикселей, попавших в рассматриваемую область.

5. Подсчитать новые параметры эллипса по полученным статистикам

$$v_{k+1} = \left(\mu_x, \mu_y, a \tan \left(\frac{\mu_{11}}{\mu_{20} - \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}} \right), a_{k+1}, b_{k+1} \right).$$

6. $k=k+1$, перейти на 2.

Новые длины осей эллипса a_{k+1}, b_{k+1} вычисляются как

$$a_{k+1} = \frac{1}{1+\beta} (a_k + d_k a_k \beta),$$

$$b_{k+1} = \frac{1}{1+\beta} (b_k + d_k b_k \beta),$$

где d_k – среднее расстояние Махаланобиса от центра эллипса лица до точек цвета кожи, находящихся внутри эллипса увеличенного размера; β – коэффициент, регулирующий скорость адаптации (по результатам экспериментов принятый равным 0,19).

Обычно двух-трех итераций пересчета параметров эллипса достаточно для вычисления нового положения лица, даже при достаточно больших перемещениях между кадрами.

Сравнение предложенных методов с существующими показало, что предлагаемые методы обладают заметно более высокой устойчивостью к зашумленным исходным данным и ошибкам цветовой сегментации, сохраняя простоту программной реализации и высокую скорость работы, позволяющую использовать разработанные алгоритмы в режимах реального времени.

Простые схемы алгоритмов гарантируют возможность получения быстрой программной реализации. Несмотря на простоту, они обладают высокой устойчивостью к шумным и неточным исходным картам вероятности кожи. Предложенные методы также обладают важным свойством устойчивости к выбору начального приближения, что выгодно отличает их от многих методов распознавания, основанных на деформируемых моделях. Описанная схема обладает большой гибкостью и позволяет с легкостью встраивать дополнительные условия и ограничения на возможную форму выделяемых областей (пропорции, положение, ориентации эллипса).

Учитывая устойчивость алгоритмов к точности начального приближения на основе описанной схемы, строится следующий порядок обнаружения лиц на изображениях:

- 1) провести сегментацию кожи, создать карту вероятности кожи;

- 2) выделить связные компоненты пикселей с высокой кожной вероятностью или подсчитать интегральные проекции карты кожи;
- 3) инициализировать модель в центрах масс связанных компонент или пиках интегральных проекций;
- 4) произвести обнаружение эллиптических областей пикселей кожи;
- 5) оценить их размеры, пропорции, форму и зафиксировать их как потенциальные лица, если проверка дает положительный результат;
- 6) убрать обнаруженные области из карты вероятностей;
- 7) повторять процедуру с шага 2, пока обнаруживаются новые области.

СПИСОК ЛИТЕРАТУРЫ

1. Визильтер Ю.В., Желтов С.Ю., Ососков М.В. Система распознавания и визуализации характерных черт человеческого лица в реальном времени на персональной ЭВМ с использованием web-камеры // Графикон: Труды конф. – Н. Новгород, 2002. – С. 251–254.
2. Rabiner L.R. A Tutorial on hidden Markov models and selected applications in speech recognition // IEEE scientific works compilation. – 1989. – V. 77. – № 2. – P. 257–285.
3. Jordao L., Perrone M., Costeira L., Santos L. Active face and feature tracking // Proc. Inter. Conf. on lineage analysis and processing. – Lisbon, 1999. – V. 9. – P. 572–576.
4. Albiol A., Torres L., Delp E.J. Optimum color spaces for skin detection // Proc. Inter. Conf. on image processing. – Valencia, 2001. – V. 1. – P. 122–124.
5. Baskan S., Bulut M.A., Atalay V. Projection based method for segmentation of human face and evaluation // Pattern Recognition Letters. – 2002. – V. 20. – P. 1623–1629.

Выводы

Предложены алгоритмы выделения лиц на статических RGB изображениях и в видеопотоке, работающие в реальном времени и удовлетворяющие следующим критериям:

- работа с изображениями низкого разрешения (порядка 320×240 пикселей) и высокого уровня шума, характерного для бытовых фото- и видеокamer;
- устойчивость к ошибкам и неточностям цветовой сегментации;
- быстрая работа методов;
- возможность гибкой настройки и задания ограничений на форму эллиптических моделей.

6. Brand J., Mason J. A comparative assessment of three approaches to pixellevel human skin – detection // Proc. Inter. Conf. on pattern recognition. – Swansea, 2002. – V. 1. – P. 1056–1059.
7. Craw I., Ellis H., Ashman J. Automatic extraction of face features // Pattern recognition letters. – 1987. – V. 5. – P. 183–187.
8. Edwards G., Taylor C., Cootes T. Interpreting face images using active appearance models // Proc. Conf. on automatic face and gesture recognition. – Manchester, 1998. – V. 7. – P. 300–305.
9. Hansen D.W., Hansen J.P., Nielsen M. Eye typing using Markov and active appearance models // Workshop on applications of computer vision. – 2002. – V. 12. – P. 132–136.
10. Feraud R., Bernier O., Viallet J., Collobert M. A fast and accurate face detector based on neural networks // Transactions on pattern analysis and machine intelligence. – 2001. – V. 34. – P. 42–53.

Поступила 10.12.2011 г.

УДК 004.93

ПРОСТОЙ АЛГОРИТМ КОНТРОЛЯ БИТРЕЙТА КОДЕРА СТАНДАРТА JPEG

Д.В. Сидоров, А.Н. Осокин

Томский политехнический университет

E-mail: rauco@mail.ru

Разработан алгоритм контроля коэффициента сжатия полутоновых реалистичных изображений для кодера стандарта JPEG, отличающийся простотой программной реализации и невысокой вычислительной сложностью. Проведено его исследование на наборе полутоновых изображений Calgary Corpus Gray Set 2. Сделаны выводы о пригодности предложенного алгоритма к использованию в системах видеонаблюдения и видеорегистрации при ограниченности вычислительных ресурсов и пропускной способности каналов передачи данных.

Ключевые слова:

Стандарт JPEG, дискретное косинусное преобразование, алгоритмы контроля битрейта, сжатие полутоновых изображений с потерями.

Key words:

JPEG coder, bit rate control algorithms, discrete cosine transform, compressing half-tone images.

Введение

В связи с появлением и продвижением стандарта JPEG2000, который предусматривает возможность контроля битрейта сжатых изображений, а также с развитием и внедрением новых стандартов связи, обеспечивающих большую пропускную способность, тема разработки алгоритмов контроля битрейта для кодера стандарта JPEG практически перестала освещаться как в зарубежной, так и в отечественной литературе [1–5]. Однако существует ряд прикладных задач, при решении которых использование современных стандартов сжатия (JPEG2000 вместо широко распространенного JPEG), каналов широкополосной связи, либо невозможно, либо сопряжено с рядом трудностей. Решение подобных прикладных задач актуализирует проблему разработки простого и нересурсоёмкого алгоритма контроля битрейта сжатых изображений кодера стандарта JPEG.

Алгоритмы контроля битрейта для кодера JPEG

Теоретическая база для разработки алгоритмов контроля битрейта при сжатии с потерями изображений и видео была заложена достаточно давно и основана на формуле отношения битовых затрат на передачу или хранение сигнала к величине вносимых в него искажений или ошибок [6]. На практике, в отношении кодера стандарта JPEG, данная модель часто сводилась к простой аппроксимации кривой отношения затрат к ошибке по нескольким точкам [7]. На основании полученной кривой выбирался тот вариант сжатия изображения, который при требуемом коэффициенте сжатия или битрейте давал наилучшее качество распакованного изображения. Очевидно, что данный алгоритм обеспечивает хорошие результаты, но обладает высокой вычислительной сложностью и, как следствие, малоприменим для решаемой задачи.

Более современные алгоритмы, основанные на минимизации функции Лагранжа [4, 5] (функ-

ция представляет собой взвешенную сумму ошибки и затрат), обычно позволяют сократить количество сжатий одного и того же изображения до 2–5 раз, но они все еще обладают достаточной вычислительной сложностью. Последний предложенный в работах [1–3] алгоритм, решающий данную проблему и признанный исследователями одним из лучших [2], основан на предположении о том, что существует линейная зависимость между количеством нулей, образовавшихся в ходе квантования коэффициентов после дискретного косинусного преобразования (ДКП), и размером сжатого изображения. Алгоритм позволяет, по сравнению со всеми ранее упомянутыми алгоритмами, значительно упростить процедуру поиска необходимых параметров сжатия (коэффициента квантования): исключить предварительный анализ несжатого изображения, сократить количество сжатий изображения до двух и свести расчет необходимого параметра (таблицы) квантования до простых линейных операций.

В ходе использования этого алгоритма при JPEG-сжатии его авторами [1] было обнаружено значительное отклонение полученных результатов от линейной зависимости, а именно: линейное отношение размера сжатого изображения к количеству нулей нарушалось при высоком коэффициенте сжатия, причём часто в большую сторону (размер сжатого изображения превышал необходимый).

Исходя из вышесказанного, авторами статьи была сформулирована следующая цель работы – разработать алгоритм контроля битрейта, который бы основывался на аналогичных алгоритму [1] принципах, отличался простотой программной реализации по сравнению с исходным, требовал менее двух полных процедур сжатия, легко вписывался в стандартную схему кодера JPEG и обеспечивал приемлемые результаты.

Описание предложенного алгоритма контроля битрейта

Основываясь на работах [1–3, 7], с учётом предъявленных ранее требований к алгоритму (простота программной реализации, малая ресурсоёмкость, менее двух полных сжатий изображения, простота интеграции в стандартную схему кодера JPEG), было сформулировано предположение о существовании близкой к линейной зависимости между количеством бит, необходимых для представления всего сжатого изображения, и количеством бит, необходимых для представления заданного количества сжатых ДКП-блоков изображения. Другими словами, во время поиска необходимого коэффициента квантования (качества) Q при JPEG-сжатии, например, методом дихотомии на отрезке $Q_n \in [5, 100]$, можно ограничиться перекодированием (сжатием) небольшого предварительно заданного количества ДКП-блоков, тем самым значимо снизив время поиска и вычислительную сложность алгоритма (количество обработанных пикселей).

В процессе такого поиска также должны выполняться следующие условия: *во-первых*, заданные ДКП-блоки должны быть равномерно распределены по всему изображению (необходимо учесть его особенности) и составлять некий заданный процент, далее обозначаемый как n , от их общего числа; *во-вторых*, при их сжатии не должны учитываться символы конца блока, показывающие нули в конце зигзаг-цепочки; *в-третьих*, максимальное количество сжатий одного и того же изображения должно быть менее двух, т. е. должно удовлетворяться неравенство $8(n/100)+1=n/12,5+1 < 2$ (дробная часть показывает, что была сжата только часть ДКП-блоков; 8 – это максимально возможное количество шагов в методе дихотомии во время поиска нужного коэффициента квантования на интервале $Q_n \in [5, 95]$), отсюда процент перекодируемых ДКП-блоков должен быть в интервале $n \in (0, 12,5)$ и, *в-четвертых*, ДКП-блоки необходимо после квантования упаковывать с помощью статических таблиц Хаффмана, т. к. это значительно повышает быстродействие.

Обобщая вышесказанное, авторами статьи был сформулирован следующий алгоритм контроля битрейта (алгоритм приведен для полутоновых изображений).

Начало.

Шаг 1. Задать битрейт (коэффициент сжатия) K .

Шаг 2. Согласно стандарту JPEG произвести ДКП, например, при помощи binDCT (быстрая реализация ДКП) [8].

Шаг 3. Из общего числа ДКП-блоков равномерно выбрать $n\%$ блоков, которые затем проквантовать с высоким качеством $Q_1 \in [90, 95]$ (используется стандартная рекомендованная таблица квантования) и сжать с помощью статических таблиц Хаффмана (используются рекомендованные в стандарте таблицы) [9].

Шаг 4. Используя подсчитанное в ходе выполнения *Шага 3* количество бит, которое необходимо для представления $n\%$ сжатых блоков, линейно вычислить предполагаемый конечный размер сжатого изображения; если полученный размер меньше необходимого при заданном битрейте K , то осуществляется переход на *Шаг 7*, иначе *Шаг 5*.

Шаг 5. Используя заданный битрейт K , линейно вычислить необходимое количество бит для представления $n\%$ сжатых ДКП-блоков (далее обозначено как S).

Шаг 6. Методом дихотомии произвести поиск необходимого коэффициента квантования (качества) из интервала $Q_n \in [5, Q_1]$, обеспечивающего при сжатии заданных $n\%$ ДКП-блоков размер, близкий к вычисленному на *Шаге 5* размеру S , поиск производить до тех пор, пока $|Q_n - Q_{n-1}| \geq 2$, а при сжатии ДКП-блоков использовать стандартные таблицы квантования и статические таблицы Хаффмана [10].

Шаг 7. На основании найденного коэффициента квантования Q_n (коэффициент всегда берется с левого конца отрезка) происходит квантование и сжатие всего изображения с применением динамических таблиц Хаффмана.

Конец.

Исследование алгоритма контроля битрейта

Цель исследования – получить количественные характеристики предложенного алгоритма контроля битрейта для кодера стандарта JPEG, т. е. определить, насколько точно алгоритм соблюдает заданный битрейт.

Исходные данные для исследования:

- стандартный, широкоиспользуемый и хорошо зарекомендовавший себя при решении подобных задач набор «Calgary Corpus Gray Set 2» из 12 полутоновых изображений размером 512×512 пикселей (глубина 8 бит на пиксель) [10];
- коэффициент сжатия $K \in [4, 30]$, шаг 1;
- максимальная ширина интервала коэффициента квантования $Q \in [4, 90]$, шаг 1;
- количество кодируемых ДКП-блоков $n=7,5\%$.

Минимальные и максимальные значения параметров K , Q и n были определены исходя из следующих соображений. Максимальное значение коэффициента сжатия K ограничено 30-ю, т. к. при дальнейшем его увеличении качество распакованных изображений набора становится неприемлемым. Диапазон значений коэффициента Q также ограничен соображением визуального качества распакованного изображения. Процент сжимаемых ДКП-блоков был предварительно подобран эмпирическим путем.

Часть результатов исследования приведена в таблице (показаны результаты для изображения «barb_1.bmp»). Полные результаты исследования для всех изображений набора доступны в работе [11]. Следует заметить, что определение качества

распакованного изображения, например, по метрике PSNR, SSIM [12], не имеет смысла, т. к. кодер полностью соответствует стандарту JPEG: используется рекомендованная в стандарте таблица квантования и не происходит отбрасывание каких-либо ДКП-коэффициентов.

Таблица. Результаты исследования алгоритма контроля битрейта для изображения «barb_1.bmp» из набора «Calgary Corpus Gray Set 2»

К	Изменение размера, байт/%	Q	Количество сжатий
4	-1590/-2,43	88	1,52
5	-928/-1,77	82	1,52
6	-1043/-2,39	75	1,52
7	-804/-2,15	67	1,37
8	-557/-1,70	58	1,44
9	-541/-1,86	49	1,59
10	399/1,52	44	1,52
11	382/1,60	38	1,52
12	-450/-2,06	32	1,52
13	-234/-1,16	29	1,52
14	252/1,35	27	1,52
15	-623/-3,56	23	1,52
16	-652/-3,98	21	1,52
17	312/2,02	21	1,52
18	-633/-4,35	18	1,52
19	133/0,96	18	1,52
20	-517/-3,94	16	1,52
21	107/0,86	16	1,52
22	675/5,67	16	1,52
23	-1577/-13,84	12	1,52
24	-1102/-10,09	12	1,52
25	-665/-6,34	12	1,52
26	-262/-2,60	12	1,52
27	111/1,14	12	1,52
28	-893/-9,54	10	1,52
29	-570/-6,31	10	1,52
30	-269/-3,08	10	1,52

Согласно полученным в ходе исследования результатам работы предложенного алгоритма можно отметить, что: *во-первых*, обеспечивается приемлемая погрешность, в среднем менее 10 %, контроля битрейта сжатого изображения; *во-вторых*, значительно снижается вычислительная сложность процесса сжатия по сравнению с существующими аналогами [1–3, 7], а именно – количество обработанных

пикселей не более чем в полтора раза превышает количество исходных (менее двух полных процедур сжатия) и, *в-третьих*, не происходит модификаций стандартной схемы сжатия алгоритма JPEG.

Предложенный алгоритм контроля битрейта, в отличие от распространенных аналогов, обладает малой ресурсоемкостью, приемлемой точностью контроля битрейта, не модифицирует схему JPEG-сжатия и, как следствие, может быть с успехом применен при решении прикладных задач по организации систем видеонаблюдения и видеорегистрации в условиях ограниченной производительности вычислительной аппаратуры и пропускной способности каналов передачи данных. Например, при разработке «умных» видеокамер для проходческих угледобывающих комбайнов на недорогих контроллерах ARM mx233 фирмы Freescale.

Данные камеры, благодаря невысоким требованиям предложенного алгоритма к производительности контроллера, могут быть дополнены, в частности, следующими функциями: автоматической фокусировки/позиционирования камеры на рабочий орган комбайна с целью определения комбайнером его состояния во время добычи; архивирования состояния рабочего органа в случае аварии комбайна и/или линий передачи данных вследствие взрыва газопылевой смеси; контроля комбайнером границ разрабатываемого пласта угля и т. д.

Выводы

1. Предложен алгоритм контроля коэффициента сжатия полутоновых изображений для кодера стандарта JPEG, основанный на использовании зависимости между размером всего сжатого изображения и размером, необходимым для заданного количества сжатых дискретно-косинусных блоков преобразования. Это позволяет в ходе контроля битрейта обрабатывать заметно меньшее количество пикселей и реализовывать алгоритм на простых контроллерах при сохранении свойств стандартного JPEG.
2. Алгоритм исследован на наборе полутоновых изображений Calgary Corpus Gray Set 2, погрешность контроля битрейта не превышает 10 %.
3. Показана возможность использования предложенного алгоритма в системах видеонаблюдения и видеорегистрации при ограниченности вычислительных ресурсов и пропускной способности каналов передачи данных.

СПИСОК ЛИТЕРАТУРЫ

1. Nguyen R.D. Rate control and bit allocation for JPEG transcoding: Master's thesis. – Cambridge: Massachusetts Institute of Technology, 2007. – 51 p.
2. He Z., Mitra S.K. A linear source model and a unified rate control algorithm for DCT video coding // IEEE Trans. on Circuit and System for Video Tech. – 2002. – V. 12. – № 11. – P. 970–982.
3. He Z., Mitra S.K. Optimum bit allocation and accurate rate control for video coding via p-domain source modeling // IEEE Trans. on Circuit and System for Video Tech. – 2002. – V. 12. – № 10. – P. 840–849.
4. Морозов В.В. R-D проблема и эффективность систем сжатия изображения // Автоматика. Автоматизация. Электротехнические комплексы и системы. – 2006. – Т. 2. – № 18. – С. 8–12.
5. Умняшкин С.В. Схема RD-оптимизированной компрессии для обработки видеоданных в реальном масштабе времени // Известия вузов. Сер. Электроника, 2002. – № 2. – С. 60–65.
6. Berger T. Rate Distortion Theory. Englewood Cliffs. – N.J.: Prentice-Hall, 1971. – 311 p.

7. Wu S.W., Gersho A. Rate-constrained picture-adaptive quantization for JPEG baseline coders // IEEE Intern. Conf. on Acoustic, Speech and Signal Processing. – 1993. – V. 5. – P. 390–392.
8. Liang J. Fast Multiplierless approximations of DCT with the lifting scheme // IEEE Trans. on Signal Processing. – 2001. – V. 49. – № 12. – P. 3032–3044.
9. ITU-CCITT Recommendation T.81 Information technology. – Digital compression and coding of continuous-tone still images. – ISO/IEC 10918-1. – 1992. – 182 p.
10. Набор изображений Calgary Corpus Gray Set 2. 2002. URL: <http://cdb.paradise-insight.us/?corpus=10> (дата обращения: 16.03.2011).
11. Модификация JPEG сжатия. 2012. URL: <http://jpg.vt.tpu.ru> (дата обращения: 10.03.2012).
12. Wang Z., Bovik A.C. Modern image quality assessment. – N.Y.: Morgan & Claypool, 2006. – 157 p.

Поступила 11.04.2012 г.

УДК 519.71+004

ПОСТРОЕНИЕ МНОГОКОМПОНЕНТНЫХ ВИЗУАЛЬНЫХ 3D-МОДЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ РАЗНОРОДНЫХ ИСТОЧНИКОВ ИНФОРМАЦИИ, НА ПРИМЕРЕ СОЗДАНИЯ ГЕОЛОГИЧЕСКИХ МОДЕЛЕЙ

А.А. Захарова, А.В. Шкляр

Томский политехнический университет
E-mail: shklyarAV@mail.ru

Показана необходимость разработки и совершенствования визуальных моделей данных, способов их создания и использования. Представлен подход к построению многокомпонентных визуальных 3D-моделей, позволяющих проводить проверку, интерпретацию и анализ пространственной информации на примере моделей геологических данных.

Ключевые слова:

3D-моделирование, способы визуализации, визуальные образы, технология обработки данных.

Key words:

3D-modelling, visualization, data processing, visual form.

Использование визуальных способов анализа и интерпретации информации имеет многочисленные преимущества, одним из которых является высокая скорость понимания общего смысла данных и обнаружения содержащихся в них особенностей. Отдельным направлением в визуальном анализе данных является создание 3D-моделей – трехмерных графических образов, способных иметь разнообразную информационную основу [1].

Целью работы является обоснование актуальности использования трехмерных визуальных моделей при анализе сложных разнородных данных. Эффективность использования визуальных 3D-моделей обусловлена тем, что к анализу информации подключается ассоциативное образное мышление человека. Во многих случаях, создание подобной модели позволяет решать многочисленные задачи, для которых любое другое решение является дорогостоящим, технически сложным или вообще не существует. Важной особенностью 3D-моделирования является возможность создания моделей, как реальных пространственных объектов, так и несуществующих. Отдельной задачей, в общем случае не имеющей однозначного решения, является создание моделей при отсутствии исчерпывающей информации о пространственных характеристиках исходных объектов.

Трехмерное пространство, в котором создается модель, имеет интуитивно понятную ассоциативную связь с привычным для наблюдателя координатным представлением. Поэтому 3D-модели являются удобным способом представления информации о реальных пространственных объектах, позволяя легко распознавать и интерпретировать. Во многих задачах, созданные 3D-модели можно рассматривать как компактный и эффективный способ представления и, возможно, хранения информации.

Кроме того, пространство моделирования способно выполнять важную роль объединения пространств представления моделей, имеющих совершенно разное информационное содержание. Это позволяет дополнять восприятие каждой из таких субмоделей новой информацией, а также значительно ускорить интерпретацию предлагаемых наблюдателю данных. Как следствие, при создании визуальных 3D-моделей часто возникает стремление использовать пространство моделирования для объединения моделей всех данных, которые имеют отношение к исследуемой задаче. Одним из оценочных параметров визуальной модели является ее информационная ценность, которая определяется как количеством подключенных к модели информационных источников, так и скоростью интерпретации данных [2].

Вопросы, которые возникают при разработке визуальных 3D-моделей выглядят следующим образом:

- Общая постановка задачи, для решения которой создается модель.
- Источники данных, участие которых в формировании модели обязательно.
- Способы представления данных разных типов в едином трехмерном пространстве создания модели.
- Способы дополнительной обработки исходных данных и варианты включения результатов в состав модели при условии увеличения информационной ценности модели.
- Оценка необходимой для правильной визуализации точности (детальности) исходной информации, а также возможных вариантов упрощения модели.
- Определение способов визуализации данных, позволяющих интерпретировать их максимально эффективно, учитывая все имеющиеся компоненты модели.

В качестве показательного примера рассмотрим создание 3D-модели геологической среды. В общем случае такая модель включает в себя сразу несколько субмоделей: модель, описывающую геологическое строение моделируемого объекта, модель поверхности, объекты инфраструктуры (дороги, коммуникации, строения и т. д.), возможно, выделение в обособленные субмодели водных объектов, растительности, определенных по какому-либо критерию наборов искусственных сооружений и т. д. В силу целого ряда причин, каждая из таких субмоделей может иметь свой источник исходной для моделирования информации. Эта информация, в свою очередь, может иметь разный тип (цифровые данные, фотоизображения, схемы, чертежи и т. д.), а также различную детальность или достоверность.

Построение подобной модели может преследовать несколько целей: сопоставление информации, полученной из разных источников, но имеющей пересекающиеся области, моделирование технологических процессов, поиск визуальных решений для некоторых задач и, наконец, формирование у наблюдателя полноценного представления о сложной системе пространственных объектов, которой является любая геологическая система. В последнем случае требования к точности данных не столь высоки, как при моделировании для решения каких-то технологических задач. В такой ситуации имеет смысл оценить качество всех информационных источников, подключаемых к модели, чтобы определить максимально точный и при построении модели считать его опорным. Иными словами, возникает возможность на этапе моделирования уточнять и, при необходимости, корректировать некоторые данные в сторону увеличения их достоверности. Кроме того, оценивая подключаемые информационные источники с точки зрения их полноты и достоверности, мы опре-

деляем очередность их участия в моделировании. В большинстве ситуаций субмодели, имеющие наименьшее информационное наполнение (обычно, это разного рода схемы или, например, словесные описания), создаются в последнюю очередь. Это позволяет в качестве корректирующей информации или граничных условий для таких частей модели использовать уже созданные субмодели.

На следующем этапе происходит разработка механизмов использования исходной информации, в которую входит возможная конвертация форматов данных, определение соответствующих процедур моделирования 3D-геометрии, а также вариантов использования данных, не соответствующих визуальной модели по размерности. Ответы на эти вопросы могут быть получены после определения программного обеспечения, участвующего в создании модели. В настоящий момент, существует большое количество программных комплексов предназначенных для 3D-моделирования, как специализированных, обладающих средствами трехмерного моделирования только для определенных задач, так и универсальных, обладающих очень большим, постоянно пополняющимся набором инструментов. В ситуации, когда модель содержит компоненты разных типов, отличающихся друг от друга способами построения, анимации, взаимодействия с окружением и т. д., имеет смысл создавать 3D-модель, используя именно универсальные средства. Кроме того, использование универсальных средств является хорошим решением, предоставляющим возможность в случае необходимости в качестве субмоделей использовать результаты моделирования, полученные в специализированных программах. Это позволяет намного эффективнее использовать вычислительные мощности и время.

Вместе с вопросами импорта внешних данных, необходимо определиться с выбором наиболее правильного, с точки зрения общей цели создания модели, и информативного их представления. Поиск адекватной метафоры представления является отдельной самостоятельной задачей моделирования в силу того, что присутствие в модели нескольких информационных источников, может привести к созданию субмоделей, одновременное восприятие которых в едином визуальном пространстве становится затрудненным. Кроме того, моделирование и, возможно, анимация некоторых объектов может быть чрезвычайно трудоемким процессом как с технологической, так и с вычислительной точек зрения. Обычно, это касается моделирования сложных многочастичных систем, жидкостей или органических объектов. К тому же, модель может содержать данные, визуальное представление которых является условным. Например, варианты визуализации скалярных или векторных массивов данных могут быть определены исходя из их собственных свойств или происхождения, а также с позиции их наиболее эффективного визуального представления в рамках создаваемой модели.

Задачи выбора представления на этапе моделирования могут решаться по-разному. В одних случаях, используются возможности, предоставляемые средствами моделирования. Например, текстурирование является мощным средством, иногда позволяющим избежать, излишне детального воссоздания геометрии объекта. В других ситуациях, имеет смысл разработать для данных собственный визуальный образ, позволяющий наблюдателю правильно понимать заключенную в них информацию, но не создающий непреодолимых сложностей при моделировании. Одним из распространенных решений является использование цветового кодирования, позволяющее с хорошей степенью наглядности визуализировать скалярные данные на трехмерных поверхностях. Если же информация, на основании которой создается субмодель, имеет более сложную природу, то вопрос о соответствующем ей визуальном образе может иметь самые разные ответы. В любом случае, необходимо помнить о том, что основным приоритетом является не визуальный образ отдельной информацион-

ной составляющей модели, а общая цель ее создания и определяемая этой целью информационная ценность модели.

Наконец, помимо разрешения всех конструктивных вопросов, необходимо разработать еще одну составляющую часть любой 3D-модели. Речь идет об определении способов наблюдения модели, ее визуального исследования и, возможно, изменения параметров, управляющих вариантами представления модели. Существует достаточно большое количество различных приемов, позволяющих эффективно использовать преимущества трехмерного представления данных в визуальных моделях. К ним относятся, например, панорамный обзор, наблюдение за моделью из любой удобной точки или с выбранного ракурса, упрощенное (с уменьшенной детализацией) представление, возможность исключения из визуального образа модели отдельных компонент и т. д.

Этапы описанной выше технологии представлены на рис. 1.



Рис. 1. Этапы технологии создания визуальных моделей

В задаче создания модели геологической системы выбор источников информации сделан следующим образом. Собственно геологическая часть строится на основании данных, полученных из специализированных геоинформационных пакетов – так называемые *grid*-данные. Данные этого типа описывают геологическую структуру как набор непересекающихся поверхностей, являющихся геометрическим местом узловых точек с трехмерными координатами, определенными относительно заданной системы отсчета. Кроме этого, в качестве источников исходной информации выступают результаты аэрофотосъемки, схемы расположения наземных сооружений, коммуникаций и других объектов. Все эти данные имеют различное происхождение, форматы хранения и точность. Однако, с учетом схемы моделирования (рис. 1), при создании 3D-модели вся эта информация может быть использована.

В качестве среды моделирования выбран пакет 3Ds Max компании Autodesk, который предоставляет возможности трехмерного моделирования и соответствуют требованиям решаемой задачи. Данный пакет является достаточно универсальным средством, что позволяет эффективно использовать его для задач создания трехмерных геометрических объектов разных типов в едином пространстве многокомпонентной модели. Кроме того, 3Ds Max имеет инструменты взаимодействия (в том числе на уровне обмена форматами данных) со многими другими программными продуктами, а также для работы с текстами. Все это позволяет значительно упростить создание моделей объектов, имеющих сложную геометрию, не имеющих симметрии или регулярных структур. Именно с такими объектами приходится иметь дело при моделировании большинства природных объектов.

Исходя из специфики данных, описывающих геологическое строение набором N поверхностей, процесс моделирования предполагает создание $(N-1)$ трехмерных объектов, ограниченных сверху и снизу соответствующими поверхностями. В общем случае, значение N может быть достаточно большим, поэтому процедура создания каждого такого объекта должна содержать минимально возможное количество действий и зависеть лишь от определенного набора численных или иных параметров, полученных на основании исходных данных. Чтобы реализовать такой процесс, необходимо к одному из стандартных геометрических примитивов, имеющихся в 3Ds Max, последовательно применить серию преобразований с необходимыми параметрами.

Наиболее подходящим геометрическим примитивом для описываемой задачи является параллелепипед (*Vox*). Поскольку все трансформации, которые произойдут с исходным объектом при моделировании будут действовать в направлении вертикальной оси координат (Z), то вертикальные поверхности *Vox* могут иметь лишь минимально возможное разбиение. По этой же причине горизон-

тальные поверхности *Vox*: верхнюю (*Top*) и нижнюю (*Bottom*) следует разбить на полигоны, получив таким образом точки поверхностей, которые будут непосредственно участвовать в преобразованиях. Частота разбиения по каждой из горизонтальных осей (X, Y) для одного параллелепипеда не может быть больше 200, а кроме этого зависит от числа точек в исходных для построения *grid*-данных. Так как целью создания визуальной модели для рассматриваемой задачи является предоставление наблюдателю всей необходимой исходной информации в виде набора визуальных образов, способных сформировать правильное представление об объекте моделирования даже у неподготовленного наблюдателя, то указанное выше разбиение является вполне приемлемым. Поскольку частота разбиения напрямую связана с достижимой точностью построения модели, то в случаях, когда детальность исходных данных оказывается более высокой, необходимо определять иные способы моделирования.

Размеры создаваемого объекта, который является образом одного из моделируемых геологических слоев, в горизонтальной плоскости задаются равными размерам реального объекта. Это позволяет в дальнейшем достаточно легко контролировать расположение и размеры любых субмоделей на поверхности. Для усиления выразительности ко всем вертикальным размерам может быть применен масштабирующий коэффициент, приемлемое значение которого определяется визуально. Расположение точки отсчета системы координат, общей для всего пространства моделирования определяется исходя из формата исходных данных, а также из соображений удобства моделирования.

Итак, процедура создания модели отдельного слоя происходит следующим образом. Исходный объект (ИО), имеющий необходимые размеры в плоскости XY и минимальный размер по высоте (Z), например 1 м (учитывая, что по X и Y размеры имеют значения порядка нескольких тысяч метров), позиционируется в пространстве. Расположение по Z относительно нижней поверхности (*Bottom*). Затем к ИО применяется модификатор растяжения (*Stretch*) в направлении Z на величину расстояния между нижними точками моделируемых поверхностей, соответствующих нижней и верхней поверхностям ИО.

$$K_n = z_{\min, n+1} - z_{\min, n}.$$

Остается только трансформировать эти поверхности в соответствии с исходными данными моделирования. Для трансформации нижней поверхности к узлам, расположенным на ней, применяется модификатор смещения (*Displace*). В качестве параметра трансформации выступает карта смещений (*Displacement Map*), полученная на основании данных, описывающих геологическую модель в пакете Petrel (рис. 2). Карта смещений представляет собой растровое монохромное изображение, значение яркости которой в точке, соответствующей узлу поверхности пропорционально величине мак-

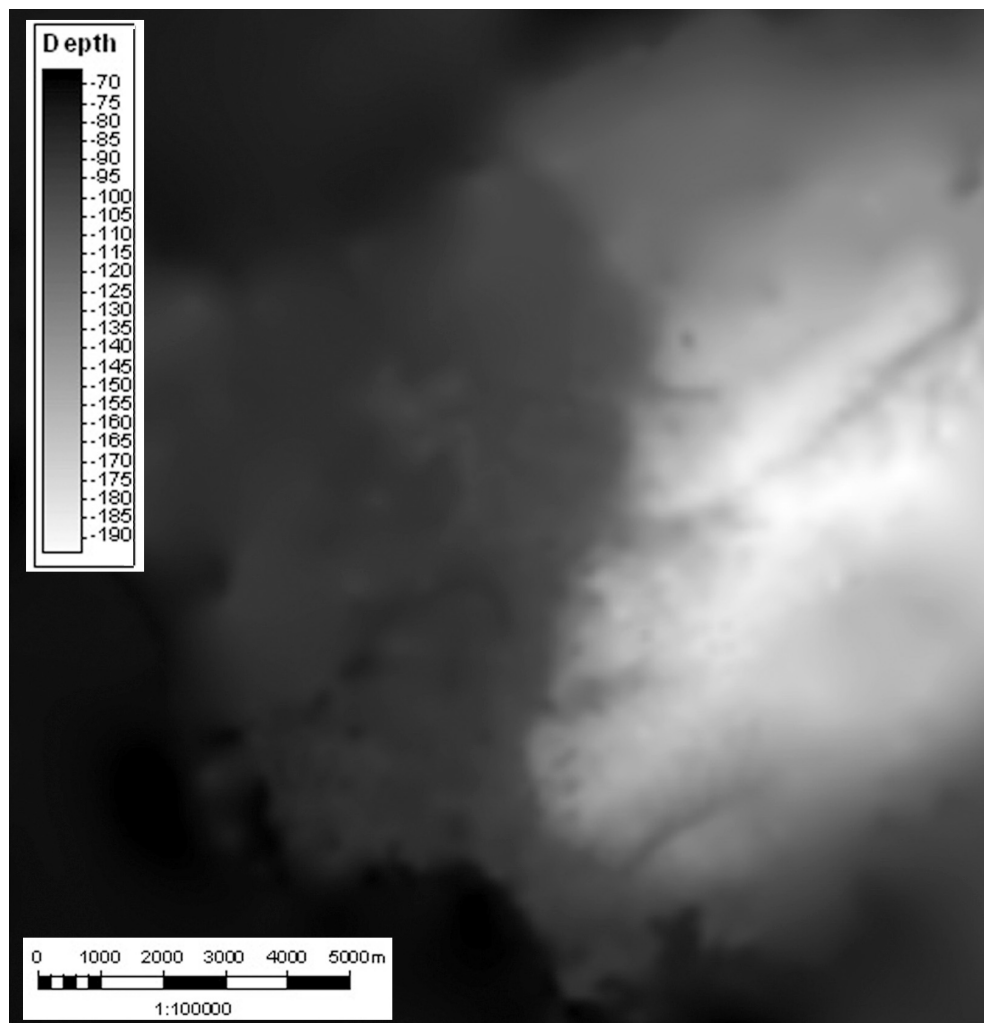


Рис. 2. Исходные данные для создания карты смещений

симального смещения $S_{D,n}$ для этой поверхности. Коэффициентом смещения *Strength* для модификатора *Displace* является расстояние между верхней и нижней точками поверхности n :

$$S_{D,n} = Z_{\max,n} - Z_{\min,n}$$

К узлам верхней поверхности будет применяться та же процедура, но с данными, соответствующими поверхности $n+1$.

В результате мы получаем 3D-модель отдельного слоя – независимый объект, ограниченный со всех сторон построенными поверхностями, спозиционированный в пространстве и соответствующий данным, исходным для моделирования (рис. 3).

Преобразования, примененные к ИО, являются независимыми и управляемыми, поэтому полученный объект свободно поддается изменениям и кор-

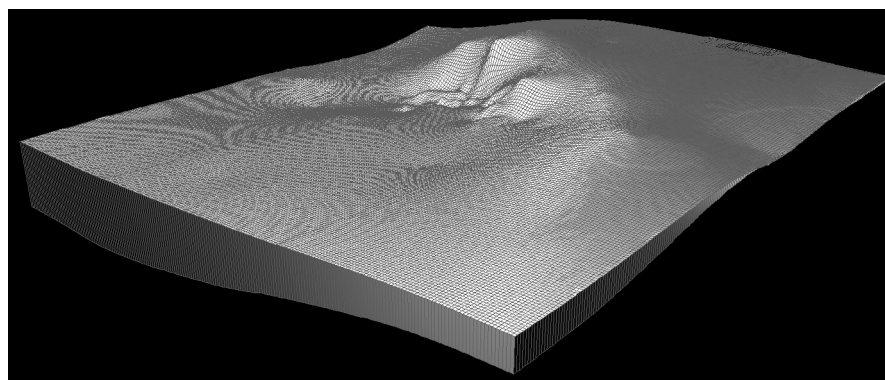


Рис. 3. 3D-модель отдельного слоя

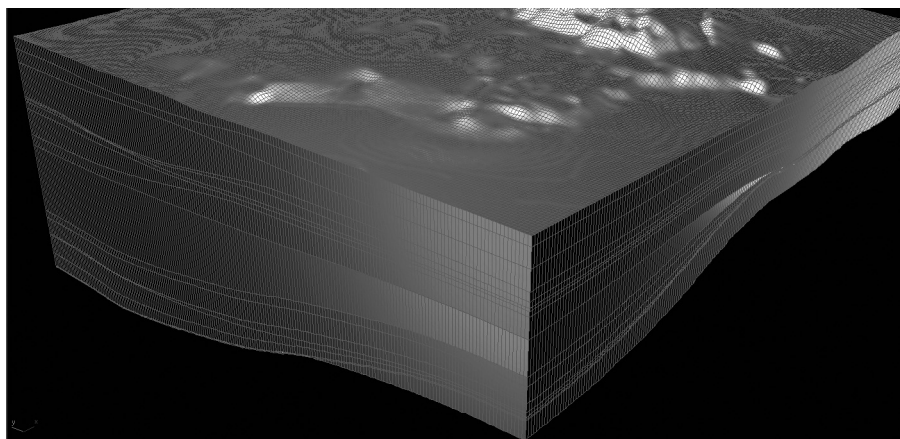


Рис. 4. Полный набор слоев модели

ректировке. Таким образом, для получения уникальной модели слоя геологической системы к ИО, являющемуся стандартным примитивом 3Ds Max, применена последовательность трансформаций, определенная и зависящая лишь от набора управляющих ею параметров: линейные размеры, положение в пространстве, карты смещений и соответствующие им коэффициенты растяжений. Все эти параметры могут быть получены из соответствующих информационных источников. Для построения моделей всех остальных слоев, присутствующих в геологической модели, необходимо в копиях уже построенного объекта заменить управляющие данные на соответствующие другим слоям (рис. 4).

Полученный набор объектов уже позволяет наблюдателю получить представление о геометрических свойствах моделируемой системы. Чтобы облегчить эту задачу, необходимо решить вопросы текстурирования и наиболее эффективной визуализации модели.

Инструменты текстурирования могут быть использованы в этой задаче двумя способами (рис. 5). С одной стороны, возможности 3Ds Max позволяют подобрать текстуры, соответствующие реальным породам, составляющим тот или иной слой, и получить, таким образом, возможность наблюдения действительной картины геологического строения. Этот вариант имеет смысл использовать лишь в случае особой необходимости, так как возникают сложности правильного визуального восприятия

объектов разного масштаба, которыми являются образцы пород и сама геологическая система.

С другой стороны, есть смысл сосредоточить усилия на подборе текстур, передающих наблюдателю информацию о функциональной или какой-то другой роли объекта, но не затрудняющий визуальное восприятие остальной информации. Этот вариант является разновидностью цветового кодирования, интуитивно понятен и потому имеет достаточно широкое применение.

Использование результатов аэрофотосъемки создает отдельный информационный объект, существенно дополняющий визуальную информативность модели (рис. 6). Происходит это из-за появления большого количества деталей, содержащихся в растровых источниках, а также за счет возникновения визуально распознаваемых масштабных элементов. В то же время, уже созданная рельефная 3D-модель дополняет информативность исходных материалов аэрофотосъемки, представляющих собой, по сути, обычную плоскую картографическую проекцию [3].

Построенная к этому моменту 3D-модель уже представляет собой достаточно объемный источник визуальной информации, но может, при необходимости, дополняться субмоделями любых, интересующих наблюдателя, объектов. Например, самостоятельными объектами могут быть модели коммуникаций, дорожной сети, инженерных сооружений и т. д. — все зависит от постановки основной задачи моделирования (рис. 7).

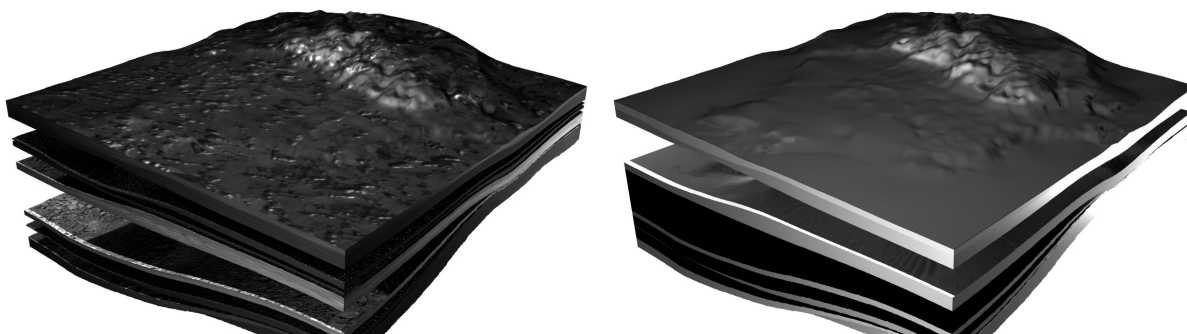


Рис. 5. Примеры различных вариантов текстурирования

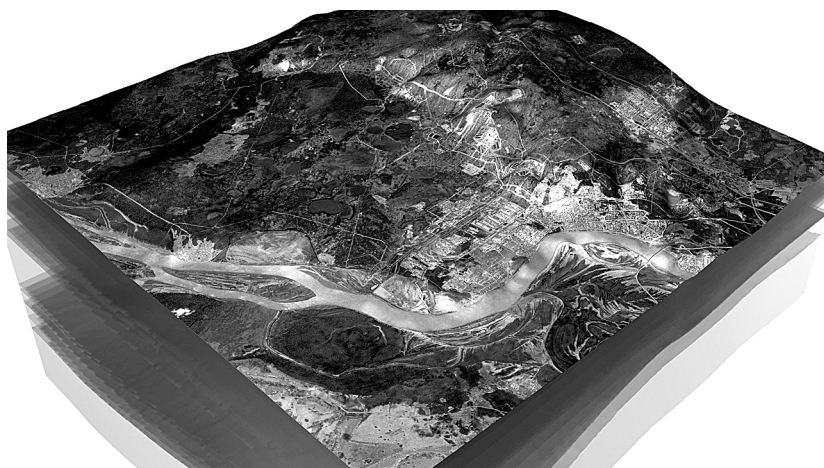


Рис. 6. Модель, содержащая данные аэрофотосъемки

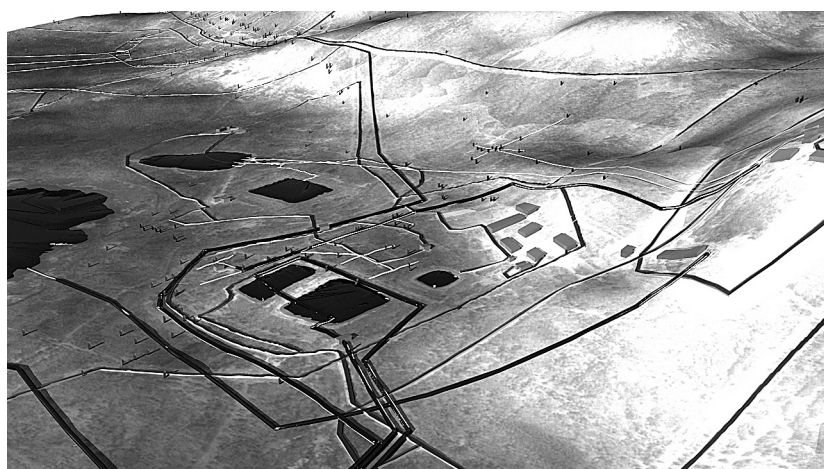


Рис. 7. Субмодели (дополнительные объекты модели)

Если модель включает в себя большое число субмоделей или визуальные образы некоторых из них слишком сложны, может возникнуть ситуация, когда восприятие содержащейся в них информации становится затрудненным. В таких случаях необходимо либо находить способы упрощения визуального представления субмоделей (или модели в целом), либо, если постановка задачи позволяет это сделать, временно исключать некоторые из субмоделей из общего визуального образа.

Одним из самых эффективных способов получения наблюдателем информации, заложенной в 3D-модель, является возможность наблюдения за ней из любой точки пространства, выбирая ту, которая обеспечивает максимальную информативность (точки наблюдения) [4]. Для достижения этой цели используют всевозможные перемещения точки наблюдения (виртуальной камеры), а также ее вращения. Кроме этого, очень часто в ситуациях, когда исследуемый объект имеет сложную геометрию или является совокупностью большого числа более мелких объектов, получение наблюдателем полноценного представления о модели возможно только с использованием подвижной камеры. Создание замкнутых траекторий движения ка-

меры различной сложности позволяет наблюдателю плавно менять положение точки наблюдения, использовать при необходимости эффекты масштабирования, заменяя статичное наблюдение на динамичный содержательный процесс исследования.

Заключение

Обоснована эффективность использования визуальных 3D-моделей для интерпретации и анализа сложной информации. Построена визуальная 3D-модель геологического комплекса, имеющая необходимые средства наблюдения и изучения. Информационное содержание модели объединяет в едином визуальном пространстве все имевшиеся при создании модели источники информации. Обеспечена возможность дополнения и расширения модели при подключении новых данных, а также механизмы бесконфликтного объединения разнородной информации в рамках единой модели. Описанная технология в общем случае может быть применена при создании визуальных 3D-моделей объектов самой разной природы, степени сложности и использующих источники информации разных типов.

СПИСОК ЛИТЕРАТУРЫ

1. Могиленских Д.В., Павлов И.В., Федоров В.В., Мельникова С.Н., Сапожникова Е.Э. Принципы построения и функциональное содержание системы визуализации для анализа скалярных и векторных полей, заданных на двумерных регулярных сетках / Препринт РФЯЦ – ВНИИТФ. – Снежинск, 2000. – № 172. – 26 с.
2. Бондарев А.Е., Галактионов В.А. Анализ развития концепций и методов визуального представления данных в задачах вычислительной физики / Препринты ИПМ им. М.В. Келдыша. – 2009. – № 53. – 28 с. URL: http://www.keldysh.ru/papers/2009/prep53/prep2009_53 (дата обращения: 28.03.2012).
3. Андреев С.В., Галактионов В.А., Денисов Е.Ю., Кирилов Н.Е. Синтез фотореалистичных трехмерных изображений в современных системах презентаций // Программные продукты и системы. – 2007. – № 3. – С. 37–40.
4. Ванданов В.Г., Долговесов Б.С., Исламов Р.Р., Казанский И.П., Клименко С.В., Ковальков М.А., Мазурок Б.С., Морозов Б.Б. Программно-аппаратный комплекс 3D презентаций на основе виртуальной студии и виртуального окружения // Трехмерная визуализация научной, технической и социальной реальности. Кластерные технологии моделирования: Труды I Международ. конф. – Ижевск: УдГУ, 2009. – Т. 1. – С. 73–77.

Поступила 28.03.2012 г.

УДК 550.8.053

ИССЛЕДОВАНИЕ РАЗРЕШАЮЩЕЙ СПОСОБНОСТИ МЕТОДОВ ФАЗОЧАСТОТНОГО ПРОСЛЕЖИВАНИЯ СЕЙСМИЧЕСКИХ СИГНАЛОВ

В.П. Иванченков, А.И. Кочегуров, О.В. Орлов

Томский политехнический университет
E-mail: kai@cc.tpu.edu.ru

Приведены результаты исследования разрешающей способности методов фазочастотного прослеживания сейсмических сигналов на моделях волновых полей и реальных данных. Показано, что методы фазочастотного прослеживания сигналов с равновесной и неравновесной обработкой обладают высокой разрешающей способностью и позволяют выделять слабые отражения в волновом поле. Результаты обработки полевых материалов подтверждают выводы математического моделирования и практическую ценность разработанных методов.

Ключевые слова:

Сейсмические сигналы, фазочастотные характеристики, разрешающая способность, функция правдоподобия.

Key words:

Seismic signals, phase-frequency characteristic, resolving ability, likelihood function.

Прослеживание сейсмических волн и границ занимает одно из ведущих положений при решении задач структурной сейсморазведки. В общей постановке в задачу прослеживания входит обнаружение волн, их идентификация и оценка параметров. Основными неизвестными параметрами при прослеживании волн являются моменты их прихода. Варианты прослеживания фиксированных волн (границ) связаны с назначением участков, в пределах которых исследуется поведение искомого объекта. На основе результатов прослеживания выходных данных осуществляется построение структурных карт и палеопостроения.

