

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

ЯНВАРЬ-ФЕВРАЛЬ 2012

ВЫПУСК 1 (77)



ISSN 2226-1494

ОПТИЧЕСКИЕ И ОПТИКО-ЭЛЕКТРОННЫЕ СИСТЕМЫ. ОПТИЧЕСКИЕ ТЕХНОЛОГИИ

Принципы формирования структуры интерференционных спектроределителей

Выбор параметров многоволнового аэрозольного лидара для дистанционного

зондирования атмосферы

Инвариантная модель для оценки коэффициента пропускания атмосферы при мониторинге объектов в оптическом диапазоне спектра

ФОТОНИКА И ОПТОИНФОРМАТИКА

Исследование спектра молекулы ДНК в терагерцовой области частот

Комбинированный метод подсчета эритроцитов на изображениях мазков крови

Установка квантовой криптографии с источником одиночных фотонов,

основанным на явлении спонтанного параметрического рассеяния света

АНАЛИЗ И СИНТЕЗ СЛОЖНЫХ СИСТЕМ

Идентификация линейно меняющейся частоты синусоидального сигнала

Усовершенствованный алгоритм анализа сигналов малозаметных радиолокационных станций на основе оконного преобразования Фурье

Использование условий качественной экспоненциальной неустойчивости для оценки динамических процессов

Система управления следящего электропривода с нежесткой исполнительной осью

Анализ энергетических показателей многоуровневых полупроводниковых преобразователей систем электропривода

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Среда имитационного моделирования агентных систем реального времени

Эффективный логический вывод в продукционной модели с помощью баз данных с вертикальной архитектурой

Извлечение онтологий из Wiki-систем

Извлечение ключевых словосочетаний

Некоторые аспекты исследования систем распознавания речи

в телефонных службах поддержки

ТЕХНОЛОГИИ ПРОГРАММИРОВАНИЯ

Выбор функции приспособленности особей генетического алгоритма с помощью обучения с подкреплением

Автоматические методы модификации решений для тестирования проверяющих программ

Применение деревьев для реализации массовых операций на многомерных массивах данных

Использование генетического алгоритма для поиска оптимальной траектории наблюдателя

Применение методов решения задачи о выполнимости булевой формулы для построения управляющих конечных автоматов по сценариям работы

Применение адаптивного генетического алгоритма для генерации клеточных автоматов

МЕХАНИКА И МЕХАТРОНИКА

Установка для формирования аэродинамического профиля газовых опор

МАТЕРИАЛОВЕДЕНИЕ И НАНОТЕХНОЛОГИИ

Микрофлюидные чипы для исследования биологических объектов

методами микроскопии высокого разрешения

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ

Разработка стеганоалгоритма на базе форматных и пространственных принципов сокрытия данных

Методика оценки вероятности обнаружения несанкционированного проникновения оптико-электронным извещателем

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

Автоматизированная система проектирования и исследования деревянных конструкций на металлических зубчатых пластинах

Система поддержки принятия проектных решений для документирования научно-технической информации

НОВЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

Методика оценивания компетенций выпускника

Организация обучения студентов технологии производства электронного устройства в единой информационной среде

Комплексы визуальных и расчетных лабораторных моделей по основам цифровой электронной обработки информации

КРАТКИЕ СООБЩЕНИЯ

Метод понижения коэффициента отражения волоконных брэгговских решеток с помощью эффекта фотохромизма

Параметрическая идентификация тензоров инерции тел на сферических движениях с медленным собственным вращением

Идентификация тензора инерции тела на реверсивно-симметричных в прецессиях в ограниченном угловом интервале

SUMMARY

Губанова Л.А., Путилин Э.