Не менее важное значение имеет прослеживание волн при решении задач прогноза геологического разреза, в том числе прогноза нефтегазоносности. При решении этих задач требуется проводить как детальное изучение геометрических параметров и формы выделяемых локальных геологических объектов, так и осуществлять оценку комплекса кинематических параметров, используемых при прогнозе. При прослеживании волн может также производиться определение ряда динамических параметров выделяемых отражений, напри-

мер, их амплитудных, энергетических и фазовых спектров.

На практике довольно часто необходимо решать задачи прослеживания волн в условиях их интенсивной интерференции. Такая ситуация наиболее характерна при исследованиях тонкослоистых сред. При этом, как правило, приходится осуществлять прослеживание волн в условиях, когда форма сигналов неизвестна. В этой связи реализуемые алгоритмы прослеживания сейсмических волн должны обладать повышенной разрешающей способностью и позволять надежно осуществлять оценку их кинематических параметров в условиях неполной априорной информации о свойствах выделяемых сигналов.

Для решения задач прослеживания волн нами был предложен ряд спектральных методов, использующих только информацию о фазовых спектрах регистрируемых сигналов [1–3]. Так в [1] рассмотрен оптимальный метод прослеживания фиксированных волн, наблюдаемых на фоне гауссовых помех, который реализуется в виде процедуры поиска максимума функции правдоподобия:

$$L(t) = \sum_{k=1}^m \delta(\omega_k) \cos(\Delta\phi(\omega_k) - \omega_k t). \quad (1)$$

Здесь $\Delta\phi(\omega_k)=\phi_x(\omega_k)-\phi_s(\omega_k)$ – отклонение фазового спектра сигнала от фазового спектра смеси сигнала и шума, $\delta(\omega_k)=A(\omega_k)/\sigma(\omega_k)$ – пиковое отношение сигнала к шуму на частоте ω_k , m – число анализируемых частотных компонент.

Практическое применение оптимальных фазочастотных алгоритмов вида (1) наталкивается на ряд проблем, в частности связанных с оценкой распределения отношений $\delta(\omega_k)$ в исследуемом диапазоне частот. Действительно, форма регистрируемых сейсмических сигналов, как правило, неизвестна, а, следовательно, неизвестны $\delta(\omega_k)$, $k=\overline{1,m}$. При этом также обычно оказываются неизвестными точные значения фазового спектра выделенных сигналов. В этой связи в [3] предложены так называемые фазочастотные алгоритмы с равновесной и неравновесной обработкой. Эти алгоритмы могут быть получены из оптимального метода путем замены в (1) весовой функции $\delta(\omega_k)$ на другие, специально подобранные функции. В общем случае функция правдоподобия (критерий оценки временного положения сигналов) для таких алгоритмов может быть записана в виде:

$$L(t) = \sum_{k=1}^m w(\omega_k) \cos[\phi(\omega_k, t)], \quad (2)$$

где $w(\omega_k)$ – частотная весовая функция, вид которой зависит от реализуемого фазочастотного алгоритма; $\phi(\omega_k, t)$ – текущий фазовый спектр участка трассы, вычисляемый в скользящем окне анализа [3].

Для равновесного алгоритма весовая функция $w(\omega_k)$ принимается равной единице во всей полосе частот. Для алгоритма с неравновесной обработкой $w(\omega_k)$ может быть задана в треугольной, синусоидальной или экспоненциальной форме. Так, при дальнейших исследованиях $w(\omega_k)$ задавалась в виде треугольной функции [3]:

$$w(\omega) = \frac{4}{3\omega_c} \begin{cases} 0, & \omega \leq \omega_H \\ \frac{2}{\omega_c}(\omega - \omega_H), & \omega_H < \omega \leq \omega_C, \\ -\frac{1}{\omega_c}(\omega - \omega_B), & \omega_C < \omega \leq \omega_B \end{cases}$$

где ω_B и ω_H – соответственно верхние и нижние частоты, определяющие $w(\omega_k)$; ω_C – частота максимума $w(\omega_k)$. При этом $\omega_C=2\omega_H$; $\omega_B=2\omega_C$.

Важно отметить, что в основу синтеза алгоритмов с равновесной и неравновесной обработкой положено установленное ранее свойство стационарности фазовых спектров импульсов однократно-отраженных волн. Это свойство заключается в том, что при совмещении начала отсчета с центром окна анализа фазовые спектры сейсмоимпульсов принимают постоянное, не зависящее от частоты значение [2].

Исходя из анализа (2) можно также показать определенную аналогию между процедурой оценки временного положения сигналов в принятых алгоритмах отслеживания и их низкочастотной

фильтрацией, а именно выражение функции правдоподобия (2) является обратным дискретным преобразованием Фурье от результата фильтрации исходного процесса цифровым фильтром с частотной характеристикой вида [4],

$$H(\omega_k) = \frac{w(\omega_k)}{|X(\omega_k)|}, \quad k = \overline{1,m},$$

где $|X(\omega_k)|$ – амплитудно-частотная характеристика сигнала.

Рассмотрим влияние действия фильтра. Прежде всего, отметим, что данный фильтр сначала выравнивает амплитудный спектр исследуемого колебания, а затем взвешивает его с помощью заданных весовых коэффициентов. При этом фазовые соотношения в исходной записи не изменяются. Известно [5], что выравнивание амплитудно-частотной характеристики при линейной фазочастотной характеристике приводит к сжатию сигнала, а, следовательно, появляется реальная возможность увеличить разрешение сигналов на записи. Кроме того, при реализации такого фильтра, с помощью задания весовых коэффициентов можно управлять его частотной характеристикой, тем самым, усиливая или ослабляя различные частотные составляющие сигнала.

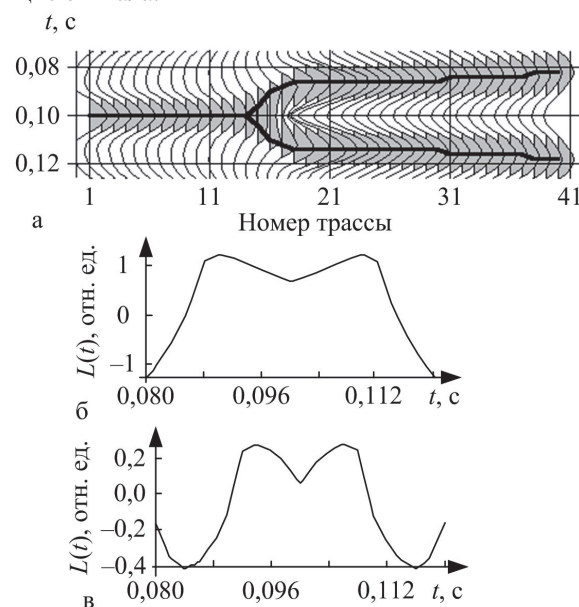


Рис. 1. Пример, иллюстрирующий разрешающую способность методов ФЧП. Модель волнового поля (а). Функция отслеживания для алгоритмов ФЧП с равновесной обработкой с разрешением 16 мс (б) и с неравновесной обработкой с разрешением 11 мс (в). Длительность окна отслеживания 40 мс; полоса частот 20...60 Гц

Для оценки разрешающей способности предложенных фазочастотных алгоритмов проведено их исследование на синтезированных моделях сейсмического волнового поля [3]. Модель задавалась в форме

$$s_{\Sigma}^k(t) = s\left(t - t_0 + \frac{\Delta T_k}{2}\right) + s\left(t - t_0 - \frac{\Delta T_k}{2}\right).$$

Здесь k – номер сеймотрассы; $t_0=0,1$ с; ΔT_k – временной интервал между импульсами, мс. Каждый из импульсов $s(t)$ является сигналом с колоколообразной огибающей.

Исследование разрешающей способности алгоритмов фазочастотного прослеживания (ФЧП) проводилось следующим образом. Выбирались параметры прослеживания, при которых разрешение являлось наилучшим. Далее непосредственно осуществлялось ФЧП. Полученное двумерное изображение исследовалось с целью определения трассы

с минимальным номером, на которой сигналы еще разрешались (имеется два явно выраженных экстремума, соответствующих теоретическому местоположению интерферируемых сигналов). Номер данной трассы, равный временному сдвигу между сигналами, выраженному в мс, соответствует предельному разрешению сигналов при прослеживании волн для исследуемого алгоритма.

На рис. 1, а, в качестве примера показана модель волнового поля и функции прослеживания для алгоритмов ФЧП с равновесной (рис. 1, б)

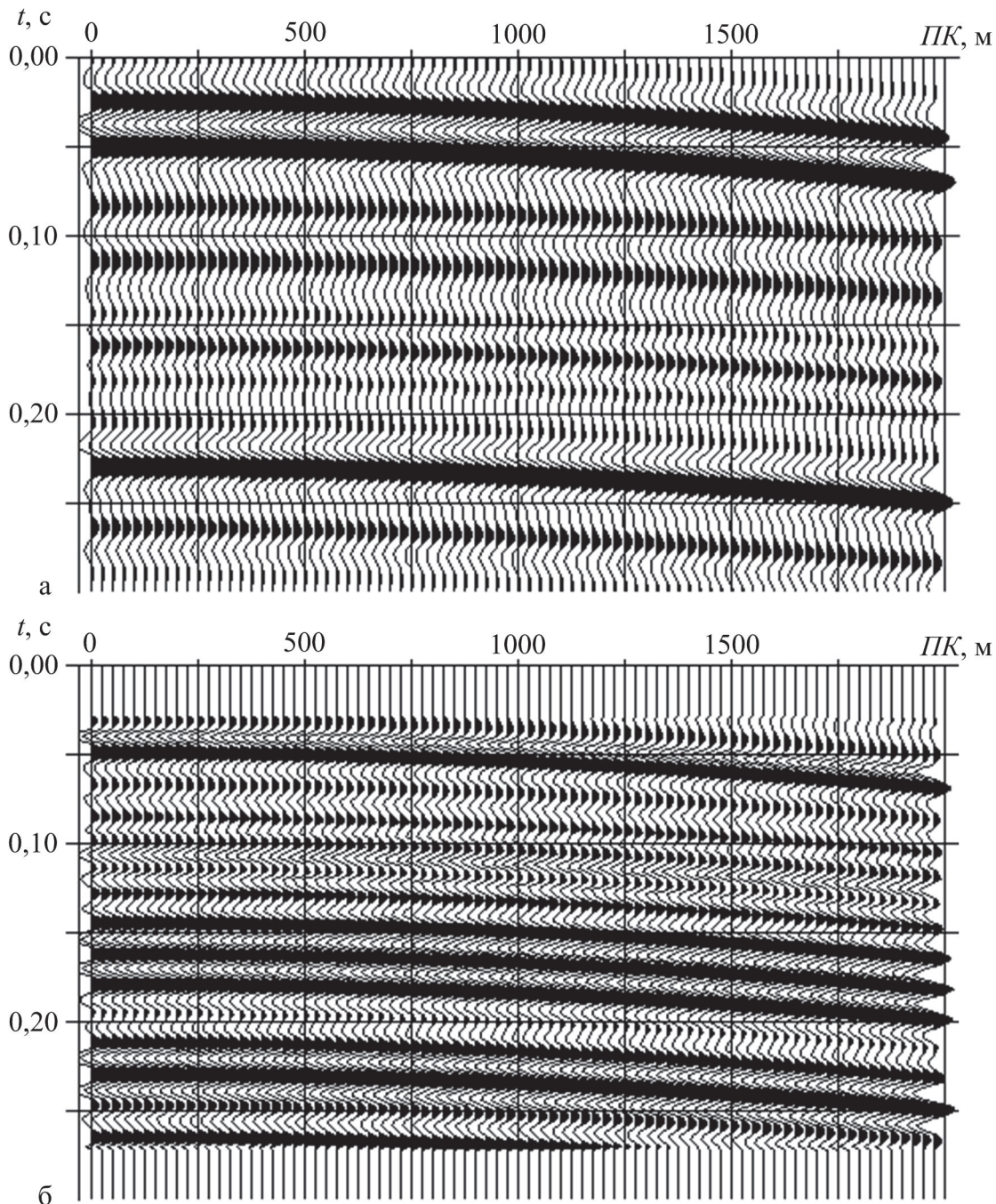


Рис. 2. ФЧП сигналов на модели сейсмограммы (алгоритм с неравновесной обработкой). Модель волнового поля (сигнал – импульс с колоколообразной огибающей; $f_0=34$ Гц, $\beta=51$) (а). Результат прослеживания (б). Длительность окна прослеживания 40 мс; центральная частота весовой функции $f_c=70$ Гц

и неравновесной обработкой (рис. 1, в) для трассы, которая в данном случае определяет предельное разрешение сигналов.

Приведенный пример показывает высокую разрешающую способность фазочастотных алгоритмов. Это позволяет говорить о возможности определения временного местоположения отражающих границ при ФЧП с высокой точностью.

На рис. 2, а приведена модель сейсмограммы, а на рис. 2, б – результат ФЧП (алгоритм с неравновесной обработкой), когда интерференционная картина слагается из 14 волн. Каждая волна состоит из совокупности импульсов с колоколообразной огибающей, основная частота которых $f_0=34$ Гц, а коэффициент затухания $\beta=51$.

Первая трасса модели сейсмограммы изображена на рис. 3, а. Видно, что в полученной интерференционной картине не удается выделить все действительно присутствующие в модели волны. Результаты ФЧП для данной трассы приведены на рис. 3, б, где уже явно все волны идентифицированы.

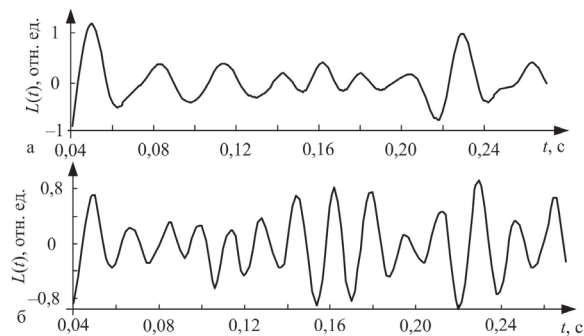


Рис. 3. ФЧП сигналов на модели сейсотрассы. Модель волнового поля (а). Функция прослеживания (б)

Количественные результаты исследования разрешающей способности алгоритма ФЧП с неравновесной обработкой при различной длительности окна прослеживания T приведены на диаграмме, рис. 4.

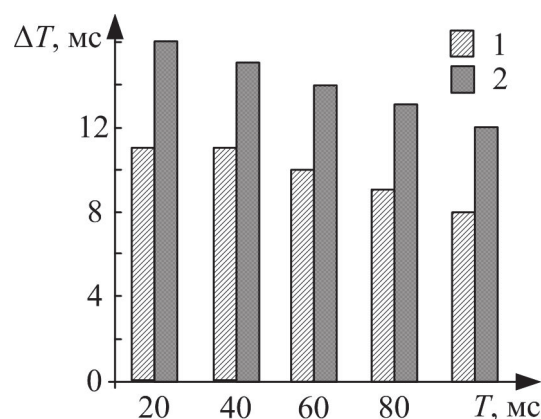


Рис. 4. Разрешающая способность алгоритма с неравновесной обработкой для сигналов с частотами: 1) 24; 2) 34 Гц

Из анализа диаграммы следует, что методы ФЧП сигналов с неравновесной обработкой позволяют разрешать сигналы, сдвинутые друг относительно друга не более чем на треть видимого пе-

риода. Такое разрешение дает возможность на практике получать надежные оценки кинематических параметров сейсмических сигналов при исследовании тонкослоистых сред.

На рис. 5 приведены фрагменты разреза (Крапивинское месторождение, Томская обл.) после стандартной обработки (рис. 5, а) и ФЧП (рис. 5, б).

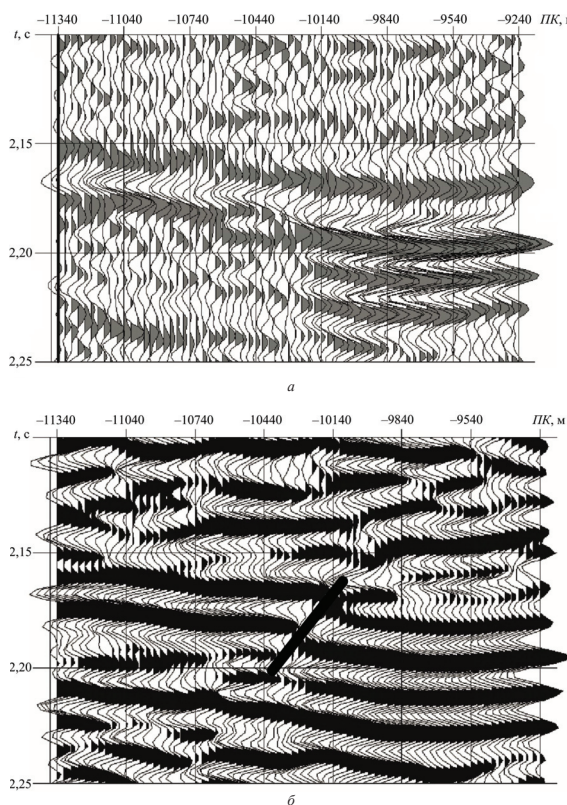


Рис. 5. Фрагменты разреза после стандартной обработки (а) и ФЧП (б)

Из рисунка видно, что в результате проведения ФЧП удалось выделить амплитудно-выраженный сброс, которого на фрагменте разреза после стандартной обработки не наблюдается.

Еще более убедительный пример приведен на рис. 6, где удалось не только выделить двухступенчатое нарушение, но и увидеть зону предполагаемого водонасыщенного коллектора.

Выводы

Исследования на моделях волновых сейсмических полей и результаты обработки реальных данных показали, что:

1. Методы фазочастотного прослеживания сигналов позволяют существенно увеличить разрешение сигналов на исходных сейсмических записях, по крайней мере, до одной трети видимого периода.
2. Результаты фазочастотного прослеживания дают возможность выделять предполагаемые тектонические нарушения, необходимые как для решения задач структурной сейсмозаведки, так и для прогноза нефтегазоносности.

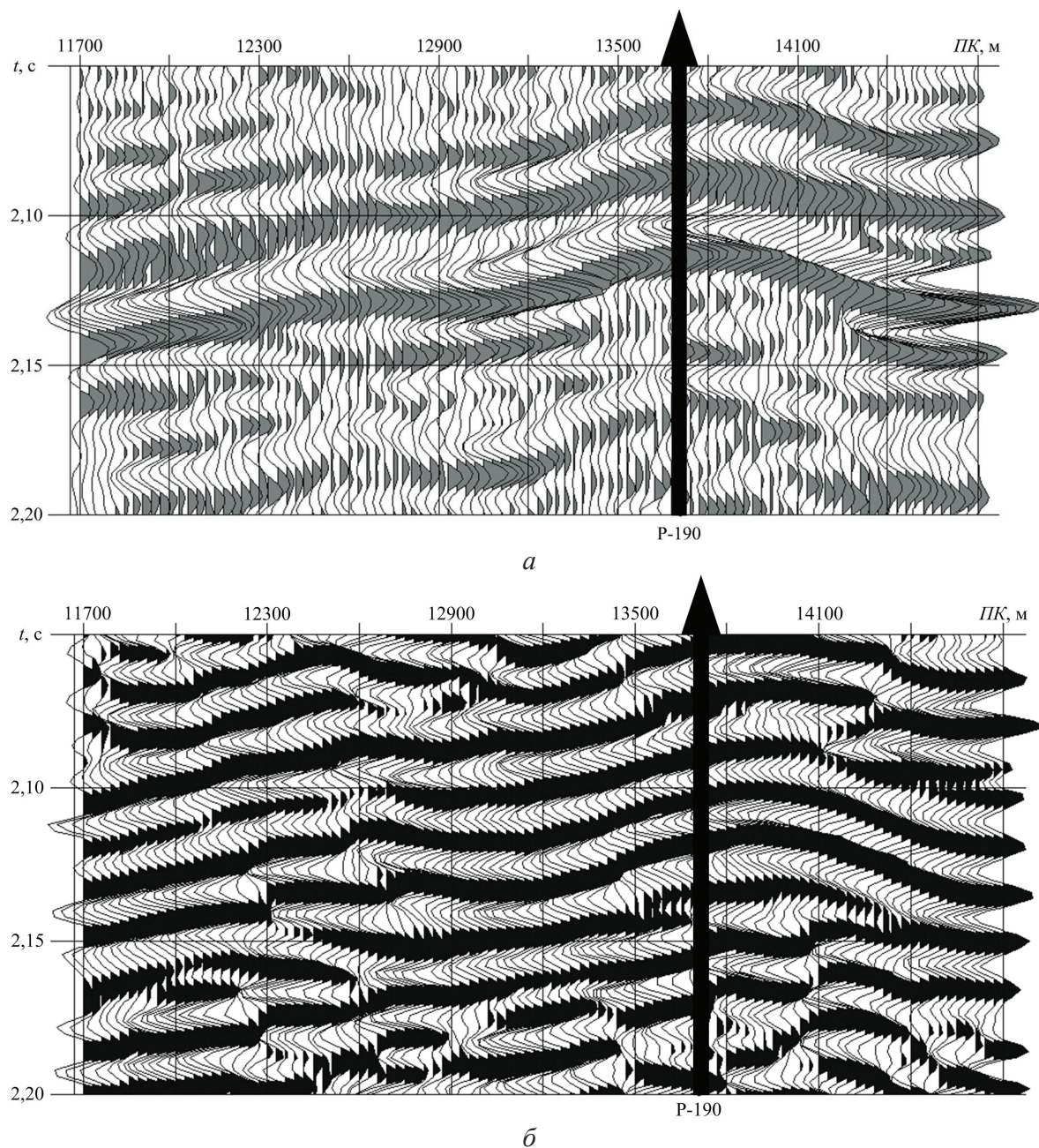


Рис. 6. Фрагменты разреза в области скважины P-190 после стандартной обработки (а) и ФЧП (б)

СПИСОК ЛИТЕРАТУРЫ

1. Иванченков В.П., Кочегуров А.И. Определение временного положения сейсмических сигналов по оценкам их фазочастотных характеристик // Геология и геофизика. – 1988. – № 9. – С. 77–83.
2. Иванченков В.П., Кочегуров А.И., Орлов О.В. О точности определения временного положения сейсмических сигналов по оценкам их фазочастотных характеристик // Известия Томского политехнического университета. – 2009. – Т. 315. – № 5. – С. 49–53.
3. Иванченков В.П., Вылегжанин О.Н., Орлов О.В., Козлов А.А. и др. Методы фазочастотного анализа волновых полей и их применение в задачах обработки данных сейсморазведки // Известия Томского политехнического университета. – 2006. – Т. 309. – № 7. – С. 65–70.
4. Кочегуров А.И. Исследование функции правдоподобия, формируемой при реализации методов фазочастотного прослеживания сейсмических волн // Наука и образование – ведущий фактор стратегии Казахстан-2030. (Сагиновские чтения № 2): Сб. трудов Междунар. научн. конф. – Караганда, 24–26 июня 2010. – Ч. 3. – Караганда: КарГТУ, 2010. – С. 364–365.
5. Тяпкин Ю.К. Оптимальная линейно-фазовая полосовая фильтрация сейсмических записей // Геология и геофизика. – 1984. – № 3. – С. 99–105.

Поступила 18.04.2012 г.

УДК 681.5

ПАРАМЕТРИЧЕСКИЙ СИНТЕЗ ПИ-РЕГУЛЯТОРА ЛИНЕЙНОЙ САУ НА ОСНОВЕ КОЭФФИЦИЕНТНЫХ ОЦЕНОК СТЕПЕНИ УСТОЙЧИВОСТИ И ЗАДАННОЙ ДОБРОТНОСТИ

М.И. Пушкарёв, С.А. Гайворонский

Томский политехнический университет

E-mail: pushkarev@tpu.ru

Разработана методика выбора настроек линейного пропорционально-интегрального регулятора, обеспечивающих квазимаксимальную степень устойчивости и заданную точность системы автоматического управления в установившихся статических режимах. В основу методики положены коэффициентные оценки показателей качества стационарных систем и базирующиеся на них достаточные условия заданной степени устойчивости.

Ключевые слова:

Регулятор, максимальная степень устойчивости, показатели качества, коэффициентные оценки, точность.

Key words:

Controller, maximal degree of stability, quality indexes, coefficient estimation, accuracy.

Введение

К настоящему времени разработано большое число методов синтеза линейных систем автоматического управления (САУ), позволяющих осуществлять обоснованный выбор структуры и параметров регулятора для обеспечения в системе заранее заданных требований к ее качеству. Показатели качества можно разбить на четыре группы: частотные, временные, корневые и коэффициентные, определяемые набором коэффициентов передаточной функции. Анализ методов синтеза САУ показывает, что для выбора настроек регулятора желательно иметь простые аналитические или графические зависимости, позволяющие легко перейти от показателей качества САУ к искомым параметрам регулятора. За исключением простых случаев (для систем первого и второго порядков) такие зависимости сложно получить при использовании частотных или временных показателей качества. Значительно проще задача параметрического синтеза регулятора решается на основе корневого подхода или при использовании коэффициентных методов.

Коэффициенты передаточной функции линейной системы или некоторые их комбинации наиболее удобно иметь в качестве показателей работоспособности системы, поскольку они наиболее просто и непосредственно связаны с физическими параметрами системы, выбираемыми при ее проектировании. Это обстоятельство является одной из причин интереса к коэффициентным методам оценки устойчивости и качества динамических систем [1].

Коэффициентные методы позволяют получить пусть и приближенные в некотором смысле, но простые соотношения, позволяющие связать (обычно некоторыми неравенствами) показатели качества САУ произвольного порядка и искомые параметры регулятора.

Одним из широко используемых при проектировании САУ критериев является максимальная степень устойчивости системы. Известно, что системы, синтезированные по этому критерию, при прочих равных условиях, обладают более высоким

быстродействием, меньшим перерегулированием и большим запасом устойчивости [1].

Синтезу линейных регуляторов, обеспечивающих максимальную степень устойчивости в стационарных САУ, посвящены работы [2–6]. В большинстве из них используется подход, предложенный в [3, 4], где используются полиномы, как правило, нормированные по определенному параметру. Для решения задачи обеспечения максимальной степени устойчивости представляет интерес также применение метода нелинейного программирования [5]. Однако получаемые в [5] системы уравнений позволяют аналитически находить максимальную степень устойчивости и обеспечивающие ее параметры регулятора только для систем низкого порядка, а для сложных САУ приходится применять численные методы синтеза.

Таким образом, представляет интерес решение на основе коэффициентного метода задачи максимизации степени устойчивости САУ произвольного порядка выбором соответствующих параметров стандартного пропорционально-интегрального (ПИ) регулятора. Заметим, что наряду с максимизацией степени устойчивости САУ необходимо также гарантировать в ней и требуемую точность управления, определяемую величиной добротности системы.

Постановка задачи

Пусть структурная схема САУ, включающая объект управления W_{OY} и регулятор W_p , имеет вид, показанный на рис. 1, где g – входное воздействие; ε – ошибка системы; u – управляющее воздействие; y – выходной сигнал системы. Объект управления описывается передаточной функцией

$$W_{OY}(s) = \frac{k_{OY}}{\sum_{i=0}^z d_i s^i},$$

где d_i – коэффициенты характеристического полинома; k_{OY} – коэффициент передачи объекта управления, s – оператор Лапласа; $z=1,2,3...$

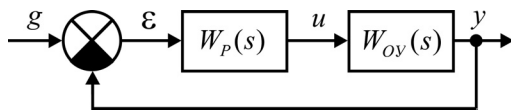


Рис. 1. Структурная схема САУ

Передаточная функция ПИ-регулятора имеет вид

$$W_p(s) = \frac{k_1 s + k_0}{s},$$

где k_0 и k_1 – настраиваемые параметры регулятора.

Тогда характеристический полином САУ может быть представлен в виде

$$A(s) = s \sum_{i=0}^z d_i s^i + k_{ov}(k_1 s + k_0).$$

Необходимо выбрать такие значения параметров k_0 и k_1 , которые на основе достаточных условий коэффициентов метода [1] обеспечивают близкую к максимальной (квазimaxимальную) степень устойчивости η^* и заданную добротность D системы управления. Полосы передаточной функции замкнутой САУ должны лежать левее вертикальной прямой, проходящей через точку $(-\eta^*, j0)$, рис. 2.

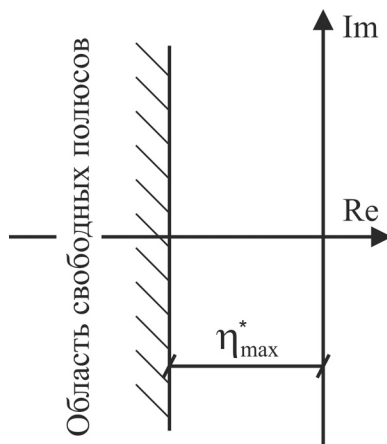


Рис. 2. Расположение полюсов САУ

Коэффициентные оценки устойчивости САУ

Рассмотрим линейную стационарную непрерывную систему с характеристическим полиномом

$$A(s) = a_n s^n + a_{n-1} s^{n-1} + \dots + a_0, \quad a_n > 0. \quad (1)$$

Введем вспомогательные параметры λ_i , образуемые четверками рядом стоящих коэффициентов (1):

$$\lambda_i = \frac{a_{i-1} a_{i+2}}{(a_i a_{i+1})}, \quad i = \overline{1, n-2}. \quad (2)$$

В соответствии с (2) $\lambda_1 = \frac{a_0 a_3}{a_1 a_2}$; $\lambda_2 = \frac{a_1 a_4}{a_2 a_3}$ и т. д.

Указанные параметры λ_i , называются показателями устойчивости [1]. Из [1] известно, что для устойчивости системы с характеристическим полиномом (1) достаточно, чтобы выполнялись неравенства

$$\lambda_i < \lambda^* \approx 0,465, \quad \forall i = \overline{1, n-2}, \quad (3)$$

$$\lambda_i + \lambda_{i+1} < \lambda^{**} \approx 0,89, \quad \forall i = \overline{1, n-3}. \quad (4)$$

Достаточные условия устойчивости являются основным математическим аппаратом при обеспечении устойчивости в методах синтеза на основе характеристического уравнения. Их простота позволяет строить хорошо алгоритмизируемые процедуры синтеза, а некоторая избыточность способствует получению устойчивости с запасом, всегда необходимым при проектировании реальных систем.

Оценка снизу максимальной степени устойчивости

В [1] предложены достаточные условия заданной степени устойчивости η , использующие коэффициенты характеристического полинома САУ

$$\left\{ \begin{array}{l} \frac{a_{i-1} a_{i+2}}{[a_i - a_{i+1}(n-i-1)\eta][a_{i+1} - a_{i+2}(n-i-2)\eta]} < \lambda^*, \\ k = \overline{1, n-2}; \\ a_i - a_{i+1}(n-l-1)\eta \geq 0, \quad l = \overline{1, n-1}; \\ a_0 - a_1 \eta + \frac{2a_2 \eta^2}{3} \geq 0. \end{array} \right. \quad (5)$$

Выполнение этих условий гарантирует расположение корней характеристического полинома (1) левее вертикальной прямой, проходящей через точку $(-\eta, j0)$. Очевидно, что увеличение η в указанных условиях позволяет найти его максимальное значение, которое будем рассматривать как оценку снизу степени устойчивости системы. Обозначим ее через η^* .

Показатель η^* предлагается использовать для параметрического синтеза линейного регулятора САУ коэффициентным методом. Задачей синтеза в этом случае является выбор настроек ПИ-регулятора k^* , при которых достигается максимум η^* . Обозначим его через η^*_{max} . Таким образом, $\eta^*_{max} = \max_k \eta^*$, где

η^*_{max} – оценка снизу максимальной степени устойчивости. Она является своего рода квазimaxимальной степенью устойчивости САУ.

Очевидно, что определить не оценочное, а реальное значение максимальной степени устойчивости можно, например, в результате непосредственного вычисления корней полинома (1) САУ с синтезированным регулятором.

Введем следующие обозначения

$$\lambda_i(\eta) = \frac{a_{i-1} a_{i+2}}{[a_i - a_{i+1}(n-i-1)\eta][a_{i+1} - a_{i+2}(n-i-2)\eta]}, \quad k = \overline{1, n-2}; \quad (6)$$

$$f_l(\eta) = a_l - a_{l+1}(n-l-1)\eta, \quad l = \overline{1, n-1};$$

$$g(\eta) = a_0 - a_1 \eta + \frac{2a_2 \eta^2}{3}.$$

Перепишем систему неравенств (5) в следующем виде

$$\begin{cases} \lambda_i(\eta) < \lambda^*, & i = \overline{1, n-2}; \\ f_l(\eta) \geq 0, & l = \overline{1, n-1}; \\ g(\eta) \geq 0. \end{cases} \quad (7)$$

При $\eta=0$ условия (7) соответствуют достаточным условиям устойчивости (3).

Таким образом, для определения квазикаксимальной степени устойчивости достаточно ($n-2$) раз решить следующую систему

$$\begin{cases} \lambda_i(\eta) = \lambda^*, & i = \overline{1, n-2}; \\ \lambda_j(\eta) < \lambda^*, & j = \overline{1, n-2}, \quad j \neq i; \\ f_l(\eta) \geq 0, & l = \overline{1, n-1}; \\ g(\eta) \geq 0. \end{cases} \quad (8)$$

находя на каждом шаге максимальное значение η^* , и после этого выбрать из них максимальное.

Поскольку найденное значение η_{\max}^* является квазикаксимальной степенью устойчивости САУ, следует заметить, что все корни синтезированной системы будут лежать левее вертикальной прямой, проведенной через точку $(-\eta_{\max}^*, j0)$.

Алгоритм параметрического синтеза регулятора

Для заданной передаточной функции объекта управления $W_{ov}(s)$ и функции ПИ-регулятора $W_p(s)$ величина добротности по скорости будет определяться выражением

$$D = \frac{k_0 k_{ov}}{d_0}. \quad (9)$$

Обозначим вектор настроечных параметров ПИ-регулятора $\bar{k} = [k_0, k_1]$. Таким образом, система (8) примет вид:

$$\begin{cases} \lambda_i(\bar{k}, \eta) = \lambda^*, & i = \overline{1, n-2}; \\ \lambda_j(\bar{k}, \eta) < \lambda^*, & j = \overline{1, n-2}, \quad j \neq i; \\ f_l(\bar{k}, \eta) \geq 0, & l = \overline{1, n-1}; \\ g(\bar{k}, \eta) \geq 0. \end{cases} \quad (10)$$

Для ее решения авторами разработана блок-схема алгоритма параметрического синтеза ПИ-регулятора, рис. 3.

Пример

Пусть объект управления задан передаточной функцией

$$W_{ov}(s) = \frac{k_{ov}}{d_3 s^3 + d_2 s^2 + d_1 s + d_0}, \quad (11)$$

где безразмерные коэффициенты имеют следующие значения: $d_0=1, d_1=0,61, d_2=0,056, d_3=0,005, k_{ov}=1$.

Необходимо выбрать параметры регулятора

$$W_p(s) = \frac{k_1 s + k_0}{s}, \quad (12)$$

обеспечивающего системе квазикаксимальную степень устойчивости и заданную точность. Пусть в соответствии с требованиями к точности необходимо обеспечить в системе добротность $D=10$.

На основании (11) и (12) характеристический полином системы будет иметь вид

$$d_3 s^4 + d_2 s^3 + d_1 s^2 + (k_{ov} k_1 + d_0) s + k_{ov} k_0 = 0, \quad (13)$$

Из (9) определим первый коэффициент регулятора $k_0=10$ и, подставляя в (13) численные значения коэффициентов передаточной функции объекта управления, получим следующие коэффициенты характеристического полинома САУ: $a_0(k_1)=10, a_1(k_1)=k_1+1, a_2=0,61, a_3=0,056, a_4=0,005$.

В соответствии с (6) сформируем выражения для показателей устойчивости $\lambda_i(k, \eta)$:

$$\begin{aligned} \lambda_1(k_1, \eta) &= \frac{10 \cdot 0,056}{[(k_1 + 1) - 0,61(4 - 1 - 1)\eta][0,61 - (4 - 1 - 2)\eta]} = \\ &= \frac{0,56}{[(k_1 + 1) - 1,22\eta][0,61 - \eta]}, \\ \lambda_2(k_1, \eta) &= \frac{0,005(k_1 + 1)}{[0,61 - 0,056(4 - 2 - 1)\eta][0,056 - (4 - 2 - 2)\eta]} = \\ &= \frac{0,005(k_1 + 1)}{0,056[0,61 - 0,056\eta]}, \end{aligned}$$

где обозначим

$$(k_1 + 1) - 1,22\eta = f_1(\bar{k}, \eta), \quad 0,61 - 0,056\eta = f_2(\bar{k}, \eta).$$

Согласно (10), для рассматриваемой задачи запишем систему

$$\begin{cases} \frac{0,56}{[(k_1 + 1) - 1,22\eta][0,61 - \eta]} = \lambda^*; & (14) \\ \frac{0,005(k_1 + 1)}{0,056[0,61 - 0,056\eta]} < \lambda^*; & (15) \\ (k_1 + 1) - 1,22\eta \geq 0; & (16) \\ 0,61 - 0,056\eta \geq 0; & (17) \\ 10 - (k_1 + 1)\eta + 0,407\eta^2 \geq 0. & (18) \end{cases}$$

Из (14) находим выражение $k_1(\eta)$ и подставляем его в (15)–(18). В результате получаем систему из четырех неравенств с одним неизвестным параметром η . В результате решения данной системы получаем значение $\eta_1^*=2,29$ и соответствующее ему значение параметра регулятора $k_1=4,294$. Из условия $\lambda_2(\bar{k}, \eta)=0,465$ аналогично находим $\eta_2^*=0,325$ и соответствующее ему значение параметра регулятора $k_1=29,821$. Очевидно, что искомыми параметрами будут максимальное значение $\eta_1^*=2,29$ и $k_1=4,294$.

На рис. 4 изображена реакция системы на линейно-нарастающее входное воздействие $g(t)$ и переходные характеристики системы с регуляторами (рис. 5).

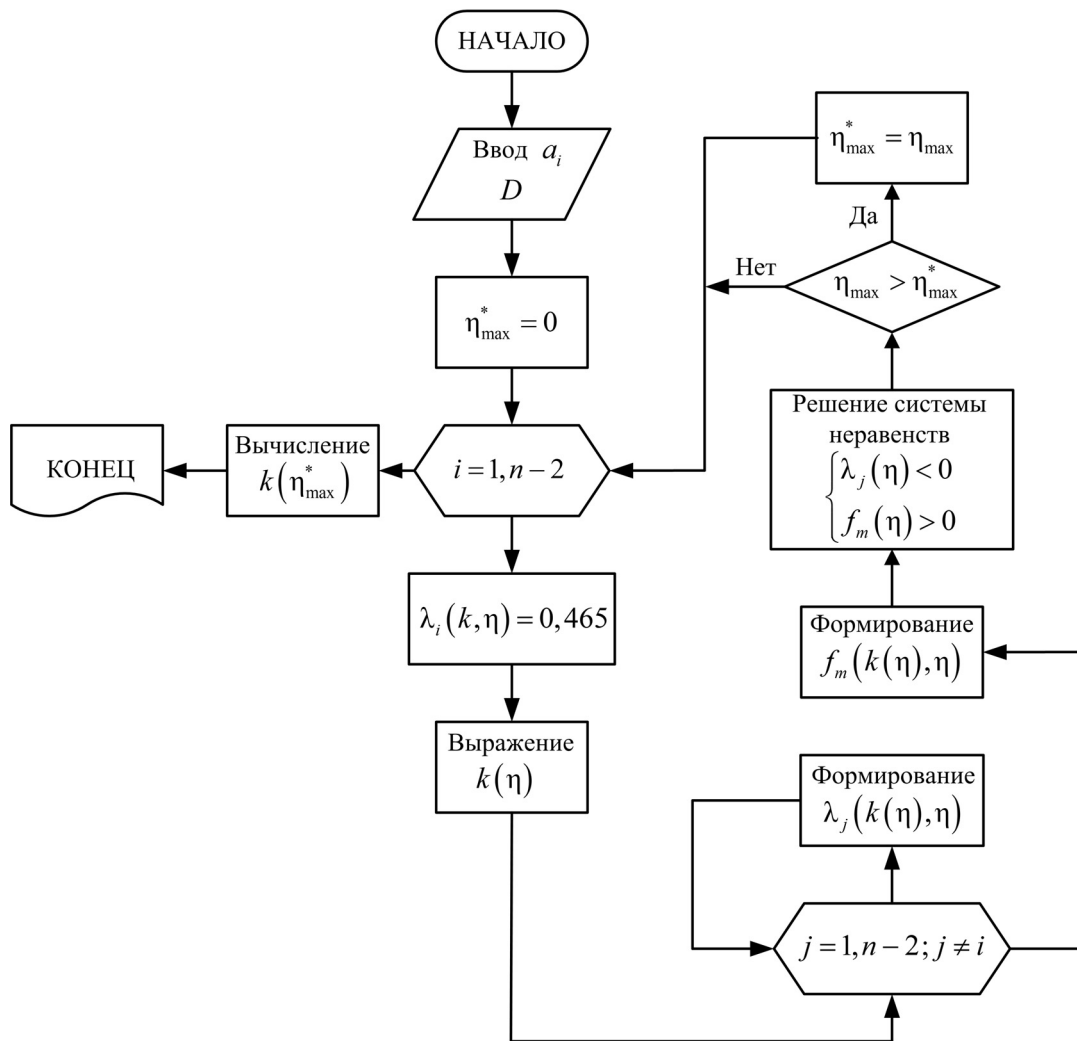


Рис. 3. Блок-схема алгоритма параметрического синтеза ПИ-регулятора

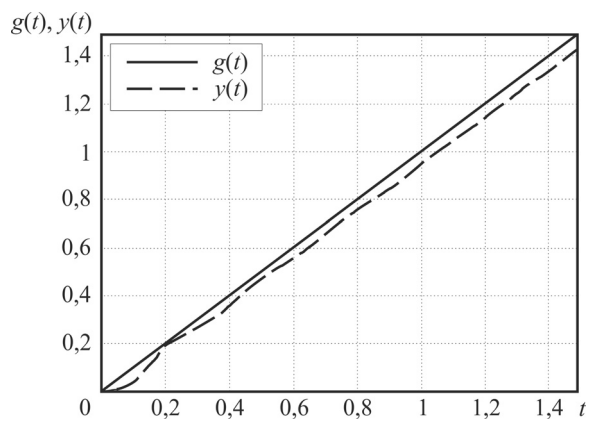


Рис. 4. Графики изменения входного $g(t)$ и выходного $y(t)$ сигналов системы

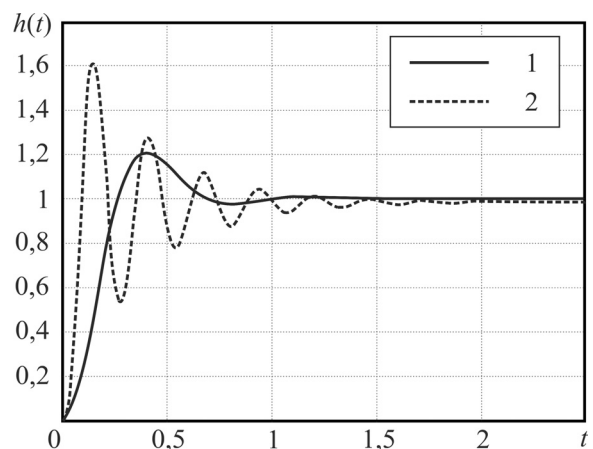


Рис. 5. Переходные характеристики системы при: 1) $\eta_1^* = 2,29$; 2) $\eta_2^* = 0,325$

Заметим, что при обоих значениях k_1 в системе гарантируется устойчивость и требуемая точность функционирования при реакции на линейно-растающее входное воздействие $g(t)$, рис. 4. Однако регулятор с коэффициентами, определенными из условия $\eta_1^* = 2,29$ обеспечивает большее быстродействие, а также меньшее перерегулирование в системе по сравнению с регулятором, обеспечивающим $\eta_2^* = 0,325$, что видно из рис. 5.

Выводы

1. Представлена методика параметрического синтеза ПИ-регулятора, обеспечивающего квазикасиальную степень устойчивости и заданную добротность системы автоматического управления.

СПИСОК ЛИТЕРАТУРЫ

1. Петров Б.Н., Соколов Н.И., Липатов А.В. и др. Системы автоматического управления объектами с переменными параметрами: Инженерные методы анализа и синтеза. – М.: Машиностроение, 1986. – 256 с.: ил.
2. Волков А.Н., Загашвили Ю.В. Метод синтеза систем автоматического управления с максимальной степенью устойчивости при наличии ограничений // Известия РАН. Сер. Теория и системы управления. – 1997. – № 3. – С. 12–19.
3. Шубладзе А.М. Способы синтеза систем управления максимальной степени устойчивости // Автоматика и телемеханика. – 1980. – № 1. – С. 28–37.
4. Шубладзе А.М. Методика расчета оптимальных по степени устойчивости ПИ-законов. I // Автоматика и телемеханика. – 1987. – № 4. – С. 16–25.

2. Устойчивость системы автоматического управления достигается путем расположения ее полюсов левее максимальной оценки снизу степени устойчивости системы, что обеспечивает максимальное быстродействие и снижает колебательность.
3. Алгоритм синтеза использует коэффициентные оценки показателей качества системы, на основе которых формируются упрощенные алгебраические соотношения между оценкой снизу степени устойчивости системы и коэффициентами характеристического полинома.
4. Полученные результаты рассмотрены и апробированы на численном примере.
5. Татаринов А.В., Цирлин А.М. Задачи математического программирования, содержащие комплексные переменные, и предельная степень устойчивости линейных динамических систем // Известия РАН. Сер. Теория и системы управления. – 1995. – № 1. – С. 28–33.
6. Воронина Н.О., Татаринов А.В., Цирлин А.М. Предельная степень аperiodической устойчивости и соответствующие ей настройки для типовых систем регулирования // Известия вузов. Сер. Приборостроение. – 1989. – № 3. – С. 26–32.

Поступила 28.03.2012 г.

УДК 681.5.015

ИДЕНТИФИКАЦИЯ ОБЪЕКТОВ УПРАВЛЕНИЯ В ФОРМЕ ДИСКРЕТНЫХ ПЕРЕДАТОЧНЫХ ФУНКЦИЙ НА ОСНОВЕ ВЕЩЕСТВЕННОГО ИНТЕРПОЛЯЦИОННОГО МЕТОДА

В.А. Рудницкий, А.С. Алексеев, В.В. Курганкин

Томский политехнический университет
E-mail: aleksejev@tpu.ru

Для решения задачи параметрической идентификации линеаризуемых объектов использован вещественный интерполяционный метод. На основе аппарата чисел обусловленности рассмотрена возможность повышения точности определения структуры модели в виде дискретной передаточной функции.

Ключевые слова:

Идентификация, объект управления, числа обусловленности, вещественный интерполяционный метод.

Key words:

Identification, controlled object, condition numbers, real interpolation method.

Введение

В современной технике управления широко применяются различные цифровые устройства. Прогресс в развитии микроэлектроники привел к улучшению основных качественных и количественных показателей микроконтроллеров, что сделало возможным эффективно решать задачи адаптации,

оптимизации и диагностики систем управления. В то же время успешное их решение невозможно без получения и использования адекватного математического описания объектов и систем.

Методы идентификации цифровых объектов во многих случаях удобно классифицировать по области описания математической модели:

в области времени [1, 2] и в области изображений, в частности, в области z -изображений [4] и частотной области [3]. Методы временной группы обладают сравнительно невысокой помехоустойчивостью и в общем случае требуют значительных вычислительных затрат. В случае представления модели системы с помощью частотных характеристик число операций оказывается сравнительно высоким из-за необходимости вычисления амплитудных и фазовых частотных характеристик объектов по довольно сложному алгоритму. Наиболее широкое распространение получило использование аппарата z -преобразования для идентификации моделей цифровых объектов.

В связи с этим целью настоящего исследования является разработка эффективного алгоритма идентификации объектов управления, поведение которых можно достаточно адекватно описать с помощью линейных математических моделей, представленных в дискретном времени.

Задачу идентификации линеаризуемых объектов управления в форме дискретных передаточных функций (ПФ) можно решать двумя путями – на основе непрерывной и цифровой модели. В первом случае используется косвенный путь: по экспериментальным отсчетам входного X и выходного Y сигналов находится ПФ $W_0(p)$, которая в последующем переводится в класс дискретных. Во втором случае определяется непосредственно дискретная ПФ $W_0(z)$ идентифицируемого объекта, структурная схема которого представлена на рис. 1.

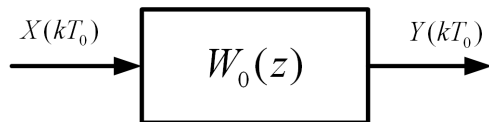


Рис. 1. Модель объекта управления

При использовании любого из этих подходов задача идентификации объекта формулируется следующим образом. Известны входной $X(kT_0)$ и выходной $Y(kT_0)$ сигналы, где T_0 – период квантования; k – номер отсчета. Необходимо определить параметры ПФ вида

$$W_0(z) = \frac{Y(z)}{X(z)} = \frac{b_m z^m + b_{m-1} z^{m-1} + \dots + b_0}{a_n z^n + a_{n-1} z^{n-1} + \dots + 1}, \quad n \geq m. \quad (1)$$

Обычно структурные показатели m, n известны или заданы, что соответствует параметрической постановке задачи.

Общей проблемой различных алгоритмов идентификации остается довольно сложная процедура получения математических моделей исследуемых объектов, требующая для своей реализации существенных вычислительных затрат. По этой причине разработка более эффективных численных способов решения задачи идентификации дискретных объектов и систем остается актуальной. Определенные перспективы в реализации этих возможностей дает применение вещественного интерполяционного метода (ВИМ).

1. Вещественный интерполяционный метод

В основе использования ВИМ для решения задачи идентификации цифровых объектов лежит вещественное дискретное преобразование

$$F^*(\delta) = \sum_{k=0}^{\infty} f(kT_0) e^{-kT_0 \delta}, \quad c \leq \delta \leq \infty, \quad c \geq 0, \quad (2)$$

в котором функции-оригиналу $f(kT_0)$ соответствует изображение $F^*(\delta)$ [5]. В частном случае, когда рассматриваются устойчивые объекты и используются их импульсные переходные характеристики, параметр c может быть принят равным нулю [5, 6].

Для использования ВИМ в реальных задачах рекомендуется переходить от преобразования (2) к более простому в вычислительном отношении v -преобразованию. С этой целью достаточно выполнить подстановку $v = \exp(T_0 \delta)$ в выражении (2), что приводит к соотношению

$$F(v) = \sum_{k=0}^{\infty} f(kT_0) v^{-k}, \quad v \in [C, \infty], \quad C \geq 1. \quad (3)$$

В формуле (3) переменная v не может принимать значения меньше единицы, что соответствует области устойчивости на плоскости z . Основные свойства дискретного v -преобразования во многом близки свойствам z -преобразования [5].

Понятие вещественного изображения распространяется на описание линейных импульсных систем и протекающих в них процессов. Если на некоторую линейную импульсную систему действует входной сигнал $X(kT_0)$ и известна реакция системы $Y(kT_0)$, то на основании (2) и (3) можно записать уравнение «вход-выход объекта» в вещественной области

$$Y(v) = W(v) \cdot X(v). \quad (4)$$

Для выполнения численных действий над вещественными изображениями необходимо перейти от непрерывных функций $F(v)$ к их дискретным аналогам в виде численных характеристик. Под ними, как и в случае дискретного вещественного преобразования, понимают совокупность значений

$$\{F(v)\}_\eta = \{F(v_1), F(v_2), \dots, F(v_\eta)\}, \quad \eta = m + n + 1, \quad (5)$$

заданных на сетке $1 \leq v_1 < v_2 < \dots < v_\eta$. Значения узлов интерполирования $v_i, i \in \overline{1, \eta}$, могут назначаться по известным рекомендациям, изложенным в [5, 6].

Рассмотрим приложение аппарата вещественных функций-изображений и численных характеристик к решению задачи идентификации. Положим в (4): $X(v)$ – единичный входной сигнал; $Y(v)$ – реакция объекта на это входное воздействие. По формуле (3) для принятого входного сигнала можно найти изображение $X(v) = v/(v-1)$, где $v \in [1, \infty)$. Выходной сигнал реакции объекта $Y(v)$ по определению будет представлять собой переходную функцию $h(v)$. В этом случае при нулевых начальных условиях ур. (4) примет вид

$$h(v) = W(v) \frac{v}{v-1}, \quad v \in [C, \infty]. \quad (6)$$

Положим далее, что рассматривается класс объектов управления с самовывравниванием, что

соответствует абсолютному большинству практических задач. Тогда в соотношении (6) можно принять $C=1$ [7]. Приведенные сведения позволяют перейти непосредственно к решению задачи параметрической идентификации.

2. Алгоритм параметрической идентификации цифровых объектов

Задача заключается в определении коэффициентов вещественной ПФ вида

$$W(v) = \frac{b_m v^m + b_{m-1} v^{m-1} + \dots + b_0}{a_n v^n + a_{n-1} v^{n-1} + \dots + 1}, \quad (7)$$

по переходной характеристике, представленной своим вещественным изображением $h(v)$. Для решения задачи используются численная характеристика переходной функции $\{h(v_i)\}_\eta$ и входного сигнала $\left\{ \begin{matrix} v_i \\ v_i - 1 \end{matrix} \right\}_\eta$. Размерность η характеристик (5)

принимается равной числу неизвестных коэффициентов ПФ (7). В этом случае равенство (6) можно развернуть в систему η линейных алгебраических уравнений (СЛАУ), которая имеет решение и оно единственно [5].

Техника формирования СЛАУ включает в себя получение численной характеристики единичного ступенчатого сигнала и переходной функции

$\{h(v_i)\}_\eta$. Элементы $X(v_i) = \frac{v_i}{v_i - 1}$, $v_i > 1$ находятся

расчетным путем, а получение значения $\{h(v_i)\}_\eta$ можно осуществить по формуле (3), зная последовательность $h(kT_0)$, где $k=0,1,2,\dots,N$:

$$h(v_i) = \sum_{k=0}^N h(kT_0) \cdot v_i^{-k}, \quad v_i \in [1, \infty]. \quad (8)$$

Параметр N в (8) определяется временем наблюдения переходного процесса. Далее на основании (6) определяется численной характеристикой искомой ПФ:

$$W(v_i) = \frac{v_i - 1}{v_i} \cdot h(v_i), \quad i = \overline{1, \eta}, \quad \eta = m + n + 1.$$

Теперь имеются все необходимые данные для вычисления коэффициентов функции (7). Задача свелась к решению СЛАУ

$$W(v_i) = \frac{b_m v_i^m + b_{m-1} v_i^{m-1} + \dots + b_0}{a_n v_i^n + a_{n-1} v_i^{n-1} + \dots + 1} \cdot \frac{v_i - 1}{v_i}, \quad i = \overline{1, \eta}, \quad \eta = m + n + 1. \quad (9)$$

Систему уравнений (9) при решении практических задач целесообразно представлять в матричной форме

$$D \cdot K = W, \quad (10)$$

где

$$K = [b_m, b_{m-1}, \dots, b_0, a_n, a_{n-1}, \dots, a_1]^T, \\ W = [W(v_1)v_1, W(v_2)v_2, \dots, W(v_\eta)v_\eta]^T,$$

$$D = \begin{bmatrix} v_1^{m+1} - v_1^m & v_1^m - v_1^{m-1} & \dots & v_1 - 1 & -W(v_1)v_1^{n+1} & \dots & -W(v_1)v_1 \\ v_2^{m+1} - v_2^m & v_2^m - v_2^{m-1} & \dots & v_2 - 1 & -W(v_2)v_2^{n+1} & \dots & -W(v_2)v_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ v_{\eta-1}^{m+1} - v_{\eta-1}^m & v_{\eta-1}^m - v_{\eta-1}^{m-1} & \dots & v_{\eta-1} - 1 & -W(v_{\eta-1})v_{\eta-1}^{n+1} & \dots & -W(v_{\eta-1})v_{\eta-1} \\ v_\eta^{m+1} - v_\eta^m & v_\eta^m - v_\eta^{m-1} & \dots & v_\eta - 1 & -W(v_\eta)v_\eta^{n+1} & \dots & -W(v_\eta)v_\eta \end{bmatrix}$$

Поиск решения может осуществляться известными методами, т. к. СЛАУ не содержит каких-либо особенностей. Трудности могут возникнуть на этапе формирования СЛАУ, в частности, с получением первого уравнения – для $v=v_1$. Рассмотрим особенности этого уравнения. Ранее принято ограничение $v_1 \geq 1$, как условие существования v -преобразования и сходимости ряда (8), что однако не указывает на конкретное значение v_1 . Так, в случае неудачного выбора значения v_1 матрица D может оказаться плохо обусловленной. Для устранения этого недостатка существует подход, который снимает трудности выбора первого узла интерполирования [5].

С этой целью рассмотрим взаимную связь конечных значений оригиналов с их изображениями. Для них справедливо равенство

$$\lim_{k \rightarrow \infty} f(kT_0) = \lim_{v \rightarrow 1} \frac{v-1}{v} f(v),$$

если эти пределы существуют. Узел интерполирования v_1 и значение $f(v_1)$ обеспечивают сохранение информации о поведении функции $f(kT_0)$ преимущественно на бесконечности. Эту информацию можно задать также и в области изображений по ПФ $W(v)$ (7). Поэтому можно записать

$$\lim_{k \rightarrow \infty} h(kT_0) = \lim_{v \rightarrow 1} \frac{v-1}{v} h(v) = \\ = \lim_{v \rightarrow 1} \frac{v-1}{v} W(v) \frac{v}{v-1} = W(1).$$

В большинстве случаев значение $h(nT_0)$ в установившемся режиме известно из исходных данных, поэтому преобразуем предыдущее выражение к виду

$$h(kT_0)_{k \rightarrow \infty} = \frac{b_m + b_{m-1} + \dots + b_0}{a_n + a_{n-1} + \dots + 1}, \quad (12)$$

обеспечивающему совпадение установившихся значений выходов идентифицируемого объекта и его модели. Включение последнего уравнения в состав СЛАУ (9) снимает проблему выбора первого узла интерполирования v_1 . Узел v_2 не оказывает критического влияния на обусловленность матрицы СЛАУ даже при равномерном распределении узлов интерполирования [5].

Для моделей объектов с самовыравниванием в форме (7), полученных при начальных нулевых условиях, возможно осуществить переход к z -форме $W_0(z)$ (1) путем формальной замены $v \rightarrow z$, что является важной практической чертой ВИМ и вещественного дискретного преобразования.

Последним этапом в решении задачи идентификации цифрового объекта или системы является

оценивание погрешности полученного решения и, если это необходимо, его коррекция, обычно сводящаяся к минимизации погрешности. В этой связи необходимо сформировать критерий близости в виде

$$\begin{aligned} \Delta h &= \max_k |\Delta h(kT_0)| = \\ &= \max_k \left| 100 \cdot \frac{h_0(kT_0) - h_m(kT_0)}{h_0(kT_0)} \right| \rightarrow \min, \end{aligned} \quad (11)$$

где $h_m(kT_0)$ – переходная функция идентифицированной модели, а также описать процедуру минимизации Δh .

При решении задачи идентификации цифровых объектов на основе ВИМ может возникнуть ситуация, когда отсутствует достаточная априорная информация о порядке ПФ исследуемого объекта. С целью формализации выбора порядка модели возможно привлечение аппарата обусловленности матрицы D . В виду важности и практической значимости проблемы поиска подходящей структуры модели рассмотрим этот подход подробнее.

3. Применение чисел обусловленности для повышения точности решения задачи идентификации

Изменение значений степеней полиномов числителя m и знаменателя n ПФ модели объекта означает изменение размерности соответствующих матриц, описывающих модели сигналов и их взаимосвязи. Высокая размерность матриц может привести к результатам, которые могут не соответствовать физическому содержанию задачи. Свойства идентифицированной модели могут весьма значительно отличаться от свойств самого объекта. Для минимизации негативного влияния этого обстоятельства и получения результатов, имеющих практическое значение, применяются различные методы регуляризации. Эти методы нуждаются в наличии определенного объема априорных данных о свойствах объекта, в целом ряде практических задач такие сведения отсутствуют [8]. В связи с этим возникает вопрос о необходимости введения в алгоритм решения задачи идентификации некоторого механизма, который позволил бы контролировать корректность задачи.

В практических задачах целесообразно использовать математический аппарат чисел обусловленности, применяемый для апостериорной оценки ошибок при решении СЛАУ (10) [9, 10]. Этот подход рассматривает числа обусловленности как количественную оценку меры обусловленности задачи, т. е. оценку того, насколько она чувствительна к ошибкам исходных данных.

Выделяют три типа чисел обусловленности [9]. Естественное число обусловленности $K_1(D)$ определено выражением

$$K_1(D) = \frac{\|D^{-1}\| \|W\|}{\|K\|}.$$

Оно может быть легко получено, если применить общую теорию обусловленности к частному случаю СЛАУ.

Число обусловленности Эйрда–Линча определено зависимостью

$$K_2(D) = \frac{\|D^{-1}W\|}{\|K\|}.$$

Важное свойство числа $K_2(D)$ состоит в том, что позволяет получать верхнюю и нижнюю границы ошибки при обращении матрицы D .

Числа обусловленности $K_1(D)$, $K_2(D)$ непосредственно зависят от нормы $\|K\|$, а само точное решение в задаче идентификации не известно. Поэтому перспективы применения чисел $K_1(D)$ и $K_2(D)$ в рассматриваемой задаче весьма ограничены. С этой точки зрения более перспективно использование стандартного числа обусловленности $K_3(D) = \|D^{-1}\| \cdot \|D\|$, которое может быть найдено на основе вычисления векторных норм для матриц D^{-1} и D [9]:

$$\|D\| = \max_j \sum_{i=1}^n |\lambda_{ij}|,$$

$$\|D^{-1}\| = \max_j \sum_{i=1}^n |\lambda_{ij}^{-1}|,$$

где λ_{ij} , λ_{ij}^{-1} – элементы матриц D и D^{-1} соответственно.

Формулы для $\|D\|$ и $\|D^{-1}\|$ свободны от отмеченных недостатков первых двух стандартных чисел обусловленности и потому использование числа $K_3(D)$ более перспективно.

Однако общим недостатком каждой из вышеприведенных оценок остается необходимость нахождения нормы $\|D^{-1}\|$, поскольку найти норму матрицы без вычисления самой обратной матрицы D^{-1} удается чрезвычайно редко. Вычисление элементов обратной матрицы D^{-1} с заранее гарантированной достаточной точностью в некорректных задачах, к которым относится задача идентификации, практически невозможно. В этой связи большие преимущества для применения идентификации в рамках ВИМ может дать использование подхода, предложенного в [8].

Известно, что любая норма матрицы не меньше ее наибольшего собственного значения, т. е. $\|D\| \geq \max |\lambda_D|$. Кроме того, собственные значения матриц D и D^{-1} взаимно обратны и поэтому справедливо соотношение:

$$\|D^{-1}\| \geq \max \frac{1}{|\lambda_D|} = \frac{1}{\min |\lambda_D|}.$$

Отсюда выражение для нахождения стандартного числа обусловленности $K_3(D)$ принимает вид:

$$K_3(D) = \frac{\max |\lambda_D|}{\min |\lambda_D|}.$$

Отслеживая изменения величины стандартного числа обусловленности в каждом конкретном слу-

чае можно найти наилучшее соотношение параметров структуры модели. Значение $K_3(D)=1$ свидетельствует о наиболее благоприятном случае, когда оценка относительной погрешности решения задачи идентификации совпадает с оценкой относительной погрешности исходных данных. К сожалению такая ситуация редко возникает в практических задачах [9, 10]. Зачастую $K_3(D) \gg 1$ и имеет порядок сотен, тысяч и более.

Числа обусловленности $K_3(D)$ позволяют делать некоторые выводы относительно структуры модели объекта идентификации. При первоначальных малых значениях параметров модели m и n значение $K_3(D)$ будет наиболее близким к наилучшему случаю $K_3(D)=1$, хотя величины относительной ошибки $\Delta h(kT_0)$ временных динамических характеристик объекта идентификации и его модели могут быть значительными. С целью уменьшения ошибки $\Delta h(kT_0)$ следует увеличивать m и n . При этом будет происходить рост величины $K_3(D)$. На некотором этапе может сложиться ситуация, когда дальнейшее уменьшение относительной ошибки $\Delta h(kT_0)$ за счет усложнения структуры модели окажется не в состоянии компенсировать погрешность, вносимую погрешностями вычислений и некорректностью задачи. Это проявляется в виде резкого, иногда лавинообразного нарастания величины $K_3(D)$, что показывает возрастание ошибки при решении СЛАУ. Полученная информация позволяет обоснованно выбрать значения m и n , при которых модель объекта или системы будет иметь наименьшую погрешность в соответствии с критерием (11). Для подтверждения сказанного рассмотрим численный пример.

4. Численный пример

Пусть известна переходная функция объекта управления, аналитическое выражение которой имеет вид

$$h_0(t) = 1 - 2,3 \cdot 0,368^t + 1,5 \cdot 0,135^t - 0,2 \cdot (1,832 \cdot 10^{-2})^t.$$

Известно время квантования $T_0=0,5$ с и время наблюдения $t_n=12$ с. Найдем модель объекта в форме дискретной ПФ

$$W(z) = \frac{b_m z^m + b_{m-1} z^{m-1} + \dots + b_0}{a_n z^n + a_{n-1} z^{n-1} + \dots + 1}, \quad n \geq m$$

для различных значений m и n . Для каждой из полученных моделей найдем огибающие их переходных характеристик, оценим близость полученных решений к $h_0(nT_0)$ по критерию (11). Значения числа $K_3(D)$ для различных m и n используем для выбора модели, имеющей наибольшую точность по (11). Расчеты по приведенному ранее алгоритму идентификации дискретных объектов выполнялись в среде программного пакета MathCAD. Результаты расчетов сведены в таблицу.

На рис. 2 построены графики переходных функций объекта $h_0(t)$ и модели $h_m(kT_0)$, для которой получено наименьшее значение критерия (11).

Таблица. Результаты идентификации объекта управления

m	n	$K_3(D)$	ПФ идентифицированных моделей	$\Delta h, \%$
0	1	1,67	$W(z) = \frac{-0,75}{-1,75z + 1}$	21
1	1	44,41	$W(z) = \frac{9,81 \cdot 10^{-3} z - 0,69}{-1,69z + 1}$	14
1	2	294,29	$W(z) = \frac{-0,15z - 0,53}{-1,39z^2 - 0,28z + 1}$	7
2	2	$8,32 \cdot 10^3$	$W(z) = \frac{-2,19 \cdot 10^{-3} z^2 - 0,16z - 0,59}{-1,61z^2 - 0,14z + 1}$	5
2	3	$6,78 \cdot 10^8$	$W(z) = \frac{-12,11z^2 - 0,88z + 2,41}{-34,28z^3 + 32,73z^2 - 10,04z + 1}$	1

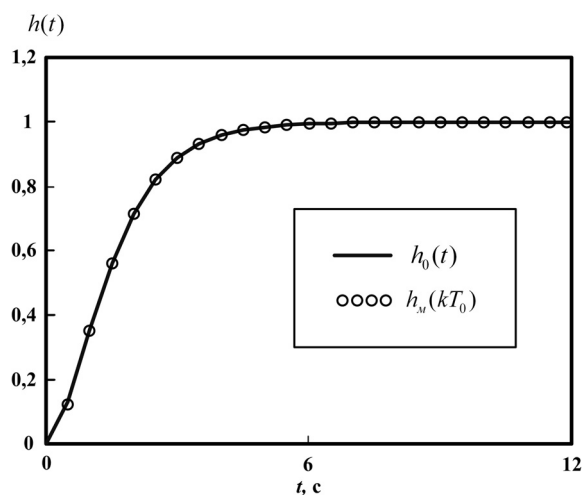


Рис. 2. Графики переходных характеристик

Из таблицы видно, что при возрастании параметров m и n точность идентификации увеличивается. При этом число $K_3(D)$ имеет минимальное значение 1,67 при $m=0, n=1$. Последующее увеличение m и n приводит к снижению относительной ошибки Δh и росту числа $K_3(D)$, например, при $m=2, n=3$ значение $K_3(D)$ возрастает до $6,78 \cdot 10^8$, а относительная ошибка Δh не превышает 1%. Дальнейшее увеличение значений m и n модели даже на единицу приводит к тому, что получить модель объекта в классе устойчивых систем уже не удастся. Наиболее вероятная причина – резкий рост вычислительной погрешности, обусловленной некорректностью задачи. На основании полученных результатов можно определить структуру модели с параметрами $m=2$ и $n=3$. Дальнейшее снижение погрешности идентификации возможно на основе применения методов регуляризации, но эти вопросы в данной статье не рассматриваются.

Выводы

Применение вещественного интерполяционного метода для идентификации дискретных объектов и систем управления позволяет получить эффективный в вычислительном плане алгоритм идентификации. Для получения моделей, в наибольшей степени удовлетворяющих условиям кон-

кретной задачи по точности, сложности математического описания и т. д., целесообразно использовать математический аппарат стандартных чисел обусловленности, который достаточно просто вписывается в структуру алгоритма идентификации. Эти обстоятельства позволяют рассматривать ве-

щественный интерполяционный метод как эффективный инструмент для решения практических задач адекватного математического описания цифровых объектов и систем.

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

СПИСОК ЛИТЕРАТУРЫ

1. Балакирев В.С., Дудников Е.Г., Цирлин А.М. Экспериментальное определение динамических характеристик промышленных объектов управления. – М.: Энергия, 1967. – 232 с.
2. Ордынцев В.М. Математическое описание объектов автоматизации. – М.: Машиностроение, 1965. – 360 с.
3. Гроп Д. Методы идентификации систем. – М.: Мир, 1979. – 302 с.
4. Цыпкин Я.З. Основы информационной теории идентификации. – М.: Наука, 1984. – 320 с.
5. Гончаров В.И. Синтез электромеханических исполнительных систем промышленных роботов. – Томск: Изд-во ТПУ, 2002. – 100 с.
6. Алексеев А.С., Антропов А.А., Гончаров В.И., Замятин С.В., Рудницкий В.А. Вещественный интерполяционный метод

7. Дёч Г. Руководство к практическому применению преобразования Лапласа и z-преобразования. – М.: Наука, 1971. – 288 с.
8. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. – М.: Бином. Лаборатория Знаний, 2003. – 630 с.
9. Райс Дж. Матричные вычисления и математическое обеспечение. – М.: Мир, 1984. – 265 с.
10. Трофимов А.И., Егупов Н.Д., Дмитриев А.Н. Методы теории автоматического управления, ориентированные на применение ЭВМ. Линейные стационарные и нестационарные модели. – М.: Энергоатомиздат, 1997. – 651 с.

Поступила 06.04.2012 г.

УДК 62–533.65

СИНТЕЗ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ПОДСИСТЕМЫ УПРАВЛЕНИЯ ПРОГРЕВОМ ВАКУУМНОЙ КАМЕРЫ КАЗАХСТАНСКОГО МАТЕРИАЛОВЕДЧЕСКОГО ТОКАМАКА КТМ

А.Г. Коровиков, В.М. Павлов*, Д.А. Ольховик

Институт атомной энергии НЯЦ Республики Казахстан, г. Курчатов

*Томский политехнический университет

E-mail: Korovikov@nnc.kz

Исследован процесс прогрева вакуумной камеры токамака КТМ как объект управления, разработана математическая модель процесса как объекта управления. Сделан вывод о том, что для управления технологическим процессом прогрева необходимо и достаточно использовать систему управления, построенную на основе ПИ-регуляторов. Определены коэффициенты параметров настройки регулятора для каждой зоны прогрева.

Ключевые слова:

Система управления, математическая модель, вакуумная камера, нагрев.

Key words:

Control system, mathematical model, vacuum chamber, heat.

В настоящее время в г. Курчатов, Республика Казахстан, ведутся работы по реализации проекта строительства термоядерной установки Казахский материаловедческий токамак КТМ.

На этапе подготовки установки к эксперименту вакуумная камера токамака КТМ и патрубки должны прогреваться до температуры 200 °С с целью обезгаживания и уплотнения вакуумных конструкций. Процесс должен проходить с обеспечением контроля равномерности прогрева. Техническая структура системы управления подробно рассмотрена в статье [1].

Синтез систем автоматического регулирования (САР) является главной задачей теории управле-

ния, в результате определяются состав, структура САР и параметры всех ее устройств из условия соответствия заданному комплексу технических требований [2, 3]. Основной задачей, которая должна быть решена на данном этапе, является получение модели объекта управления, его математического описания, которое представляет собой, систему обычных линейных/нелинейных дифференциальных или алгебраических уравнений.

При построении математической модели подсистемы омического и индукционного нагрева воспользуемся структурным методом. На первом этапе исходная принципиальная схема заменяется упрощенной функциональной схемой. Для

этого исследуемая система разбивается на функциональные элементы, каждый из которых в дальнейшем будет представлен типовым структурным звеном, т. е. описан дифференциальным уравнением не выше второго порядка. При построении считаем систему абсолютно управляемой, то есть все возмущающие воздействия в системе, примем равными нулю, а также предполагаем, что САР температуры вакуумной камеры (ВК) является стационарной. Предположение о стационарности является некоторой идеализацией, так как не учитывается изменение элементов системы со временем.

Функциональные схемы подсистемы омического и индукционного нагрева представлены на рис. 1 и 2.

Источник питания ленточных нагревателей, уложенных на внешней поверхности вакуумной камеры токамака, представляет собой шестиканальный регулятор переменного напряжения с широтно-импульсным управлением. Все каналы независимые. Выходное напряжение ~ 220 В (50 Гц). Контроллер формирует управляющие сигналы на открытие тиристорных ключей в зависимости от заданной мощности прогрева. Напряжение на выходе пропорционально коэффициенту заполнения управляющих сигналов источника питания:

$$U_{\text{ип}}(t) = K_{\text{ип}} U_y,$$

где $K_{\text{ип}}$ – коэффициент пропорциональности источника питания.

Передаточная функция источника питания:

$$W_{\text{ип}}(p) = K_{\text{ип}}.$$

Источник питания – типовое безынерционное звено.

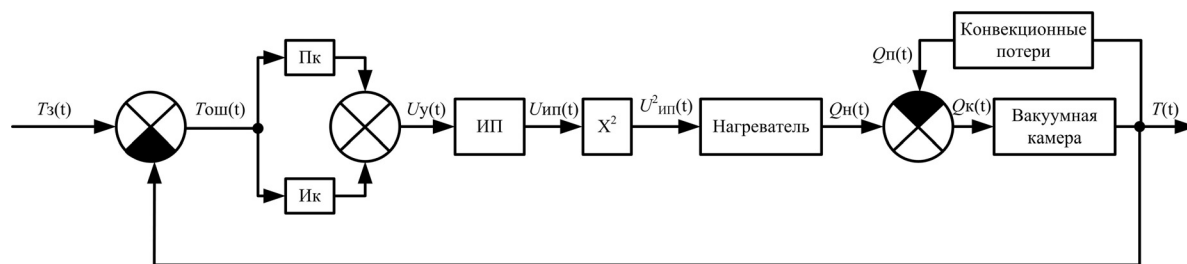


Рис. 1. Функциональная схема подсистемы омического нагрева. Пк – пропорциональный канал регулирования; Ик – интегральный канал регулирования; ИП – источник питания; $T_z(t)$ – заданная температура; $Toш(t)$ – сигнал рассогласования (ошибка); X^2 – блок возведения в квадрат; $U_y(t)$ – сигнал управления; $U_{ип}(t)$ – сигнал с выхода источника питания; $Q_n(t)$ – количество теплоты, вложенное нагревателем; $Q_k(t)$ – количество теплоты, разогрева вакуумной камеры; $Q_p(t)$ – конвективные потери; $T(t)$ – температура вакуумной камеры

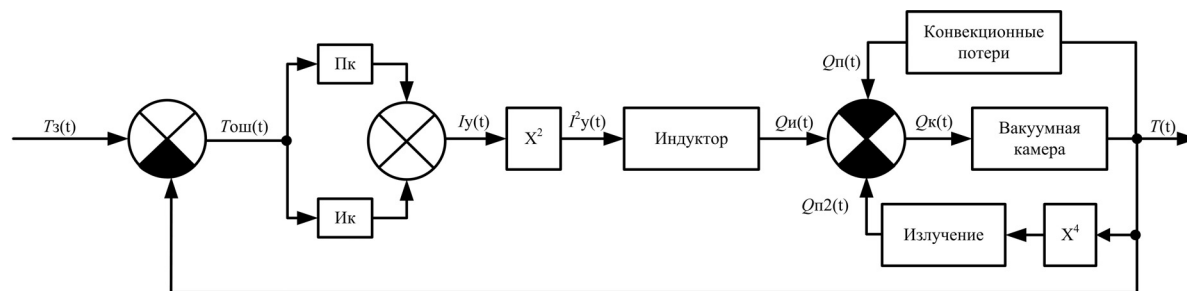


Рис. 2. Функциональная схема подсистемы индукционного нагрева. $I_y(t)$ – сигнал управления; $Q_{и}(t)$ – количество теплоты, выделяемое индуктором; $Q_{п2}(t)$ – потери от излучения тепла

Определим действующее значение напряжения источника питания, разделив максимальное значение напряжения на корень из 2. Учитывая, что сигнал управления принимает значение в диапазоне 0...100 %, коэффициент усиления источника будет рассчитываться как:

$$K_{\text{ип}} = \frac{U_{\text{ип}}}{U_y} = \frac{220}{100\sqrt{2}} = 1,556.$$

Количество теплоты, выделенное в нагреваемом материале, определяется по закону Джоуля–Ленца:

$$Q_n = I^2 R \Delta t = \frac{U_{\text{ип}}^2}{R} \Delta t.$$

Отсюда передаточная функция ленточного нагревателя:

$$W_{\text{нар}}(p) = \frac{Q_n}{U_{\text{ип}}^2} = \frac{1}{Rp}.$$

Таким образом, нагреватель может быть описан как интегрирующее звено.

Характеристики зон вакуумной камеры сведены в табл. 1.

Для определения температуры каждой зоны вакуумной камеры используется зависимость вложенного количества теплоты от свойств материала камеры, начальной и конечной температуры.

$$Q_k = cm(T_{\text{кон}} - T_{\text{нач}}),$$

где m – масса металла, кг; c – удельная теплоемкость металла, Дж/(кг·град);

Передаточная функция, описывающая процесс нагрева ВК, имеет вид:

$$W_{\text{вк}}(p) = \frac{T_{\text{кон}}}{Q} = \frac{1}{cm}.$$

Таблица 1. Характеристики зон вакуумной камеры

Нагреваемая зона ВК	Площадь нагреваемой зоны, м ²	Вес зоны, кг	Сопротивление нагревателя, Ом	Длина нагревателя, м	Номер зоны нагрева
Внутренний цилиндр	3,8	198	–	–	1
Патрубки	17,8	2590	7,2	16	2
Крышка	8,1	1910	15,66	17,4	3
Конус	8,6	335	14,625	13	4
Внешний цилиндр	4,2	590	7,2	8	5
Днище	3,1	2208	8,55	9,5	6

Теплообмен между поверхностью вакуумной камеры и потоком воздуха с другой температурой осуществляется за счет конвективной теплоотдачи. Тепловой поток при конвективном теплообмене рассчитывают по формуле Ньютона:

$$\Phi = \lambda F \Delta T,$$

где Φ – тепловой поток, Вт; λ – коэффициент теплоотдачи, Вт/(м²·град); F – площадь поверхности теплоотдачи, м².

Количество теплоты, переданное от внешней поверхности ВК при конвективном теплообмене, будет равно:

$$Q_n = \Phi t = \lambda F \Delta T t.$$

Передаточную функцию, описывающую процесс конвективных потерь, можно записать в виде:

$$W_{\text{конв}}(p) = \frac{Q_n}{T} = \frac{\lambda F}{p}.$$

Коэффициент теплоотдачи ВК через теплоизоляцию равен 0,1 Вт/(м²·град). Коэффициент конвективной теплоотдачи зоны ВК «Внутренний цилиндр» зависит от скорости воздуха и направления потока по отношению к вертикальной оси ВК, диаметра внутреннего цилиндра, теплофизических характеристик воздуха. При ламинарном режиме движения воздуха (критерий Рейнольдса Re меньше 1000) коэффициент теплоотдачи будет равен, Вт/(м²·град):

$$\lambda = \frac{0,43 \beta_v Re^{0,5} \lambda_v}{D} = 0,115;$$

$$Re = \frac{v \beta_v D}{\nu_v} = 109,629,$$

где v – скорость движения воздуха (при расчетах задано 0,01 м/с); ν_v – кинематическая вязкость воздуха (при расчетах задано $1,51 \cdot 10^{-5}$ м²/с); λ_v – теплопроводность воздуха (при расчетах задано 0,02485 Вт/(м·град)); β_v – поправочный коэффициент на скорость потока воздуха (при расчетах задано 0,55); D – диаметр зоны ВК «Внутренний цилиндр» (при расчетах задано 0,534 м); β_v – поправочный коэффициент, учитывающий высоту расположения ВК и характер местности (при расчетах задано 0,31).

Количество теплоты, переданное от ВК при лучистой теплоотдаче, зависит от температуры возду-

ха, температуры поверхности ВК, а также от степени черноты поверхности ВК.

$$Q_{п2} = \Phi_{из} t = \epsilon_n C_0 F \left(\left(\frac{T_{ВК}}{100} \right)^4 - \left(\frac{T_{возд}}{100} \right)^4 \right) t,$$

где C_0 – коэффициент излучения абсолютно черного тела ($C_0=5,67$ Вт/(м²·К⁴)); ϵ_n – коэффициент лучистой теплоотдачи (при расчетах принято $\epsilon_n=0,9$ [4]); $T_{ВК}$ – текущая температура вакуумной камеры, К; $T_{возд}$ – температура воздуха, К (при расчетах задано 290 К).

Передаточную функцию, описывающую процесс лучистой теплоотдачи, можно записать в виде:

$$W_{\text{луч}}(p) = \frac{Q_{п2}}{T_{ВК}^4} = \frac{1}{p} \left(19,3914 \cdot 10^{-8} - \frac{1371,517}{T_{ВК}^4} \right).$$

Индукционный нагрев металлов основан на законах электромагнитной индукции Фарадея–Максвелла и Джоуля–Ленца. Внутренний цилиндр вакуумной камеры находится в переменном магнитном поле, которое возбуждает в нем вихревое электрическое поле. Под действием ЭДС индукции в камере протекают вихревые (замкнутые внутри тел) токи, выделяющие теплоту по закону Джоуля–Ленца.

Определим расчетный диаметр зоны ВК «Внутренний цилиндр» при нагреве цилиндрической плоскости изнутри, м:

$$D'_2 = D_2 + \xi = D_2 + M \Delta_k = D_2 + M \frac{0,5}{\sqrt{f}} = 0,7531,$$

где D_2 – диаметр зоны ВК «Внутренний цилиндр», м; f – частота тока, Гц; ξ – эффективная глубина проникновения тока, м; M – коэффициент. $M=0,27$ – определяется из табл. 3, 4 [5].

Активное и реактивное сопротивления нагреваемого слоя равны, Ом:

$$r_2 = \pi D'_2 2,81 \cdot 10^{-6} K \sqrt{f} \frac{\cos \varphi}{a} = 0,00136.$$

где $K=3,45$ и $\cos \varphi=0,77$ – определяются из [5, 6, рис. 3–7] соответственно; $a=0,092$ м – ширина индуктора.

$$x_{m2} = r_2 \frac{\sin \varphi}{\cos \varphi} = 0,001125.$$

Реактивное сопротивление, Ом:

$$x_e = \frac{x_{10} k_1}{1 - k_1} = 1,217 \cdot 10^{-4},$$

где $x_{10} = \frac{\omega \mu_0 S_1}{a} = 3,13 \cdot 10^{-4}$ Ом; $k_1=0,28$ – определяется из [5, рис. 6–2]; S_1 – площадь боковой поверхности индуктора. $S_1=0,23$ м².

Реактивное сопротивление рассеяния, Ом:

$$x_s = \frac{\omega \mu_0 (S_2 - S_1)}{a} = 2,651 \cdot 10^{-4},$$

где S_2 – площадь боковой поверхности зоны ВК «Внутренний цилиндр». $S_2=0,423$ м².

Коэффициент приведения активного сопротивления материала ВК (сталь 12Х18Н10Т), Ом:

$$c = \frac{1}{\left(\frac{r_2}{x_e}\right)^2 + \left(1 + \frac{x_s + x_{m2}}{x_e}\right)^2} = 1,079 \cdot 10^{-4}.$$

Приведенное активное сопротивление равно, Ом:

$$r'_2 = cr_2 = 1,4674 \cdot 10^{-7}.$$

Расчет приведенного активного сопротивления произведен для одновиткового индуктора ($w=1$), в индукторе токамака КТМ $w=423$, отсюда следует:

$$r'_{423} = r'_2 w^2 = 0,02626 \text{ Ом}.$$

Мощность, передаваемая в оболочку вакуумной камеры, равна:

$$Q_n = I_y^2 r'_{423} t.$$

Передаточная функция индуктора имеет вид:

$$W_{\text{инд}}(p) = \frac{Q_n}{I_y^2} = \frac{r'_{423}}{p}.$$

В результате синтеза математической модели были получены выражения для всех динамических звеньев. Модель, в том числе блоки возведения в квадрат и четвертую степень, реализованы в приложении Simulink инженерного математического пакета Matlab. Проведенные исследования разгонных характеристик (реакция на ступенчатое воздействие), логарифмическая амплитудно-частотная характеристика (ЛАЧХ) и логарифмическая фазово-частотная характеристика (ЛФЧХ) моделей объекта управления для зон нагрева, показали (рис. 3, а), что подсистемы омического и индукционного нагрева обладают свойством самовыравнивания, и являются устойчивыми, т. к. годограф (рис. 3, б) не охватывает точку с координатой $(-1j, 0)$ (критерий Найквиста).

ЛАЧХ и ЛФЧХ подсистемы управления прогревом ВК без регуляторов представлены на рис. 3, в.

В задаче стабилизации наиболее важны свойства установившегося режима. Для следящих систем в первую очередь нужно обеспечить высокое качество переходных процессов при изменении задающего сигнала (уставки) [6]. Можно выделить три основных требования:

- точность;
- устойчивость;
- качество переходных процессов.

Для управления системой омического нагрева наиболее рационально использовать алгоритм с фиксированной структурой, в состав которой входят пропорциональная и интегральная составляющие, т. е. ПИ-алгоритм, т. к. система нагрева обладает высокой инерционностью.

В результате синтеза САР были получены параметры настройки ПИ-регуляторов для каждой зоны ВК (табл. 2). Для настройки применен метод СНР, который основан на использовании критерия максимальной скорости нарастания при отсутствии перерегулирования или при наличии не более чем 20%-го перерегулирования [7].

Таблица 2. Параметры настройки ПИ-регуляторов

Зона ВК	Параметр регулятора		Номер зоны нагрева
	K	T, с	
Внутренний цилиндр	89 А/град	7,8	1
Патрубки	1,004 1/град	0,0191	2
Крышка	0,843 1/град	0,0197	3
Конус	0,998 1/град	0,0183	4
Внешний цилиндр	0,735 1/град	0,0195	5
Днище	0,4723 1/град	0,0225	6

Подтверждением результатов проектирования любой системы регулирования, оценкой ее характеристик соответствию техническим требованиям, предъявляемым к проектируемой системе, являются модельные и/или экспериментальные исследования.

К основным техническим характеристикам системы относятся следующие:

- диапазон регулируемой температуры от 0 до 250 °С;
- максимальная динамическая ошибка регулирования (этап входа на стационарный режим) не более 5 %;
- максимальная статическая ошибка поддержания заданного уровня температуры (режим поддержания температуры ВК на уровне 200 °С в течение длительного времени, необходимого для проведения дегазации) не более 2 %.

Анализ результатов исследований показал, что в подсистеме управления прогревом ВК перерегулирование в переходных режимах не превышает 5 %, а в режиме поддержания температуры не более 2 % (рис. 4). Система полностью устойчива.

Оценивая быстродействие системы управления по времени регулирования, получаем, что время регулирования не превышает $1,25 \cdot 10^5$ с для 5 % отклонения выходной величины от установившегося значения (рис. 4). Минимальный темп роста температуры на начальном участке прогрева – не менее 0,1 град/мин.

Запас устойчивости системы с регуляторами составляет:

- по амплитуде – бесконечность;
- по фазе – от 160 до 167° в зависимости от зоны нагрева.

Выводы

Произведен синтез подсистемы управления прогревом вакуумной камеры токамака КТМ. Определен тип регулятора и параметры его настройки. Показано, что для управления технологическим процессом прогрева необходимо и достаточно использовать систему управления, построенную на основе ПИ-регуляторов. Подсистема управления позволяет вывести вакуумную камеру на оптимальную рабочую температуру с минимальными затратами времени. Перерегулирование не превышает 5 %, а в режиме стабилизации температуры – не более 2 %.

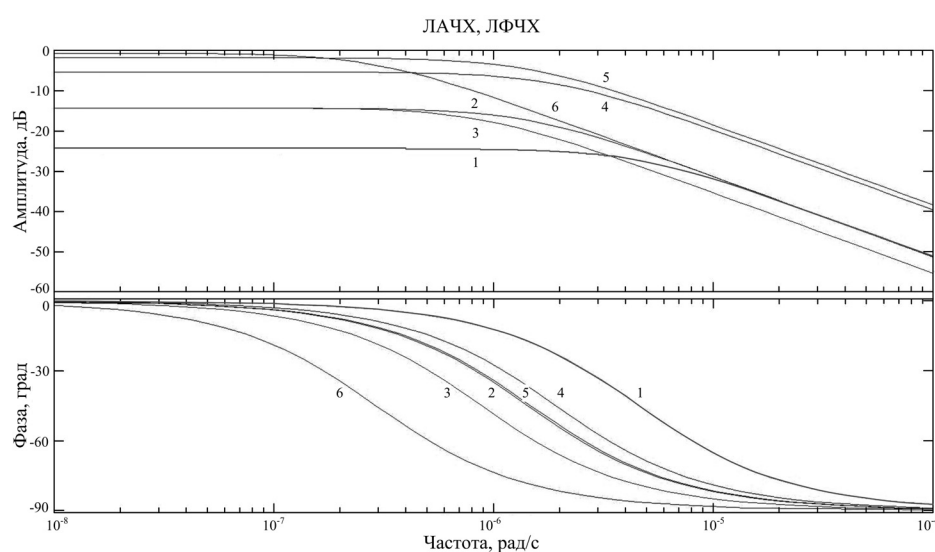
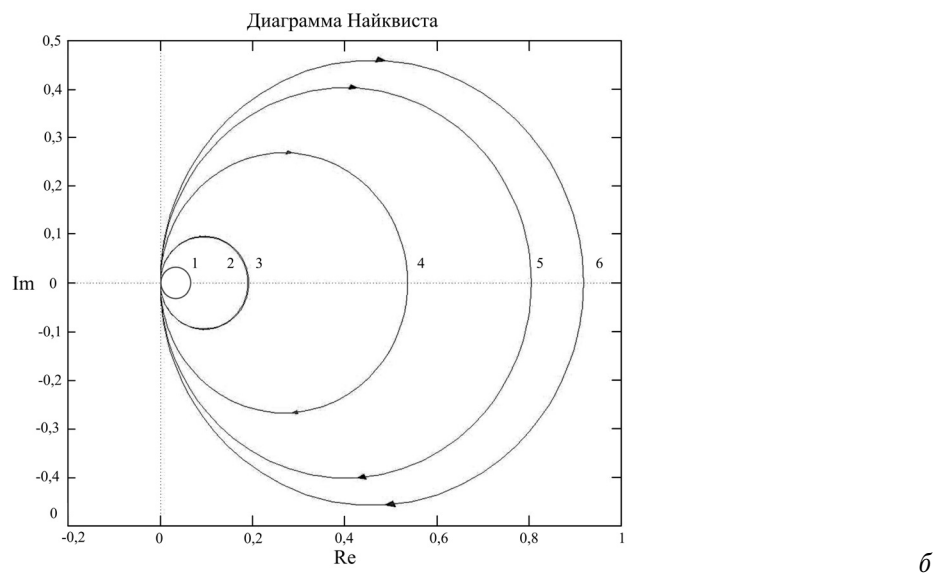
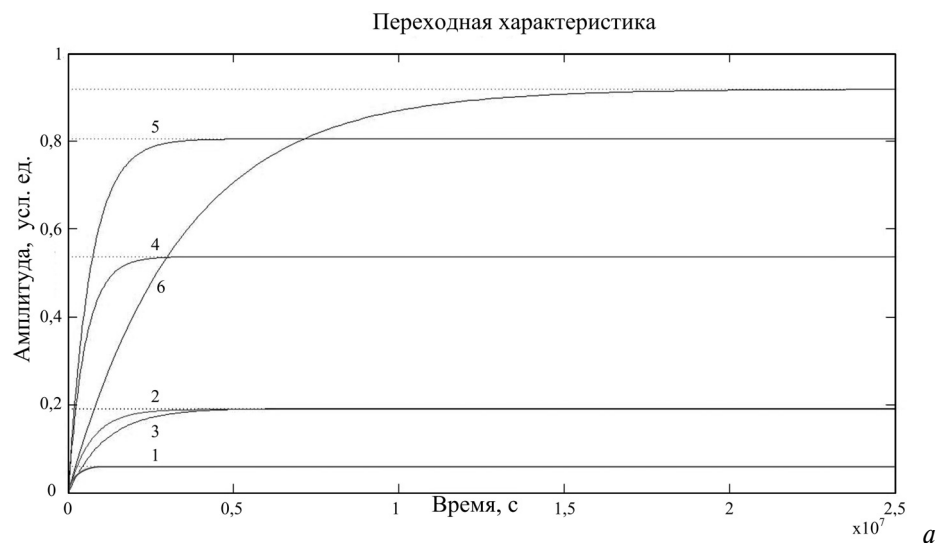


Рис. 3. Подсистемы управления прогревом ВК: а) переходные процессы подсистемы управления прогревом ВК; б) годографы Найквиста; в) ЛАЧХ и ЛФЧХ. Зона: 1) «Внутренний цилиндр»; 2) «Патрубки»; 3) «Крышка»; 4) «Конус»; 5) «Внешний цилиндр»; б) «Днище»

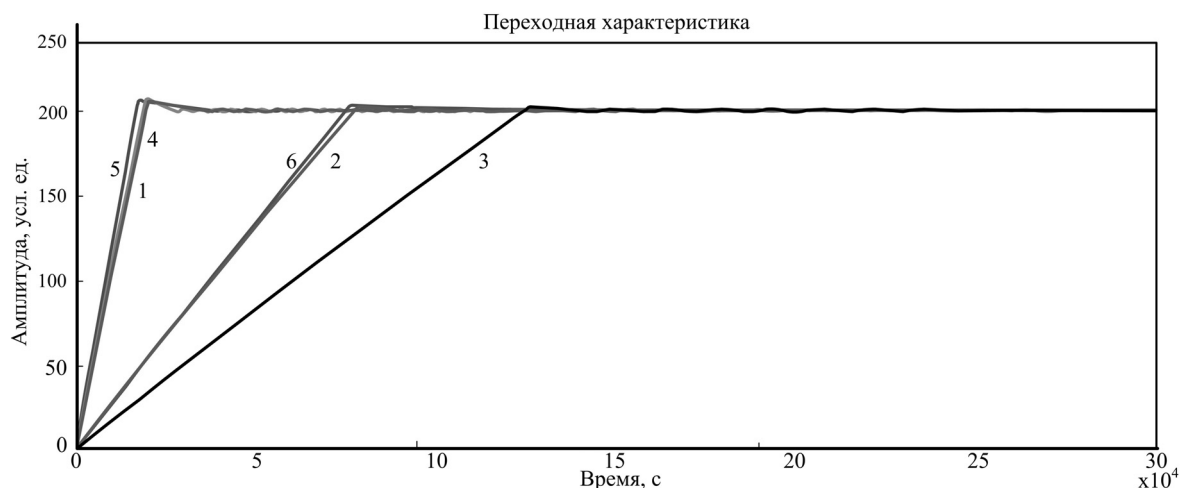


Рис. 4. Реакция подсистемы нагрева с ПИ-регуляторами на ступенчатое воздействие. Зона: 1) «Внутренний цилиндр»; 2) «Патрубки»; 3) «Крышка»; 4) «Конус»; 5) «Внешний цилиндр»; 6) «Днище»

СПИСОК ЛИТЕРАТУРЫ

1. Коровиков А.Г., Павлов В.М., Ольховик Д.А. Алгоритмическое обеспечение подсистемы управления прогревом вакуумной камеры токамака КТМ // Известия Томского политехнического университета. — 2011. — Т. 318. — № 5. — С. 116–119.
2. Методы классической и современной теории автоматического управления. В 3-х т. / под ред. Н.Е. Егупова. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2000. — Т. 2: Синтез регуляторов и теория оптимизации систем автоматического управления. — 736 с.
3. Кулаков Г.Т. Анализ и синтез систем автоматического регулирования. — М.: Технопринт, 2003. — 134 с.
4. Нащокин В.В. Техническая термодинамика и теплопередача. — М.: Высшая школа, 1975. — 496 с.
5. Сухоцкий А.Е. Установки индукционного нагрева. — Л.: Энергоиздат, 1981. — 330 с.
6. Коновалов Б.И., Лебедев Ю.М. Теория автоматического управления. — Томск: ТУСУР, 2010. — 162 с.
7. Chien K.L., Hrones J.A., Reswick J.B. On automatic control of generalized passive systems // Transactions of the ASME. — 1972. — № 74. — P. 175–185.

Поступила 20.01.2012 г.

УДК 681.516.75

СИСТЕМА УПРАВЛЕНИЯ ПРОМЫШЛЕННОЙ УСТАНОВКОЙ «КО-МКР»

О.В. Пащенко, А.М. Соловьев, С.В. Юдаков

Томский политехнический университет
E-mail: ysv@tpu.ru

Описана система управления промышленной установки «КО-МКР». Кратко отражены конструкция и технологические возможности установки. Обоснован выбор структуры системы управления и центрального контроллера. Описан принцип построения управляющей программы контроллера. Отражены конструктивные особенности системы управления.

Ключевые слова:

Система управления, контроллер ХРАС, рефлектор.

Key words:

Control system, controller ХРАС, reflector.

Введение

Технология ионно-плазменного нанесения многослойных радиоотражающих покрытий на изделия сложной геометрии и большой площади весьма сложна. Она требует взаимодействия многих систем и агрегатов – высоковакуумной системы, системы термического отжига, магнетронов и ионных источников, напуска рабочего газа, механических приводов, систем контроля и т. д. – в реальном масштабе времени. Система должна быстро реагировать на изменения параметров процессов, не допускать ухудшения качества покрытий и порчи весьма дорогостоящих изделий, предотвращать аварийные ситуации.

Вышесказанное обуславливает необходимость создания автоматической системы управления установкой, формулирует требования к ней.

Цель работы – создание системы управления промышленной вакуумной установкой «КО-МКР» (комплект оборудования для металлизации крупногабаритных рефлекторов), снабжённой источниками плазмы магнетронного типа и ускоренных ионов.

1. Постановка задачи. Описание технологического процесса

В последнее время углепластик получает широкое распространение в различных областях техники. Прочность данного материала позволяет заменить им многие металлические детали, а его легкость значительно уменьшает их массу. Одной из таких деталей можно назвать рефлекторы параболических антенн. В связи с этим возникает задача по увеличению отражающей способности поверхности углепластика, которая решается осаждением тонкопленочного покрытия. Обычно в качестве материала отражающего слоя этого покрытия применяется алюминий, для предотвращения повреждения отражающий слой защищают пленкой из диоксида кремния или оксида алюминия. В работе [1] авторами был обоснован выбор структуры и материального состава покрытия вида {Ni-Cr} (0,2 мкм) – {Al} (до 3 мкм) – {SiO₂} (до 0,2 мкм).

Перед нанесением радиотражающего покрытия необходимо выполнить обезгаживание рефлектора для обеспечения адгезионной прочности покрытий. Обезгаживание проводится при давлении в вакуумной камере менее 1 Па при температуре рефлектора 100 °С в течении 10 ч. Скорость нагрева рефлектора 1...3 град./мин. Точность поддержания температуры ±3 °С.

Процедура осаждения радиотражающего покрытия на поверхность рефлектора на готовой к работе установке включает в себя следующие операции:

- тепловая дегазация рефлектора;
- очистка поверхности рефлектора с помощью ионного пучка;
- осаждение адгезионного слоя из сплава {Ni-Cr};
- осаждение плёнки алюминия;
- осаждение защитного слоя {SiO₂}.

Под эту технологическую схему в Физико-техническом институте ТПУ была спроектирована и изготовлена установка «КО МКР». Для неё была разработана специальная система управления, отвечающая условиям, изложенным выше.

2. Основные элементы установки как объекты управления

Установка «КО МКР» состоит из следующих связанных между собой подсистем, которыми необходимо управлять:

- вакуумная;
- напуска рабочего газа;
- электропитания магнетронов и ионных источников;
- транспортная (перемещение тележки и вращение рефлектора);
- нагрева;
- охлаждения;
- электромеханических приводов магнетронов;
- пневматическая.

Подсистемы вакуумная и напуска рабочего газа (рис. 1) предназначены для поддержания рабочего давления (разрежения) и состава газовой среды в вакуумной камере.

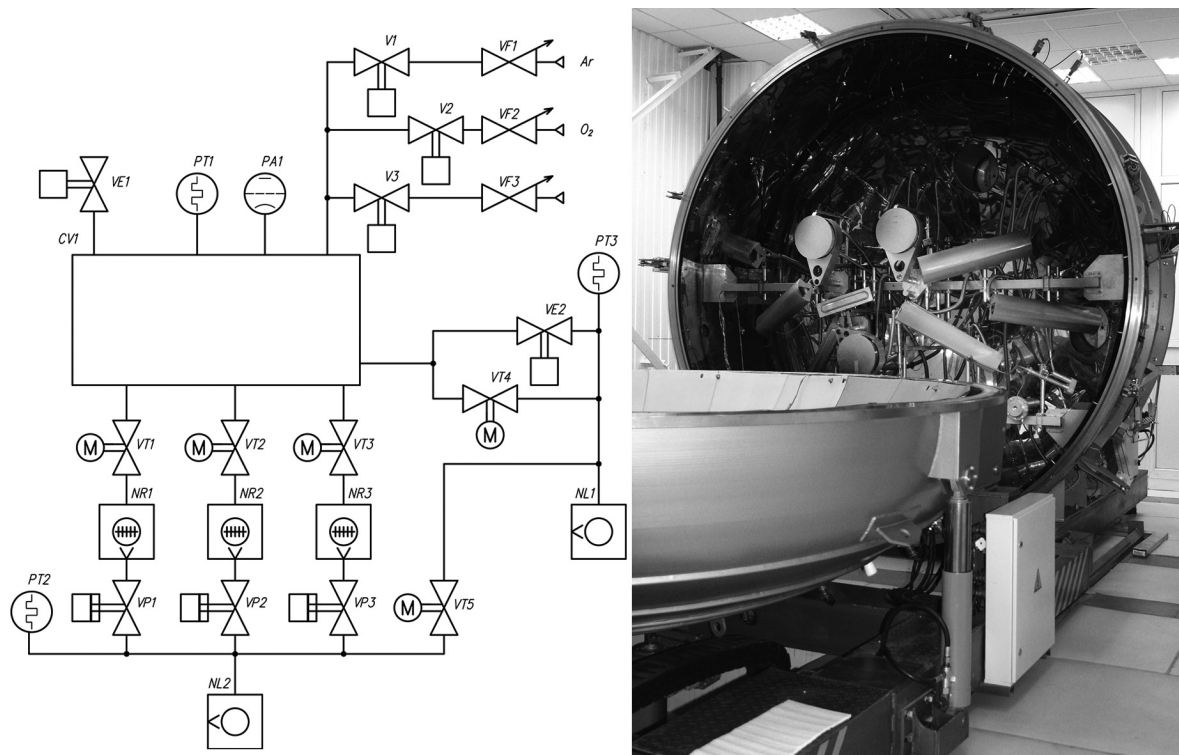


Рис. 1. Вакуумная схема установки «КО МКР» слева и фотография вакуумной камеры справа (CV1 – вакуумная камера, VT1–VT4 – вакуумные затворы, NL1, NL2 – форвакуумные насосы, NR1–NR3 – турбомолекулярные насосы, V1–V3 – на- текатели рабочего газа, VE1, VE2 – вакуумные клапаны, VF1–VF3 – расходомеры, PT1–PT3 – термопарные вакуумные лампы, PA1 – ионизационная вакуумная лампа)

Вакуумная подсистема включает в себя вакуумную камеру, вакуумные насосы (форвакуумные «Kashiya-та» MU-1203 и турбомолекулярные «Shimadzu» TMP-4203 LMC), затворы тарельчатые 23ВЭ-400, клапаны, трубопроводы, а также датчики и приборы контроля вакуума (АТВ-2 и АИВ-51). Подсистема напуска рабочего газа содержит баллоны со сжатым газом (Ar, O₂), клапаны (НМБ-1), трубопроводы и расходомеры (устройства, контролирующие величину потока рабочего газа, поступающего в вакуумную камеру) типа РРГ12 фирмы «Элтотприбор».

В качестве источников плазмы и ионных пучков используются соответственно планарные магнетроны (9 шт.) на постоянном токе [2] и протяжённые источники ионов (3 шт.) с замкнутым дрейфом электронов [3]. Подсистема электропитания магнетронов и ионных источников включают в себя блоки питания, а также сопутствующие защитные и распределительные устройства. Она предназначена для поддержания необходимых для технологического цикла значений тока и напряжения в этих диодных системах.

Транспортная подсистема предназначена для перемещения подвижной тележки, вращения фермы с рефлектором и для управления гидравлической системой подъема крышки вакуумной камеры.

Подсистема нагрева предназначена контролируемого разогрева и поддержания заданной температуры рефлектора. Включает в себя нагреватели, переизлучатель, схему управления нагревателями

и системы измерения температуры поверхности рефлектора.

Подсистема охлаждения установки состоит из клапанов, датчиков потока и системы охлаждения рабочей жидкости – воды. Охлаждению подлежат вакуумные насосы, магнетроны, ионные источники, стенки вакуумной камеры. Мощность отводимого потока тепла может достигать 100 кВт.

Все магнетроны снабжены управляемыми заслонками, а центральные магнетроны (3 шт.) подвижны и выполняют сканирование при нанесении покрытия на рефлектор. Подсистема электро-механических приводов магнетронов предназначена для управления заслонками магнетронов и сканированием. Для повышения равномерности тонкопленочных покрытий система способна изменять мощность магнетронов при сканировании по задаваемому технологом закону.

Пневматическая подсистема предназначена для управления пневмозамками передней крышки камеры и пневматическими клапанами вакуумной системы (VP1–VP3). Состоит из блока подготовки воздуха, пневмоцилиндров и пневмоклапанов.

3. Структура системы управления

Можно выделить три основных функции системы управления установкой:

- управляющая (все виды автоматизированного управления, автоматическое регулирование, технологические защиты и блокировки);

- информационная (сбор, обработка, хранение информации о технологических параметрах и управляющих воздействиях, решение вычислительных задач, формирование отчетов, визуализация);
- сервисная (диагностика датчиков, сетевых связей).

Оптимальным способом построения системы управления, как показывает практика, является децентрализованный.

Это означает, что некоторые компоненты установки, такие как натекатели рабочего газа, электроприводы, блоки питания магнетронов и ионных источников являются «интеллектуальными» и имеют в своем составе специализированные контроллеры, позволяющие этим устройствам работать автономно. Центральный контроллер осуществляет координацию работы всех устройств путем непосредственного управления исполнительными механизмами и задания параметров для работы «интеллектуальных» устройств. В качестве центрального контроллера мы используем программируемый логический контроллер (ПЛК) серии XPC фирмы «ICPDAS» (Тайвань). Опыт эксплуатации подобных контроллеров в системах управления вакуумных установок, разработанных ранее в лаборатории [1, 4], свидетельствует о надежности выбранного ПЛК.

Традиционно в системе управления мы выделяем четыре уровня:

- 1) датчиков и исполнительных устройств (вакуумные датчики давления, натекатели рабочего газа, вакуумные и форвакуумные насосы, вакуумные затворы, клапаны и т. д.);
- 2) модулей сопряжения с объектами управления (специализированные оптоизолированные блоки с транзисторными выходами, реле и т. п.);
- 3) ПЛК, который обеспечивает контроль и управление всей установкой;
- 4) оператора/технолога.

Для обеспечения надежности третий и четвертый уровни были совмещены и реализованы на ПЛК.

В системе управления «КО МКР» имеются блокировки на всех четырех уровнях, которые при сбое в работе оборудования (в том числе ПЛК) или из-за некорректных действий оператора не дадут возможности выйти из строя отдельным устройствам, таким как вакуумные затворы, приводы перемещения крышки вакуумной камеры.

4. Программное обеспечение

Программное обеспечение центрального контроллера написано на языке C++ и реализует бесконечный управляющий цикл (рис. 2) с периодом не более 100 мс, обеспечивающий:

- опрос состояния устройств установки. Опрос устройств, коммутируемых по медленным каналам связи, таким как RS-232 или RS-485, реализуется специальным контроллером в отдельных потоках, что способствует уменьшению периода цикла и, как следствие, уменьшению времени реакции управляющей программы;
- посылку управляющих сигналов и команд устройствам установки, определяемых логикой алгоритмов работы системы и командами оператора;
- реализацию макрокоманд-алгоритмов (например, «создать высокий вакуум в рабочей камере»), заключающихся в последовательном и параллельном условном выполнении команд устройствами;
- блокировку недопустимых в данной ситуации команд оператора;
- автоматическое, без участия оператора, выполнение макрокоманд-алгоритмов по выведению установки из аварийных или опасных состояний;
- автоматическое выполнение заданного оператором режима – тепловой дегазации или ионно-плазменной обработки.

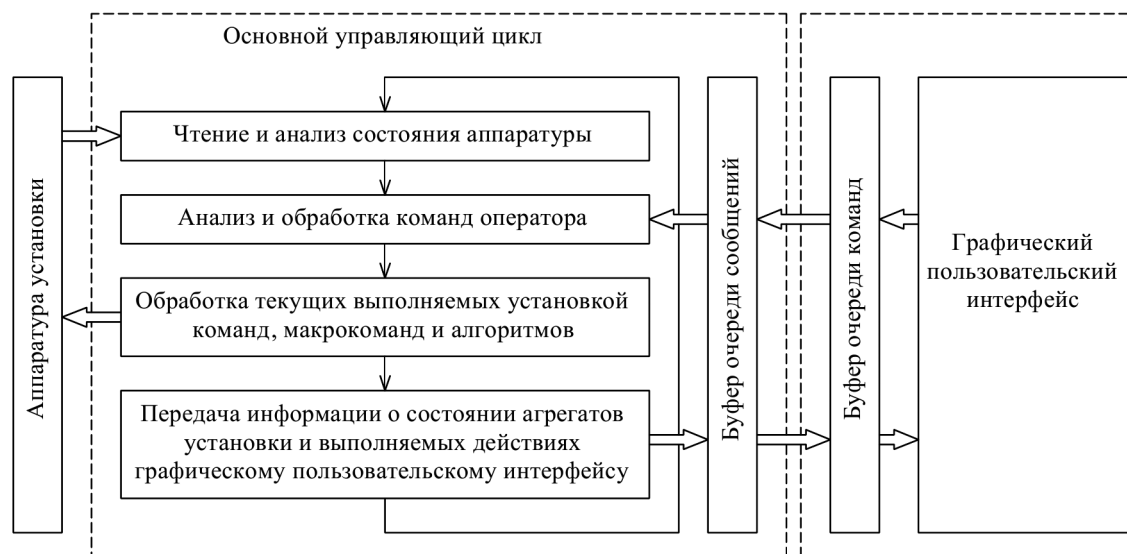


Рис. 2. Структура программы центрального контроллера

Графический пользовательский интерфейс программы обеспечивает:

- визуализацию состояния и активности систем и агрегатов установки в виде мнемосхемы и таблицы состояний объектов;
- протоколирование состояния систем и агрегатов и действий оператора;
- визуализацию основных параметров технологической обработки (параметры облучения, вакуумные условия), их протоколирование для последующего анализа;
- выдачу команд оператора;
- создание и хранение библиотеки режимов облучения для нанесения покрытий различных оптических свойств.

5. Конструктивные особенности

В составе установки имеется 3 ионных источника, 9 магнетронов, 3 турбомолекулярных насоса, 2 форвакуумных вакуумных насоса, 5 вакуумных затвора, и 9 вакуумных клапанов, 5 гидроклапанов, 15 датчиков потока воды, 3 расходомера, 4 пирометра и контроллер измерения температуры.

Для управления вакуумной установкой требуется 55 дискретных выходных каналов ПЛК, 66 дискретных входных каналов ПЛК и 21 интерфейсных канала RS-485 – это более, чем способен обеспечить один контроллер. Поэтому было решено ис-

пользовать два контроллера серии XPAC фирмы «ICPDAS», соединенных между собой через интерфейс Ethernet. Первый, ведущий контроллер состоит из базового модуля XP-8741-EN, четырех модулей дискретного ввода I-8040PW-G (32 канала), четырех модулей дискретного вывода I-8041PW-G (32 канала) и одного модуля дискретного ввода-вывода I-8042W-G (16 входных и 16 выходных каналов). Для работы с интеллектуальным оборудованием используется второй контроллер, состоящий из базового модуля XP-8741-EN, модуля I-8114iW/D2 (четыре порта RS-232) и шести модулей I-8144IW (четыре порта RS-485).

Структурная схема системы управления изображена на рис. 3.

Для контроля давления в вакуумной системе были выбраны активные вакуумметры АТВ-2 и АИВ-51 фирмы «Инситек» (г. Томск). Обмен информации между ПЛК и вакуумметрами осуществляется через интерфейс RS-485.

Каждый магнетрон и ионный источник имеют собственный источник питания. В системе используются два типа источников питания магнетронов MPS-15/10-PC (7 шт.) (максимальный ток 15 А, максимальное напряжение 700 В) и MPS-9/5-PC (3 шт.) (максимальный ток 9 А, максимальное напряжение 600 В) фирмы «ИПС» (г. Томск). Кроме того два источника питания

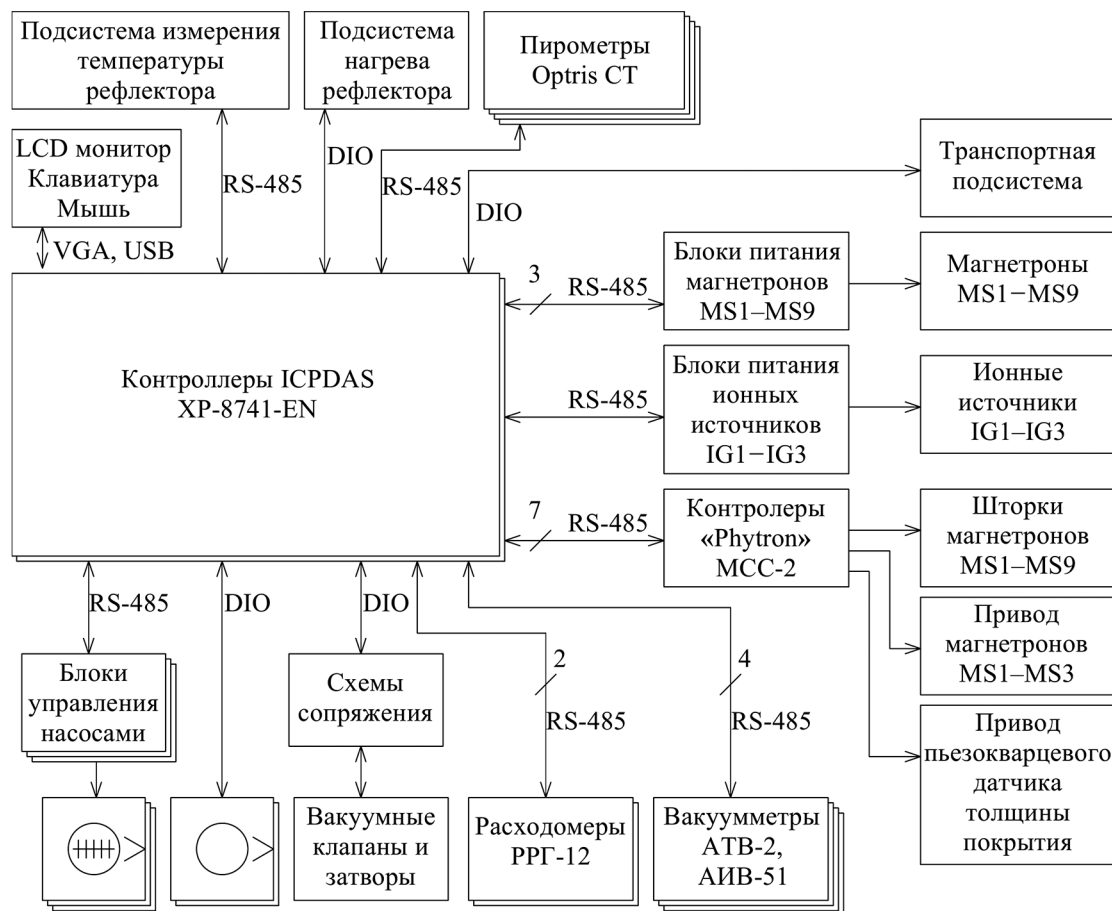


Рис. 3. Структурная схема системы управления установки «КО-МКР»

MPS-15/10-PC объединены и подключены к одному магнетрону, что позволяет увеличить ток магнетрона до 30 А. Источник питания ионного источника – PS-3/5-DC (3 шт.) (максимальный ток 2 А, максимальное напряжение 3 кВ). Возможны три режима работы блоков питания: стабилизация тока, напряжения и мощности. Управление блоками осуществляется через интерфейс RS-485. Для уменьшения времени опроса состояния блоков они разделены на четыре группы по три блока. Каждая группа подключена к индивидуальному порту RS-485 ПЛК.

Для перемещения магнетронов и шторок используются комплектные электроприводы фирмы «Phytron» (Германия), включающие в себя шаговые двигатели с редукторами VSS43.200.2,5 в вакуумном исполнении (14 шт.) и двухкоординатные блоки управления шаговыми двигателями MCC-2 (7 шт.). Несмотря на то, что блоки управления подключаются к контроллеру через интерфейс RS-485, объединить в сеть блоки не удалось. Каждый блок подключен к отдельному порту ПЛК.

Нагрев рефлектора в вакууме осуществляют керамические инфракрасные нагревательные элементы фирмы «Ceramic» (Ирландия). На крышке вакуумной камеры под переизлучателем равномерно расположены 120 нагревателей FFE мощностью 750 Вт и 10 нагревателей HFE – 250 Вт. Для нагрева используются 79 нагревательных элемента, остальные – резерв. Все нагреватели сгруппированы в 6 колец. ПЛК может управлять подачей напряжения на каждое кольцо через соответствующее ему твердотельное реле, осуществляя контролируемый нагрев рефлектора.

Контроль температуры рефлектора осуществляют две независимые подсистемы измерения – это 24 канальная основная подсистема, использующая платиновые термосопротивления Pt100 фирмы «Негаeus» (Германия), и дополнительная – четыре

пирометра «Optris CT LT» (Германия). Связь с подсистемами осуществляется по интерфейсу RS-485.

Конструктивно система управления состоит из нескольких крупных узлов:

- стойка высотой 1600 мм (контроллеры, 17 дюймовый LCD монитор, полка с клавиатурой и манипулятором типа «мышь»);
- стойка высотой 2200 мм (силовая коммутация, блоки питания иконных источников и блоки питания турбомолекулярных насосов);
- стойка с системой кондиционирования (10 источников питания магнетронов). В стойке выполнена теплоизоляция, проведены воздуховоды. Температурный режим создается и поддерживается холодильным агрегатом RITTAL SK 3387.140;
- на раме вакуумной камеры расположены шкафы: управления двигателями шторок магнетронов, пневмоклапанами;
- на подвижной тележке вместе с крышкой камеры расположены: шкаф управления движением, два шкафа подсистемы нагрева рефлектора. Для укладки шлангов системы охлаждения крышки, силовых кабелей и кабелей управления использованы гибкие кабельные каналы «CPS».

Выводы

Разработана система управления промышленной вакуумной установкой «КО МКР», которая позволяет:

- автоматизировать контроль и управление технологическим процессом нанесения модифицирующих плазменных покрытий;
- оперативно выявлять предаварийные ситуации, блокировать их развитие и предупреждать об этом оператора;
- автоматически формировать архивы накопленных данных для дальнейшего анализа и совершенствования технологического процесса и оборудования.

СПИСОК ЛИТЕРАТУРЫ

1. Асаинов О.Х., Баинов Д.Д., Кривококов В.П., Романенко С.Е., Чернятина А.А. Плазменная технология осаждения отражающего покрытия на поверхность углепластика // Известия вузов. Сер. Физика. – 2011. – Т. 54. – № 11/2. – С. 158–161.
2. Данилин Б.С., Сырчин В.К. Магнетронные распылительные системы. – М.: Радио и связь, 1982. – 72 с.
3. Жуков В.В., Кривококов В.П., Янин С.Н. Распыление мишени при ассистировании магнетронного разряда ионным пучком //

Известия Томского политехнического университета. – 2004. – Т. 307. – № 7. – С. 40–45.

4. Асаинов О.Х., Баинов Д.Д., Кривококов В.П., Юдаков С.В. Обработка поверхности материалов плазмой магнетронного разряда и ионными пучками // Известия вузов. Сер. Физика. – 2009. – Т. 52. – № 11/2. – С. 172–175.

Поступила 22.02.2012 г.

УДК 51-73

ИСПОЛЬЗОВАНИЕ ИМИТАЦИОННОГО КОМПЛЕКСА ДЛЯ ОТЛАДКИ СИСТЕМЫ УПРАВЛЕНИЯ ПРОЦЕССОМ ПЕРЕРАБОТКИ НАТРИЯ

П.Ю. Кривцов, В.М. Павлов*

Национальный ядерный центр Республики Казахстан, г. Курчатов

*Томский политехнический университет

E-mail: Krivtsov@nnc.kz

Показана возможность использования методов полунатурного имитационного моделирования для отладки системы управления пожаро- и взрывоопасным технологическим процессом переработки натриевого теплоносителя. Дана структура имитационного комплекса, математической модели технологического процесса, программного тренажера, определены функциональные характеристики системы управления, подлежащие проверке на имитационном комплексе. Представлены результаты испытаний системы управления.

Ключевые слова:

Натриевый теплоноситель, установка по переработке натрия, имитационный комплекс, модельные исследования, система управления.

Key words:

Sodium coolant, sodium process facility, simulation complex, modeling researches, control system.

Введение

Развитие ядерной энергетики связано с решением комплекса задач, направленных на повышение эффективности и безопасности энергетических объектов, методов утилизации, переработки и захоронения ядерных отходов. Начатая в Республике Казахстан программа по выводу реактора БН-350 из эксплуатации затрагивает широкий комплекс проблем, одной из которых является переработка и захоронение натрия, который использовался в качестве теплоносителя [1, 2].

Важным этапом работ по утилизации теплоносителя является изучение технологии переработки, которая включает в себя решение задач, связанных с синтезом оптимальных алгоритмов управления технологическим процессом, проектированием информационного обеспечения, анализом аварийных ситуаций и обучением оперативного персонала методам безаварийного управления установкой. Решение поставленного комплекса задач на реальной установке не представляется возможным ввиду того, что процесс переработки относится к взрыво- и пожароопасной категории работ. Одним из современных методов изучения и исследования поведения динамических систем или физических процессов является имитационное (компьютерное) моделирование, основу которого составляет математическое описание моделируемого объекта или динамической системы.

Важным направлением в области имитационного моделирования является полунатурное моделирование (Hardware In the Loop Simulation /HILS/). Суть метода заключается в том, что часть агрегатов или устройств реального объекта включается в состав имитационного комплекса, а моделируется только технологическая составляющая, отражающая физические особенности работы моделируемой системы [3, 4]. Так, представленный в работе полунатурный имитационный комплекс

установки по переработке натрия (УПН) содержит контроллерное оборудование автоматизированной системы управления технологическим процессом (АСУ ТП) и рабочие станции операторов установки. Математическая модель позволяет имитировать показания средств измерения технологических параметров и индикаторов состояния исполнительных элементов установки.

Структура имитационного комплекса

Имитационный комплекс, рис. 1, кроме оборудования системы управления включает в себя модельный комплекс, состоящий из станции моделирования и контроллера имитации сигналов нижнего уровня.

Программно-технический комплекс АСУ ТП содержит контроллер сбора данных и управления, выполненный на платформе ControlLogix фирмы Allen Bradley, а также рабочие станции управления технологическим процессом, которые предоставляют информацию о технологическом процессе при помощи графического интерфейса, выполненного в SCADA системе RSView32.

Модельный комплекс представляет собой набор программно-технических средств фирмы National Instruments. Аппаратная часть построена на архитектуре PXI и включает в свой состав шасси, контроллер реального времени и набор плат аналогового и дискретного ввода-вывода. Программная часть разработана с использованием пакета LabVIEW, позволяющего создавать распределенные системы моделирования, работающие на ядре реального времени, либо под управлением ОС Windows. Для разработки программного обеспечения моделирования, а также для настройки и конфигурации модельного комплекса, используется станция моделирования. Подключение станции моделирования к контроллеру имитации осуществляется посредством сети Ethernet.

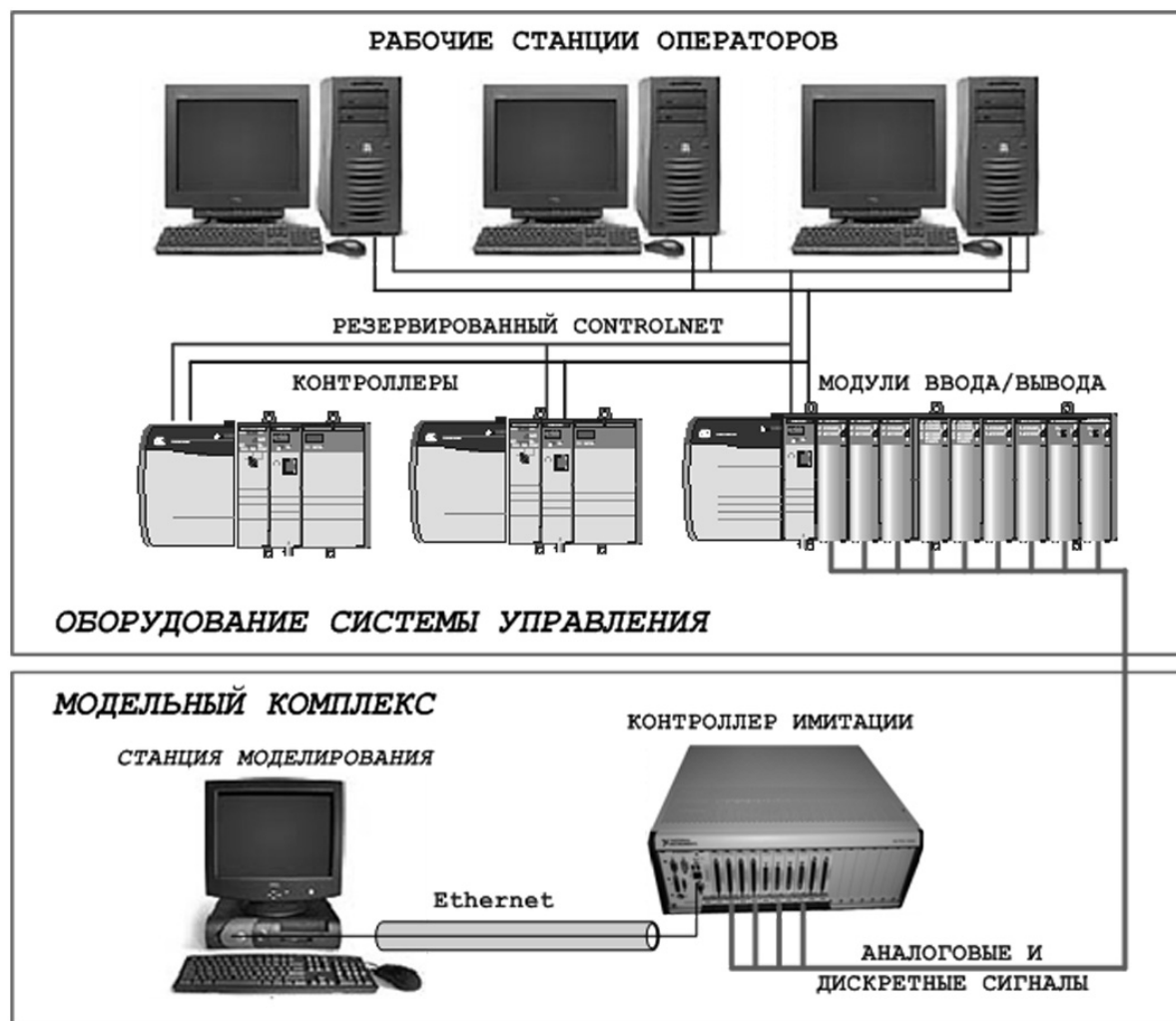


Рис. 1. Структурная схема имитационного комплекса

Модельный комплекс обеспечивает имитацию показаний 200 датчиков и обработку 90 сигналов управления.

Реализация математической модели процесса

Основное назначение модельного комплекса – имитация изменения технологических параметров основных систем УПН при формировании управляющих воздействий со стороны системы управления.

Для удобства разработки, тестирования и отладки, программное обеспечение, реализующее модель технологического процесса, иерархически структурировано. Все структурные элементы программы реализуют либо модели технологического оборудования УПН, либо – физического процесса. Структура программы приведена на рис. 2.

Подсистема имитации строится на основе математического описания, отражающего основные характеристики объекта и процесса. Для УПН моделируются:

- химическая реакция натрия с водой в растворе гидроксида в химическом реакторе;

- гидродинамические процессы течения жидкостей, изменение уровня в емкостях;
- термодинамические процессы нагрева оборудования и рабочих сред (жидких и газовых);
- термодинамические процессы в теплообменниках;
- изменение объема газов и жидкостей при изменении температуры и плотности;
- потоки газов с учетом конфигурации оборудования и физических параметров.

При реализации математического описания УПН учитывалась конфигурация и состав оборудования, свойства материалов и рабочих сред, природа протекающих процессов и их взаимосвязь. Математическое описание процессов взаимодействия натрия с водой в химическом реакторе установки представлены в работе [5]. Рассмотрим основные уравнения, выражающие характер изменения технологических параметров во времени.

В случае, когда необходимо моделировать изменение уровня жидкости в емкости, используется выражение:

$$A_i \frac{dL_i(t)}{dt} = \sum_{i=1} F_{Vin_i}(t) - \sum_{i=1} F_{Vout_i}(t),$$

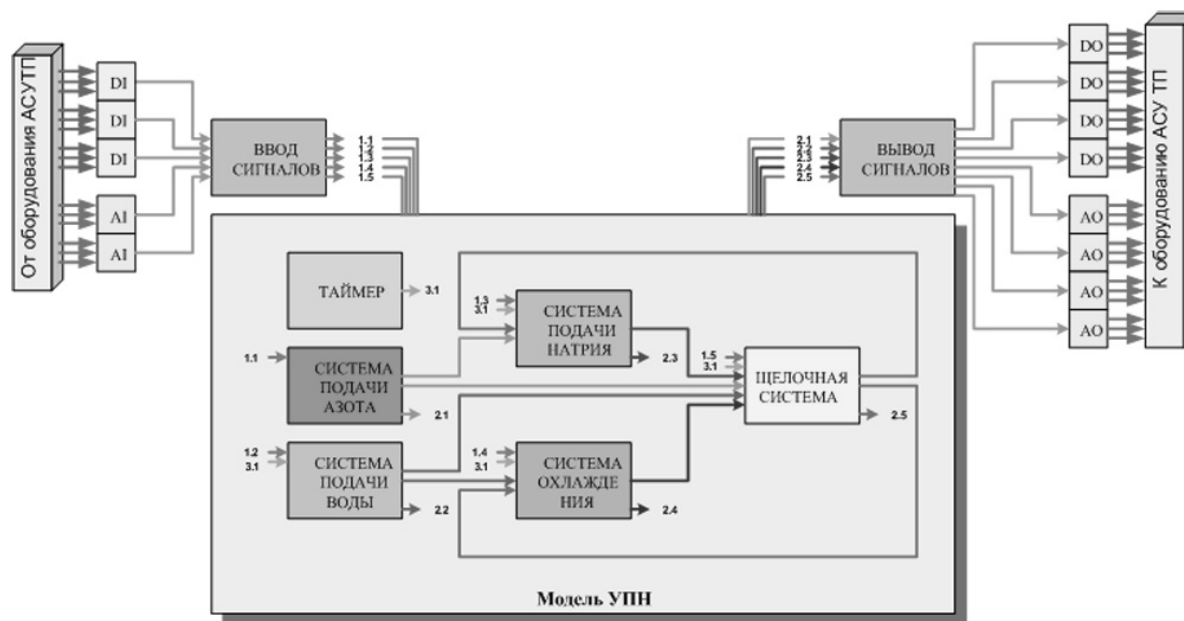


Рис. 2. Структурная схема программного обеспечения модели УПН: 1.1–1.5 сигналы управления; 2.1–2.5 моделируемые сигналы; 3.1 – время таймера; DI – модули дискретного ввода; DO – модули дискретного вывода; AI – модули аналогового ввода; AO – модули аналогового вывода

где A_i – площадь дна моделируемой емкости, m^2 ; $L_i(t)$ – уровень жидкости (натрий, вода или гидроксид натрия), m/c ; $F_{vin_i}(t)$, $F_{vout_i}(t)$ – объемный входной/выходной расход жидкости в емкость, m^3/c .

Выходной расход жидкости моделируется при помощи следующего выражения:

$$F_{vout_i}(t) = a_p \sqrt{2gL_i(t) + \frac{2P_i(t)}{\rho_i}},$$

где a_p – площадь поперечного сечения трубы, m^2 ; g – ускорение свободного падения, m/c^2 ; ρ_i – плотность жидкости, kg/m^3 ; $P_i(t)$ – перепад давления на линии подачи жидкости, Pa/c ; t – время, c .

При моделировании выхода газа из емкости используется выражение

$$F_{gm}(t) = \sqrt{\frac{P_{gm}^2(t) - P_{gm_out}^2(t)}{kR_{gm}T_{gm}(t)}},$$

где $F_{gm}(t)$ – выходной расход газовой смеси, kg/c ; $P_{gm}(t)$ – давление газовой смеси в моделируемой емкости, Pa/c ; $P_{gm_out}(t)$ – давление на последующем элементе, Pa/c ; k – общий коэффициент гидравлического сопротивления, $1/m^4$; R_{gm} – газовая постоянная смеси, $Dж/(кг \cdot K)$; $T_{gm}(t)$ – температура газовой смеси, K/c .

Уравнения материального баланса газовой смеси представляет собой выражение

$$\frac{dP_{gm}(t)}{dt} = \frac{R_{gm}T_{gm}(t)}{V_{gm}} \left(\sum_{i=1} F_{gin_i} - F_{gm}(t) \right) + \frac{R_{gm}M_{gm}(t)}{V_{gm}(t)} \frac{dT_{gm}(t)}{dt} - R_{gm}T_{gm}(t)M_{gm}(t) \frac{dV_{gm}(t)}{dt},$$

где $V_{gm}(t)$ – объем газовой смеси, m^3/c ; $F_{gin_i}(t)$ – массовый входной расход газа (водород, азот или вода-

ной пар), kg/c ; $M_{gm}(t)$ – масса газовой смеси в моделируемой емкости, kg .

Тепловой баланс газовой смеси описывается как

$$C_{gm}M_{gm}(t) \frac{dT_{gm}(t)}{dt} = Q_{gm_in}(t) + V_{gm} \frac{dP_{gm}(t)}{dt},$$

где C_{gm} – теплоемкость газовой смеси, $Dж/(кг \cdot K)$; $Q_{gm_in}(t)$ – количество энергии, вносимой газами в моделируемую емкость, $Dж/c$.

Для всех подсистем УПН моделируется состояние запорной и регулирующей арматуры, а также показания датчиков температуры, давления, расхода. Модели исполнительных элементов представляют собой нормировочные логические выражения, которые в зависимости от сигнала управления возвращают текущее состояние устройства. Так для модели запорного клапана используется выражение

$$DV_i = \begin{cases} U_{\max} & \rightarrow 1 \\ U_{\min} & \rightarrow 0 \end{cases},$$

где U_{\min} и U_{\max} – минимальное и максимальное напряжения на клапане; DV_i – коэффициент, характеризующий состояния запорного клапана.

Состояние регулирующего клапана определяется как:

$$CV_i = \lim \{ I_{\min}; I_{\max} \} (I_c - I_{\min}) / (I_{\max} - I_{\min}),$$

где I_{\min} и I_{\max} – минимальный и максимальный токовые сигналы; CV_i – коэффициент, характеризующий состояния регулируемого клапана; I_c – токовый управляющий сигнал.

Датчики температуры, давления и расходомеры описываются с помощью выражения вида:

$$I_i = \lim \{ h_{\min}; h_{\max} \} \frac{x}{h_{\max}} (I_{\max} - I_{\min}) + I_{\min},$$

где h_{\min} и h_{\max} – нижний/верхний пределы измерения; x – измеряемый параметр; I_i – выходной токовый сигнал.

В модельном комплексе нет жесткой привязки к временной оси, что позволяет проводить испытания системы управления в ускоренном масштабе времени. При этом секунда модельного времени соответствует минуте реального, что позволяет значительно быстрее привести виртуальную установку в состояние готовности. На реальном оборудовании процесс разогрева элементов конструкции УПН и заполнение основных емкостей натрием, щелочью и водой продолжается порядка 12 ч.

Испытания системы управления

Целью проведения исследований является проверка соответствия характеристик разработанной конфигурации программно – технических средств системы управления как функциональным и техническим требованиям, так и требованиям эргономики и технической эстетики. Помимо представленных выше требований выполняется проверка исправности технических средств и правильность подключения каналов аналогового и дискретного ввода/вывода. Исследования программно – технических средств АСУ ТП на полунатурном имитационном комплексе позволили оценить качество

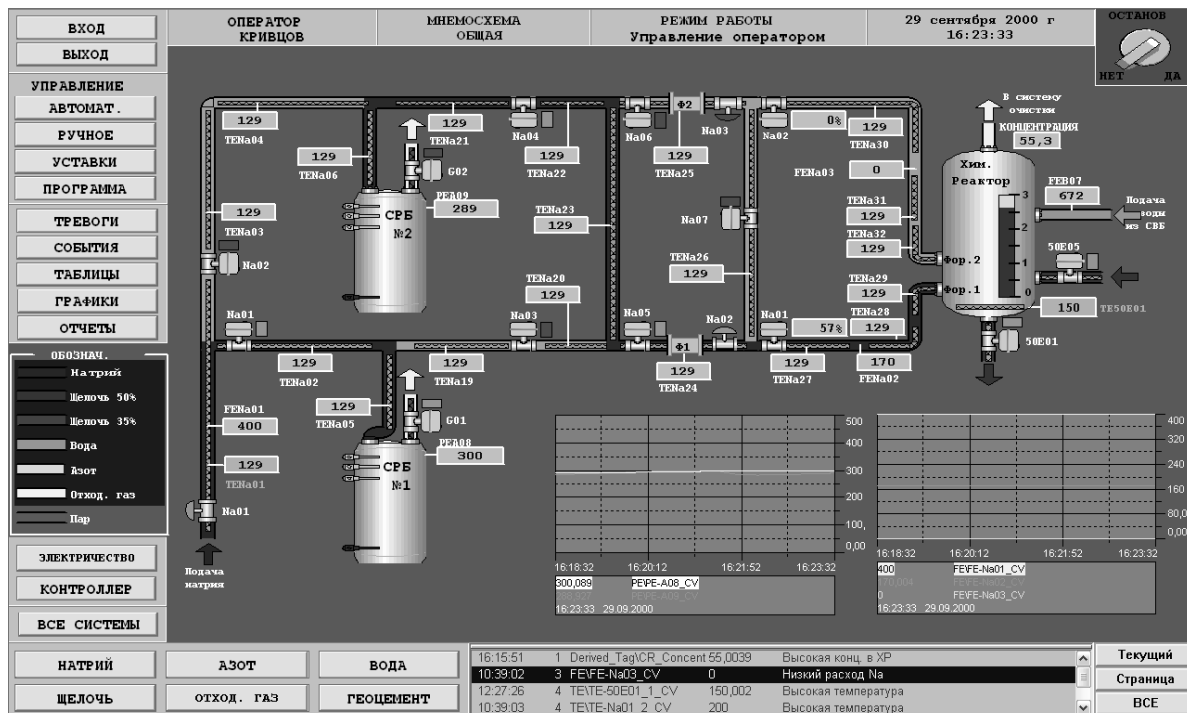


Рис. 3. Процесс переработки натрия на имитационном комплексе

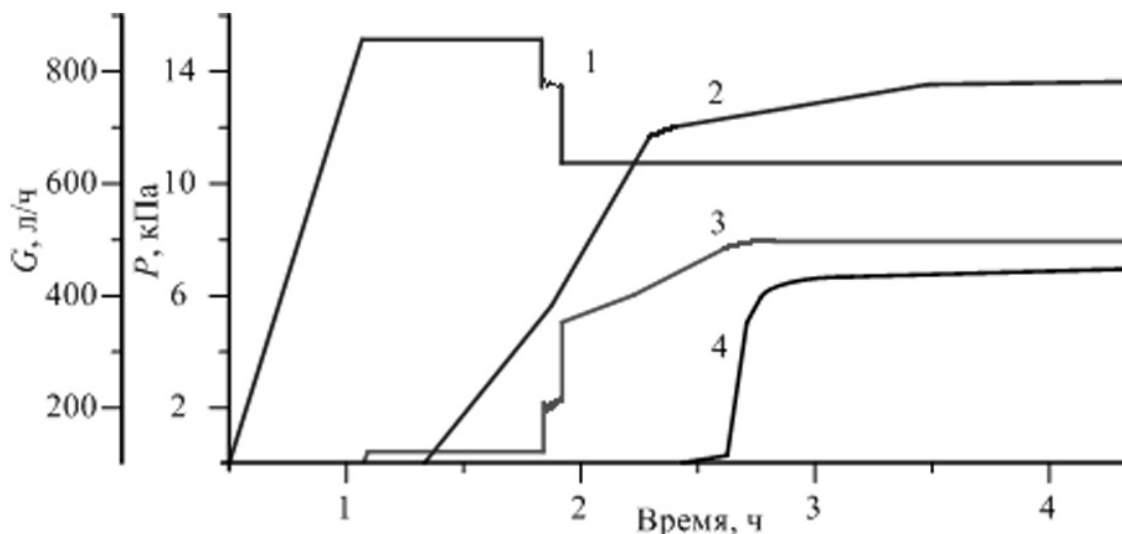


Рис. 4. Контур управления уровнем в химическом реакторе 1) давление в нижней части реактора; 2) расход воды; 3) давление в верхней части реактора; 4) расход щелочи

управления технологическим процессом и определить степень законченности проекта в целом.

В ходе проведения исследований системы управления, оценке подлежат следующие качественные (функциональные) характеристики системы:

- измерение, регистрация и отображение измеряемых технологических параметров, таких как температура, давление, расход, уровень, концентрация;
- оперативный контроль, обеспечение цветовой индикации и звуковой сигнализации отклонений технологических параметров относительно заданных значений (уставок);
- контроль положения и состояния запорной арматуры, мощности выделяемой на электрических нагревателях;
- управление исполнительными устройствами (задвижка, клапан, насос, регулятор мощности нагревателя) в соответствии с заданным алгоритмом управления;

За время существования полунатурного имитационного комплекса была проведена серия исследований, в результате которых был переработан объем «виртуального» натрия, в 2 раза превышающий его реальные запасы на БН-350, рис. 3. Прове-

дена оценка функциональных характеристик системы управления, которая полностью подтвердила правильность технических решений принятых на этапах проектирования и разработки АСУ ТП. На рис. 4 показаны результаты работы контура управления уровнем в химическом реакторе.

Программный тренажер

Имитационные комплексы, которые в максимально возможной степени приближены к реальным установкам, играют большую роль в подготовке операторов сложных систем и позволяют тренирующимся приобрести правильные и устойчивые навыки управления процессом [6].

После проведения комплексных испытаний, оборудование АСУ ТП было передано в монтаж на УПН, а полунатурный имитационный комплекс был преобразован в программный тренажер, который используется для обучения операторов установки, рис 5.

Программное обеспечение системы управления исключает вмешательство оператора в технологический процесс в режиме прямого ручного управления исполнительными механизмами, но управление процессом не обходится без участия чело-

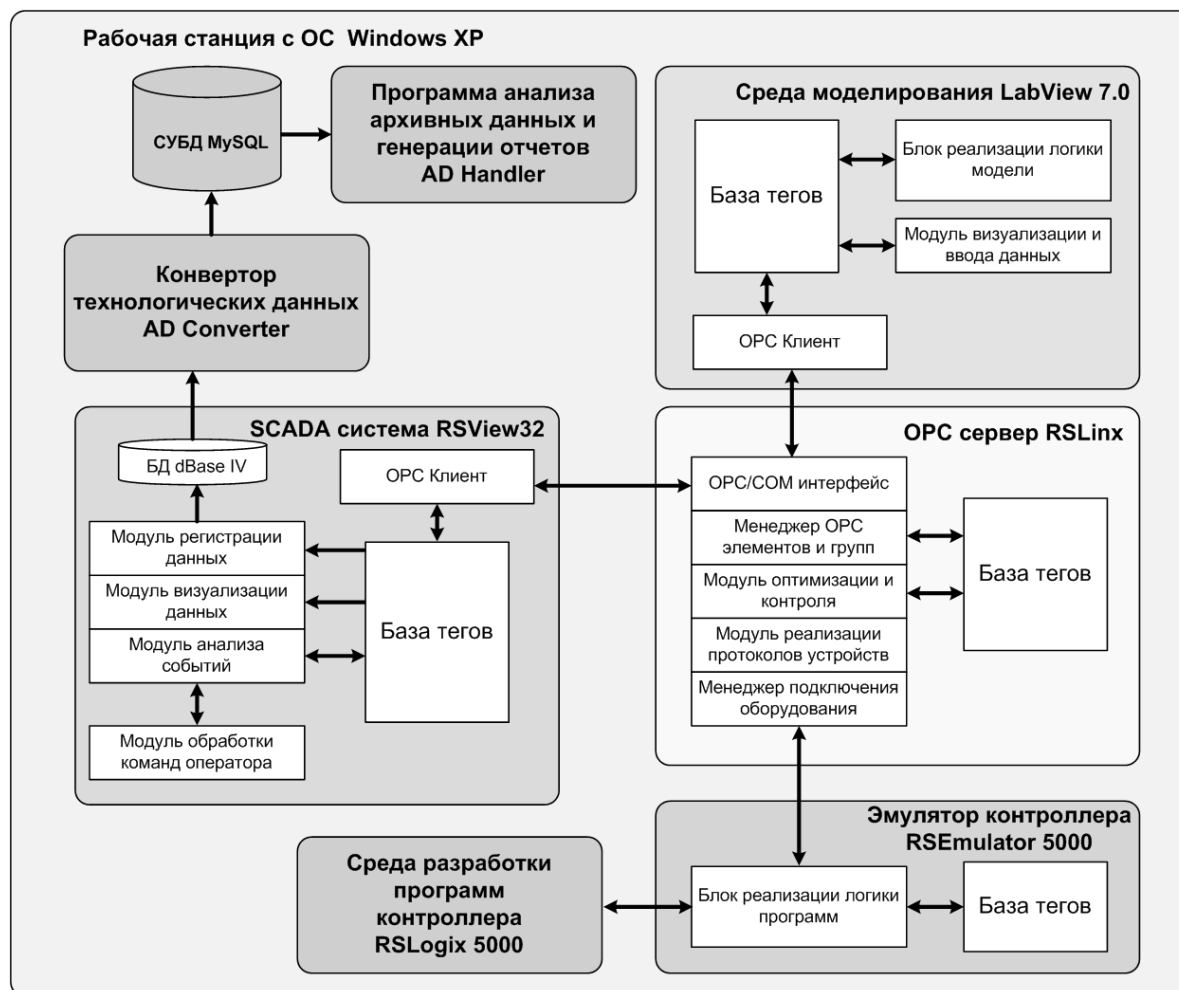


Рис. 5. Структура программного тренажера

века. Операторы установки участвуют в настройке технологического процесса и управляют им в тех ситуациях, когда требуется разрешение нештатных и неплановых ситуаций, в том числе запроектных аварий. Программный тренажер обеспечивает:

- выработку навыков безопасного и эффективно-го управления в пусковых, переходных и стационарных режимах;
- обучение операторов своевременному обнаружению и устранению отклонений в ходе технологического процесса;
- проведение компьютерных экспериментов для исследования различных условий работы оборудования, в том числе критических режимов и аварийных ситуаций;
- ведение архива действий операторов;
- объективный контроль уровня знаний операторов.

Заключение

Применение современных методов проектирования систем управления технологическими про-

цессами, связанных с использованием компьютерного имитационного моделирования, позволило обеспечить оптимальное решение задач отработки, отладки технологии переработки, разработкой алгоритмов управления, сбора, обработки и представления технологической информации, отладки и апробации операционных схем функционирования технологического процесса. Математическое описание процессов, происходящих на установке по переработке натрия, дало возможность провести изучение поведения элементов системы в различных режимах работы установки и выполнить анализ реакции объекта управления на управляющие воздействия. Экономический эффект комплексного подхода к созданию системы проявляется в виде решения поставленных задач без натурных испытаний и исследований, имеющих высокую стоимость и опасных в реализации. Разработанная система управления прошла комплексные испытания в г. Актау, Казахстан; во время проведения пусконаладочных работ потребовалась лишь «тонкая» настройка отдельных контуров управления.

СПИСОК ЛИТЕРАТУРЫ

1. Tazhibayeva I., Herrick A. Sodium Coolant Handling Project of BN-350 Fast Breeder Reactor. – Almaty: Glory Ltd., 2010. – 304 p.
2. Garcia H.E. Modeling and Control of a Sodium Conversion Process Applied to Nuclear Decommissioning Activities // Nuclear Technology. – 1998. – V. 123. – № 8. – P. 54–60.
3. Максимей И.В. Имитационное моделирование на ЭВМ. – М.: Радио и связь, 1988. – 232 с.
4. Дорф Р. Современные системы управления. – М.: Лаборатория базовых знаний, 2002. – 832 с.
5. Кривцов П.Ю., Павлов В.М. Разработка модели химического реактора установки по переработки натриевого теплоносителя // Известия Томского политехнического университета. – 2011. – Т. 318. – № 3. – С. 41–46.
6. Мамросенко К.А. Имитационно-тренажерные и обучающие распределенные системы // Программные продукты и системы. – 2008. – № 3. – С. 35–40.

Поступила 20.01.2012 г.

УДК 004.75

ПЕРЕДАЧА ДАННЫХ В БЕСПРОВОДНЫХ СЕНСОРНЫХ СЕТЯХ С ПРИОРИТЕТАМИ НА ОСНОВЕ АГРЕГИРОВАНИЯ ПРЕДПОЧТЕНИЙ

С.В. Муравьев, Е.В. Тараканов

Томский политехнический университет

E-mail: tark@tpu.ru

Предложен способ управления передачей данных в беспроводной сенсорной сети, основанный на использовании отношений консенсуса для назначения приоритетов передаваемым данным. Описана статистическая аналитическая модель для анализа поведения предложенного протокола передачи в условиях ограниченной пропускной способности сети, позволяющая обеспечить на необходимом уровне качество передачи измерительных данных.

Ключевые слова:

Беспроводная сенсорная сеть, перегрузка сети, агрегирование предпочтений, отношение консенсуса.

Key words:

Wireless sensor network, network congestion, preference aggregation, consensus relation.

Введение

Беспроводные сенсорные сети (*wireless sensor networks*, WSN) все чаще используются для организации различных видов мониторинга: параметров окружающей среды [1], состояния конструкций, зданий и сооружений [2], в системах безопасности (пожарной, сейсмической, экологической и др.) [3], для отслеживания целей в процессе ведения боевых действий [4] и т. п. В таких системах разнородные данные собираются мультисенсорами, входящими в состав узлов, расположенных в подлежащих мониторингу точках определенной географической области, и передаются по беспроводной сети в центральный узел (*sink*) для обработки и принятия решений. Обычно сеть имеет иерархическую (древовидную) структуру, в которой на каждом уровне данные могут передаваться от узлов-источников к одному или нескольким узлам-приемникам.

При передаче данных в беспроводных сенсорных сетях возникают существенные проблемы, связанные с ограниченной полосой пропускания (*bandwidth*) используемых в качестве линий связи радиоканалов. В частности, в ситуациях, когда много узлов-источников одновременно инициируют передачу данных, может возникнуть перегрузка или даже коллапс сети, в результате чего ее пропускная способность (*throughput*), выражаемая в количестве проходящих от источника к центральному узлу пакетов данных в единицу времени, падает практически до нуля. Кроме того, отдельные сенсорные узлы могут как добавляться в сеть, так и, по разным причинам, выходить из состава сети. Изменение в конфигурации сети, как правило, приводит к необходимости изменять маршруты пакетов данных, что также снижает пропускную способность сети [5].

Одним из возможных подходов к решению проблемы является назначение приоритетов передаваемым по сети пакетам и организация первоочередной доставки пакетов с более высоким приорите-

том. Этот подход согласуется с концепцией качества обслуживания (*Quality of Service*, QoS), являющейся общепринятой для сенсорных сетей, основанных на обнаружении событий. Основными показателями QoS являются малая задержка передачи данных от источника к центральному узлу и низкие потери данных о событиях [6].

Для обеспечения бесперебойной работы сенсорной сети требуются специальные методы передачи данных, позволяющие решать проблему перегрузки сети и гарантировать доставку пакетов данных до центрального узла. Для этой цели в [7] предложен способ организации передачи пакетов в WSN путем назначения интервала ожидания на основе динамического вычисления соотношения текущей и максимальной пропускной способности сети. В этом способе каждый узел сообщает требуемый ему интервал ожидания центральному узлу, после чего центральный узел назначает интервалы ожидания узлу пропорционально затребованному. Назначенные интервалы ожидания передают узлам WSN с помощью пакетов обратной связи. При переполнении очереди передачи вновь поступающие пакеты отбрасываются. Недостатком этого способа является необходимость принятия каждым узлом решения о требуемом ему интервале ожидания.

В данной статье предлагается способ назначения приоритетов узлам беспроводной сенсорной сети с использованием некоторого бинарного отношения консенсуса, получаемого в результате процедуры агрегирования предпочтений. Предпочтения в форме m ранжирований n узлов формируются на основе показаний мультисенсоров. Впервые предлагаемый подход описан в [8]. Он позволяет динамически назначать приоритеты узлам сети (и, следовательно, передаваемым ими пакетам), формировать очередь передачи пакетов и распределять пропускную способность сети в зависимости от содержания (семантики) передаваемых данных.

Агрегирование предпочтений и отношение консенсуса

Пусть $A=\{a_1, a_2, \dots, a_n\}$ – множество n узлов сенсорной сети, расположенных в заданных точках распределенного сложного контролируемого объекта. Будем рассматривать в качестве примера систему пожарной сигнализации некоторого здания. Пусть в каждом помещении этого здания расположен узел, содержащий сенсоры, контролирующие m разнородных свойств: температуру воздуха, скорость изменения температуры, задымленность, уровень СО, срабатывание датчика пламени и присутствие персонала (таблица). По данным измерений этих свойств можно получить m ранжирований n узлов. Каждое из m ранжирований $\lambda=\{a_1 \succ a_2 \succ \dots \sim a_s \sim a_t \succ \dots \sim a_n\}$ может включать как строгое отношение предпочтения \succ , так и отношение безразличия \sim . Множество полученных таким образом бинарных отношений $\Lambda=\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ будем называть *профилем предпочтения* для данных m свойств и n узлов.

Например, для представленных в таблице результатов измерений шести свойств, полученных с семи узлов, профиль предпочтения имеет следующий вид:

$$\begin{aligned} \lambda_1 : a_5 \succ a_2 \succ a_3 \succ a_6 \sim a_7 \succ a_4 \succ a_1; \\ \lambda_2 : a_5 \succ a_3 \succ a_2 \succ a_4 \succ a_6 \succ a_1 \succ a_7; \\ \lambda_3 : a_3 \succ a_6 \succ a_5 \succ a_2 \succ a_7 \succ a_4 \succ a_1; \\ \lambda_4 : a_5 \succ a_2 \succ a_3 \succ a_6 \succ a_4 \succ a_1 \succ a_7; \\ \lambda_5 : a_2 \sim a_3 \sim a_4 \sim a_6 \succ a_1 \sim a_5 \sim a_7; \\ \lambda_6 : a_3 \sim a_6 \succ a_1 \sim a_2 \sim a_4 \sim a_5 \sim a_7. \end{aligned} \quad (1)$$

Задача *агрегирования предпочтений* заданного профиля Λ , состоящего из m ранжирований n альтернатив, заключается в нахождении единственного строгого ранжирования – *отношения консенсуса*.

Следуя [9], ранжирование λ представим $(n \times n)$ матрицей отношения $R=[r_{ij}]$, строки и столбцы которой соответствуют узлам a_i и $r_{ij}=1$ если $a_i \succ a_j$; $r_{ij}=0$ если $a_i \sim a_j$; $r_{ij}=-1$ если $a_i \prec a_j$. Тогда *расстояние Кемени* $d(\lambda_k, \lambda_l)$ между двумя ранжированиями λ_k и λ_l определяется формулой

$$d(\lambda_k, \lambda_l) = \sum_{i < j} |r_{ij}^k - r_{ij}^l|. \quad (2)$$

и может рассматриваться как число различий между двумя ранжированиями. В формуле (2) суммируются только элементы верхней треугольной подматрицы r_{ij} , $i < j$, матрицы R .

Расстояние между произвольным ранжированием λ и профилем Λ можно теперь определить следующим образом:

$$D(\lambda, \Lambda) = \sum_{k=1}^m d(\lambda, \lambda_k) = \sum_{i < j} \sum_{k=1}^m |r_{ij}^k - r_{ij}| = \sum_{i < j} \sum_{k=1}^m d_{ij}^k.$$

Из (2), с учетом $r_{ij}=1$ для всех $i < j$, что соответствует естественному линейному порядку $a_1 \succ a_2 \succ \dots \succ a_n$, следует, что для любого $k=1, \dots, m$ имеем: $d_{ij}^k=|1-1|=0$ если $a_i^k \succ a_j^k$; $d_{ij}^k=|0-1|=1$ если $a_i^k \sim a_j^k$ и $d_{ij}^k=|-1-1|=2$ если $a_i^k \prec a_j^k$.

Таблица. Данные с узлов системы пожарной сигнализации (модельный пример)

Узел	Результаты измерений					
	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
	Температура, °C	Скорость изменения температуры, °C/c	Задымленность, ДБ/м	Уровень СО, мг/м³	Срабатывание датчика пламени	Наличие людей
a_1	42	8	0,03	0,31	-	-
a_2	270	15	0,08	0,86	+	-
a_3	80	35	0,13	0,83	+	+
a_4	50	12	0,03	0,58	+	-
a_5	300	38	0,10	0,90	-	-
a_6	62	10	0,11	0,70	+	+
a_7	62	7	0,04	0,18	-	-

Тогда можно определить $(n \times n)$ матрицу профиля $P=[p_{ij}]$, где

$$p_{ij} = \sum_{k=1}^m d_{ij}^k, \quad i, j = 1, \dots, n$$

и отношение консенсуса β определяется как такое ранжирование узлов из A , что расстояние $D(\beta, \Lambda)$ от β до профиля Λ минимально, то есть

$$\beta = \arg \min_{\lambda \in \Pi} D(\lambda, \Lambda) = \arg \min_{\lambda \in \Pi} \sum_{i < j} p_{ij}, \quad (3)$$

где Π – множество всех $n!$ строгих предпочтений \succ на A . Каждая перестановка объектов из A соответствует перестановке соответствующих строк и столбцов матрицы. Поэтому задача (3) заключается в нахождении такой перестановки строк и столбцов матрицы профиля, что сумма элементов ее верхней треугольной подматрицы минимальна.

Заметим, что агрегирование предпочтений в других терминах представляет собой проблему *голосования* или *группового решения*, где A – это множество n альтернатив или кандидатов, которые ранжируются группой из m индивидуумов (избирателей, экспертов и т. д.). В рамках этой проблематики, Ж.-А. Кондорсе еще в 1785 г. [10], предложил естественное правило нахождения отношения консенсуса для профиля Λ , описываемого матрицей P : в каждом сравнении альтернатив предпочтительной является альтернатива, предпочитаемая большинством избирателей, т. е. $a_i \succ a_j$ если и только если $p_{ij} > p_{ji}$. Однако, ранжирование Кондорсе не обязательно транзитивно, т. е. может быть, что $a_i \succ a_j$ и $a_j \succ a_k$, тогда как $a_k \succ a_i$. Единственным корректным способом преодоления указанной трудности является нахождение отношения консенсуса по правилу Кемени (3) [11].

Для профиля (1) ранжирование консенсуса имеет следующий вид:

$$\beta_1 = \{a_3 \succ a_2 \succ a_5 \succ a_4 \succ a_1 \succ a_7\}.$$

Оно позволяет выстроить n узлов в порядке важности измеряемых ими проявлений свойств. Этот порядок может служить основанием для назначения приоритетов передаваемых узлами пакетов, так как ранжирования исходного профиля

формируются исходя из актуальных данных измерений, характеризующих обстановку в окружении каждого конкретного сенсорного узла.

Далее будет рассмотрен механизм поддержания производительности WSN с использованием ранжирования консенсуса.

Механизм обеспечения QoS сенсорной сети

WSN, как правило, имеет древовидную топологию (рис. 1), в которой только определенные узлы (будем называть их *узлами-предками*) могут взаимодействовать с другими узлами (*потомками*). *Кластером* будем называть узел-предок со всеми его ближайшими потомками.

Пусть каждый узел использует для передачи сообщений буфер в виде очереди. Если сеть перегружена или ее пропускная способность не позволяет отправить поступившее сообщение немедленно, то пакет отправляется в очередь сообщений. По достижении максимального размера очереди поступление нового пакета приводит к необходимости либо удалить уже существующий пакет из очереди, либо отбросить поступивший пакет. Назначение приоритетов узлам сети внутри кластера в соответствии с ранжированием консенсуса позволяет создавать очередь передачи пакетов с сортировкой по убыванию важности узла-источника пакета. Так, приведенное в конце предыдущего раздела ранжирование обеспечивает наивысший приоритет пакету, отправляемому узлом 3, а наименьший – пакету узла 7.

Приоритет узла-источника входящего пакета сравнивается с приоритетами узлов-источников уже находящихся в очереди пакетов, начиная с первого пакета. После обнаружения пакета, узел-источник которого имеет меньший приоритет, чем приоритет узла-источника входящего пакета, будет произведена вставка поступившего пакета перед пакетом с менее приоритетного узла. Если же окажется что узел-источник входящего пакет имеет приоритет меньший, чем приоритет узлов-источников всех пакетов в очереди, то поступивший пакет будет отброшен.

Предположим, некоторый узел-предок собирает информацию с n узлов-потомков. При этом пропускная способность сети позволяет передать не все n , а только $(n-k)$ пакетов за один такт, то есть узел-предок должен на каждом такте отбросить k пакетов. Ясно, что отбрасывание пакетов будет приводить к потере определенной части измерительных данных.

Ниже будет введена статистическая аналитическая модель, которая позволит выбирать и оценивать значения n и k с учетом изменения уровня согласованности приоритетов при отбрасывании пакетов.

Определение относительной несогласованности приоритетов при удалении пакетов

Для решения проблемы воспользуемся основным на статистических соображениях коэффициентом ранговой корреляции Кендалла τ [12], который принимает значения в диапазоне от -1 (абсолютно несовместимые ранжирования) до $+1$ (полное совпадение ранжирований). Совместимость двух ранжирований β_1 и β_2 определяется по формуле:

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} r_{ij}^1 r_{ij}^2, \tag{4}$$

где r_{ij}^1 и r_{ij}^2 – элементы матрицы отношений для отношения β_1 и β_2 соответственно.

Как коэффициент ранговой корреляции Кендалла (4), так и расстояние Кемени (2) являются мерами совместимости ранжирований. Как показано в [13], между ними существует связь:

$$\tau(\beta_1, \beta_2) = 1 - \frac{d(\beta_1, \beta_2)}{n(n-1)},$$

которая позволяет установить связь между параметрами n и k .

Действительно, рассмотрим два ранжирования β_1 и β_2 для n узлов некоторого кластера. Естественным требованием для них является то, что они должны быть согласованы между собой. Соответствующий уровень согласованности обозначим че-

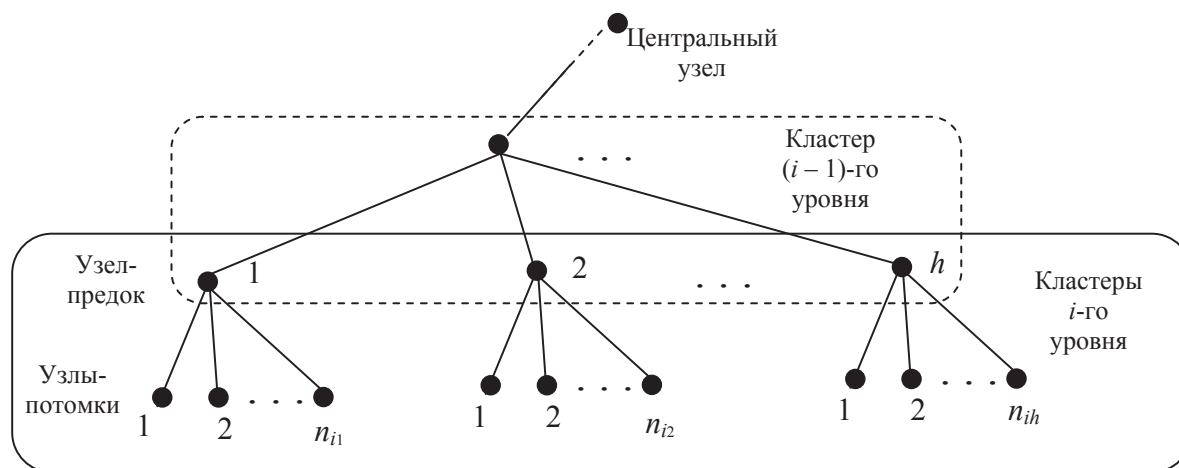


Рис. 1. Топология древовидной сети

рез τ_1 . Для $n-k$ узлов, где k – число отброшенных пакетов, получатся два других отношения консенсуса β_1' и β_2' , с уровнем согласованности τ_2 . Стремясь сохранять согласованность при любом k , примем, что $\tau_2 = \tau_1$.

Обозначив $d_1 = d(\beta_1, \beta_2)$ для случая n узлов и $d_2 = d(\beta_1', \beta_2')$ для случая $n-k$ узлов, получаем выражения для соответствующих коэффициентов ранговой корреляции:

$$\tau_1 = 1 - \frac{d_1}{n(n-1)} \quad \text{и} \quad \tau_2 = 1 - \frac{d_2}{(n-k)(n-k-1)}.$$

Тогда $\frac{d_1}{n(n-k)} = \frac{d_2}{(n-k)(n-k-1)}$, откуда

$$\frac{d_2}{d_1} = \frac{(n-k)(n-k-1)}{n(n-k)}. \quad (5)$$

Введем параметр $\theta(n, k)$ как *относительную несогласованность приоритетов* при удалении k из n пакетов:

$$\theta(n, k) = \frac{d_1 - d_2}{d_1} = 1 - \frac{d_2}{d_1}.$$

Тогда, с учетом (5), получаем

$$\theta(n, k) = 1 - \frac{(n-k)(n-k-1)}{n(n-1)}. \quad (6)$$

Ясно, что $0 \leq \theta(n, k) \leq 1$. Задаваясь значением параметра $\theta(n, k)$, можно выбирать обоснованное количество k отбрасываемых пакетов для известных значений n кластера (рис. 2). Например, при $n=30$ удаление четырех пакетов соответствует относительной несогласованности 0,25, а удаление восьми пакетов увеличивает несогласованность до 0,47.

Рассмотрим применение параметра θ для обоснования выбора параметров n и k .

Обоснование выбора количества кластеров в сенсорной сети

Предположим, на некотором уровне i сенсорной сети существует h узлов-предков, каждый из которых получает данные с n_i узлов-потомков (см. рис. 1). Полагая, что кластеры содержат одинаковое число узлов, определим среднее число n_i узлов-потомков в каждом кластере на этом уровне:

$$n_i = \frac{n}{h}, \quad (7)$$

где n – общее число узлов i -го уровня.

Аналогичным образом можно определить среднее число отбрасываемых пакетов k_i в каждом узле-предке на этом уровне:

$$k_i = \frac{k}{h}. \quad (8)$$

Подставляя (8) в выражение (6), получаем зависимость θ от количества кластеров и количества отбрасываемых пакетов на i -м уровне сети:

$$\theta(h, k_i) = 1 - \frac{(n - k_i h)(n - k_i h - 1)}{n(n-1)}. \quad (9)$$

При организации некоторого уровня сенсорной сети можно для фиксированного количества узлов n выбирать количество кластеров (узлов-предков) h и количество отбрасываемых пакетов k_i . Например, на рис. 3 для $n=30$ приведен график зависимости θ от h и k_i .

Из графика видно, что для заданных n и θ существует множество вариантов выбора h и k_i . Очевидно, что при выборе конкретной пары значений h и k_i следует стремиться к обеспечению максимальной пропускной способности сети.

Сформулируем *условие обеспечения максимальной пропускной способности сети*: среднее число пакетов $n_i - 1$, получаемых центральным узлом от h кластеров, должно быть равно среднему числу

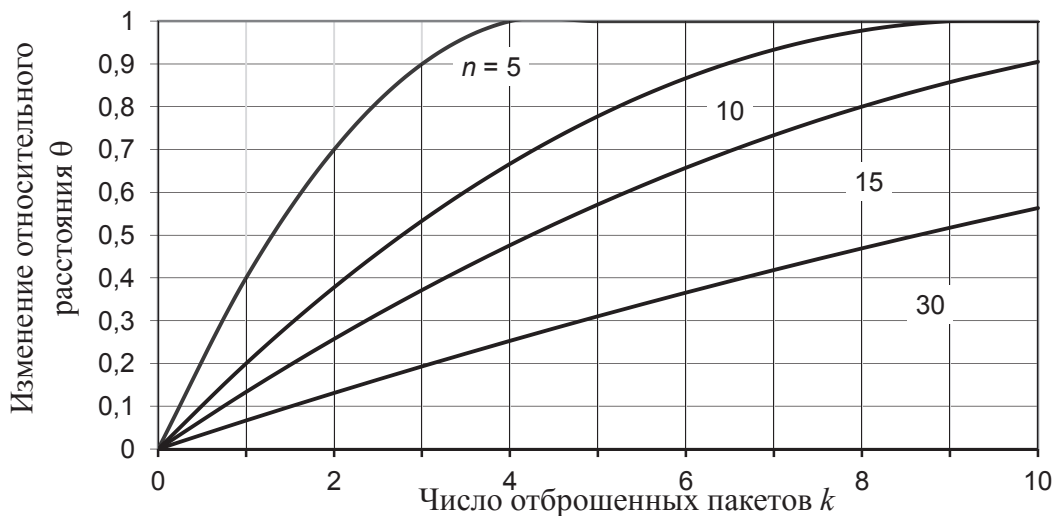


Рис. 2. Зависимость параметра θ от n и k

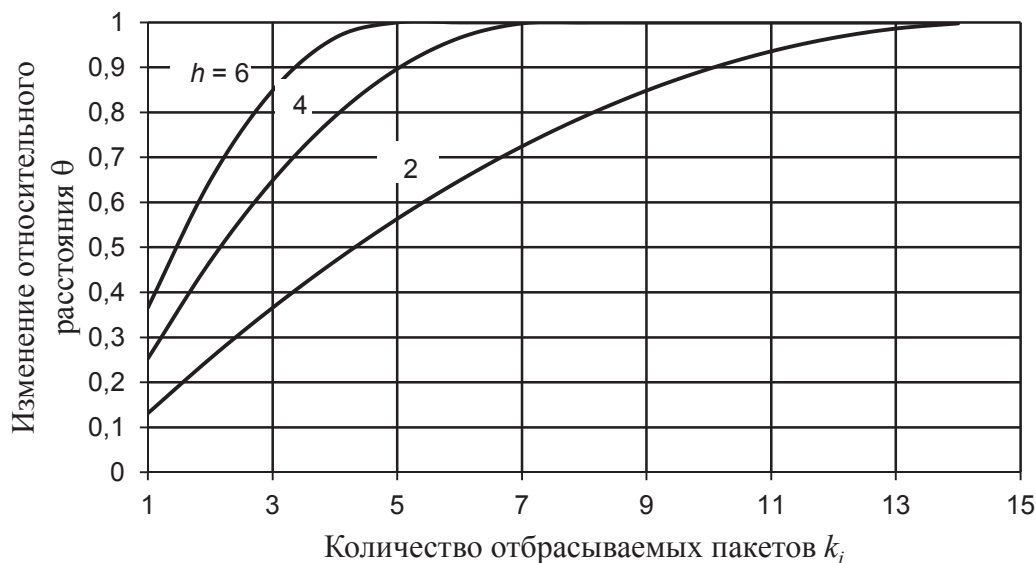


Рис. 3. Изменение θ в зависимости от h и k_i при $n=30$

$h(n-k_i)$ пакетов, переданных узлами-потомками, то есть

$$n_i - 1 = h(n_i - k_i).$$

С учетом (7), получаем:

$$\frac{n}{h} - 1 = n - k_i h. \quad (10)$$

Из выражения (10) следует, что число отбрасываемых на каждом уровне пакетов может быть выражено через общее число узлов n и число кластеров h рассматриваемого уровня сети:

$$k_i = \frac{nh + h - n}{h^2}. \quad (11)$$

Используя (7) и (11), можно переписать выражение (14) для параметра θ как функцию от n и h :

$$\begin{aligned} \theta(n, h) &= 1 - \frac{\left(n - \frac{nh + h - n}{h}\right) \left(n - \frac{nh + h - n}{h} - 1\right)}{n(n-1)} = \\ &= 1 - \frac{(n-h)(n-2h)}{nh(n-1)}. \end{aligned} \quad (12)$$

Зная число узлов сети и задавая параметр $\theta(n, h)$, можно, используя выражение (12), выбирать число кластеров h . Зная n , h и пропускную способность сети V , можно вычислить n_i , k_i и $\theta(n, h)$:

$$n_i = V + 1, \quad k_i = \frac{n-V}{h}, \quad \theta(n, h) = 1 - \frac{V(n-2h)}{n(n-1)}.$$

Например, для сети из $n=30$ узлов, обладающей и пропускной способностью $V=9$ пакетов за такт, при заданном $\theta=0,8$ получаем $n_i=10$, $h=3$, $k_i=7$.

Заключение

В статье описан механизм использования агрегирования предпочтений для организации переда-

чи данных в беспроводной сенсорной сети, позволяющий решить проблему ограниченной пропускной способности радиоканалов и обеспечивающий минимальную задержку при передаче пакетов данных, несущих информацию о событиях.

Предложенный подход был использован при совместной с Сингапурским национальным университетом разработке протокола Pritrans [14], реализующего динамическое назначение приоритетов узлов с использованием отношений консенсуса, формирование очереди передачи сообщений и распределение пропускной способности сети с учетом приоритетов. Полученные результаты показали эффективность разработанного протокола при превышении максимальной нагрузки на сеть, что позволило повысить надежность доставки приоритетных пакетов данных.

Следует отметить, при практическом применении отношения консенсуса следует принимать во внимание особенности правила Кемени. *Во-первых*, задача поиска отношения консенсуса по правилу Кемени является *NP*-полной [11] и, в общем случае, требует значительного времени для поиска решения (следовательно, требует значительных затрат вычислительных ресурсов узла, что приводит к снижению времени его автономной работы). *Во-вторых*, ранжирование Кемени далеко не всегда единственно, и мощность множества ранжирований Кемени может быть весьма большой [15]. Поэтому для проектирования протоколов передачи данных в реальном времени лучше использовать отношение консенсуса, найденное по правилу Борда [16]. Последнее представляет собой приближенное решение задачи Кемени и, как показывают эксперименты, в большом количестве практических случаев может служить хорошей заменой ранжирования Кемени.

СПИСОК ЛИТЕРАТУРЫ

- Hart J., Martinez K. Environmental Sensor Networks: A revolution in the earth system science? // *Earth-Science Reviews* – 2006. – № 78. – P. 177–191.
- Ададуров С.Е., Шабельников В.А. Интеллектуальная система мониторинга искусственных сооружений // *Мир транспорта*. – 2009 – № 3 – С. 32–37.
- Szewczyk R., Mainwaring A., Polastre J., Anderson J., Culler D. An analysis of a large scale habitat monitoring application // *Proc. 2nd Intern. Conf. on Embedded Networked Sensor Systems*. – Baltimore, MD, USA, 2004. – P. 214–226.
- Parker J.R. Multiple sensors, voting methods and target value analysis // *Proc. VI SPIE Conf. on Signal Processing, Sensor Fusion and Target Recognition*. – Orlando, Florida, USA, 1999. – P. 305–335.
- Hull B., Jamieson K., Balakrishnan H. Mitigating Congestions in Wireless Sensor Networks // *Proc. 2nd Intern. Conf. on Embedded Networked Sensor Systems*. – Baltimore, MD, USA, 2004. – P. 134–147.
- Chen D., Varshney P.K. QoS Support in Wireless Sensor Networks: A Survey // *Proc. Intern. Conf. on Wireless Sensor Networks*. – Las Vegas, Nevada, USA, 2004. – P. 227–233.
- Paek J., Govindan R. RCRT: Rate-controlled reliable transport for wireless sensor network // *Proce. ACM Conf. on Embedded Networked Sensor Systems*. – Sydney, Australia, 2007. – P. 305–319.
- Muravyov S., Chan M.C., Khomyakova M. Prioritizing sensed data transmission by consensus relation in wireless sensor network // *Proc. 12th IMEKO TC1 & TC7 Joint Symp. on Man Science & Measurement*. – Annecy, France, 2008. – P. 277–282.
- Кемени Дж., Снелл Дж. Кибернетическое моделирование. – М.: Советское радио, 1972. – 192 с.
- Muravyov S.V. Rankings as ordinal scale measurement results // *Metrology and Measurement Systems*. – 2007. – V. 13. – № 1. – P. 9–24.
- Muravyov S.V., Savolainen V. Special interpretation of formal measurement scales for the case of multiple heterogeneous properties // *Measurement*. – 2001. – V. 29. – P. 209–223.
- Кендэл М. Ранговые корреляции. – М.: Статистика, 1975. – 216 с.
- Кузьмин В.Б., Овчинников С.В. Об измерениях в порядковых шкалах // *Автоматика и телемеханика*. – 1974. – № 11. – С. 106–112.
- Shao Tao, Chan M.C., Muravyov S.V., Tarakanov E.V. Pritrans: A Prioritized Converge-cast Scheme using Consensus Ranking in Wireless Sensor Networks // *Proc. of SAS-2010*. – Limerick, Ireland, 2010. – P. 251–256.
- Muravyov S.V., Tarakanov E.V. Multiple solutions of an exact algorithm for determination of all Kemeny rankings: preliminary experimental results // *Proc. of ICIMCS*. – Hong Kong, 2011. – V. 1. – N.Y.: ASME Press, 2012. – P. 17–20.
- Литвак Б.Г. Экспертная информация: Методы получения и анализа. – М.: Радио и связь, 1982. – 184 с.

Поступила 23.03.2012 г.

УДК 621.398.725:621.317.727.1

ЦИФРОВАЯ ОБРАБОТКА ИМПУЛЬСНЫХ СИГНАЛОВ ДЛЯ ОПРЕДЕЛЕНИЯ ЧАСТОТНЫХ ХАРАКТЕРИСТИК ПРЕОБРАЗОВАТЕЛЕЙ ТОКА

А.И. Заревич, С.В. Муравьев, Е.В. Бедарева, О.Н. Величко*

Томский политехнический университет

E-mail: antonzarevich@ngs.ru

*ГП «Укрметртестстандарт», г. Киев, Украина

Предложен метод определения амплитудно- и фазочастотных характеристик линейных преобразователей тока, основанный на совместной цифровой обработке входного короткого импульсного воздействия и отклика на него. Частотная зависимость коэффициента передачи преобразователя рассчитывается с помощью компонентов комплексных спектральных преобразований напряжений с выходов исследуемого и эталонного преобразователей тока. Предложены пути повышения точности метода, основанные на усреднении по ансамблю спектральных составляющих сигналов.

Ключевые слова:

Цифровая обработка сигналов, преобразование Фурье, функция когерентности, токовый преобразователь, динамические характеристики.

Key words:

Digital signal processing, Fourier transform, coherence function, current transducer, dynamic behaviour.

Введение

Для преобразователей импульсных токов большой амплитуды (например, токовых шунтов) [1, 2] актуальна задача определения динамических метрологических характеристик, которая традиционно требует для своего решения наличия дорогостоящих источников гармонического или импульсного тока большой амплитуды и труднореализуемых методов измерения малых импедансов. Аль-

тернативой этим подходам являются методы определения амплитудно- и фазочастотной характеристик преобразователей, основанные на формировании короткого импульсного воздействия на исследуемый преобразователь и последующей цифровой обработки его входных и выходных сигналов [3].

Целью статьи является описание алгоритма цифровой обработки импульсных сигналов, получае-

мых при определении амплитудно- и фазочастотной характеристик и результатов его экспериментального применения для анализа динамического поведения тестовой RC-цепи.

Метод определения динамических характеристик преобразователей тока

В исследуемом методе определения динамических характеристик на вход преобразователя подается короткий импульс тока, входной и выходной сигналы сохраняются в памяти цифрового осциллографа и затем подвергаются спектральным преобразованиям. Искомые амплитудно- и фазочастотные характеристики определяются по арифметическому отношению спектральных плотностей входных и выходных сигналов на соответствующих частотах.

При практическом применении метода зондирующий сигнал может иметь любую форму и должен удовлетворять условиям Дирихле: на исследуемом участке быть кусочно-непрерывным, ограниченным и иметь конечное число экстремумов. Практическое выполнение данных условий является сложной задачей. Так, требование кусочной непрерывности сигнала подразумевает отсутствие разрывов первого рода, которые неизбежно возникают при естественном ограничении длительности числовой последовательности, соответствующей сигналу. Повышение детализации анализа приводит к необходимости уменьшения шага дискретизации и повышения разрядности квантования, что, в свою очередь, способствует зашумленности сигнала и, следовательно, увеличивает число локальных экстремумов.

Таким образом, в полной мере метод применим только для цепей, которые в исследуемом частотном диапазоне описываются дифференциальными уравнениями первого порядка. Для цепей более высокого порядка применимость метода ограничена минимальной шириной резонансов, которые могут быть различимы в результате спектрального анализа.

Получаемые в эксперименте входные и выходные сигналы шунта представлены последовательностью из N элементов $x(t_k)$, где t_k – время регистрации k -го элемента в последовательности ($1 \leq k \leq N$); $x(t_k)$ – входной сигнал, а $y(t_k)$ – выходной сигнал. Дискретное преобразование Фурье, примененное сигналам, даст их спектральный состав в виде соответствующих векторов комплексных чисел $X(\omega_k)$ и $Y(\omega_k)$:

$$X(\omega_k) = \sum_{j=1}^N x(t_j) \omega_N^{(j-1)(k-1)}, \quad Y(\omega_k) = \sum_{j=1}^N y(t_j) \omega_N^{(j-1)(k-1)},$$

$$\omega_N = e^{-2\pi i/N},$$

где ω_k – вектор частот, соответствующий комплексным векторам $X(\omega_k)$ и $Y(\omega_k)$, полученный исходя из длины выборки и частоты Найквиста.

Комплексный коэффициент передачи шунта $K(\omega_k)$ будет определяться по выражению

$$K(\omega_k) = \frac{Y(\omega_k)}{X(\omega_k)}. \quad (1)$$

Модуль комплексного коэффициента передачи есть амплитудно-частотная характеристика шунта, а его аргумент – фазочастотная характеристика.

В условиях физического эксперимента реальный импульсный сигнал всегда зашумлен и характеризуется нестабильностью амплитуды, длительности и формы огибающей. Средства измерения, применяемые для регистрации таких сигналов, также вносят свою погрешность. Влияние этих случайных факторов снижают путем многократных измерений и усреднения их результатов по ансамблю. Однако указанные особенности сигнала снижают эффективность использования временного сглаживания данных, а нестабильность длительности импульсов и условий запуска развертки цифрового осциллографа существенно ограничивают практическую применимость усреднения сигналов по ансамблю.

При постоянном размере выборки N и неизменности длительности развертки цифрового осциллографа, частоты компонент ω_k дискретного преобразования Фурье остаются неизменными и не зависят от смещения сигнала относительно начала выборки. Следовательно, шумы и искажения измеряемых сигналов приведут к случайной погрешности спектральных компонент на фиксированных частотах.

В предположении, что случайные погрешности подчинены нормальному закону, повышение точности практического применения метода может быть достигнуто путем равномерного усреднения по ансамблю реализаций спектральных компонент сигналов $X(f_k)$ и $Y(f_k)$ на фиксированной частоте f_k . При выполнении n измерений входного импульсного сигнала и соответствующего отклика шунта, усредненные спектральные компонент сигналов $\bar{X}(f_k)$ и $\bar{Y}(f_k)$ находятся по выражениям:

$$\bar{X}(f_k) = \frac{1}{n} \sum_{j=1}^n X_j(f_k), \quad \bar{Y}(f_k) = \frac{1}{n} \sum_{j=1}^n Y_j(f_k). \quad (2)$$

Тогда, с учетом (2), коэффициент передачи (1) может быть записан как:

$$\bar{K}(f_k) = \frac{\bar{Y}(f_k)}{\bar{X}(f_k)}. \quad (3)$$

Погрешность предложенного метода определения амплитудно- и фазочастотных характеристик токового преобразователя может быть оценена путем определения функции когерентности γ_{xy} между спектральными плотностями сигнала-отклика и тестового сигнала, как предложено в [4]:

$$\bar{\gamma}_{xy}^2(f_k) = \frac{\left| \sum_{j=1}^n \bar{X}_j^*(f_k) \bar{Y}_j(f_k) \right|^2}{\sum_{j=1}^n |\bar{X}_j(f_k)|^2 \sum_{j=1}^n |\bar{Y}_j(f_k)|^2}. \quad (4)$$

Функция когерентности γ_{xy} , определенная по выражению (4), в идеальном случае равна 1, что соответствует полной корреляции выходного $Y(f)$ и входного $X(f)$ сигналов. Дополнительные шумы, вносимые исследуемой цепью и средствами измерения, а также возможные нелинейные явления уменьшают γ_{xy} . В пределе γ_{xy} стремится к нулю, что соответствует отсутствию причинно-следственной связи между сигналами.

Среднее квадратическое отклонение σ , а, следовательно, и относительная погрешность определения коэффициента передачи $K(f_k)$ могут быть получены из выражения (4) следующим образом:

$$\sigma(\bar{K}(f_k)) = \frac{1}{\sqrt{2n}} \frac{\sqrt{1 - \bar{\gamma}_{xy}^2(f_k)}}{|\bar{\gamma}_{xy}(f_k)|}. \quad (5)$$

Полная погрешность измерения должна включать в себя относительную погрешность определения коэффициента передачи, полученную из выражения (5), и погрешности, вносимые используемыми средствами измерения.

На основании изложенных рассуждений был предложен алгоритм цифровой обработки импульсных сигналов, получаемых при исследовании преобразователей тока, и позволяющий определить амплитудно- и фазочастотные характеристики.

Алгоритм реализован на языке технических вычислений MATLAB [5], позволяющем осуществлять векторную обработку табличных данных и операции с вещественными числами разрядности 64. Таким образом, заметно сокращается время расчетов и снижается погрешность вычисления. Однако это приводит к нерациональному использованию памяти ЭВМ. Поэтому одновременная загрузка в память всех экспериментальных данных невозможна. Для обхода указанного ограничения в алгоритме применяются последовательная загрузка каждого файла данных. Далее данные подвергались быстрому преобразованию Фурье, умножались на весовой коэффициент усреднения и суммировались с результирующими данными для дальнейших вычислений.

Алгоритм быстрого преобразования Фурье, реализованный в MATLAB, предполагает, что длина исходной последовательности данных кратна степени 2. Для меньшей длины последовательности ее длина увеличивается до значения, кратного степени 2:

$$N_2 = 2^{\lceil \log_2 N \rceil}.$$

Алгоритм экстраполирует данные нулями, что эквивалентно оконному преобразованию Фурье с прямоугольным окном [6]. В то же время получаемые в физическом эксперименте сигналы зачастую содержат постоянную составляющую, а значения данных в начале и конце последовательности отличаются от нуля и, в общем случае, не равны друг другу. Поэтому в алгоритме использована экстраполяция входного и выходного сигналов полиномом нулевой степени. Значение экстраполяции, в отличие от предлагаемого в MATLAB, опре-

деляется путем усреднения данных из последней четверти последовательности:

$$x(t_k) = \frac{4}{N} \sum_{j=\frac{3}{4}N}^N x(t_j), \quad y(t_k) = \frac{4}{N} \sum_{j=\frac{3}{4}N}^N y(t_j),$$

$$k = N + 1, \dots, N_2.$$

Подобная экстраполяция допустима только для случая, если в исследуемой цепи на участке усреднения уже завершены переходные процессы, что имеет место при импульсном входном сигнале.

Алгоритм. Определение АЧХ и ФЧХ преобразователя тока

- 1: Пусть:
 - N – длина последовательности мгновенных отсчетов входного и выходного сигналов;
 - n – число импульсов в ансамбле;
 - t_k – временные отсчеты;
 - $x(t_k)$ и $y(t_k)$ – мгновенные, \bar{x} и \bar{y} – усредненные отсчеты входного и выходного сигналов;
 - N_2 – длина последовательности отсчетов, кратная степени 2;
 - f_k – частоты спектральных компонент сигналов;
 - $X(f_k)$, $Y(f_k)$, $\bar{X}(f_k)$ и $\bar{Y}(f_k)$ – спектральные компоненты сигналов и их усредненные значения
- 2: **for** $j=1, n$ **do**
- 3: Загрузка мгновенных отсчетов входного $\langle t_k, x(t_k) \rangle$ и выходного $\langle t_k, y(t_k) \rangle$ сигналов
- 4: Определение длины последовательности отсчетов N
- 5: Разделение загруженных отсчетов на три последовательности: t_k , $x(t_k)$ и $y(t_k)$
- 6: ► Проверка N на кратность степени 2
if $N \leftarrow 2^{\lceil \log_2 N \rceil}$ **then**
- 7: ► Поиск усредненного значения отсчетов последовательностей $x(t_k)$ и $y(t_k)$

$$\bar{x} \leftarrow \frac{4}{N} \sum_{j=\frac{3}{4}N}^N x(t_j), \quad \bar{y} \leftarrow \frac{4}{N} \sum_{j=\frac{3}{4}N}^N y(t_j)$$

для $k = \left[\frac{3}{4}N \right], \dots, N$
- 8: ► Определение новой длины последовательности отсчетов
 $N \leftarrow 2^{\lceil \log_2 N \rceil}$
- 9: Увеличение длин последовательностей $x(t_k)$ и $y(t_k)$ до N_2
- 10: ► Экстраполяция мгновенных отсчетов
 $x(t_k) = \bar{x}, \quad y(t_k) = \bar{y}$ для $k = N + 1, \dots, N_2$
- 11: **end if**
- 12: $N \leftarrow N_2$
- 13: Определение шага квантования t_s как усредненной разности между соседними элементами в последовательности t_k
- 14: Применение к последовательностям отсчетов x и y процедуры быстрого преобразования Фурье и определение спектральных компонент X и Y сигналов

- 15: ► Определение частот f_k спектральных компонент сигналов

$$f_k \leftarrow k/2t, \text{ для } k = 0, \dots, N/2 + 1$$

- 16: ► Усреднение спектральных компонент сигналов

$$\bar{X}(f_k) \leftarrow \bar{X}(f_k) + \frac{1}{n} X_j(f_k);$$

$$\bar{Y}(f_k) \leftarrow \bar{Y}(f_k) + \frac{1}{n} Y_j(f_k)$$

17: end for

- 18: Определение АЧХ $|K(f_k)|$ и ФЧХ $\arg(K(f_k))$ по выражениям (2) и (3)

- 19: Определение функции когерентности $\gamma_{xy}(f_k)$ и среднего квадратического отклонения $\sigma(f_k)$ по выражениям (4) и (5)

- 20: Визуализация и сохранение результатов

Исследование динамического поведения тестовой RC-цепи

Экспериментальное тестирование метода проведено для интегрирующей RC-цепи. Для такой цепи коэффициент передачи по току определяется как:

$$K_I(f) = \frac{1}{j2\pi fC}, \quad (6)$$

где R – сопротивление резистора; C – емкость конденсатора; f – частота.

Принципиальная схема эксперимента приведена на рис. 1. RC-цепь состояла из резистора сопротивлением $R=47,92$ Ом и конденсатора емкостью $C=47$ нФ. На вход цепи с генератора сигналов Актаком AWG-4110 подавался сигнал в виде последовательности коротких импульсов близкой к прямоугольной формы с частотой следования 23 кГц, скважностью 10 и амплитудой 10 В. Для измерения тока, использовался установленный на входе цепи эталонный трансформатор тока 13W0100 производства фирмы Lilco Ltd [7]. Сигналы с выхода генератора, трансформатора тока и конденсатора поступали на входы цифрового осциллографа LeCroy WaveSurfer 62Xs.

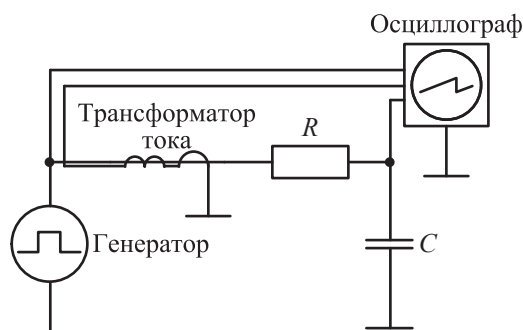


Рис. 1. Принципиальная схема эксперимента по определению коэффициента передачи RC-цепи

С целью получения ансамбля реализаций были зафиксированы 50 импульсов тока и напряжения на выходе и входе RC-цепи. Типичные формы сиг-

налов, наблюдаемые на экране осциллографа, приведены на рис. 2.

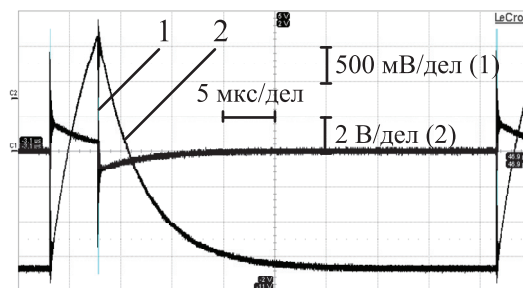


Рис. 2. Экспериментально измеренные сигналы на экране осциллографа: с выхода трансформатора тока (кривая 1) и с выхода RC-цепи (кривая 2)

В соответствии с предложенным алгоритмом, сигналы достраивались до завершения полного периода, в результате, длительности соответствующих сигналам числовых последовательностей стали кратными степени 2 и состояли из 524288 временных отсчетов. Результаты обработки данных показаны на рис. 3, 4. На рис. 3 сплошными линиями также представлены результаты теоретического расчета коэффициентов передачи по напряжению и по току, выполненные по выражению (6).

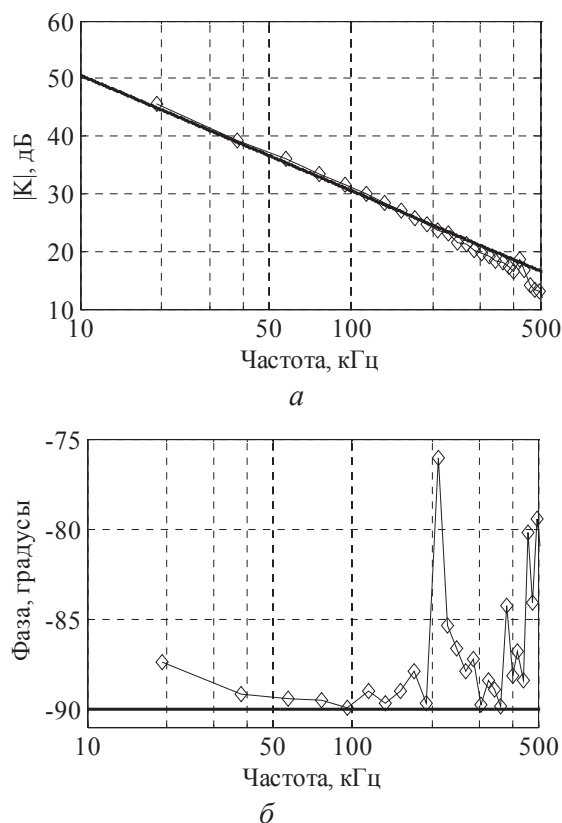


Рис. 3. Амплитудно- и фазочастотные характеристики RC-цепи: а) АЧХ; б) ФЧХ

Из представленных графиков видно, что в полосе частот до 180 кГц погрешность определения коэффициента передачи не превышает 0,1%. Результаты обработки экспериментальных результа-

тов весьма близки к результатам расчета. Рост погрешности при увеличении частоты связан с резонансными явлениями в соединительных проводах и шумами в использованном оборудовании. На частоте 300 кГц коэффициент передачи по напряжению уменьшается на 20 дБ. Измерения высокочастотных компонент сигнала при таком ослаблении крайне затруднительно и ограничено чувствительностью осциллографа.

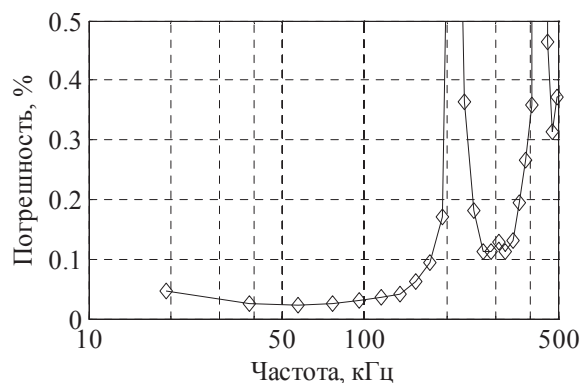


Рис. 4. Погрешность определения коэффициента передачи RC-цепи

Метод не требует использования дорогостоящего оборудования для формирования гармонических или импульсных токов большой амплитуды, позволяет избежать необходимости измерения малых импедансов. Использование цифровой обработки позволяет существенно сократить трудоем-

кость определения метрологических характеристик сильноточного оборудования.

Выводы

Предложен и экспериментально апробирован алгоритм цифровой обработки импульсных сигналов, получаемых при определении амплитудно- и фазочастотной характеристик преобразователей тока, описываемых моделью линейной электрической цепи. Алгоритм основан на совместной цифровой обработке усредненных по ансамблю мгновенных отсчетов входного короткого импульсного воздействия и отклика на него. Частотная зависимость коэффициента передачи рассчитывается с применением быстрого преобразования Фурье напряжений с выходов исследуемого и эталонного преобразователей. Работоспособность метода продемонстрирована на примере RC-цепи с известными параметрами. Погрешность определения не превышает 0,1 %.

Исследования выполнены в рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 гг.» по теме: «Программно-аппаратный комплекс для автоматизированных испытаний сильноточных преобразователей» (Государственный контракт № 11.519.11.6026) и в соответствии с грантом НК-566П/13 по направлению «Создание электронной компонентной базы» в рамках мероприятия 1.2.1 «Проведение научных исследований научными группами под руководством докторов наук» федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг.

СПИСОК ЛИТЕРАТУРЫ

1. Векслер М.С., Теплинский А.М. Шунты переменного тока. — Л.: Энергоатомиздат, 1987. — 120 с.
2. Muravyov S.V., Borikov V.N., Natalinova N.M. A computer system: measurement of welding surge currents // Measurement and Control. — 2009. — V. 42. — № 3. — P. 44–47.
3. Cherbaucich C., Crotti G., Kuljaca N., Novo M. Evaluation of the dynamic behaviour of heavy current shunts // Metrology in the 3rd Millennium: Proc. XVII IMEKO World Congress. — 22–27 June, 2003. — Dubrovnik, Croatia, 2003. — P. 586–589.
4. Бендат Дж.С., Пирсол А.Дж. Применения корреляционного и спектрального анализа. — М.: Мир, 1983. — 312 с.
5. Мэтьюз Д.Г., Финк К.Д. Численные методы. Использование MATLAB, 3-е изд.: Пер. с англ. — М.: Издательский дом «Вильямс», 2001. — 720 с.
6. Сергиенко А.Б. Цифровая обработка сигналов. — СПб.: Питер, 2002. — 608 с.
7. Cordingley B., Chamund D.J. Some observations on the performance of modern wideband current transformers in pulse current measurement applications // Proc. of 5th Modulator-Klystron Workshop for Future Linear Colliders CERN. — 26–27 April 2001. — Geneva, Switzerland, 2001. URL: <http://mdk2001.web.cern.ch/mdk2001/Proceedings/SessionPoster/sessionpos.pdf> (дата обращения: 19.04.2012).

Поступила 20.04.2012 г.

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ CUDA ДЛЯ УСКОРЕНИЯ ВЫЧИСЛЕНИЙ В НЕЙРОННЫХ СЕТЯХ

В.В. Парубец, О.Г. Берестнева, Д.В. Девятых

Томский политехнический университет
E-mail: VParubets.mail@gmail.com

Рассматривается применение искусственных нейронных сетей в задачах медицинской диагностики. Описывается возможность ускорения вычислений в сетях с технологией NVIDIA CUDA для реализации параллельных вычислений в обучении и распознавания образов сетью. Показано увеличение производительности сети.

Ключевые слова:

Искусственная нейронная сеть, NVIDIA CUDA.

Key words:

Artificial Neural Network, NVIDIA CUDA.

Введение

В современной медицине всё больше возрастает роль компьютерной диагностики. Особенно ярко это выявляется в случаях, когда необходимо произвести классификацию большого количества разнородных данных, чтобы выявить те или иные общие признаки для, например, разделения больных на группы. Искусственные нейронные сети изначально хорошо себя зарекомендовали в таких областях применения.

Существенную роль при анализе медико-биологической информации играет её особенности: описательный характер, использование формализмов, подверженность крайней вариабельности. Данные, даже описанные с помощью чисел, в большинстве случаев не могут быть хорошо упорядочены и классифицируемы, так как изменяются в зависимости от клинических традиций различных школ, геосоциальных особенностей регионов и даже отдельных учреждений, а также от времени.

Все задачи, решаемые человеком, с позиций нейроинформационных технологий можно условно классифицировать на две группы [1]:

1. Задачи, имеющие известный и определенный набор условий, на основании которого необходимо получить четкий, точный, недвусмысленный ответ по известному и определенному алгоритму.
2. Задачи, в которых не представляется возможным учесть все реально имеющиеся условия, от которых зависит ответ, а можно лишь выделить приблизительный набор наиболее важных условий. Так как часть условий при этом не учитывается, ответ носит неточный, приблизительный характер, а алгоритм нахождения ответа не может быть выписан точно.

Для решения задач первой группы с большим успехом можно использовать традиционные компьютерные программы. Как бы ни был сложен алгоритм, ограниченность набора условий (входных параметров) дает возможность составления алгоритма решения и написания конкретной программы, решающей данную задачу.

При решении задач второй группы применение нейротехнологии оправдывает себя по всем параметрам, при выполнении, однако, двух условий: *во-первых*, наличия универсального типа архитектуры и единого универсального алгоритма обучения (отсутствие необходимости в их разработке для каждого типа задач), *во-вторых*, наличия примеров (предыстории, фиксированного опыта), на основании которых производится обучение нейронных сетей.

Практически вся медицинская и биологическая наука состоит именно из задач, относящихся ко второй группе, и в большинстве этих задач достаточно легко набрать необходимое количество примеров для выполнения второго условия [2]. Это задачи диагностики, дифференциальной диагностики, прогнозирования, выбора стратегии и тактики лечения и др. Медицинские задачи практически всегда имеют несколько способов решения и «нечеткий» характер ответа, совпадающий со способом выдачи результата нейронными сетями.

Нейронные сети часто используются для классификации в области компьютерной диагностики. Создание оптимальной нейронной сети включает:

- 1) выбор активационной функции нейронов скрытого слоя;
- 2) выбор топологии сети;
- 3) выбор метода обучения;
- 4) обучение сети.

На сегодняшний день представлен широкий спектр нейронных сетей: многослойный перцептрон, самоорганизующиеся карты Кохонена, рекуррентные нейронные сети, свёрточные нейронные сети. Все они имеют свои отличительные особенности в топологии, функциях активации и методах обучения.

Известным ограничением развития нейросетевых алгоритмов следует признать высокие вычислительные затраты на реализацию таких методов [3]. К традиционным способам решения данной проблемы относят организацию параллельных и распределенных вычислений на специализированном аппаратном обеспечении.

Описание технологии CUDA

Особенностью оборудования, поддерживающего технологию *CUDA* (Compute Unified Device Architecture – унифицированная архитектура вычислительного устройства), является возможность обеспечивать на порядок большую (по сравнению с кластерами) пропускную способность при работе с памятью [4].

В графических ускорителях *NVIDIA*, начиная с восьмой серии, реализована архитектура параллельных вычислений *CUDA*, которая предоставляет специализированный программный интерфейс для не графических вычислений.

Логически графический процессор с поддержкой *CUDA* можно рассматривать как набор многоядерных процессоров. Основными вычислительными блоками таких видеочипов являются мультипроцессоры, которые состоят из восьми ядер, нескольких тысяч 32-битных регистров, 16 Кбайт общей памяти, текстурного и константного кэшей.

До официального появления технологии *CUDA* проводились эксперименты по использованию графических карт настольных систем для реализации потоковых вычислений. Так, с помощью графических программно-аппаратных интерфейсов и представления данных в качестве массивов текстур, удалось добиться трехкратного увеличения производительности в экспериментах. Стоит отметить также и одно из ключевых достоинств технологии *CUDA* – отсутствие необходимости в разработке программ следовать графическим «метафорам» – типам данных и принципам построения вычислений, характерным исключительно для обработки вершин и пикселей при построении кадра.

Реализация нейронной сети с сочетанием *CUDA* и *OpenMP* и параллелизма центральных процессоров на уровне языка программирования и программно-аппаратного интерфейса создает еще один уровень прироста производительности.

Литература, касающаяся не только *CUDA*, но и предыдущие аппаратные реализации генетических алгоритмов и нейронных сетей демонстрируют ускорение между 2× и 15× [4–6].

Основным источником информации по применению данной технологии может послужить портал разработчика компании *NVIDIA*, где можно найти как наиболее свежую версию программных библиотек, так и основную информацию о непосредственном внедрении графических процессоров для общих расчетов в своём проекте.

Несмотря на большую информативность портала компании *NVIDIA*, непосредственные оценки производительности и некоторые тонкости применения графических процессоров для задач численного интегрирования, опирались на работы [4–6].

Таким образом, применение описанных выше алгоритмов в связке друг с другом имеет право на существование и возможность развития.

Математическая постановка задачи

Поставим задачу диагностики пациента как задачу распознавания образов [7].

Пусть X – множество описаний объектов, Y – множество номеров (или наименований) классов. Существует неизвестная целевая зависимость значения которой известны только на объектах конечной обучающей выборки. Требуется построить алгоритм способный классифицировать $x \in X$.

Математически исходная информация о лабораторных показателях представляется в виде матрицы:

$$X_{N \times P} = \|x_{ij}\|,$$

где $i=1,2,\dots, n$ – номер элемента; $j=1,2,\dots, p$ – номер показателя; x_{ij} – значение j -го показателя для i -го элемента; n – размер выборки; p – число показателей.

Также есть вектор значений индивидуальных интегральных показателей здоровья для каждого ребенка:

$$Y_{N \times 1} = \{y_i\}.$$

Запишем функцию вида:

$$F(X \cdot W) = Y_{out},$$

где $W_{p \times 1}$ – вектор весовых коэффициентов синапсов; $Y_{out, N \times 1}$ – вектор выходных значений.

Пусть имеется нейронная сеть, выполняющая преобразование $F: X \rightarrow Y$ матрицы X из признакового пространства входов X в матрицу – столбец Y выходного пространства Y . Нейронная сеть находится в состоянии W из пространства состояний W . Пусть далее имеется обучающая выборка (X^α, Y^α) , $\alpha=1\dots p$. Рассмотрим полную ошибку E , получаемую сетью в состоянии W .

$$E = E(W) = \sum_{\alpha} \|F(X^\alpha; W) - Y^\alpha\| = \\ = \sum_{\alpha} \sum_i [F_i(X^\alpha; W) - Y^\alpha]^2.$$

Необходимо найти такую совокупность адаптивных весов нейронной сети W^* , которая доставляет минимум некоторому квадратичному функционалу: $\min E(W) \rightarrow 0$.

Задача кластеризации тогда будет представлена следующим образом. Пусть m – размерность входного пространства, а входной вектор выбирается из этого пространства случайно и обозначается так:

$$x = [x_1, x_2, \dots, x_m]^T.$$

Вектор синаптических весов каждого из нейронов имеет ту же размерность, что и входное пространство. Обозначим синаптический вес нейрона следующим способом:

$$w_j = [w_{j1}, w_{j2}, \dots, w_{jm}]^T, j = 1, 2, \dots, l,$$

где l – общее количество нейронов сети. Для того чтобы подобрать наилучший вектор w_j , соответ-

ствующий входному вектору x , нужно сравнить скалярные произведения $w_j^T x$ для $j = 1, 2, \dots, l$ и выбрать наибольшее значение, таким образом, мы определяем местоположение, которое должно стать центром топологической окрестности возбужденного нейрона.

Критерий соответствия основанный на максимизации скалярного произведения математически эквивалентен минимизации Евклидова расстояния между векторами x и w . Если использовать индекс $i(x)$ для идентификации нейрона, то сущность кластеризации проявится в данном выражении:

$$i(x) = \arg \min_j \|x - w_j\|, j = 1, 2, \dots, l.$$

Необходимо реализовать процесс обучения искусственной нейронной сети, распараллелив для применения на графических процессорах.

Выбор структуры нейронной сети обуславливается спецификой и сложностью решаемой задачи [7]. Для решения некоторых типов задач разработаны оптимальные конфигурации. В большинстве случаев выбор структуры нейронной сети определяется на основе объединения опыта и интуиции разработчика. Однако существуют основополагающие принципы, которыми следует руководствоваться при разработке новой конфигурации:

- возможности сети возрастают с увеличением числа ячеек сети, плотности связей между ними и числом выделенных слоев;
- введение обратных связей наряду с увеличением возможностей сети поднимает вопрос о динамической устойчивости сети;

Вопрос о необходимых и достаточных свойствах сети для решения того или иного рода задач представляет собой целое направление нейрокомпьютерной науки. Так как проблема синтеза нейронной сети сильно зависит от решаемой задачи, дать общие подробные рекомендации затруднительно. Очевидно, что процесс функционирования нейронной сети, то есть сущность действий, которые она способна выполнять, зависит от величин синаптических связей, поэтому, задавшись определенной структурой нейронной сети, отвечающей какой-либо задаче, разработчик сети должен найти оптимальные значения всех переменных весовых коэффициентов.

Из теории нейронных сетей известно [7], что количество нейронов во входном слое предпочтительно брать равным количеству входных сигналов. В выходном же слое также предпочтительно брать количество нейронов соответственно равное количеству выходных сигналов. Для нейронной сети, имеющей один скрытый слой, количество синаптических весов этого слоя можно приближенно оценить по формуле:

$$\frac{n_y n_p}{1 + \log_2 n_p} \leq n_w \leq n_y \left(\frac{n_p}{n_y} + 1 \right) (n_x + n_y + 1) + n_y,$$

где n_w – необходимое число синаптических весов создаваемой искусственной нейронной сетью; n_x и

n_y – размерность входного и выходного сигналов; n_p – число элементов обучающей выборки.

В свою очередь, получив n_w , можно найти количество нейронов во внутреннем (скрытом) слое:

$$n = \frac{n_w}{n_x + n_y},$$

где n_w – приблизительное количество нейронов в скрытом внутреннем слое.

В исследованиях [8] показана зависимость точности выполнения операции распознавания образов в медицинской диагностике от числа нейронов в скрытом слое. Выявлено, что сравнительно малое количество нейронов позволяет достичь большей точности.

В нашем случае число пациентов в каждой выборке равно 87, количество входных сигналов и, соответственно количество нейронов во входном слое, равно 16, а количество нейронов в выходном слое – 2. Получим, что приблизительное количество нейронов в скрытом слое лежит в пределах от 1 до 91.

В качестве активационной функции для нейронов скрытого слоя была выбрана сигмоидальная функция [7]:

$$F(x) = \frac{1}{1 + e^{-\beta x}},$$

а для нейронов выходного слоя – линейная. Здесь β – коэффициент, контролирующий «пологость» активационной функции.

Прямое распространение сигнала

Прямое распространение сигнала в нейросети описывается следующими формулами:

$$S_j^1 = \sum_{i=1}^m W_{ij}^1 x_i + W_0^1;$$

$$u_j = F(S_j^1), j = 1, \dots, n;$$

$$S_j^2 = \sum_{i=1}^n W_{ij}^2 u_i + W_0^2;$$

$$y_l = S_l^2, l = 1, \dots, n.$$

Здесь x_i – внешний сигнал, поступающий на i -й нейрон входного слоя; u_j – сигнал, вырабатываемый j -м нейроном скрытого слоя; y_l – сигнал, вырабатываемый l -м выходным нейроном. Верхние индексы у весовых коэффициентов обозначают номер слоя. Отметим, что матрицы весов можно объединить с соответствующими матрицами смещений (W_0^1, W_0^2). В дальнейшем под матрицей весов будем понимать именно объединенную матрицу.

Алгоритм обучения на графическом процессоре

Для того, чтобы в полной мере использовать возможности параллельной системы, которой является видеокарта, алгоритм настройки нейросети должен быть соответствующим образом преобразован. Для уменьшения числа обменов с глобальной памятью видеокарты и увеличения отношения

объема вычислений к объему загружаемых из нее данных по возможности все массивы данных должны быть записанными в двумерные массивы и обрабатываться на двумерной решетке. Для этого, например, всю обучающую выборку, т. е. совокупность пар (\vec{X}, \vec{D}) , поместим в две матрицы с размерами $P \times M$ и $P \times K$ соответственно. С этой же целью избавляемся от матриц, хранящих ошибки обучения в текущей эпохе, заменив их одним массивом для хранения весов $W_{N \times K}^2$, рассчитанных на предыдущей операции.

Естественно, что если число обучающих примеров очень велико, то эти матрицы придется разбивать на блоки. В этом случае пакетная обработка обучающей выборки позволяет снизить частоту обменов между центральным и графическим процессорами, которая приводит к существенному замедлению всего алгоритма. В результате этого при прямом распространении сигнала операция умножения матрицы весов на вектор, получаемый с входа нейронов, заменяется на операцию перемножения двух матриц:

$$Y_{P \times K} = F(X_{P \times M} W_{M \times N}^1) W_{N \times K}^2.$$

Наконец, изменив порядок вычислений, выделим ядра потоковой обработки, выполняющиеся на видеокарте, по возможности преобразуя схему вычислений так, чтобы во время обработки данных оперировать блоками нужного размера и обеспечить загруженность процессорных элементов, способную скрыть задержку при доступе к глобальной памяти видеокарты. Это также позволит ускорить обработку за счет такой особенности параллелизма видеокарты, как векторность вычислений на графической карте.

Таким образом, процесс обучения сети сводится к следующему алгоритму:

- 1) получение сигнала, поступающего на скрытый слой,

$$U_{P \times N} = F(X_{P \times M} W_{M \times N}^1);$$

- 2) вычисление выходного сигнала сети

$$Y_{P \times K} = U_{P \times N} W_{N \times K}^2;$$

- 3) итеративное изменение весовых коэффициентов – элементов матрицы $W_{N \times K}^2$

$$W_{jk}^{2*} = W_{jk}^2 - \alpha \sum_{p=1}^P (Y_{pk} - D_{pk}) U_{pj};$$

- 4) получение вспомогательной матрицы $U'_{P \times N}$

$$U'_{pj} = \sum_{k=1}^K (Y_{pk} - D_{pk}) W_{jk}^2;$$

- 5) изменение весовых коэффициентов $W_{M \times N}^1$ на новом шаге обучения

$$W_{ij}^{1*} = W_{ij}^1 - \beta \alpha \sum_{p=1}^P [U_{pj} (1 - U_{pj}) X_{pi} U'_{pj}].$$

Основным приемом перевода последовательно алгоритма в параллельный для реализации

в среде CUDA является замена циклов параллельно выполняющимися командами вида SIMD (Single Instruction, Multiple Data – множественный набор данных, обрабатываемых одной инструкцией). При этом ускоряется выполнение операций умножения векторов и матриц. Вначале нужно разбить задачу на вычислительные блоки и потоки [10], после чего, если необходимо, внести изменения в способ их взаимодействия и записать алгоритм в среде CUDA. Помимо этих преобразований при реализации алгоритмов, рассмотренных в настоящей работе были использованы специальные приемы работы с различными видами видеопамати [10]:

- использование общей памяти видеокарты для хранения часто используемых значений;
- планирование порядка исполнения вычислительных потоков, чтобы избежать невыровненного и непоследовательного (некогерентного) доступа к глобальной памяти видеокарты;
- использование текстур для осуществления невыровненного доступа к глобальной памяти;
- изменение порядка обработки данных во избежание конфликтов доступа к отдельным участкам памяти.

Реализация алгоритма

В качестве входных данных была использована база лабораторных показателей: общего анализа крови, общего анализа урины, анализ каленинкреиновой системы 87 детей обоего пола в возрасте от 1 до 36 мес. с патологией выделительной системы. Выборка содержала 31 представителя контрольной группы, которые были здоровы, у остальных был диагностирован пиелонефрит различных степеней тяжести. Каждый пациент описывается 16 входными показателями.

Смоделированы нейронные сети со следующими параметрами:

- топология сети – многослойный перцептрон;
- количество нейронов во входном/скрытом/выходном слое – 16/64/2;
- активационная функция – сигмоидальная;
- метод обучения – обратное распространение ошибки;
- ошибка, при которой пример считается распознанным верно, 0,1.

В первом случае, операции над матрицами производились на центральном процессоре. Во втором случае, помимо ресурсов центрального процессора применялся CUDA-совместимый графический процессор.

Сравнение производительности нейронных сетей производилось на рабочей станции с процессором Intel Pentium E5200, графической картой NVIDIA GeForce GTX550TI и объёмом оперативной памяти 4 Гб. Время выполнения обучения/тестирования алгоритмом, реализованным для центрального процессора – 712/26 с, с применением CUDA – 88/4 с.

Выводы

1. Рассмотрены приемы повышения производительности искусственных нейронных сетей в задачах медицинской диагностики при реализации параллельных алгоритмов с применением CUDA-устройств.

СПИСОК ЛИТЕРАТУРЫ

1. Галушкин А.И. О методике решения задач в нейросетевом логическом базисе // Нейрокомпьютер. – 2006. – № 2. – С. 49–70.
2. Мызников А.В., Россиев Д.А., Лохман В.Ф. Нейросетевая экспертная система для оптимизации лечения облитерирующего тромбангиита и прогнозирования его непосредственных исходов // Ангиология и сосудистая хирургия. – 1995. – № 2. – С. 100–103.
3. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. – Новосибирск: Наука, 1996. – 276 с.
4. Harris M. Mapping Computational Concepts to GPUs // GPU Gems 2. – 2006. – № 2. – P. 493–508.
5. Hall J.D., Carr N.A., Hart J.C. Cache and Bandwidth Aware Matrix Multiplication on the GPU. Technical Report. – UIUCDCS-R-2003–2328. 2012. URL: <http://graphics.cs.uiuc.edu/~jch/papers/UIUCDCS-R-2003–2328.pdf> (дата обращения: 17.01.2012).
6. Horn D. Stream Reduction Operations for GPGPU Applications // GPU Gems 2. – Addison Wesley, 2006. – P. 573–589.
7. Хайкин С. Нейронные сети полный курс. – М.: ООО «И.Д. Вильямс», 2006. – 1104 с.
8. Rossiev D.A., Savchenko A.A., Borisov A.G., Kochenov D.A. The employment of neural-network classifier for diagnostics of different phases of immunodeficiency // Modelling, Measurement & Control. – 1994. – V. 42. – № 2. – P. 55–63.
9. Горбань А.Н., Дунин-Барковский В.Л., Кардин А.Н. Нейроинформатика / Отв. ред. Е.А. Новиков. – Новосибирск: Наука, 1998. – 296 с.
10. NVIDIA CUDA Programming Guide Version 2.3.1 // NVIDIA – World Leader in Visual Computing Technologies. 2011. URL: http://www.nvidia.com/object/cuda_develop.html (дата обращения: 17.01.2012).

Поступила 24.01.2012 г.

УДК 004.652.3

ПРИМЕНЕНИЕ МЕТОДОВ ВИЗУАЛИЗАЦИИ ПРИ ИССЛЕДОВАНИИ СТРУКТУРЫ ЭКСПЕРИМЕНТАЛЬНЫХ МНОГОМЕРНЫХ ДАННЫХ

В.А. Воловоденко, О.Г. Берестнева, Е.В. Немеров*, И.А. Осадчая

Томский политехнический университет

*Сибирский государственный медицинский университет, г. Томск

E-mail: ogb@tpu.ru

Рассмотрены методы структурного анализа многомерных данных, представлены различные подходы к визуализации результатов экспериментальных исследований. Приведены примеры решения прикладных задач с использованием NovoSparkVisualizer.

Ключевые слова:

Кластерный анализ, методы визуализации, когнитивная графика, спектральные представления.

Key words:

Cluster analysis, imaging, cognitive graphics, spectral representation.

В настоящее время накоплен обширный арсенал средств анализа многомерных данных. Наиболее полное изложение применяемых здесь подходов, сопровождающееся подробными ссылками на ключевые работы, содержится в [1]. В [2] приведена классификация основных методов анализа структуры многомерных данных:

- 1) визуализация данных: линейные методы снижения размерности, нелинейные отображения, многомерное шкалирование, заполняющие пространство кривые;
- 2) автоматическое группирование: факторный и кластерный анализ объектов и признаков, иерархическое группирование, определение «точек сгущения».

В основу приведенной классификации положен признак, отображающий степень участия экспериментатора в выделении особенностей взаимоотношений между исследуемыми объектами и признаками. Применение методов визуализации данных нацелено на поиск наиболее выразительных изображений совокупности исследуемых объектов для последующего максимального задействования потенциала зрительного анализатора экспериментатора.

Компьютерная обработка данных предполагает некоторое математическое преобразование данных с помощью определенных программных средств. Для этого необходимо иметь представление как о математических методах обработки данных, так и о соответствующих программных средствах [2].

Целью работы являлось стремление визуализировать образы, представленные численными данными, и показать возможность выявления отношений между ними на основе наблюдений за совокупностью полученных образов.

Применение предложенного метода, на наш взгляд, приводит к значительному упрощению процедуры визуализации и способствует выявлению скрытых связей в очевидной форме. Это позволяет не только использовать свойства метода, но и способствует учету опыта группы экспертов, рассматривающих конкретную проблему. Последнее обстоятельство повышает актуальность предложенного подхода.

Методы визуализации позволяют исследователю одним взглядом обнаружить особенности, выявить закономерности и аномалии в больших объемах информации. Основной задачей визуализации данных является задача получения визуального образа, однозначно соответствующего набору данных [3]. В данной статье рассмотрен подход [4, 5], который позволяет провести визуализацию основных линейных конструкций: отрезок, ломаная линия, симплекс в многомерных пространствах. Основой визуализационного подхода является линейное преобразование значений многомерного наблюдения A в двумерную кривую $f_A(t)$, т. е. $A \leftrightarrow f_A(t)$, при этом гарантируется, что близким по значениям наблюдениям A и B будут соответствовать визуально близкие образы-кривые $f_A(t)$ и $f_B(t)$; для сильно различающихся по значениям наблюдений их образы-кривые будут заметно отличаться.

В рассматриваемом случае наиболее общей формой представления данных является вектор конечномерного пространства R_n

$$A = (a_0, a_1, a_2, \dots, a_{n-1}) \in R_n. \quad (1)$$

Для перехода от данного вектора к визуальному образу будет использоваться базис ортонормированных функций $\{\varphi_i(\tau)\}_{i=0}^{\infty}$. В качестве такого базиса можно использовать известные функции, в частности ортонормированные полиномы Лежандра на отрезке $[0, 1]$, множество которых обозначим через $\{l_i(\tau)\}_{i=0}^{\infty}$.

В таком случае точке с координатами $A=(a_0, a_1, a_2, \dots, a_{n-1})$ можно поставить в соответствие функцию:

$$F_A(\tau) = \sum_{i=0}^{n-1} a_i l_i(\tau). \quad (2)$$

Формирование вектора A связано с преобразованием данных. Для характеристики наблюдаемого многомерного объекта большую роль играют значения его координат. В большинстве случаев каждый показатель имеет свою единицу измерения, и его значение будет влиять на вид функции $F_A(\tau)$. Для того чтобы исключить влияние разноименности показателей на вид функции $F_A(\tau)$, необходимо перейти к безразмерным единицам одним из известных способов. Следует отметить, что порядок

включения показателей в вектор A , также будет влиять на вид функции $F_A(\tau)$. Разница между формулами (1) и (2) заключается в том, что для вектора A из (1) возможно только аналитическое представление, в то время как для функции $F_A(\tau)$ возможно представление в виде графика этой функции. Между (1) и (2) устанавливается однозначная связь в обе стороны, т. е. взаимно-однозначная связь. Если ввести в рассмотрение второй вектор

$$B = (b_0, b_1, b_2, \dots, b_{n-1}) \in R_n,$$

то ему ставится в соответствие функция

$$F_B(\tau) = \sum_{i=0}^{n-1} b_i l_i(\tau).$$

Будем считать, что функции $F_A(\tau)$ и $F_B(\tau)$ являются визуальными образами точек A и B , принадлежащих пространству R_n (рис. 1).

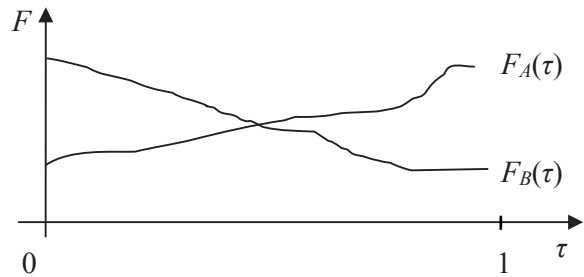


Рис. 1. Визуальные образы точек A и B , принадлежащих пространству R_n

Чем больше кривые неотличимы друг от друга, тем идентичнее наблюдения, которые они представляют, т. е. метод устанавливает взаимно-однозначное соответствие между строками в наборе данных и их кривыми.

Если отобразить кривые наблюдений в трехмерном пространстве, используя третье измерение, называемое также « Z -измерением», как расстояние в многомерном пространстве или промежуток времени между двумя наблюдениями, можно обнаружить много интересных свойств. Ось Z — это ось следования образов. В простейшем случае значения координаты Z совпадает с номером образа, но этой координате можно придать значение расстояния в пространстве признаков от начала координат до образа (объекта).

Представленный подход реализован в пакете *NovoSpark Visualizer* [6], на базе которого авторами был успешно решен ряд прикладных задач анализа и интерпретации многомерных данных в социальной сфере [7] и медицине [5].

Рассмотрим результаты решения двух прикладных задач, иллюстрирующих эффективность представленного метода визуализации многомерных экспериментальных данных.

В качестве исходной информации первой задачи имеем выборочные сведения о беременных женщинах и соответствующие им результаты лабораторных и инструментальных исследований (подробное описание экспериментальных данных представлено в [8–10]). Экспериментальная вы-

борка состоит из трех групп беременных женщин (30 человек) в возрасте от 18 до 44 лет.

Первую (контрольную) группу составили 10 практически здоровых беременных женщин в возрасте от 20 до 42 лет, не получавших оздоровительных мероприятий. Вторую (группу сравнения) составили 10 беременных женщин в возрасте от 18 до 40 лет, имевших соматические заболевания, не получавших комплекса оздоровительных мероприятий. Третью (основную группу) составили 10 беременных женщин в возрасте от 19 до 43 лет, имевших соматические заболевания, которым проводился различный комплекс оздоровительных мероприятий.

Изменение их состояние оценивалось в I и III триместрах беременности по набору некоторых медико-биологических показателей. В табл. 1 представлен фрагмент таблицы исходных данных.

В качестве эталонного набора выбрана контрольная группа. Выберем по одному наблюдению из каждой группы:

- 1 – контрольная;
- 2 – сравнения;
- 3 – основная.

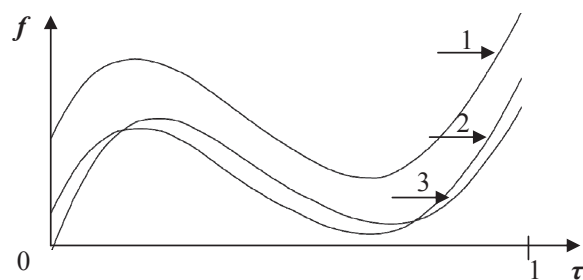


Рис. 2. Сравнение наблюдений I триместра

Как видно из рис. 2, наблюдения основной группы и группы сравнения очень похожи и значительно отличаются от наблюдений контрольной группы, что свидетельствует о том, что исходный набор переменных состояния беременных «подходит» для оценки динамики их состояния при проведении комплекса оздоровительных мероприятий.

Рассмотрим кривые наблюдений в III триместре.

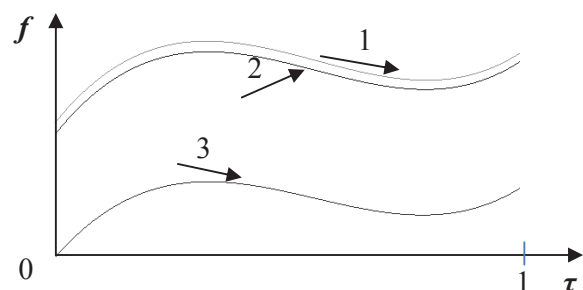


Рис. 3. Сравнение наблюдений III триместра

Из рис. 3 и 4 видно, что наблюдение из основной группы сильно приблизилось к наблюдению контрольной группы, а наблюдение группы сравнения практически не изменилось.

Это означает, что к III триместру при получении комплекса оздоровительных мероприятий показате-

ли беременных женщин, имевших соматические заболевания, значительно приближаются к показателям практически здоровых беременных женщин.

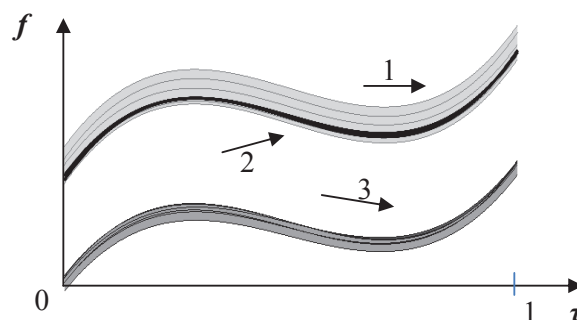


Рис. 4. Сравнение группы эталонов

Рассмотрим реализацию подхода, предложенного в [6], позволяющего отображать многомерные объекты в виде кривых или «спектров». «Спектральные представления» в данном методе подчеркивают отличительные характеристики каждой кривой и помогают более детально исследовать их визуальные свойства. Цветная палитра акцентирует уровни изменения значений кривых. Производя воображаемое упорядочение кривых вдоль оси и, наблюдая сверху на результат этой операции, можно получить цветные полосы, представляющие собой спектр каждого наблюдения [7]. Цветовая гамма определена по аналогии с картой, где наибольшим значениям функции соответствуют красные тона, а наименьшим значениям соответствуют холодные, синие оттенки.

Выберем 4 наблюдения из сравнительной (группа сравнения) и основной групп I триместра. Построим горизонтальные спектры выбранных наблюдений (рис. 5).

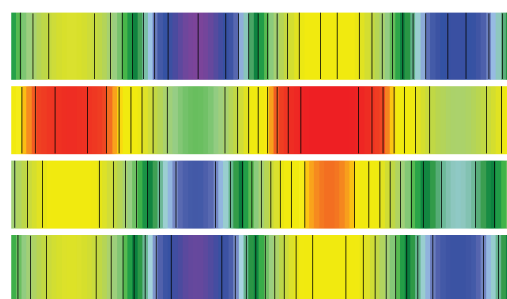


Рис. 5. Спектры выбранных наблюдений I триместра

Из рис. 5 видно, что при сопоставлении полихроматических рядов каждого из 4 горизонтальных спектров, построенных по единой технологии можно найти основания для их объективной классификации и соизмерения. Для этого могут быть использованы многоэтапные (многоуровневые) процедуры сравнения по качественным хроматическим признакам.

На первом этапе таких процедур ищется доминирующий признак различия, применяемый в качестве основания для разбиения всего множества спектров на классы. Таким дифференцирующим признаком может быть Pr_1 – наличие в спектрах

хроматических компонентов наиболее темного цвета (в нашем случае – фиолетового). Тогда, в соответствии с Pr_1 , на первом этапе из исходной совокупности индивидуальных спектров наблюдений $\{1, 2, 3, 4\}$ выделяются классы $K_{11}=\{1, 4\}$ и $K_{12}=\{2, 3\}$. На втором этапе в классе K_{12} используется признак Pr_2 – наличие компонентов менее темного цвета, чем тот, который использовался для формулирования Pr_1 (в данном случае, темно-синий). Данный признак производит разбиение класса K_{12} на два: $K_{21}=\{2\}$ и $K_{22}=\{3\}$. На данном этапе возможности дальнейшего формирования и применения признаков, аналогичных Pr_1 , Pr_2 , необходимых для продолжения классификации исчерпаны. Окончательными результатами проводимой классификации будут классы $K_{11}=\{1, 4\}$, $K_{21}=\{2\}$ и $K_{22}=\{3\}$.

Аналогично разделяем на классы выборку, которая содержит 6 наблюдений из сравнительной (группа сравнения) и основной групп III триместра.

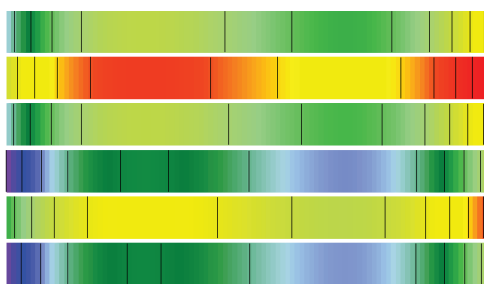


Рис. 6. Спектры выбранных наблюдений III триместра

В результате выделены 4 класса: $K_{11}=\{4, 6\}$, $K_{21}=\{1, 3\}$, $K_{31}=\{5\}$ и $K_{32}=\{2\}$.

Исходная информация второй прикладной задачи представляет собой данные о пациентах с четырьмя типами заболеваний, а именно:

- бронхиальная астма непсихогенная (*BANP*);
- бронхиальная астма сомато-психогенная (*BASP*);
- бронхиальная астма психогенно-индуцированная (*BAPI*);
- психогенная одышка (*PD*).

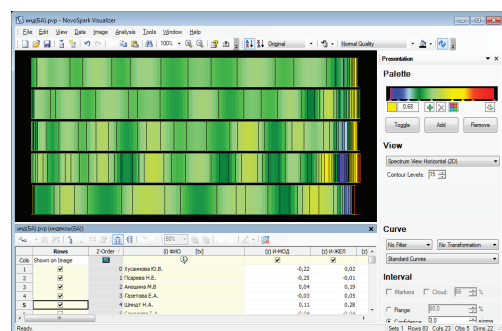
Проведем анализ показателей физиологических реакций бронхолегочной системы в ответ на психофизиологическое воздействие (аудиовизуальную стимуляцию). Технология получения экспериментальных данных подробно изложена в [10].

Сравним выборочные данные для четырех форм бронхиальной астмы по визуальной близости спектров наблюдений (рис. 7).

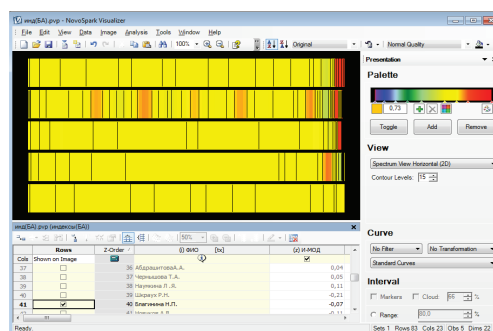
Каждая цветная полоска в спектральном виде соответствует показателю одного пациента. В нашем случае на рис. 7 представлено по пять цветных полосок, соответственно представляющих по пять пациентов с различными формами бронхиальной астмы. Цветовые «спектры» пациентов с диагнозом *BAPI* и *PD* схожи. То же можно сказать и о пациентах с диагнозом *BASP* и *BANP*.

Как видно из рис. 7, наиболее близкие показатели имеют пациенты с *BASP* и *PD*. Наиболее выраженные различия можно заметить у пациентов с *BAPI*.

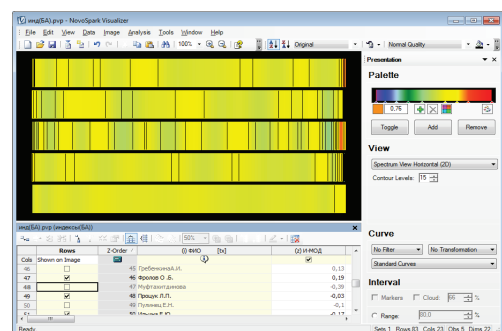
Таким образом, использование средств когнитивной графики позволило выявить некоторые



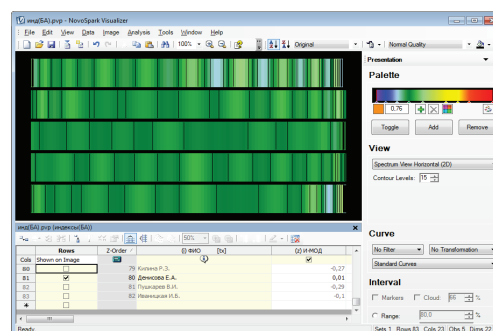
а



б



в



г

Рис. 7. Спектральные представления данных о пациентах с различными диагнозами: а) *BAPI*; б) *BASP*; в) *BANP*; г) *PD*

не известные ранее закономерности физиологических реакций бронхолегочной системы в ответ на психофизиологическое воздействие.

Наиболее ярко отражает черты многомерного анализа в классификации кластерный анализ. Кластер – объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определёнными свойствами. Главное назначение кластерного анализа – разбиение множества исследуемых объектов и признаков на однородные, в соответствующем понимании группы или кластеры. Это означает, что решается задача классификации данных и выявления соответствующей структуры в ней. Методы кластерного анализа можно применять в самых различных случаях, даже в тех случаях, когда речь идет о простой группировке, в которой все сводится к образованию групп по количественному сходству [2]. Основное достоинство кластерного анализа в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. В табл. 1 и 2 представлены результаты разбиения, полученные с помощью кластерного анализа (метод Уорда) на базе пакета *Statistica*.

В табл. 1 представлены результаты кластеризации по физиологическим показателям бронхолегочной системы, отражающим динамику их изменения после проведения сеансов аудиовизуальной стимуляции.

Таблица 1. Распределение пациентов по кластерам (степень воздействия аудиовизуальной стимуляции)

Диагноз	Количество пациентов			
	1 кластер	2 кластер	3 кластер	4 кластер
<i>BANP</i>	1	5	15	8
<i>BASP</i>	3	6	5	4
<i>BAPI</i>	8	14	2	0
<i>PD</i>	1	0	6	5

В 1-м кластере преимущественно оказались пациенты с диагнозом *BAPI*, а также 3 человека с *BASP*, 1 – с *BANP* и 1 – с *PD*.

Во второй кластер преимущественно попали также больные *BAPI*, 6 человек – с диагнозом *BASP* и 5 человек – с *PD*.

Третий кластер объединил преимущественно людей с диагнозом *BANP*. Также сюда попали 2 человека с *BAPI*, 5 человек – с *BASP* и 6 – с *PD*.

В четвертом кластере также преимущественно оказались люди с *BANP*, а также 4 – с *BASP* и 5 – с *PD*.

На основе анализа полученных результатов (значений центроидов для каждого кластера) был сделан вывод о том, что на пациентов 1, 3 и 4 кластеров аудиовизуальная стимуляция мозга оказывает аналогичное воздействие – улучшаются значе-

ния показателей вентиляции легких и механики дыхания. При этом для первого кластера улучшение значений показателей вентиляции легких менее выражено, чем для третьего кластера. Улучшение статической растяжимости легких характерно только для 4 кластера. Для пациентов, попавших во 2 кластер, аудиовизуальная стимуляция практически не оказывает никакого воздействия на изменение показателей вентиляции легких и механики легких.

В табл. 2 представлены результаты кластеризации по исходным («фоновым») физиологическим показателям.

Таблица 2. Распределение пациентов по кластерам («фоновые» показатели бронхолегочной системы)

Диагноз	Количество пациентов			
	1 кластер	2 кластер	3 кластер	4 кластер
<i>BANP</i>	6	21	0	2
<i>BASP</i>	3	6	7	2
<i>BAPI</i>	14	1	7	2
<i>PD</i>	1	8	2	1

В первый кластер попали, в основном, пациенты с диагнозом «психогенно-индуцированная астма», во второй – пациенты с непсихогенной бронхиальной астмой и с психогенной отдышкой, в третий – пациенты с психогенно-индуцированной и сомато-психогенной бронхиальной астмой. И, наконец, в четвертом оказалось практически одинаковое количество представителей каждого из заболеваний. Полученные результаты показывают, что на основе имеющихся показателей бронхолегочной системы не удалось выделить кластеры, соответствующие медицинским диагнозам.

Выводы

Показано, что методы визуализации многомерных данных и подходы на их основе перспективны для анализа и наглядного представления экспериментальных результатов.

Представление *многомерного* наблюдения в виде *двумерного* образа (кривой) гарантирует, что объектам, близким по значениям их характеристик будут соответствовать визуально близкие образы-кривые; для наиболее различающихся – их образы-кривые будут заметно отличаться. Становится возможным автоматически классифицировать наблюдения, определять наиболее важные переменные в модели, производить кластеризацию данных, визуально сравнивать индивидуальные наблюдения и целые наборы данных, а также выполнять много других задач в работе с многомерными данными.

Спектральное представление визуального образа является инструментом, наиболее подчеркивающим различия или сходства образов, чем традиционные методы, характеризующие эти свойства на уровне числовых параметров.

СПИСОК ЛИТЕРАТУРЫ

1. Симчера В.М. Методы многомерного анализа статистических данных. – М.: Финансы и статистика, 2008. – 400 с.
2. Дюк В., Эмануэль В. Информационные технологии в медико-биологических исследованиях. – СПб.: Питер, 2003. – 528 с.
3. Шаропин К.А., Берестнева О.Г., Шкатова Г.И. Визуализация результатов экспериментальных исследований // Известия Томского политехнического университета. – 2010. – Т. 316. – № 5. – С. 172–176.
4. Берестнева О.Г., Пеккер Я.С., Шаропин К.А., Воловоденко В.А. Выявление скрытых закономерностей в медицинских и социально-психологических исследованиях // Аппликативные вычислительные системы: Труды II Междунар. конф. по аппликативным вычислительным системам. – г. Москва, 29–31 октября 2010. – М.: Институт Актуального образования «ЮрИнфоР-МГУ», 2010. – С. 287–296.
5. Шаропин К.А., Берестнева О.Г., Воловоденко В.А., Марухина О.В. Визуализация медицинских данных на базе пакета NovoSpark // Известия Южного федерального университета. Сер. Технические науки. – 2010. – Т. 109. – № 8. – С. 242–249.
6. Воловоденко В.А., Эйдензон Д.В. Визуализация и анализ многомерных данных с использованием пакета «NovoSpark Visualiz-
er». 2008. URL: <http://www.tsu.ru/storage/iro/k020410/s4/s4.doc> (дата обращения: 25.11.2011).
7. Марухина О.В., Берестнева О.Г., Шаропин К.А., Осадчая И.А. Когнитивная графика в социально-психологических исследованиях // Информационные и математические технологии в науке и управлении: Труды XVI Байкальской Всерос. конф. – г. Иркутск, 2011. – С. 176–181.
8. Кривоногова Т.С., Желев В.А., Тропова Т.Е., Гергет О.М. Роль раннего комплекса оздоровительных мероприятий в охране здоровья матерей и их детей // Российский вестник перинатологии и педиатрии. – 2009. – Т. 54. – № 3. – С. 14–19.
9. Гергет О.М., Берестнева О.Г. Интеллектуальная система оценки состояния здоровья беременных женщин и детей // Информатика и системы управления. – 2010. – № 2. – С. 61–64.
10. Немеров Е.В., Языков К.Г. К вопросу изучения личностных свойств в психофизиологической реактивности больных бронхиальной астмой на аудиовизуальную стимуляцию // Вестник Томского государственного педагогического университета. – 2011. – Вып. 6 (108). – С. 134–137.

Поступила 06.04.2012 г.

УДК 004.9

ОЦЕНИВАНИЕ СОСТАВА ИЗОБРАЖЕНИЙ КЛЕТОЧНЫХ ПРЕПАРАТОВ ДЛЯ МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

М.С. Тарков, Н.В. Тихонов, В.Г. Половинкин

Институт физики полупроводников им. А.В. Ржанова СО РАН, г. Новосибирск
E-mail: tarkov@isp.nsc.ru; niknksk@list.ru; pvg@isp.nsc.ru

Предложен метод оценивания состава изображений, основанный на применении полутоновой эрозии и дилатации, преобразования «верх шляпы» и последующей пороговой сегментации. Метод реализован в виде программного продукта и используется в рамках методики цейтраферной (замедленной) компьютерной видеосъемки микроскопических объектов, разработанной для изучения барьерных свойств мембран растительных клеток. Он позволяет оценить относительную площадь наблюдаемых объектов на изображении и может быть использован в различных приложениях, когда требуется определить скорость распада клеточных структур.

Ключевые слова:

Оценивание состава изображений, клеточные препараты.

Key words:

Estimation of the image, cell preparations.

Введение

В настоящее время существует большое количество программных пакетов, предназначенных для анализа биомедицинских изображений, в том числе изображений цитологических препаратов [1, 2]. В то же время в используемых на практике программах анализ ряда характеристик клеток (периметр, площадь, количество) недостаточно автоматизирован. Данная работа ставит своей целью ликвидацию указанного недостатка существующих программных средств.

Изучение свойств мембран клеток в ответ на внешние воздействия – одна из ключевых задач современной биологии и медицины. Известно, что неблагоприятные условия окружающей среды со-

провожают нарушениями устойчивости клеточных мембран и могут приводить к развитию различных форм стресса [3]. Вот почему важно иметь технологию для оперативной оценки барьерных свойств мембран. Один из подходов к такой оценке связан с анализом изображений клеточных препаратов. При неблагоприятных внешних воздействиях (стресс) происходит разрушение клеточных мембран и гибель клеток. В результате количество клеток на изображении уменьшается (рис. 1). По динамике изменения количества клеток на серии изображений, полученной цейтраферной (замедленной) компьютерной видеосъемкой (ЦКВ) [4], можно судить об устойчивости клеточных мембран к внешним воздействиям.

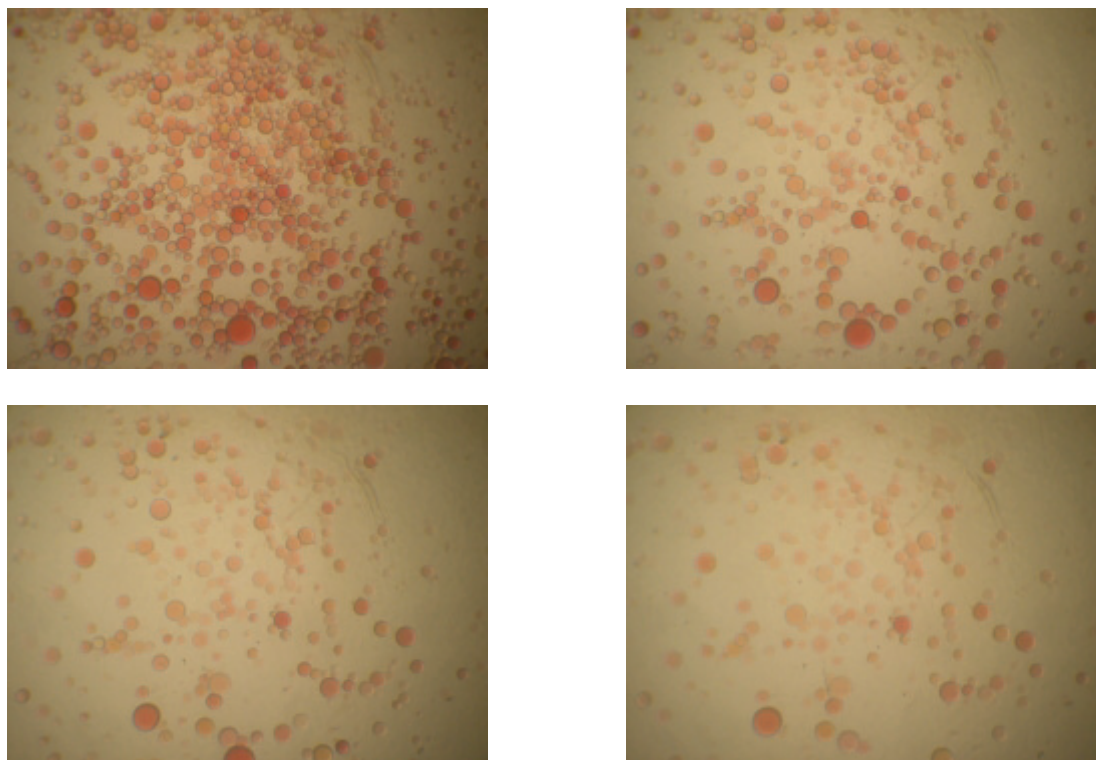


Рис. 1. Фрагмент последовательности изображений клеточного препарата, полученной цейтраферной компьютерной видеосъемкой (разрешение съемки 1600×1200)

Известные методы автоматического анализа изображений клеточных структур [5, 6] дают хорошие результаты на препаратах крови, но сложны в применении к изображениям растительных клеток. Подсчет количества растительных клеток на изображении затруднен тем, что изображения клеток существенно перекрываются. Поэтому в данной работе в качестве приближенной количественной оценки состава изображения вычисляется относительная суммарная площадь, занимаемая клетками на изображении.

Метод количественного оценивания состава изображения

Предложенный ниже алгоритм выделения объектов на изображении (рис. 2, а) и оценивания его состава основан на использовании типовых операций обработки изображений [7, 8]. Алгоритм работает с монохромным изображением и включает следующие этапы:

1. Предварительная обработка (рис. 2, б). Включает медианную фильтрацию с последующим выравниванием гистограммы изображения и построением негативного изображения. Медианный фильтр заменяет значение пикселя на значение медианы распределения яркостей всех пикселей в окрестности, включая исходный. Медианная фильтрация используется для сглаживания изображения и удаления импульсных шумов.

Последующая эквализация гистограммы изображения позволяет расширить динамический ди-

апазон уровней яркости изображения, что обычно означает большую контрастность выходного изображения. Пусть $p_r(r_j), j=1, \dots, L$ – гистограмма уровней яркости исходного изображения, тогда преобразование эквализации имеет вид

$$s_k = \sum_{j=1}^k p_r(r_j), \quad k=1, \dots, L,$$

где s_k – яркость обработанного изображения, соответствующая значению яркости r_k входного изображения.

2. Построение разомкнутого изображения (эрозия изображения с последующей дилатацией его) (рис. 2, в).

Полутонная дилатация изображения f по структурообразующему элементу b обозначается $f+b$ и определяется формулой

$$(f+b)(x, y) =$$

$$= \max \{f(x-x', y-y') + b(x', y') \mid (x', y') \in D_b\},$$

где D_b обозначает область определения b , и предполагается, что $f(x, y) = -\infty$ вне области определения f . На практике полутонная дилатация обычно выполняется с использованием плоских структурообразующих элементов, для которых $b=0$ в каждой точке, принадлежащей D_b . В этом случае операция взятия максимума полностью определяется конфигурацией нулей и единиц двоичной матрицы D_b , а уравнения для полутонной дилатации принимают более простой вид:

$$(f+b)(x, y) = \max \{f(x-x', y-y') \mid (x', y') \in D_b\}.$$

Таким образом, полутоновая дилатация является операцией взятия локального максимума, в которой максимум вычисляется по множеству пикселей окрестности, форма которой задается областью D_b .

Полутоновая эрозия изображения f по структурообразующему элементу b обозначается $f \ominus b$ и определяется формулой:

$$(f \ominus b)(x, y) = \max \{f(x - x', y - y') - b(x', y') \mid (x', y') \in D_b\},$$

где предполагается, что $f(x, y) = +\infty$ вне области определения f . Как и дилатация, полутоновая эрозия чаще всего использует плоские структурообразующие элементы. В этом случае уравнения для эрозии можно упростить и привести их к виду

$$(f \ominus b)(x, y) = \max \{f(x - x', y - y') \mid (x', y') \in D_b\},$$

то есть полутоновая эрозия является операцией взятия локального минимума, где минимум берется по множеству пикселей окрестности, форма которой задается областью D_b .

Размыкание изображения f по элементу b обозначается $f \circ b$ и определяется выражением

$$f \circ b = (f \ominus b) \oplus b.$$

Сначала выполняется эрозия f по b , за которой следует дилатация по b . Размыкание имеет простую геометрическую интерпретацию. Представим изображение $f(x, y)$ в виде трехмерной поверхности, то есть значения яркости представляются высотами над координатной плоскостью. Тогда процесс размыкания f по b представляет собой перемещение (перекатывание) структурообразующего элемента строго под трехмерной поверхностью изображения f , причем поверхность размыкания получается как огибающая высшие точки, которые достигает этот элемент. В общем случае процедура размыкания используется для удаления узких всплесков яркости при сохранении среднего полутонового фона и широких областей изменения яркости относительно неизменными. В качестве структурообразующего элемента b использовался круг с радиусом 25 пикселей.

3. Преобразование «верх шляпы» (рис. 2, г), которое сводится к вычитанию разомкнутого изображения из изображения, представленного

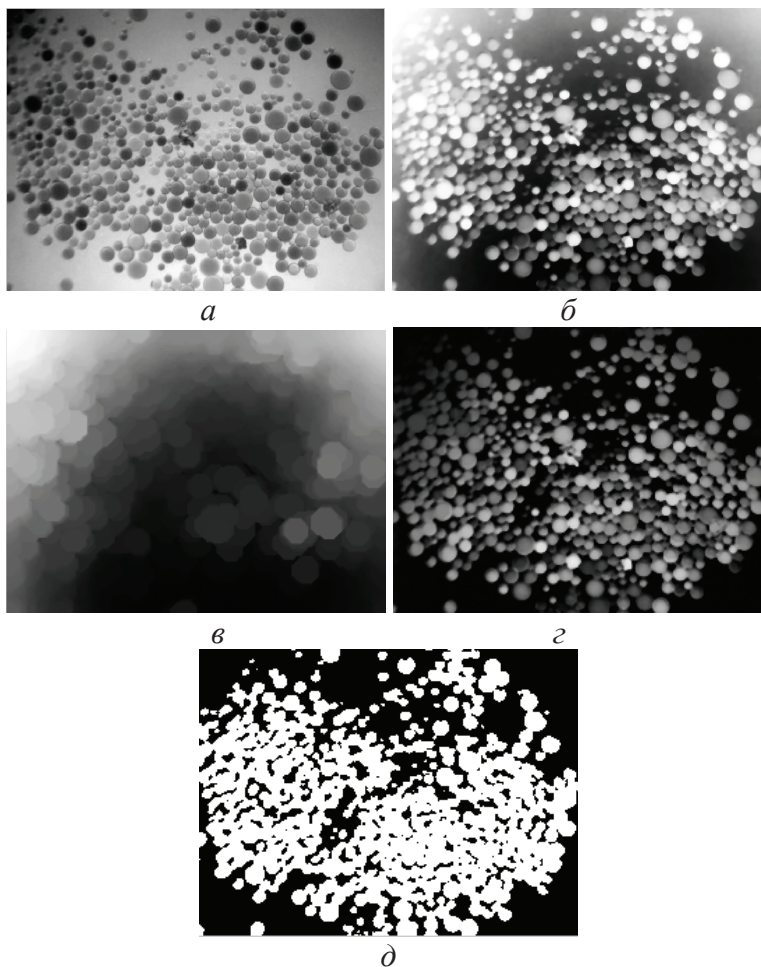


Рис. 2. Алгоритм выделения клеток на изображении: а) исходное изображение; б) результат применения медианной фильтрации, эквализации гистограммы и дополнения; в) разомкнутое изображение; г) преобразование «верх шляпы»; д) результат пороговой обработки

на рис. 2, б. Результатом преобразования «верх шляпы» является изображение с выровненным фоном.

4. Пороговая сегментация изображения. Значение порога T подбирается экспериментально. Пороговое преобразование имеет вид

$$f_b(i, j) = \begin{cases} 1 & \text{при } f(x, y) \geq T, \\ 0 & \text{при } f(x, y) < T, \end{cases}$$

где f и f_b – исходное и результирующее бинарное изображение. В результате получаем бинарное изображение f_b (рис. 2, д, порог $T=24$).

В качестве величины, характеризующей состав изображения, используется отношение

$$S = \frac{N}{|f_b|},$$

где N – количество пикселей переднего плана (т. е. тех пикселей, которым соответствует значение 1) бинарного изображения f_b , $|f_b|$ – число пикселей изображения f_b .

Программная реализация метода

Разработанный метод доведен до программного продукта и используется в рамках метода ЦКВ микроскопических объектов [4], разработанного для изучения барьерных свойств мембран. Он позволяет оценить относительную площадь наблюдаемых объектов на изображении (рис. 1). Программный продукт реализован в среде MATLAB [8] (программа CellsPro), а также в среде Visual Studio C++ с использованием процедур обработки изображений

библиотеки OpenCV [9, 10] (программа CellsProCV). Эксперименты показали, что переход от реализации предложенного метода в MATLAB к реализации его в Visual Studio C++ позволил сократить время обработки изображений в 6 раз. Время обработки одного изображения размером 1600×1200 пикселей программой CellsProCV на процессоре Intel Pentium Dual-Core CPU E 52000, 2,5 ГГц составляет 37 с.

Заключение

Предложен простой метод оценивания состава изображений, основанный на применении полутоновой эрозии и дилатации, преобразования «верх шляпы» и последующей пороговой сегментации. Метод реализован в виде программного продукта и используется в рамках методики цейтраферной (замедленной) компьютерной видеосъемки микроскопических объектов, разработанной для изучения барьерных свойств мембран растительных клеток. Метод позволяет оценить относительную площадь наблюдаемых объектов на изображении и может быть использован в различных приложениях, когда требуется определить скорость распада клеточных структур. Разработанный метод апробирован на сериях изображений препаратов, полученных с помощью метода цейтраферной компьютерной видеосъемки, и оказался эффективным в медико-биологических исследованиях клеточных структур.

Авторы выражают благодарность В.Н. Нурминскому за предоставленные изображения клеточных препаратов.

СПИСОК ЛИТЕРАТУРЫ

1. Программное обеспечение фирмы Micromed. 2012. URL: <http://www.micromed.ru/index15.htm> (дата обращения: 12.03.2012).
2. Тарков М.С., Полоз Т.Л., Полоз В.В. Нейросетевой анализатор изображений для цитологической диагностики фолликулярных опухолей щитовидной железы // Вестник Томского государственного университета. – 2006. – № 18. – С. 185–188.
3. Нурминский В.Н., Озолина Н.В., Нестеркина И.С., Колесникова Е.В., Корзун А.М., Чернышов М.Ю., Тихонов Н.В., Тарков М.С., Салаяев Р.К. Стабильность вакуолярных мембран растений при осмотическом стрессе и воздействии редокс-агентов // Биологические мембраны. – 2011. – Т. 28. – № 3. – С. 224–229.
4. Нурминский В.Н., Корзун А.М., Розинов С.В., Салаяев Р.К. Компьютерная цейтраферная видеосъемка фракции изолированных вакуолей // Биомедицинская химия. – 2004. – № 50 (Прилож. № 1). – С. 180–187.
5. Хлопов А.Н. Машинное зрение в задаче автоматического подсчета клеток крови человека // Электронный научный журнал «Исследовано в России». 2007. URL: <http://zhurnal.ape.relarn.ru/articles/2007/167.pdf> (дата обращения: 12.03.2012).
6. Ковригин А.В. Применение принципов построения систем машинного зрения в задаче анализа изображений клеточных структур // Научный журнал КубГАУ. – 2007. – № 29 (5). 2007. URL: <http://ej.kubagro.ru/2007/05/pdf/03.pdf> (дата обращения: 12.03.2012).
7. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005. – 1072 с.
8. Гонсалес Р., Вудс Р., Эдлинс С. Цифровая обработка изображений в среде MATLAB. – М.: Техносфера, 2006. – 616 с.
9. Bradski G., Kaehler A. Learning OpenCV. – O'Reilly Media, Inc., USA, 2008. – 576 p.
10. Библиотека программ обработки изображений OpenCV. 2012. URL: <http://sourceforge.net/projects/opencvlibrary/> (дата обращения: 12.03.2012).

Поступила 12.03.2012 г.

УДК 004.942:338.27:504.062

ПРЕИМУЩЕСТВА ДИФФЕРЕНЦИАЛЬНЫХ МОДЕЛЕЙ В ЭКОЛОГО-ЭКОНОМИЧЕСКОМ МОДЕЛИРОВАНИИ

А.В. Затонский

Березниковский филиал Пермского национального исследовательского политехнического университета

E-mail: zxenon@narod.ru

Показано, что использование эколого-экономических моделей в форме обыкновенных дифференциальных уравнений в задачах прогнозирования развития управляемых систем, предпочтительнее, чем полиномиальных моделей и временных трендов, в смысле адекватности и уменьшения возможного разброса прогнозов, необходимых для принятия решений.

Ключевые слова:

Принятие решений, эколого-экономическая модель, обыкновенные дифференциальные уравнения.

Key words:

Decision support, ecological-economic model, ordinary differential equation.

В практике экологического, социального и экономического моделирования часто используются модели динамики вида

$$y(\vec{x}(t), \vec{z}(t), t) = a_0 + \sum_i a_i x_i(t) + \sum_k b_k z_k(t),$$

где $\vec{x}(t) = \{x_1(t), x_2(t), \dots\}$ – вектор факторов, $\vec{z}(t) = \{z_1(t), z_2(t), \dots\}$ – вектор возмущений, $y(\cdot)$ – реакция исследуемого объекта;

$$\text{или } y(\vec{x}(t), \vec{z}(t), t) = a_0 + \prod_i a_i x_i(t) + \prod_j b_j z_j(t),$$

либо, для функций одного аргумента –

$$y(x(t), z(t), t) = a_0 + \sum_i a_i x(t)^i + \sum_k b_k z(t)^k,$$

либо модели временных рядов в форме $y(t) = \sum_{i=0}^l a_i t^i$.

Поиск в Интернет позволяет обнаружить авторефераты, в которых используются подобные модели: [1. С. 29], [2. С. 27] [3. С. 10 и 16] и т. п. Назначение моделей, обычно – исследование истории и, на ее основании, прогнозирование свойств объекта в зависимости от принятых решений $\vec{x}(t)$.

Такие модели можно упрощенно понять, например, так: если вкладывать в предприятия (отрасль) инвестиции по графику $x_i(t)$, то на выходе получим чистый дисконтированный доход (или другой показатель экономической эффективности) $y(x_i(t), z_i(t))$ с учетом спроса на продукцию (возмущающего воздействия) $z_i(t)$. Дальше обычно речь идет об идентификации a_i и b_j , об учете обратных связей, выраженных некоторой функцией $F(y)$, а точнее

$$y(\vec{x}(t), \vec{z}(t)) = a_0 + \sum_i a_i x_i(t) + \sum_k b_k z_k(t) - F(y(\vec{x}(t), \vec{z}(t))) \quad (1)$$

и т. п. При этом молчаливо принимается предположение, что существует прямая связь между факторами и значением реакции, а единственный динамический элемент в модели – чистое запаздывание (например, в моделях вида $y(\vec{x}(t)) = a_0 + \sum_i a_i x_i(t - \Delta t)$).

Однако подобное предположение не всегда близко к реальности. Например, удобряя поле по определенным правилам, можно получить рост урожая (и дальнейшие экономические или социальные бонусы). То есть достоверно, из множества наблюдений, известно, что внесенное количество удобрений x_1 *ускоряет* рост урожая в каком-то диапазоне вноса удобрений:

$$\frac{\partial y(x, z, t)}{\partial t} \approx a_0 + a_1 x(t), \quad \forall t: 0 < x(t) \leq x_{\max}, \quad a_1 > 0,$$

а снижение количества осадков в определенных условиях *снижает* скорость роста:

$$\frac{\partial y(x, z(t), t)}{\partial t} \approx a_0 + b_1 z(t), \quad \forall t: z_{\min} \leq z(t) \leq z_{\max}, \\ b_1 > 0, \quad \forall t: z(t) < 0.$$

Для сложных систем, особенно учитывающих естественные процессы в природе, массовая идентификация коэффициентов связи между $y(t)$ и $x_i(t)$ без убедительного доказательства их взаимной независимости приводит к порождению «попугайских моделей» [4], адекватно интерполирующих прошлое, но не способных к прогнозу будущего – что, собственно, и требуется при построении моделей поддержки принятия решений.

К ним же относятся попытки экстраполировать $y(t)$ вперед по данным временных рядов (трендов), особенно с учетом ошибок или ненаблюдаемых внешних возмущений, что проиллюстрировано ниже.

Возникает вопрос – что же идентифицировать при построении динамической экономико-математической модели: связь между фактором и *реакцией* или связь между фактором и *динамикой изменения* реакции под воздействием этого фактора? В теории автоматического управления, как известно, фактор (или динамика его изменения) линеаризуется, а затем исследуется его влияние на *динамику* поведения объекта. Подобные подходы к экономико-математическим системам также разработаны очень давно. Например, в [5. С. 99] формулируется модель экологического равновесия

$$\frac{\partial y(x, y, t)}{\partial t} = y - y^2 - x(t)y, \quad 0 \leq x(t) < 1, \quad (2)$$

соответствующей по форме (1), от которой недалеко как до доходности (определяемой здесь квотой вылова $x(t)$), так и до катастроф в развитии популяции, что, собственно, и рассматривается далее в книге.

Таким образом, на уровне общенаучных рассуждений получается вывод, что в эколого-экономических моделях лучше использовать в качестве основы дифференциальные, а не алгебраические уравнения динамики системы.

Попробуем проверить вывод на нескольких примерах. Добавим в ур. (2) возмущающее воздействие – например, сезонное (периодическое) влияние погодных условий на воспроизводство популяции вида $z(t) = \sin(b_2 t)$ и получим

$$\frac{\partial y(x, y, t)}{\partial t} = y - y^2 - x(t)y - b_1 \sin(b_2 t).$$

Построим в MatLAB простую модель, положив рост квоты вылова в виде линейной зависимости $x(t) = a_0 + a_1 t$ (единица измерения времени – год). В [5] аналитически доказано, что при использовании варианта модели

$$\frac{\partial y(x, y, t)}{\partial t} = y - y^2 - x$$

система теряет устойчивость при $x \geq 0,25$ (в отсутствие обратной связи, связывающей вылов с текущим значением популяции). Проверим это, включив *Manual Switch* в нижнее положение, задав *Gain*=0 (рис. 1), чтобы исключить пока возмущающие воздействия, $a_1=0$ и задавая последовательно $a_0=0,22$, $a_0=0,25$ и $a_0=0,28$. При этом начальное

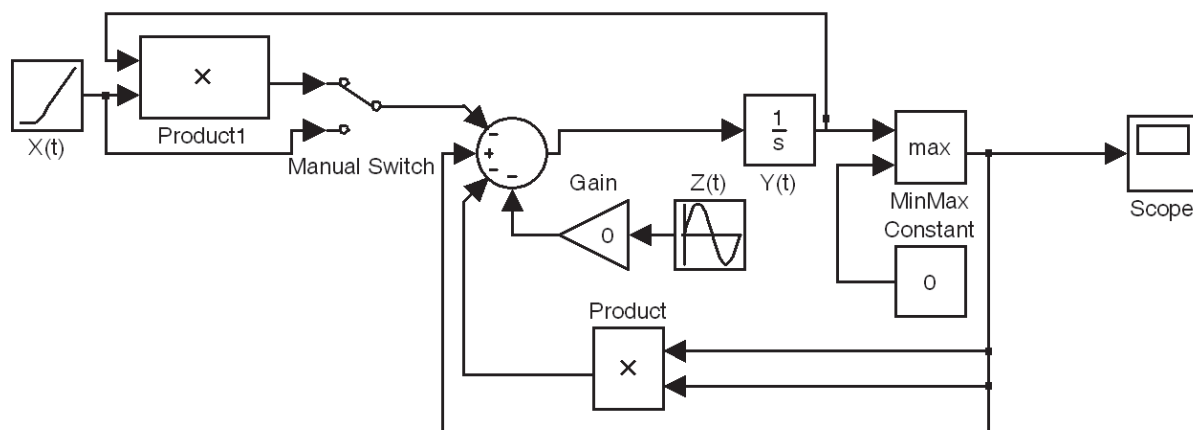


Рис. 1. MatLAB модель прогноза развития экологической системы

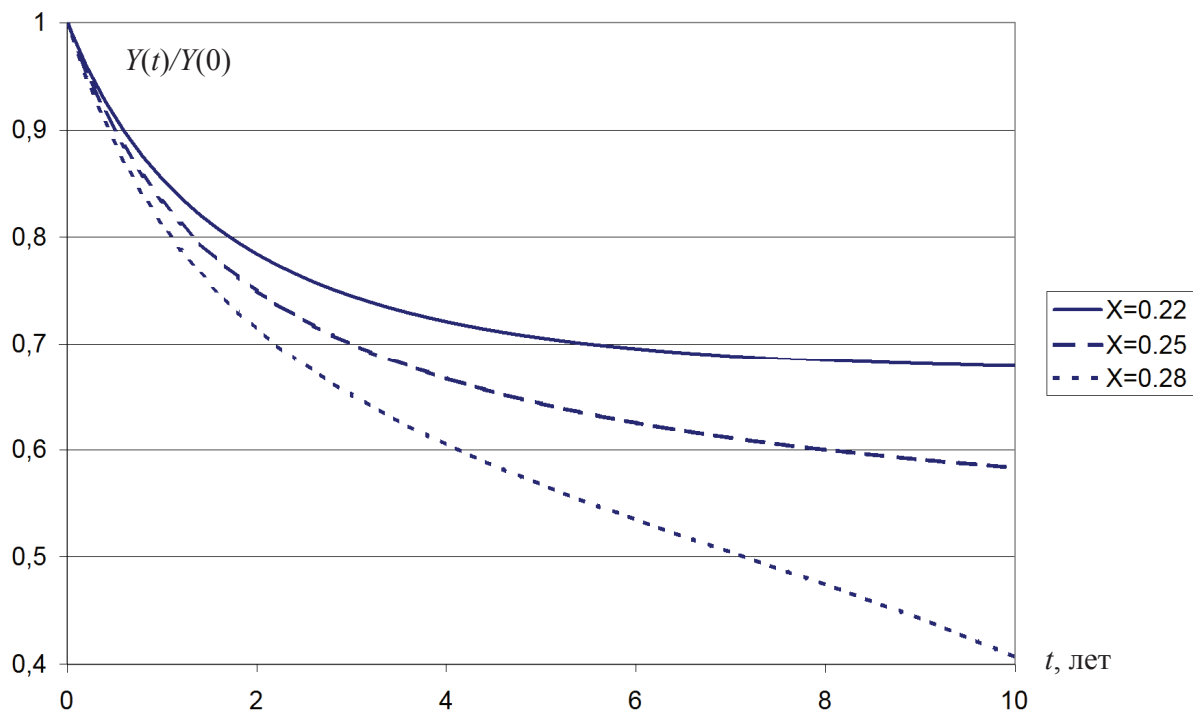


Рис. 2. Изменение по годам (ось абсцисс) динамики численности популяции (ось ординат) в зависимости от доли вылова X (ряды данных)

значение (*Initial Condition*) в блоке интегратора « $Y(t)$ » установим в 1 (полная популяция).

Блок *MinMax* в модели предназначен, чтобы не допускать падения популяции ниже 0 с последующим аварийным остановом расчетов.

Действительно, при превышении $x > 0,25$ наблюдается катастрофическое снижение популяции в конце десятилетнего периода (рис. 2).

Так как теоретический результат совпадает с расчетным, будем считать, что построенная модель адекватно отражает (2). Включим *Manual Switch* в верхнее положение (задавая, таким образом, долю от размера популяции) и установим *Gain* в произвольным образом выбранное значение 0,05; зададим $a_0 = 0,2$ и $a_1 = 0,025$, период изменения возмущения $z(t)$ установим равный 2π (один год). Получим зависимость популяции от времени, показанную на рис. 3 сплошной линией. Попробуем спрогнозировать развитие популяции по данным 1–6 годов, проведя регрессионный анализ. Используем для этого все доступные в *MS Excel* модели. Полу-

чим уравнения динамики вида $y(t) = \sum_{i=0}^l a_i t^i$:

1. $y(t) = -0,045481t + 0,927281;$ (3)

2. $y(t) = 0,008226t^2 - 0,093550t + 0,969766;$ (4)

3. $y(t) = -0,002270t^3 + 0,028255t^2 - 0,138835t + 0,987380;$ (5)

4. $y(t) = 0,000724t^4 - 0,010835t^3 + 0,060137t^2 - 0,177522t + 0,995103;$ (6)

5. $y(t) = 0,930131e^{-0,055976t}.$ (7)

Из рис. 3 очевидно, что экстраполяция в данном случае получается неудачной. Здесь и во всех следующих графиках по оси x отложено время в годах, по оси y – доля популяции от начального значения $y(0) = 1$.

Будем искать решение задачи в виде

$$y(t) = a_0 + \sum_{i=1}^m a_i x(t)^i, \text{ полагая } x(t) = 0,2 + 0,025t \text{ из-}$$

вестным (так как решение о доле вылова, принятое или планируемое лицом, принимающим решения, и должно быть известным). Для этого подготовим вспомогательную таблицу и произведем поиск решения наименьшего квадратичного отклонения

$$\bar{a} : \left(y^*(t) - \left(a_0 + \sum_{i=1}^m a_i x(t)^i \right) \right)^2, \text{ где } y^*(t) : t \in \{t_i\} - \text{полу-}$$

ченные при помощи модели (рис. 1) «экспериментальные» значения, при $m = 1, 4$ на интервале времени с 1 по 6 года. «Спрогнозируем» развитие ситуации при принятом решении $x(t) = 0,2 + 0,025t$ на период с 7 по 10 год. Получим, вне зависимости от степени полинома, неудовлетворительные по качеству прогнозы (рис. 4).

Модели, по которым произведен расчет на рис. 3 и 4, получились вполне «попугайские» (по терминологии К.С. Лосева): они неплохо интерполируют исходные данные (на интервале до 6 года включительно), но качество экстраполяции оставляет желать лучшего.

Перейдем к моделированию воздействия факторов на динамику объекта. Подберем методом наименьших квадратов коэффициенты уравнений

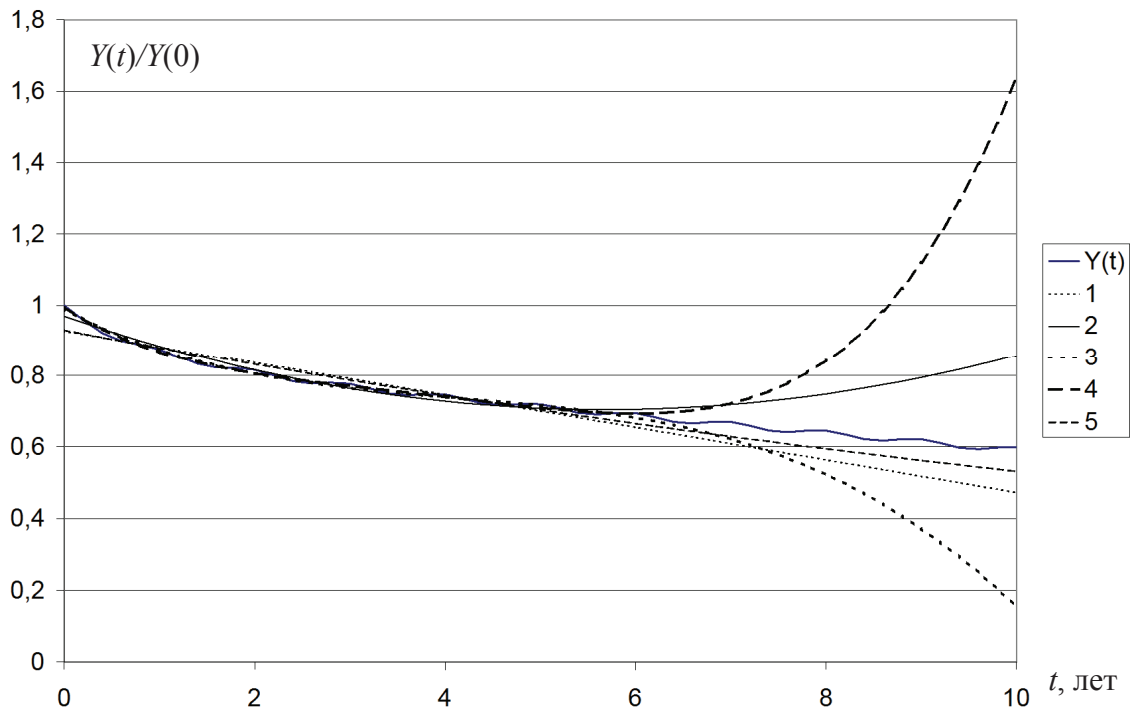


Рис. 3. Неудачные попытки экстраполяции численности популяции зависимостями $Y(t)$: исходные данные: 1) ур. (3); 2) ур. (4); 3) ур. (5); 4) ур. (6); 5) ур. (7)

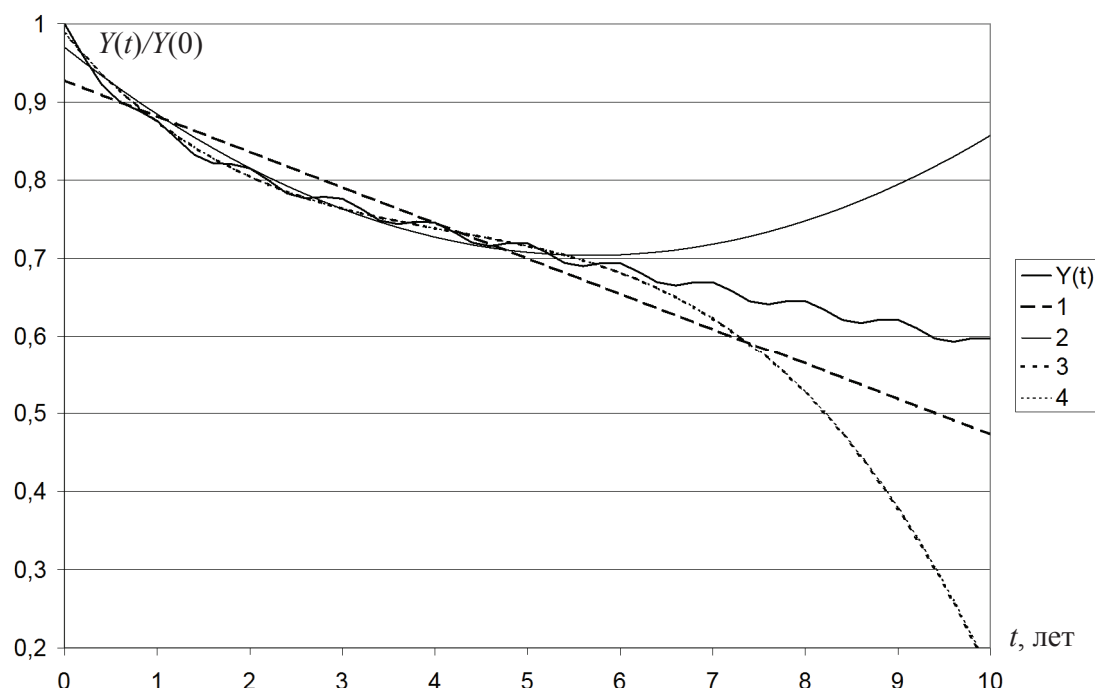


Рис. 4. Неудачные попытки экстраполяции численности популяции зависимостями: 1) $y(t)=1,291121-1,81922x(t)$; 2) $y(t)=2,245758-9,01517x(t)+13,17732x(t)^2$; 3) $y(t)=5,046407-40,798x(t)+131,2891x(t)^2-143,871x(t)^3$; 4) $y(t)=5,036115-40,6991x(t)+131,0587x(t)^2-144,083x(t)^3+0,87505x(t)^4$

$$\frac{dy(t)}{dt} = a_0 + \sum_{i=1}^m a_i x(t)^i + \sum_{j=1}^n b_j y(t)^j =$$

$$= a_0 + \sum_{i=1}^m a_i (0,2 + 0,025t)^i + \sum_{j=1}^n b_j y(t)^j$$

для $m=1,4$, $n=1,4$ и начального условия $y(0)=1$.
Развернем уравнения:

$$\frac{dy(t)}{dt} = a_0 + a_1(0,2 + 0,025t) + b_1 y(t)$$

$$\frac{dy(t)}{dt} = a_0 + a_1(0,2 + 0,025t) +$$

$$+ a_2(0,2 + 0,025t)^2 + b_1 y(t)$$

...

$$\frac{dy(t)}{dt} = a_0 + a_1(0,2 + 0,025t) + b_1 y(t) + b_2 y(t)^2$$

$$\frac{dy(t)}{dt} = a_0 + a_1(0,2 + 0,025t) +$$

$$+ a_2(0,2 + 0,025t)^2 + b_1 y(t) + b_2 y(t)^2$$

и т. д.

Используем следующий набор подпрограмм.

1. Линейная одномерная интерполяция для вычисления $x(t): t \notin \{t_j\}$
2. Численное решение обыкновенного дифференциального уравнения (ОДУ) методом Эйлера с небольшим шагом:

$$y(t + \Delta t) \approx y(t) + \Delta t \left(a_0 + \sum_{i=1}^m a_i x(t)^i + \sum_{j=1}^n b_j y(t)^j \right),$$

$$t \leq 10.$$

3. Вычисление суммы квадратов отклонений $S = \sum (y(t_i) - y^*(t_i))^2, t \neq 6$.
4. Оптимизация методом покоординатного спуска с переменным шагом $\Delta = \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$, уменьшающимся каждый раз, когда при предыдущем значении шага получен локальный минимум.

В результате подбора коэффициентов a_0, a_i, b_j путем решения задачи $S \rightarrow \min$, получили следующий набор решений (рис. 5, табл. 1).

Таблица 1. Оценка качества экстраполяционных свойств моделей на основе ОДУ

m	n	Итераций	S	y(10)	$\frac{ y(10) - y^*(10) }{y^*(10)} \cdot 100, \text{ в } \%$
0	0	17	0,11249	0,33933	43,24
0	1	8441	0,0003887	0,69345	15,98
1	0	13917	0,0090426	1,09493	83,14
1	1	70654	0,0000426	0,65809	10,07
1	2	66418	0,0000426	0,65809	10,07
2	1	58456	0,000452	0,65880	10,18
2	2	53054	0,000452	0,65880	10,18

Данные решения позволяют сделать достаточно адекватный прогноз развития системы (рис. 5). Важно даже не то, что его относительная погрешность меньше, чем раньше, а именно *адекватность*: система развивается *примерно* так, как получено в результате моделирования, тогда как на рис. 3 и 4 многие экстраполяции *неадекватны*. Кроме того, в последнем случае существует простой и понятный критерий выбора прогноза S, а в двух предыдущих случаях выбрать порядок интер-

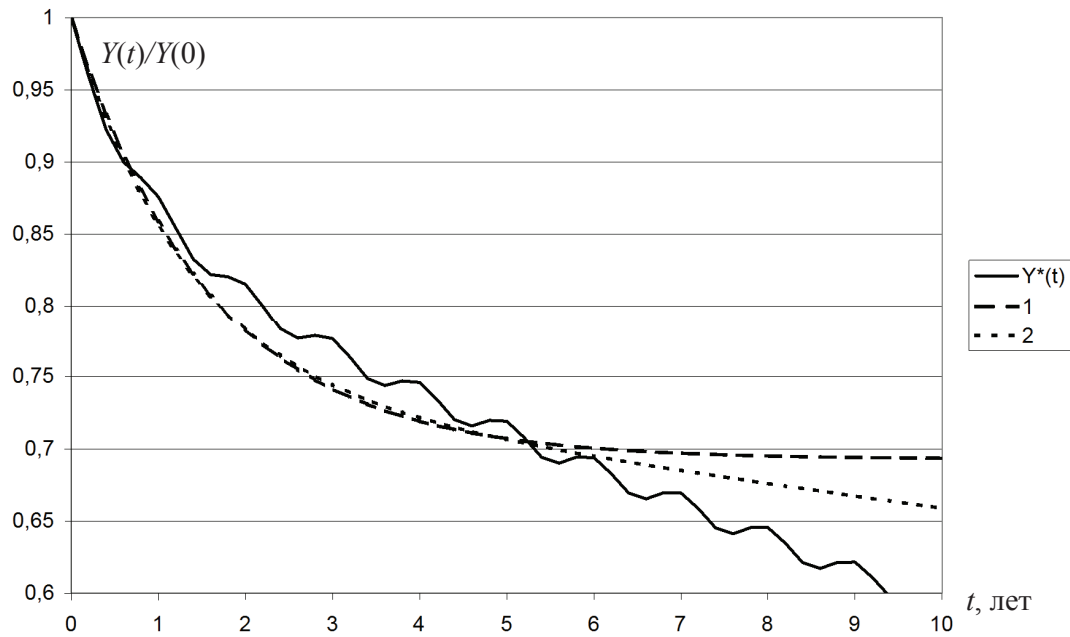


Рис. 5. Результаты экстраполяции на основе решения ОДУ без учета взаимного влияния состояния системы и управления: 1) $m=0, n=1$; 2) $m=1, n=1$

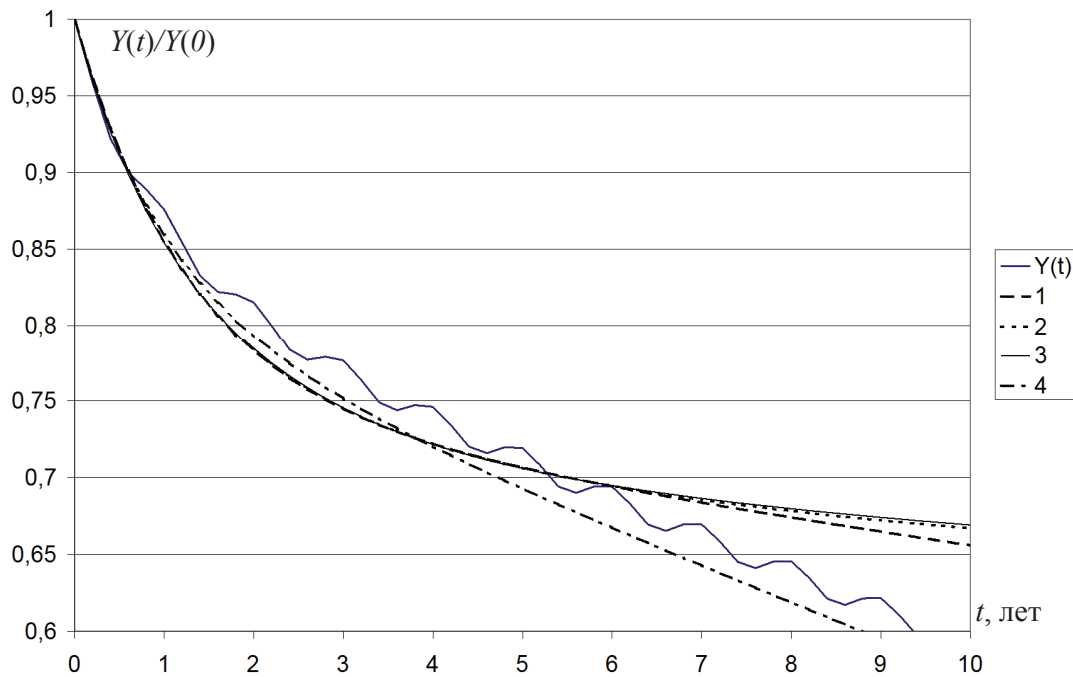


Рис. 6. Результаты экстраполяции на основе решения ОДУ с учетом взаимного влияния состояния системы и управления: 1) $n=1, m=1$; 2) $n=2, m=2$; 3) $n=1, m=2$; 4) точное решение

полирующего полинома, оценивая сумму квадратов отклонений, не удастся.

Рассмотрим модель, учитывающую взаимное влияние управления и состояния системы вида

$$\frac{dy}{dt} = \sum_{i=0}^m \sum_{j=0}^n a_{ij} x(t)^i y(t)^j.$$

Внесем соответствующие изменения в программу. Так, блок решения ОДУ примет вид

```
for (double t=0; t<tk; t+=dt) // tk - конечное время, 0...10
{
    tmp=0;
    x=xt (t);
    for (i=0; i<=im; i++)
        for (j=0; j<=in; j++)
            tmp += ak [i][j]*ipow (x, i)*ipow (y, j);
    y += dt*tmp;
}
```

```
if (y<0) y=0;
},
```

где $irow$ – функция возведения в целочисленную степень, xt – функция одномерной интерполяции.

Произведя численные эксперименты, получим результаты прогнозов на основании рассмотренных моделей (табл. 2).

Таблица 2. Оценка качества экстраполяционных свойств моделей на основе ОДУ с учетом взаимного влияния управления и состояния системы

m	n	S	$y(10)$	$\frac{ y(10) - y^*(10) }{y^*(10)} \cdot 100, \text{ в } \%$
0	0	0,11249	0,33933	43,24
1	1	0,0000369	0,65583	9,69
1	2	0,0000031	0,66706	11,57
2	2	0,0000131	0,66906	11,91

СПИСОК ЛИТЕРАТУРЫ

1. Дзюба С.А. Модели управления подсистемами предприятия в сфере среднего бизнеса и их инструментальное обеспечение: автореф. дис. ... д-ра экон. наук. – Иркутск, 2011. – 46 с.
2. Мицек Е.Б. Эконометрическое моделирование инвестиций в основной капитал экономики России: автореф. дис. ... д-ра экон. наук. – Екатеринбург, 2011. – 49 с.

Интересно, что программа не находит в качестве оптимального верное решение $a_{01}=1$, $a_{02}=a_{11}=-1$ при всех остальных $a_{ij}=0$. Квадратичная ошибка в этом случае $S=0,035783$, хотя погрешность прогноза $y(10)=0,57121$ составляет всего 4,46 %. Адекватность моделей сохраняется во всех случаях (рис. 6).

Заключение

Нельзя считать, что описанные неудачные попытки *доказывают* невозможность удачных аппроксимаций и экстраполяций временных трендов в эколого-экономическом моделировании с использованием традиционных и широко распространенных методов. Однако это хорошая иллюстрация того, что использование в качестве основы моделей обыкновенных дифференциальных уравнений может привести к качественному росту прогнозов и, следовательно, принимаемых решений.

3. Миролюбова А.А. Методология моделирования инвестиционного процесса в реальном секторе экономики региона: автореф. дис. ... д-ра экон. наук. – Иваново, 2012. – 33 с.
4. Лосев К.С. Мифы и заблуждения в экологии. – М.: Научный мир, 2010. – 224 с.
5. Арнольд В.А. Теория катастроф. – М.: Наука, 1990. – 128 с.

Поступила 26.03.2012 г.

Наши юбиляры

ПРОФЕССОРУ В.К. ПОГРЕБНОМУ – 70 ЛЕТ



Владимир Кириллович Погребной родился 24 января 1942 г. в с. Шерстобитово Парабельского района Томской области. После окончания в 1959 г. Пудинской средней школы три года работал на угольном разрезе в г. Прокопьевске Кемеровской области. В 1962 г. поступил на электромеханический факультет ТПИ и в 1967 г. с отличием окончил его по специальности «Электропривод и автоматизация промышленных установок».

Вся последующая трудовая деятельность В.К. Погребного связана с ТПУ. По распределению работал старшим лаборантом на кафедре электропривода, а в 1969 г. был переведен на вновь образованную кафедру АСУ (в настоящее время кафедра ОСУ). В 1969 г. поступил в аспирантуру ТПИ и через год в 1970 г. защитил кандидатскую диссертацию по теме: «Разработка и исследование алгоритмов автоматизации технического этапа проектирования вычислительных устройств». Научным руководителем по дипломной работе и кандидатской диссертации был В.З. Ямпольский.

В 1981 г. доцент В.К. Погребной был переведен на кафедру инженерной и вычислительной математики и избран на должность заведующего кафедрой. В 1982 г. кафедра была преобразована в выпу-

скающую кафедру автоматизации проектирования (АП), а в 2000 г. переименована в кафедру информатики и проектирования систем (ИПС). В должности заведующего кафедрой ИПС В.К. Погребной проработал до 2007 г. В настоящее время работает профессором этой кафедры. Докторскую диссертацию защитил в 1984 г. по теме «Методы построения и оптимизации модульных структур при проектировании систем реального времени».

Первые работы по автоматизации проектирования в ТПИ стали развиваться, начиная с 1967 г., когда по заказу СКБ при Томском заводе математических машин была заключена хозяйственная НИР «Автоматизация некоторых этапов проектирования вычислительных устройств». У истоков работ по постановке и выполнению этой темы стояли молодой ученый к.т.н. В.З. Ямпольский и дипломник В.К. Погребной. В дальнейшем развитие работ этого направления осуществлялось под руководством В.К. Погребного. Среди многих задач, возникающих при проектировании вычислительных устройств, особо следует выделить задачу, которая во многом определила всю научную работу В.К. Погребного. Речь шла о создании функционально полного набора модулей, из которых путем

композиции можно получать требуемое вычислительное устройство.

В последующие годы стало ясно, что принцип модульности может не менее эффективно использоваться при проектировании математического и программного обеспечения управляющих систем. К этому времени в составе Вычислительного центра ТПИ под руководством В.К. Погребного был организован научный отдел автоматизации проектирования. В 1981–1985 гг. по программе О.Ц.026 ГКНТ под руководством В.К. Погребного выполнялась НИР с целью создания и введения в действие экспериментальной автоматизированной системы проектирования математического и программного обеспечения АСУ ТП. Работа завершилась созданием новой технологии проектирования и ее инструментального комплекса САМПР-СРВ. Система в 1984 г. была сдана Госкомиссии, передана в Госфонд алгоритмов и программ, внедрена в ряде организаций страны.

В этот период кафедра имела тесные научные и хозяйственные связи с Институтом проблем управления (г. Москва), Центральным НИИ комплексной автоматизации (г. Москва), НИИ автоматики (г. Киев), Институтом кибернетики (г. Киев), СКБ «Промавтоматика» (г. Омск), рядом других организаций и ведущих вузов страны. Разработанная в эти годы «Система автоматизации модульного проектирования микропрограмм для микроЭВМ «Электроника-60» и микропроцессора К-580» удостоена в 1983 г. диплома Всероссийской выставки «Использование микропроцессоров для управления системами».

Существенное развитие научные исследования на кафедре получили с 1991 г., когда КЦ при ТПИ было поручено выполнение функций координации по разделам двух инновационных программ: «Государственной научно-технической программе», «Технологии, машины и производства будущего», раздел «Средства вычислительной техники и информатики, автоматизированные производственные системы и программно-технические комплексы» и научно-технической программе «Трансферные технологии, комплексы и оборудование», раздел «Программные системы». Научно-технические советы по данным разделам возглавил В.К. Погребной. Работа по указанным программам позволила значительно укрепить техническую базу научных лабораторий и кафедры ИПС и поставить научные исследования на более высокий уровень.

После перехода в 2007 г. на должность профессора появилось больше времени для выполнения научных исследований. Основная научная работа связана с решением проблемы формализации процесса эволюционного проектирования распределенных систем реального времени (СРВ). Другие направления исследований воспринимаются Владимиром Кирилловичем как хобби. Одно из них связано с решением проблемы получения компактных разбиений, когда на некоторой территории расположено множество объектов и нужно разбить его на компактные подмножества. В другом исследуется проблема получения инвариантного описания структур и анализа их сходства

на основе этих описаний. В научной работе участвуют сыновья, особенно по вопросам алгоритмизации и программирования.

Владимиру Кирилловичу вместе с женой Татьяной Корнильевой удалось создать большую и дружную семью: три сына, три внука и две внучки. Старший сын Дмитрий работает в программистской фирме и по совместительству доцентом ТПУ, средний сын Александр работает доцентом ТПУ, а младший Андрей учится на 3-м курсе Института кибернетики ТПУ.

Научная работа В.К. Погребного тесно связана с учебным процессом. Многие из дисциплин, которые он вел, включали результаты научных исследований. Первыми лекционными курсами были: «Основы технической кибернетики» и «Исследование операций». В последующем были разработаны и многие годы читались курсы: «Теория графов», «Математические методы в экономике», «Математические модели объектов проектирования», «Теоретические основы САПР», «Проектирование вычислительных систем». В настоящее время разработаны и читаются лекции по дисциплинам: «Моделирование автоматизированных систем», «Автоматизированное проектирование распределенных систем реального времени». В.К. Погребной опубликовал более 140 работ. Издано 8 учебных пособий. Все они написаны по результатам научных исследований. Два последних изданы с грифом УМО: «Системы реального времени. Моделирование и автоматизированное проектирование» (2007 г.), «Автоматизированное проектирование распределенных систем реального времени» (2011 г.).

В 2012 г. по предложению Международного издательского дома LAP LAMBERT Academic Publishing (Германия) В.К. Погребной с сыновьями Андреем и Дмитрием опубликовали в этом издательстве монографию «Уравнения динамики функционирования СРВ, представленной на языке SML», которая излагает основные положения языка структурного моделирования (Structural Modeling Language – SML) и соответствующей SML-технологии построения, оптимизации и эволюции модели СРВ. Эволюционный поиск приемлемого варианта модели СРВ в SML-технологии выполняется с помощью рекуррентных уравнений, описывающих в матричной форме динамику смены состояний модели. По уравнениям динамики строятся траектории работы системы, оценивается их качество и принимаются решения по внесению изменений в модель.

Профессор В.К. Погребной, доктор технических наук, Заслуженный работник высшей школы Российской Федерации, действительный член Международной академии информатизации, член-корреспондент Академии наук высшей школы, член Редколлегии журнала «Известия Томского политехнического университета». В.К. Погребной пользуется заслуженным авторитетом и уважением в коллективе кафедры и ТПУ, награжден серебряной медалью «За заслуги перед Томским политехническим университетом» и медалью II степени «За участие в развитии Томского политехнического университета».

ПРОФЕССОРУ А.М. КОРИКОВУ – 70 ЛЕТ



1 января 2012 г. отметил юбилей доктор технических наук, профессор, Заслуженный деятель науки РФ, Почетный работник высшего профессионального образования РФ и Почетный работник науки и техники РФ Анатолий Михайлович Кориков.

Трудовая биография Анатолия Михайловича началась сразу после окончания средней школы. С июня 1959 г. он стал работать в г. Каргалы Челябинской области монтером связи в 8-й дистанции сигнализации и связи Южно-Уральской железной дороги (ШЧ-8 Ю-Ур.ж.д.). Вскоре А.М. Кориков был направлен в г. Челябинск на курсы электромехаников СЦБ и связи Челябинской технической школы машинистов локомотивов (ЧТШМЛ) Южно-Уральской железной дороги. После окончания ЧТШМЛ продолжил работу электромехаником связи в ШЧ-8 Ю-Ур.ж.д. до поступления в 1961 г. в Томский государственный университет (ТГУ) на радиопизический факультет, который окончил в декабре 1966 г.

После окончания ТГУ с 1 января 1967 г. А.М. Кориков работал младшим научным сотрудником Сибирского физико-технического института (СФТИ) при ТГУ, а в сентябре 1967 г. стал аспирантом Томского государственного университета. Через 3 года в сентябре 1970 г. представил в диссертационный совет ТГУ диссертацию на соискание ученой степени кандидата физ.-мат. наук, которую защитил в ноябре 1970 г. В сентябре того же 1970 г. начал работать на только что открытом факультете прикладной математики ТГУ сначала в должности ассистента, а затем – доцента. В ТИ-

АСУР А.М. Кориков перешел в сентябре 1972 г. доцентом на кафедру ОАСУ, а в феврале 1976 г. был направлен зав. кафедрой ОАСУ В.П. Тарасенко в СФТИ при ТГУ заведовать научной лабораторией АОС для завершения летних испытаний корреляционно-экстремального координатора (КЭК). Успешные испытания КЭК стали важнейшим аргументом для присуждения В.П. Тарасенко Государственной премии РФ.

В июне 1983 г. А.М. Кориков защитил в диссертационном совете ТПУ докторскую диссертацию и в том же году ему присуждена ученая степень доктора технических наук. В ТИАСУР Анатолий Михайлович вернулся в 1984 г. на должность зав. кафедрой АСУ (апрель 1984 г.) и проректора по НР (август 1984 г.). В должности проректора по НР А.М. Кориков работал до 1 июля 1991 г. В этот период он отдает много сил научно-организационной работе. При его активном участии (А.М. Кориков – зам. руководителя программы и руководитель одного из 5 разделов программы) сформирована и успешно выполнена региональная межвузовская научно-техническая программа «Автоматизация», в которой в 12-й пятилетке участвовали все томские вузы и НИИ. В этот период много внимания Анатолий Михайлович уделяет поиску и реализации новых форм интеграции науки и производства, в частности в те же годы была сформирована и выполнена комплексная программа Минвуза и Минприбора «Контур», по которой он был зам. руководителя программы и руководителем направления.

А.М. Кориков более 25 лет возглавляет кафедру АСУ ТУСУР и известен в научном сообществе как один из лидеров научного направления интеллектуальные автоматизированные системы обработки информации и управления (АСОИУ). По этому направлению Анатолий Михайлович опубликовал более 300 научных работ, в том числе 10 монографий. За последние 5 лет по интеллектуальным АСОИУ им опубликовано более 50 научных работ и 2 монографии. По направлению АСОИУ профессор А.М. Кориков подготовил 6 докторов и более 25 кандидатов технических наук, некоторые из них были подготовлены для работы в ТПУ: д.т.н. П.Г. Яковенко работал на кафедре АТП ЭНИН, к.т.н. М.В. Момот возглавляет кафедру ЭиАСУ ЮТИ, к.т.н. Г.О. Тащиян работает доцентом на кафедре ИС ЮТИ.

Профессор А.М. Кориков разработал и прочитал в ТГУ, ТПУ и ТУСУР более 20 учебных дисциплин. В настоящее время он читает для студентов кафедр АиКС и ВТ Института кибернетики ТПУ учебный курс «Теория информационных процессов и систем».

За последние 5 лет под руководством А.М. Корикова выполнены проект по ФЦП «Интеграция», две госбюджетных и две хоздоговорных НИР, в 2008 г. им выигран грант РФФИ (№ 09-07-99029-р_офи); он также был исполнителем трех грантов РФФИ и соруководителем интеграционного проекта СО РАН. В 2010 г. под руководством Анатолия Михайловича выигран грант на НИР по ФЦП «Научные и научно-педагогические кадры инновационной России» (государственный контракт № 14.740.11.0398).

Известна общественная деятельность А.М. Корикова: 15 лет (1984–1985 гг.) – Председатель совета Дома ученых г. Томска; 19 лет (1992–2011 гг.) – главный ученый секретарь Сибирского отделения Академии наук высшей школы РФ (избран действительным членом Академии в 1993 г.); член первого состава Правления Томского профессорского собрания; в настоящее время – член Правления ТРОО «Знание»; член редколлегии научных журналов и оргкомитетов научных конференций. Приглашался в ноябре 2007 г. в качестве визит-профессора в Сибирский федеральный университет. Он – член УМК по специальности «Программное обеспечение вычислительной техники и автоматизированных систем» и член УМО по специальности «Прикладная информатика (по областям)». С 1984 г. профессор А.М. Кориков работал в нескольких диссертационных советах ТПУ, ТГУ и ТУСУР. До 2000 г. он был членом семи диссертационных советов. В последние годы А.М. Кориков – член двух диссертационных советов в ТУСУР и двух – в ТПУ.

За высокие достижения в науке и образовании в 1995 г. в составе научно-педагогического коллектива ТУСУРа А.М. Кориков стал лауреатом конкурса Томской области в сфере образования и науки, в 2002 г. он получил профессорскую стипендию ЮКОС, в 2005 г. в составе научно-педагогического коллектива, возглавляемого директором КЛИО Е.П. Гордовым, Анатолий Михайлович отмечен повторно премией конкурса Томской области в сфере образования и науки.

ПРОФЕССОРУ А.А. МИЦЕЛЮ – 65 ЛЕТ



Артур Александрович Мицель родился 21 мая 1947 г. в с. Жешарт Устьвинского района Коми АССР (ныне Республика Коми) в семье рабочих. В 1956 г. семья переехала в г. Бийск.

После семилетки учился в Бийском механико-технологическом техникуме по специальности «механик-энергетик элеваторной промышленности», который закончил в 1965 г. и по распределению уехал работать в Курганскую область электриком на элеваторе. В 1966 г. был призван в ряды Советской Армии. В 1969 г. демобилизовался и поступил на радиопизический факультет Томского государственного университета, который закончил в 1974 г. по специальности «Оптико-электронные приборы».

Научной работой Артур Александрович начал заниматься со второго курса, с 4-го курса работал по совместительству в Институте оптики атмосферы СО РАН (ИОА СО РАН), г. Томск, куда и был распределен после окончания университета. С 1975 по 1995 г. А.А. Мицель работал в ИОА СО РАН, начиная со стажера – исследователя (1975 г.) и завершая карьеру академического ученого ведущим научным сотрудником (1995 г.). В 1982 г. защитил кандидатскую диссертацию на тему «Численное моделирование лазерного зондирования газового состава атмосферы» в диссертационном Совете Томского государственного университета. В 1993 г. защитил докторскую диссертацию на тему «Прямые и обратные задачи абсорбционной газовой спектроскопии атмосферы». Диссертация защищалась

в диссертационном совете ИОА СО РАН по специальности 01.04.05 – оптика. В 2000 г. А.А. Мицель присвоено ученое звание профессора.

Работая в ИОА СО РАН, он принимал активное участие в выполнении хозяйственных и госбюджетных НИР в качестве ответственного исполнителя. По результатам этих работ А.А. Мицель в составе авторского коллектива сотрудников ИОА СО РАН был включен в энциклопедию «Лазерное оружие России». А.А. Мицель был одним из исполнителей международного проекта ISIREM, выполняемого в ИОА СО РАН в 2000–2002 гг.

С сентября 1995 г. по настоящее время Артур Александрович работает на кафедре автоматизированных систем управления (АСУ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР) в должности профессора.

А.А. Мицель выполняет обязанности заместителя заведующего кафедры АСУ ТУСУР и ответственного за НИРС. В 1998 г. он создал филиал кафедры АСУ при ИОА СО РАН, которым руководил по 2004 г. В 2003 г. А.А. Мицель создал лабораторию «Мультимедиа» в ТУСУРе и возглавлял ее до 2007 г.

С 2001 г. Артур Александрович работает по совместительству в Юргинском технологическом институте Томского политехнического университета (ТПУ) на кафедре информационных систем, а с 2004 г. – на кафедре высшей математики и математической физики ТПУ.

Научно-педагогический стаж составляет 37 лет, из них 17 лет – педагогический.

Научную карьеру А.А. Мицеля определили два человека – академик АН СССР Владимир Евсеевич Зуев и профессор Юрий Семенович Макушкин. Они были и наставниками, и научными руководителями. Слова Артура Александровича своим наставникам звучат так: «Я им бесконечно благодарен и горжусь тем, что судьба свела меня с ними. Я благодарен еще одному человеку, с которым довелось мне встретиться в 1995 г., – Анатолию Михайловичу Корикову, заведующему каф. АСУ ТУСУР, где я работаю в настоящее время».

За период работы с 1975 по 2011 гг. А.А. Мицель подготовил 21 кандидата и 3 докторов наук.

Всего Артуром Александровичем опубликовано в соавторстве 8 монографий, более 270 научных работ, из которых 174 статьи в центральной печати (в том числе 15 статей в зарубежной), три препринта, получено 4 авторских свидетельства на изобретения, издано 27 учебных и учебно-методических пособий, разработаны и внедрены несколько десятков пакетов прикладных программ для научных исследований, в том числе 4 пакета по международным контрактам (США, Южная Корея, Болгария), создано 12 мультимедийных учебника по вузовским дисциплинам.

Основным направлением исследований А.А. Мицеля является математическое и численное моделирование переноса оптического излучения в газовой атмосфере, лазерного газоанализа, оптического пассивного зондирования газового состава атмосферы наземными и аэрокосмическими приборами. Артур Александрович с 1979 по 1995 гг. руководил тематической группой в ИОА СО РАН в составе 8–10 человек. А.А. Мицель является ведущим специалистом в области прикладной абсорбционной спектроскопии газовой атмосферы. По этому направлению им опубликовано в соавторстве 3 монографии (последняя вышла в печати в 2001 г.), более 100 статей в центральной печати, из них 15 статей в зарубежной печати и подготовлено 11 кандидатов наук и 3 доктора наук. Кроме того, еще два его ученика из ИОА СО РАН защитили докторские диссертации (К.М. Фирсов, И.В. Пташник).

Двое из его учеников стали Лауреатами областного конкурса среди молодых ученых и специалистов (И.В. Пташник, 1997 г.; И.В. Бойченко, 2007 г.).

С приходом в ТУСУР наряду с учебным процессом А.А. Мицель продолжает активно заниматься научной работой, вовлекая в эту деятельность студентов и аспирантов. В рамках проекта «Интеграция науки и образования» Артур Александрович организовал в 1998 г. филиал кафедры АСУ при Институте оптики атмосферы СО РАН, где студенты получали дополнительное образование, выполняя УИРС как на кафедре, так и в стенах Института оптики атмосферы и Института мониторинга климатических и экологических систем

СО РАН. В рамках филиала кафедры АСУ А.А. Мицель обеспечивал внедрение достижений ученых ИОА СО РАН в учебный процесс ТУСУР. Его студенты регулярно участвовали в различных студенческих и молодежных конкурсах. Они неоднократно получали призовые награды во Всероссийских конкурсах студенческих работ (1998 г. – 1 медаль и 1 Диплом второй степени; 1999 г. – 1 медаль и 1 Диплом первой степени, 1 Диплом второй степени; 2000 г. – 1 Диплом первой степени, 2001 г. – 1 медаль и 3 диплома Минобразования), 2003 г. – два диплома Минобразования, 2006 г. – диплом Минобразования.

В 1999 г. его аспирант М.Н. Еременко получил стипендию Фонда Президента России для зарубежной стажировки (Парижский университет Пьера и Мари Кюри). Один из его студентов филиала каф. АСУ (А.С. Артамонов) получил Золотую медаль 12-го конкурса Российской академии наук (2000 г.). Трое его аспирантов, прошедших обучение в филиале каф. АСУ получили звание «Соросовский аспирант» (Е.С. Артамонов, А.В. Миляков, И.Г. Окладников, 2001 г.), еще трое аспирантов стали стипендиатами НК «ЮКОС» (А.В. Миляков, 2002 г.; В.В. Романенко, 2003 г.; В.В. Клыков, 2004 г.).

А.А. Мицель в 1997 г. организовал на кафедре постоянно действующий студенческий научный семинар «Информационные системы в мониторинге окружающей среды», которым руководил по 2007 г. Артур Александрович вел также активную работу по зарубежному сотрудничеству. Его аспирант проходил стажировку в парижском Университете Пьера и Мари Кюри, занимаясь разработкой программного обеспечения и моделированием по проблеме исследования газового состава атмосферы аэрокосмическими приборами.

А.А. Мицель в 1999 г. организовал совместную российско-французскую аспирантуру (ТУСУР – Университет Пьера и Мари Кюри), была выделена аспирантская стипендия французским посольством. Один из аспирантов прошел стажировку в 2002 г. в Центре по энергии и окружающей среде (Е.С. Артамонов Италия). В 2003 г. прошел трехмесячную стажировку в этом же центре другой аспирант – Н.В. Колодникова). Докторант А.А. Мицеля (М.Ю. Катаев, ныне доктор наук) успешно прошел годичную стажировку в японском научном центре исследования окружающей среды (1999–2000 гг.).

Второе научное направление, которым занимается А.А. Мицель, связано с автоматизацией разработки компьютерных обучающих программ и разработкой средств языкового анализа и интерпретации, созданием компьютерных тренажеров. С 2000 по 2007 гг. руководил студенческим семинаром «Автоматизированные системы в учебном процессе». В 2001 г. Артур Александрович создал творческий коллектив для разработки мультимедийных учебников, выполняя проекты Томского межвузовского центра дистанционного образования при ТУСУРе, а в 2003 г. создал лабораторию

«Мультимедиа». В конкурсах ТУСУРа по КУМПО (комплексного учебно-методического и программного обеспечения) коллектив А.А. Мицеля неоднократно занимал призовые места (2004 г. – 1 место, 2006 г. – 2 место, 2008 г. – 2 место, 2010 г. – 1 место, 2011 г. – 1 место). Мультимедийный учебник «Концепции современного естествознания» в составе КУМПО, представленного Новосибирской государственной академией экономики и управления на Сибирской ярмарке 2002 г. в г. Новосибирске получил Большую золотую медаль. За серию электронных обучающих курсов, включающую курсы по дисциплинам «Концепции современного естествознания» и «Вычислительная математика» получена Малая золотая медаль конкурса УЧСИБ–2004. За разработку этой серии курсов получен диплом в номинации «Электронный учебник года» конкурса УЧСИБ–2004.

Один из его учеников стал лауреатом конкурса Томской области в сфере образования, науки, здравоохранения и культуры (В.В. Романенко, 2006 г.).

По этому направлению Артуром Александровичем подготовлено 3 кандидата наук, опубликовано 14 статей, из них 12 статей в журналах из перечня ВАК, издана монография (2005 г.). В целом по этому направлению коллективом лаборатории «Мультимедиа» создано 12 мультимедийных учебника по вузовским дисциплинам, опубликовано около 40 статей в центральной печати.

Третье научное направление А.А. Мицеля связано с моделированием объектов прикладной экономики, рынка ценных бумаг, созданием моделей принятия решений социально-экономического развития муниципальных образований. В течение двух лет (1997–1999 гг.) он руководил студенческим семинаром «Информационные системы отдела ценных бумаг» при филиале коммерческого банка «Кузбассоцбанк». В 2006 г. Артур Александрович организовал студенческий семинар «Моделирование экономических процессов и систем».

Работа одного из его учеников (А.Н. Вагдаев, 2005 г.) была отмечена стипендией Международного научного фонда экономических исследований академика Н.П. Федоренко, двое из его учеников стали лауреатами конкурса Томской области в сфере образования, науки, здравоохранения и культуры (А.А. Захарова, 2007 г.; Т.Ю. Чернышева, 2009 г.).

По этому направлению подготовлено 7 кандидатов наук, опубликовано 4 монографии (2010,

2011 гг.), в том числе одна в международном издательстве, 43 статьи, в том числе 29 статей в журналах из перечня ВАК.

Награды и почетные звания

- В 1978 г. – Лауреат областного конкурса молодых ученых и специалистов.
- В 1995 г. награжден нагрудным знаком «Ветеран СО РАН».
- В 2001 г. получил нагрудный знак Министерства образования РФ «За развитие НИРС».
- В 2002 г. стал лауреатом конкурса Томской области в сфере образования и науки.
- В 2004 г. награжден медалью им. М.В. Келдыша Федерации космонавтики.
- В 2008 г. избран академиком Международной академии наук Высшей школы (МАН ВШ).
- В 2010 г. получил Премию губернатора Кемеровской области.
- В 2011 г. получил благодарственное письмо департамента науки и образования Кемеровской области.
- В 2011 г. присвоено звание «Почетный работник высшего профессионального образования».

Общественная деятельность

- С 2002 по 2003 гг. был экспертом Российского Фонда фундаментальных исследований по разделу «Информационно-вычислительные и телекоммуникационные ресурсы».
- С 2004 г. по настоящее время является экспертом-аудитором по общественно-профессиональной аккредитации образовательных программ вузов.
- Является членом Ученого совета ТУСУРа.
- А.А. Мицель входит в состав трех диссертационных советов по специальностям: 05.13.18 – математическое моделирование, численные методы и комплексы программ; 01.04.05 – оптика; 05.11.19 – методы и системы защиты информации, информационная безопасность.

А.А. Мицель пользуется заслуженным авторитетом среди коллег и студентов. Коллективы кафедры АСУ ТУСУР, кафедры информационных систем ЮТИ ТПУ и кафедры высшей математики и математической физики ТПУ поздравляют Артура Александровича с юбилеем, желают здоровья, удачи во всех его начинаниях и дальнейших успехов на благо университетов.

Summaries

UDC 004.8

Rudometkina M.N.
**THE DEVELOPMENT OF BINARY
DECOMPOSITION METHOD FOR FUNCTION PREDICATES**

The author analyzed the only sound binary decomposition method of finite predicates, the Descartes decomposition method. For the first time the author developed and substantiated the partial binary decomposition method for the function predicates based on the analysis of dependences; a number of values of the auxiliary variable was minimized. The algorithm of binary decomposition for function predicates was formed based on the proposed method.

UDC 519.872

Nazarova A.A., Semenova I.A.
**ASYMPTOTIC ANALYSIS OF QUEUING SYSTEMS
WITH INFINITE NUMBER OF SERVERS
AND SEMI-MARKOVIAN INPUT FLOW**

The article considers the queuing system with input SM flow and infinite number of servers. The investigation is carried out by the method of sifted flow and asymptotic analysis method at increasing service time. The carried out analysis allows obtaining asymptotic semi-invariants to the third order.

UDC 519.9

Nebaba S.G., Vylegzhanin O.N.
**CONSTRUCTION AND STUDY OF SEARCH ALGORITHM
OF THE FIRST EXTREME SUBSYSTEM FOR A GIVEN
CONSISTENT SYSTEM OF LINEAR INEQUALITIES**

The search algorithm of extreme subsystem for the given linear inequalities system was developed. The range of permitted values for the system was specified and the method for searching the dependent inequalities was suggested. The algorithm includes the accounting of possible equality constrains as well as reduction of initial inequalities matrix to full columns rank. The availability of suggested method was shown on the test case.

UDC 004.94

Pogrebnoy An.V., Pogrebnoy D.V.
**MATRIX ALGORITHM FOR SOLVING
THE PROBLEM OF CUTTING BIPARTITE GRAPHS**

The article introduces the algorithm for solving the problem of cutting the bipartite graphs into prescribed population of minimally connected subgraphs. The algorithm uses matrix representation of bipartite graph and considers its specific character. It allows reducing the cutting problem to the problem of linear mathematical programming. The algorithm may work both with the improved original version of cutting the bipartite graph and with the arbitrarily formed one.

UDC 519.673

Khamukhin A.A.
**ADDRESSABLE CELL OF HOMOGENEOUS COMPUTING
STRUCTURE FOR SOLVING PARTIAL
DIFFERENTIAL EQUATIONS**

The author proposes the addressable cell function diagram of homogeneous computing structure for solving the partial differential equations. The article demonstrates the possibility of solution speedup decreasing the duration of period of data load into cell registers. The article introduces the comparative results of modeling the line of addressable and non-addressable cells by the example of solving the test boundary problem with independent and dependent coefficients.

UDC 004.42

Botygin I.A., Popov V.N., Tartakovskiy V.A.
**ARCHITECTURE OF DISTRIBUTED
COMPUTING SYSTEM FOR TWO-DIMENSIONAL
ANALYSIS OF TREE DISK IMAGE**

The authors developed the architecture and algorithm of implementing the distributed computing system for processing dendroecological data. The article introduces the GPSS-model of the system for estimating the efficiency of its operation. Practical implementation of the system was carried out using the stack of free LAMP software products. The authors illustrated the system operation by such problems of two-dimensional analysis of tree disk image as computing azimuth and mean-square width of maximum increment area as well computing the values of indices of tree ring width increase.

UDC 004.4:004.89

Le Hoai, Tuzovskiy A.F.
USING ONTOLOGY IN ELECTRONIC LIBRARIES

The authors consider the use of ontologies in semantic digital libraries, give their definition and purpose. The types of ontologies in such libraries including knowledge organization systems and structural taxonomy are analyzed. The version of selecting ontologies to develop semantic digital libraries is substantiated.

UDC 681.3.06

Nguyen Ba Ngoc, Tuzovskiy A.F.
**TEXT CLASSIFICATION BASED ON ESTIMATION
OF TERMS SEMANTIC SIMILARITY**

The article considers the method for increasing text classification accuracy by the kNN algorithm applying the estimation of semantic similarity based on terms co-occurrence matrix. The authors propose the method for decreasing the size of co-occurrence matrix filtering terms by pats of speech. The influence of filtration technique on classification accuracy is tested.

UDC 004.931

**Nguyen Toan Thang, Spitsyn V.G.
ALGORITHMS AND SOFTWARE
FOR HAND SHAPE ON-LINE RECOGNITION
USING SURF DESCRIPTORS
AND NEURAL NETWORK**

The authors developed the unique algorithm for hand shape on-line recognition based on SURF descriptors and neural network. A new method of descriptor generation for the neural network was proposed. The software for hand shape on-line recognition was developed based on the proposed algorithm. The numerical experiments on hand shape on-line video sequence recognition at showed that the average accuracy of recognition amounts to 92 %.

UDC 004.932

**Bui Thi Thu Trang, Phan Ngoc Hoang, Spitsyn V.G.
FACE RECOGNITION BASED ON APPLICATION OF
VIOLA-JONES METHOD, WAVELET TRANSFORM
AND PRINCIPAL COMPONENT ANALYSIS**

Based on co-application of Viola-Jones method, wavelet transform and principal component analysis the authors have proposed a new algorithm for face in real time recognition on digital images and video sequences. The algorithm for face recognition is described and software is developed. The article introduces the example of operation and test results of the program. It is shown that the use of the proposed unique algorithm allows recognizing faces efficiently on digital images and video sequences.

UDC 004.932

**Phan Ngoc Hoang, Spitsyn V.G.
ALGORITHMS FOR FINGERPRINT CLASSIFICATION BASED
ON APPLICATION OF GABOR FILTER, WAVELET
TRANSFORM AND MULTILAYER NEURAL NETWORK**

The authors propose the algorithms for fingerprint classification by the types of papillary pictures based on application of Gabor filter, Haar, Daubechies wavelet transform, and multilayer neural network. The numerical experiments are carried out and the proposed algorithms operation results are introduced. It is shown that the use of the algorithm based on co-application of Gabor filter, five-level Daubechies wavelet transform and multilayer neural network allows classifying fingerprints efficiently.

UDC 004.931

**Druki A.A.
ALGORITHMS OF PERSON SELECTION
ON STATIC RGB IMAGES AND IN VIDEO STREAM**

The authors developed the algorithms of person selection on static images and in video stream: algorithm of person detection by color segmentation, algorithm of person selection on static RGB images by deformable elliptical models, the method of static moments for person selection in video stream.

UDC 004.93

**Sidorov D.V., Osokin A.N.
SIMPLE ALGORITHM FOR CONTROLLING BIT
RATE OF JPEG CODER**

The authors developed the algorithm for controlling the compression ratio of half-tone accurate images for JPEG coder differing in software implementation simplicity and low computational complexity. It was tested on the set of half-tone images Calgary Corpus Gray Set 2. The conclusions were made on serviceability of the proposed algorithm in video surveillance and video recording systems at limitation of computational resource and data throughput.

UDC 519.71+004

**Zakharova A.A., Shklyar A.V.
CONSTRUCTION OF MULTICOMPONENT
VISUAL 3D-MODELS USING HETEROGENEOUS
DATA SOURCES BY THE EXAMPLE
OF DEVELOPING GEOLOGICAL MODELS**

The article demonstrates the necessity for developing and improving data visual models, methods of modeling and their using. The authors introduce the approach to construction of multicomponent visual 3D-models allowing testing, interpreting and analyzing spatial information by the example of geological models.

UDC 550.8.053

**Ivanchenkov V.P., Kochegurov A.I., Orlov O.V.
STUDYING THE RESOLVING ABILITY
IN THE TECHNIQUES OF SEISMIC SIGNAL
PHASE-FREQUENCY TRACING**

The article introduces the results of studying the resolving ability in the techniques of seismic signal phase-frequency tracing by the models of wave fields and real data. It is shown that the techniques of phase-frequency tracing of signals with equilibrium and nonequilibrium processing possess high resolving ability and allow selecting weak reflections in a wave field. The results of field material processing prove the conclusions of mathematical modeling and practical effect of the developed techniques.

UDC 681.5

**Pushkarev M.I., Gayvoronskiy S.A.
PARAMETRIC SYNTHESIS OF PI-CONTROLLER
IN LINEAR AUTOMATED CONTROL SYSTEM BASED
ON COEFFICIENT ESTIMATIONS OF STABILITY DEGREE
AND PRESCRIBER Q-FACTOR**

The author developed the technique for selecting the settings of linear proportional-integral controller providing quasi-maximum stability degree and prescribed accuracy of the automated control system in steady static mode. The technique is based on the coefficient estimations of quality indices in stationary systems and sufficient conditions of the prescribed stability degree based on the latter.

UDC 681.5.015

**Rudnitskiy V.A., Alekseev A.S., Kurgankin V.V.
IDENTIFICATION OF CONTROLLED OBJECTS IN THE FORM
OF SAMPLED-DATA TRANSFER FUNCTIONS BASED
ON REAL INTERPOLATION METHOD**

Real interpolation method was used to solve the problem of parametric identification of linearized objects. Based on condition numbers apparatus the authors considered the possibility of increasing accuracy when determining the structure of a model in the form of sampled-data transfer function.

UDC 62-533.65

**Korovikov A.G., Pavlov V.M., Olkhovik D.A.
SYNTHESIS OF MATHEMATICAL
MODEL OF VACUUM CHAMBER HEAT
CONTROL SUBSYSTEM IN KAZAKHSTAN
MATERIAL TESTING TOKAMAK KTM**

The authors studied the vacuum chamber heating in tokamak KTM as the controlled object and developed the mathematical model of the process as the controlled object. The conclusion was made that it is necessary and sufficient to use the control system constructed on the base of PI-controllers to control heating process. The authors determined the coefficient for the controller adjustment parameters for each heat penetration.

UDC 681.516.75

**Pashchenko O.V., Solovyev A.M., Yudakov S.V.
THE CONTROL SYSTEM OF INDUSTRIAL UNIT «KO-MKR»**

The article describes the control system of the industrial unit «KO-MKR». The construction and engineering capabilities of the unit are briefly disclosed in the paper. The authors substantiate the selection of control system structure and master controller. The article describes the principle of constructing the controller supervisory program and discloses the control system constructional features.

UDC 51-73

**Krivtsov P.Yu., Pavlov V.M.
USE OF SIMULATION COMPLEX FOR CHECKING
THE CONTROL SYSTEM OF SODIUM PROCESS**

The paper demonstrates the possibility of using the techniques of half-sized simulation modeling for checking the control system of inflammable and explosive sodium coolant process. The authors introduce the structure of the simulation complex, mathematical model of the process, program simulator; determine control system performance being subject to checking at simulation complex. The testing results of the control system are introduced.

UDC 004.75

**Muravyev S.V., Tarakanov E.V.
DATA TRANSMISSION IN WIRELESS SENSOR NETWORKS
WITH PRIORITIES BASED ON PREFERENCE AGGREGATION**

The authors propose the method of controlling data transmission in a wireless sensor network based on application of consensus relation for prioritization to the data transmitted. The article describes the statistical analytical model for analyzing the behavior of the proposed transfer protocol at limited network capacity. The model allows supporting the quality of measuring data transfer on the desired level.

UDC 621.398.725:621.317.727.1

**Zarevich A.I., Muravyev S.V., Bedareva E.V., Velichko O.N.
DIGITAL PULSE SIGNAL PROCESSING FOR DETERMINING
FREQUENCY RESPONSE OF CURRENT TRANSDUCERS**

The authors propose the method for determining amplitude- and phase-frequency characteristics of linear current transducers based on digital co-processing of input transducer impulse action and response to it. Frequency dependence of shunt transmission coefficient is computed by the components of complex spectral transformations of voltage from the output of the test and reference current transducers. The authors propose the ways of increasing the method accuracy based on ensemble averaging of signal spectral components.

UDC 519.688;004.93'12

**Parubets V.V., Berestneva O.G., Devyatykh D.V.
APPLICATION OF CUDA TECHNOLOGY
FOR COMPUTATIONAL SPEEDUP
IN NEURAL NETWORKS**

The authors consider the application of artificial neural networks in medical diagnostic. The article describes the possibility of computational speedup in the networks with NVIDIA CUDA technology for implementing parallel computing in learning and image recognition by the network. The network speeding is shown.

UDC 004.652.3

**Volovodenko V.A., Berestneva O.G.,
Nemerov E.V., Osadchaya I.A.
IMAGING APPLICATION IN STUDYING
THE STRUCTURE OF EXPERIMENTAL
MULTIDIMENSIONAL DATA**

The article considers the methods of multidimensional data structural analysis and introduces different approaches to visualization of the experimental research results. The authors give the examples of solving the applied problems using NovoSparkVisualizer.

UDC 004.9

**Tarkov M.S., Tikhonov N.V., Polovinkin V.G.
ESTIMATION OF CELL PREPARATION IMAGE
FOR MEDICAL-BIOLOGICAL RESEARCH**

The authors propose the technique for estimating image based on application of half-tone erosion and dilatation, «hat top» transformations and following threshold segmentation. The technique is implemented as a software product and used within the method of time-lapse (slow-motion) computer video shooting of microscopic objects developed for studying barrier properties of plant cell membranes. It allows estimating the relative area of the objects under observation in the image and may be used in various applications when the decay rate of cell structure is required to be determined.

UDC 004.942:338.27:504.062

**Zatonskiy A.V.
THE ADVANTAGES OF DIFFERENTIAL MODELS
IN ECOLOGICAL-ECONOMIC MODELING**

It is shown that it is more preferable to use ecological-economic models in the form of ordinary differential equations in the problems of predicting the controlled systems development than polynomial models and time tends in terms of adequacy and decrease of probable spread in predictions required for decision support.

Сведения об авторах

- Алексеев Александр Сергеевич**, 1984 г.р., канд. техн. наук, ст. преподаватель кафедры интегрированных компьютерных систем управления Института кибернетики ТПУ. Р.т. 70-18-37. E-mail: alekseyev@tpu.ru. Область научных интересов: методы синтеза самонастраивающихся систем управления.
- Бедарева Елена Вячеславовна**, 1985 г.р., ассистент, аспирант кафедры компьютерных измерительных систем и метрологии Института кибернетики ТПУ. Р.т. 41-75-27. E-mail: blv85@sibmail.com. Область научных интересов: программное обеспечение измерительных систем, математическое моделирование измерительных систем, оценивание динамических характеристик преобразователей тока.
- Берестнева Ольга Григорьевна**, д-р техн. наук, профессор кафедры прикладной математики Института кибернетики ТПУ, профессор кафедры общей и дифференциальной психологии Сибирского государственного медицинского университета, г. Томск. Р.т. 42-61-00. E-mail: ogb6@yandex.ru. Область научных интересов: математическое моделирование, медицинские информационные технологии, когнитивная графика.
- Ботыгин Игорь Александрович**. 1947 г.р., канд. техн. наук, доцент кафедры информатики и проектирования систем института кибернетики ТПУ. Р.т. 42-04-05. E-mail: botygin@ad.cctpu.edu.ru. Область научных интересов: распределенные системы обработки данных, функционирующие в реальном масштабе времени.
- Буй Тхи Тху Чанг**, 1984 г.р., аспирант кафедры вычислительной техники Института кибернетики ТПУ. Р.т. 70-16-09. E-mail: trangbt.084@gmail.com. Область научных интересов: обработка изображений, вейвлет-преобразование, искусственные нейронные сети, распознавание образов.
- Величко Олег Николаевич**, 1958 г.р., д-р техн. наук, директор научно-производственного Института метрологического обеспечения измерений электромагнитных величин ГП «Укрметрестандарт», г. Киев, Украина. Р.т. 8-(38-044)-526-03-35. E-mail: velichko@ukrcsm.kiev.ua. Область научных интересов: методы и средства измерений переменного напряжения, тока, электрической мощности и импеданса, обработка результатов измерений.
- Воловodenko Виталий Алексеевич**, 1948 г.р., канд. техн. наук, доцент кафедры оптимизации систем управления Института кибернетики ТПУ. Р.т. 42-04-59. E-mail: wolcowvav@tpu.ru. Область научных интересов: когнитивная графика, многомерные пространства, математическое моделирование.
- Вылегжанин Олег Николаевич**, 1943 г.р., канд. хим. наук, доцент кафедры прикладной математики Института кибернетики ТПУ. Р.т. 42-61-00. E-mail: onv@am.tpu.ru. Область научных интересов: разработка математических методов обработки данных, математическая статистика, методы восстановления зависимостей, выпуклое программирование, спектральный анализ, приложение в геофизике, медицине, молекулярной спектроскопии, анализе шумов датчиков.
- Гайворонский Сергей Анатольевич**, 1961 г.р., канд. техн. наук, доцент кафедры автоматики и компьютерных систем Института кибернетики ТПУ. Р.т. 42-05-88. E-mail: sa-
- ga@tpu.ru. Область научных интересов: робастный анализ и синтез систем автоматического управления.
- Девятых Дмитрий Владимирович**, 1989 г.р., магистрант кафедры прикладной математики Института кибернетики ТПУ. Р.т. 42-61-00. E-mail: oneeighttwo3547@gmail.com. Область научных интересов: искусственные нейронные сети, анализ данных.
- Друки Алексей Алексеевич**, 1985 г.р., аспирант кафедры вычислительной техники Института кибернетики ТПУ. Р.т. 22-45-08. E-mail: druki2008@yandex.ru. Область научных интересов: информационные технологии, обработка изображений, нейронные сети.
- Затонский Андрей Владимирович**, 1969 г.р., д-р техн. наук, доцент, зав. кафедрой автоматизации технологических процессов Березниковского филиала Пермского политехнического университета. Р.т. 8-(342-4)-26-90-90. E-mail: zxenon@narod.ru. Область научных интересов: информатизация и управление в сложных системах.
- Заревич Антон Иванович**, 1977 г.р., канд. техн. наук, доцент кафедры компьютерных измерительных систем и метрологии Института кибернетики ТПУ. Р.т. 41-75-27. E-mail: antonzarevich@tpu.ru. Область научных интересов: теория измерений, связанные автоколебательные системы, нелинейные колебания.
- Захарова Алена Александровна**, 1972 г.р., д-р техн. наук, профессор, зав. кафедрой начертательной геометрии и графики, зав. лабораторией 3D-моделирования Института кибернетики ТПУ. Р.т. 42-06-98. E-mail: zaa@tpu.ru. Область научных интересов: 3D-моделирование, проектирование разработки нефтегазовых месторождений, компьютерная графика.
- Иванченко Виктор Павлович**, 1949 г.р., канд. техн. наук, доцент кафедры прикладной математики Института кибернетики ТПУ. Р.т. 42-61-00. E-mail: vpi@am.tpu.ru. Область научных интересов: цифровая обработка сигналов и изображений волновых полей.
- Коровиков Александр Геннадьевич**, 1979 г.р., начальник отдела информационных технологий ДГП «Институт атомной энергии» РГП «Национальный ядерный центр», г. Курчатов, Республика Казахстан; аспирант кафедры электроники и автоматики физических установок Физико-технического института ТПУ. Р.т. 8-(722-51)-2-35-49. E-mail: Korovikov@nnc.kz. Область научных интересов: системы автоматического управления технологическими процессами.
- Кочегуров Александр Иванович**, 1954 г.р., канд. техн. наук, доцент кафедры прикладной математики Института кибернетики ТПУ. Р.т. 42-04-63. E-mail: kai@cc.tpu.edu.ru. Область научных интересов: цифровая обработка сигналов и изображений волновых полей.
- Кривцов Павел Юрьевич**, 1978 г.р., зам. начальника отдела информационных технологий РГП «Национальный ядерный центр Республики Казахстан», г. Курчатов; аспирант кафедры электроники и автоматики физических установок Физико-технического института ТПУ. Р.т. 8-(722-51)-2-24-21. E-mail: Krivtsov@nnc.kz. Область научных интересов: разработка и исследование систем управления технологическими процессами.

- Курганкин Виктор Витальевич**, 1986 г.р., аспирант кафедры теоретической и прикладной механики Института физики высоких технологий ТПУ. Р.т. 56-45-88. E-mail: kurgankin@sibmail.com. Область научных интересов: системы управления с переменной динамикой.
- Ле Хоай**, 1984 г.р., аспирант кафедры оптимизации систем управления Института кибернетики ТПУ. Р.т. 42-14-85. E-mail: lehotomsk@yahoo.com. Область научных интересов: искусственный интеллект, семантические технологии.
- Муравьев Сергей Васильевич**, 1954 г.р., д-р техн. наук, профессор, зав. кафедрой компьютерных измерительных систем и метрологии Института кибернетики ТПУ. Р.т. 42-04-49. E-mail: muravuyov@samsam.tpu.ru. Область научных интересов: теория измерений, теория графов, алгоритмы голосования для агрегирования данных мультисенсоров, программное обеспечение измерительных информационных систем и его аттестация.
- Назаров Анатолий Андреевич**, 1947 г.р., д-р техн. наук, профессор, Почетный работник Высшего профессионального образования РФ, зав. кафедрой теории вероятностей и математической статистики Томского государственного университета. Р.т. 52-95-99. E-mail: anazarov@fpmk.tsu.ru. Область научных интересов: теория массового обслуживания, теория случайных процессов, математическое моделирование, численные методы, информационные технологии, вероятностные методы и модели, системы и сети массового обслуживания и их приложения.
- Нгуен Тоан Тханг**, 1983 г.р., аспирант кафедры вычислительной техники Института кибернетики ТПУ. Р.т. 70-16-09. E-mail: thangnt.cntt@gmail.com. Область научных интересов: обработка изображений, распознавание образов, нейронные сети.
- Нгуен Ба Нгок**, 1984 г.р., аспирант кафедры оптимизации систем управления Института кибернетики ТПУ. Р.т. 42-14-85. E-mail: nguuen_bn@hotmail.com. Область научных интересов: искусственный интеллект, параллельное программирование, информационный поиск, семантические технологии.
- Небаба Степан Геннадьевич**, 1989 г.р., магистрант кафедры прикладной математики Института кибернетики ТПУ. Р.т. 42-61-00. E-mail: stepan-lfx@mail.ru. Область научных интересов: разработка математических методов обработки данных.
- Немеров Евгений Владимирович**, 1954 г.р., канд. мед. наук, доцент Сибирского государственного медицинского университета, г. Томск. Р.т. 77-89-20. E-mail: neme-gev@mail.ru. Область научных интересов: медицинская диагностика, моделирование в медицине.
- Ольховик Дмитрий Анатольевич**, 1983 г.р., начальник группы информационно-измерительных систем отдела информационных технологий ДГП «Институт атомной энергии» РГП «Национальный ядерный центр», г. Курчатов, Республика Казахстан. Р.т. 8-(722-51)-2-35-49. E-mail: Olkховик@nnc.kz. Область научных интересов: системы автоматического управления технологическими процессами.
- Орлов Олег Викторович**, 1955 г.р., канд. техн. наук, доцент кафедры прикладной математики Института кибернетики ТПУ. Р.т. 42-61-00. E-mail: orloff@tpu.ru. Область научных интересов: обработка сейсмической информации с целью прогноза углеводородов.
- Осадчая Ирина Анатольевна**, студент кафедры прикладной математики Института кибернетики ТПУ. Р.т. 42-61-00. E-mail: Irishka_tomsk@mail.ru. Область научных интересов: математическое моделирование, медицинские информационные технологии, когнитивная графика.
- Осокин Александр Николаевич**, 1947 г.р., канд. техн. наук, доцент кафедры вычислительной техники Института кибернетики ТПУ. Р.т. 41-89-12. E-mail: osokin@vt.tpu.ru. Область научных интересов: создание алгоритмов сжатия информации, их программная и аппаратная реализация; помехоустойчивое кодирование сигналов.
- Павлов Вадим Михайлович**, 1965 г.р., канд. техн. наук, доцент кафедры электроники и автоматики физических установок Физико-технического института ТПУ. Р.т. 42-12-95. E-mail: pavlov@tpu.ru. Область научных интересов: разработка методов и средств повышения информационной эффективности систем автоматизации научных исследований для экспериментальных физических установок.
- Парубец Валерий Валерьевич**, 1989 г.р., магистрант кафедры прикладной математики Института кибернетики ТПУ. Р.т. 42-61-00. E-mail: VParubets.mail@gmail.com. Область научных интересов: математическое моделирование, информатика параллельные вычисления.
- Пашенко Олег Валентинович**, 1962 г.р., канд. физ.-мат. наук, ст. науч. сотр. лаборатории 23 Физико-технического института ТПУ. Р.т. 42-39-99. E-mail: ovp23@tpu.ru. Область научных интересов: алгоритмы управления вакуумными установками.
- Погребной Андрей Владимирович**, 1992 г.р., студент кафедры оптимизации систем управления Института кибернетики ТПУ. Р.т. 42-07-27. E-mail: avpogrebnoy@gmail.com. Область научных интересов: информационные технологии, моделирование и автоматизированное проектирование распределённых систем управления.
- Погребной Дмитрий Владимирович**, 1968 г.р., канд. техн. наук, доцент кафедры информатики и проектирования систем Института кибернетики ТПУ. Р.т. 42-07-27. E-mail: Dmitry@contek.ru. Область научных интересов: программирование, информационные системы, моделирование распределённых систем.
- Половинкин Владимир Григорьевич**, 1947 г.р., канд. физ.-мат. наук, ст. науч. сотр. лаборатории физических основ интегральной микроэлектроники Института физики полупроводников СО РАН, г. Новосибирск. Р.т. 8-(383)-333-35-06. E-mail: pvg@isp.nsc.ru. Область научных интересов: фотоприёмные устройства, моделирование физических процессов, обработка сигналов и изображений.
- Попов Владимир Николаевич**, 1978 г.р., канд. техн. наук, доцент кафедры информатики и проектирования систем института кибернетики ТПУ. Р.т. 42-07-27. E-mail: vnp@tpu.ru. Область научных интересов: разработка распределённых систем обработки и баз данных в образовании и дендрозоологии.
- Пушкарёв Максим Иванович**, 1987 г.р., ассистент кафедры автоматики и компьютерных систем Института кибернетики ТПУ. Р.т. 41-91-06. E-mail: pushkarev@tpu.ru. Область научных интересов: робастный анализ и синтез систем автоматического управления.
- Рудницкий Владислав Александрович**, 1968 г.р., канд. техн. наук, доцент кафедры интегрированных компьютерных систем управления Института кибернетики ТПУ. Р.т. 70-18-37. E-mail: vlad_rudnicki@mail.ru. Область научных интересов: идентификация и диагностика систем автоматического регулирования.
- Рудометкина Моника Николаевна**, аспирант кафедры вычислительной техники Института кибернетики ТПУ, пре-

- подаватель кафедры систем автоматизированного проектирования Института компьютерных технологий и системотехники Карагандинского государственного технического университета. Р. т. 8-(721-2)-56-44-22. E-mail: monika_7272@mail.ru. Область научных интересов: создание реляционных средств моделирования логических сетей в системах искусственного интеллекта.
- Семенова Инна Анатольевна**, аспирант кафедры теории вероятностей и математической статистики Томского государственного университета. Р.т. 53-38-25. E-mail: inna_ac@mail.ru. Область научных интересов: теория массового обслуживания, теория случайных процессов, математическое моделирование, численные методы.
- Сидоров Дмитрий Владимирович**, 1984 г.р., канд. техн. наук, инженер-электроник отдела Информационно-вычислительной техники ОАО «Томское пиво». Р.т. 41-63-80. E-mail: gauso@mail.ru. Область научных интересов: разработка алгоритмов сжатия изображений.
- Соловьев Александр Михайлович**, 1957 г.р., ведущий инженер-конструктор, лаборатории 23 Физико-технического института ТПУ. Р.т. 42-39-99. E-mail: sam@tpru.ru. Область научных интересов: аппаратное обеспечение систем управления вакуумными установками.
- Спицын Владимир Григорьевич**, 1948 г.р., д-р техн. наук, профессор кафедры вычислительной техники Института кибернетики ТПУ. Р.т. 70-16-09. E-mail: spvg@tpru.ru. Область научных интересов: обработка изображений в технических системах, распознавание образов, искусственные нейронные сети, генетические алгоритмы, рассеяние электромагнитных волн.
- Тараканов Евгений Владимирович**, 1982 г.р., ассистент кафедры компьютерных измерительных систем и метрологии Института кибернетики ТПУ. Р.т. 41-75-27. E-mail: tark@tpru.ru. Область научных интересов: агрегирование предпочтений, беспроводные сенсорные сети.
- Тарков Михаил Сергеевич**, 1955 г.р., канд. техн. наук, доцент, ст. научн. сотр. лаборатории физических основ интегральной микроэлектроники Института физики полупроводников СО РАН, г. Новосибирск. Р.т. 8-(383)-333-36-89. E-mail: tarkov@isp.nsc.ru. Область научных интересов: параллельные вычислительные системы, нейронные сети, обработка изображений.
- Тартаковский Валерий Абрамович**, 1943 г.р., д-р физ.-мат. наук, заведующий лабораторией Института мониторинга климатических и экологических систем СО РАН. Р.т. 49-26-53. E-mail: tv@imces.ru. Область научных интересов: биоинформационные технологии.
- Тихонов Николай Владимирович**, 1986 г.р., аспирант Института физики полупроводников СО РАН, г. Новосибирск. Р.т. 8-(383)-333-36-89. E-mail: niknsk@list.ru. Область научных интересов: нейронные сети, обработка изображений.
- Тузовский Анатолий Федорович**, 1952 г.р., д-р техн. наук, профессор кафедры оптимизации систем управления Института кибернетики ТПУ. Р.т. 42-14-85. E-mail: tuzovskyaf@tpru.ru. Область научных интересов: семантические технологии, онтологии, интернет технологии, технологии разработки программ.
- Фан Нгок Хоанг**, 1983 г.р., аспирант кафедры вычислительной техники Института кибернетики ТПУ. Р.т. 70-16-09. E-mail: hoangpn285@gmail.com. Область научных интересов: обработка изображений, искусственные нейронные сети, распознавание отпечатков пальца.
- Хамухин Александр Анатольевич**, 1954 г.р., канд. техн. наук, доцент кафедры информатики и проектирования систем Института кибернетики ТПУ. Р.т. 42-04-05. E-mail: aaxtru@tpru.ru. Область научных интересов: исследование и разработка компьютерных моделей и методов моделирования.
- Шкляр Алексей Викторович**, 1970 г.р., зав. лабораторией робототехнических и гибких автоматизированных систем Института кибернетики ТПУ. Р.т. 42-06-98. E-mail: shklyarav@mail.ru. Область научных интересов: 3D-моделирование, компьютерная графика.
- Юдаков Сергей Владимирович**, 1979 г.р., науч. сотр. лаборатории 23 Физико-технического института ТПУ. Р.т. 41-79-54. E-mail: ysv@tpru.ru. Область научных интересов: аппаратное обеспечение систем управления вакуумными установками.

К СВЕДЕНИЮ АВТОРОВ

Принимаются статьи, подготовленные в MS Word-2003 (файл и распечатка). Статья должна быть подписана авторами и иметь сопроводительное письмо на бланке организации.

Объем статьи до 8 стр., включая рисунки и таблицы, размещенные в тексте по упоминанию. Размер бумаги А4, поля по 25 мм. Текст в 1 интервал **без переносов, лишних пробелов и абзацных интервалов**, шрифт Times New Roman, 12 пунктов. Файлы рисунков (в градациях серого) в jpg, tif, cdr или иных форматах редакторов Photoshop, Corel Draw с разрешением 300 dpi прилагаются к статье. Рисунки и таблицы: **Рис. 1. Название; Таблица. Название.** Кавычки вида «...». Интервалы – 1,2...1,8 мм или 5–7 шт. Формулы – в MathType, настройка по умолчанию. Нумеруются только те формулы, на которые есть ссылка в тексте.

Курсивом – буквы латинского алфавита, кроме входящих в имена собственные, обозначения стандартных математических функций и химических элементов ($U_{пр.}$, Φ_i , но Al_2O_3 , $\cos\alpha_i$, \max , \lg , «BASF»). Векторы – полужирным курсивом. Список литературы – по ГОСТ Р 7.0.5-2008 (см. пример). Литература – по упоминанию: [1, 2], [2. С. 245], [3–7].

УДК 621.37 (Пример оформления статьи)

АНАЛИЗ РАБОТЫ СИСТЕМЫ АВТОМАТИЧЕСКОЙ РЕГУЛИРОВКИ

И.И. Иванов, П.П. Петров*

Томский политехнический университет

*ОАО «Центр», г. Москва

E-mail: ivanov@tpu.ru

Показана возможность расчета ... Установлено, что ... Сделан вывод о том, что ... (Аннотация, 10 кегль).

Ключевые слова (ниже ключевые слова на английском языке):

Усилительный каскад, регулировка тока.

В [1, 2] показано, что усилительный каскад с автоматической регулировкой потребляемого тока (АРПТ) позволяет получить ...

СПИСОК ЛИТЕРАТУРЫ

1. Фамилия И.О. Название книги. – М.: Издательство, 2012. – 123 с.
2. Название книги / под ред. И.О. Фамилия. – М.: Издательство, 2012. – 123 с.
3. Фамилия И.О. Название статьи // Журнал. – 2012. – Т. 316. – № 4. – С. 71–77.
4. Фамилия И.О. Название диссертации: автореф. дис. ... канд. физ.-мат. наук. – Томск, 2008. – 19 с.
5. Название изобретения: пат. 2000000 Рос. Федерация. № 2009129009/10; заявл. 27.07.10; опубл. 10.10.12, Бюл. № 4. – 3 с.
6. Фамилия И.О. Название статьи // Наименование конференции: Труды VII Междунар. научно-практ. конф. молодых ученых. – Томск, 2012. – Т. 1. – С. 226–228.
7. Фамилия И.О. Название статьи // Наименование ресурса. 2012. URL: <http://www.tpu.ru/html/izvestia.htm> (дата обращения: 25.09.2012).

Поступила 25.04.2012 г.

Сведения об авторах:

Иванов Иван Иванович, 1975 г.р., канд. техн. наук, ст. науч. сотр. кафедры автоматизации и компьютерных систем Института кибернетики ТПУ. Р.т. 22-22-22. E-mail: ivanov@tpu.ru. Область научных интересов: анализ...

Редактирование и корректура *М.А. Шустов*
Компьютерная верстка *О.Ю. Аршинова*
Перевод на англ. язык *С.В. Жаркова*

Подписано к печати 28.05.2012. Формат 60x84/8. Бумага «Снегурочка».
Печать XEROX. Усл. печ. л. 17,9. Уч.-изд. л. 16,2.
Заказ 698-12. Тираж 500 экз.



Национальный исследовательский Томский политехнический университет
Система менеджмента качества
Издательства Томского политехнического университета сертифицирована
NATIONAL QUALITY ASSURANCE по стандарту BS EN ISO 9001:2008



ИЗДАТЕЛЬСТВО  ТПУ, 634050, г. Томск, пр. Ленина, 30. Тел./факс: 8(3822) 563-291, www.tpu.ru, shustov@tpu.ru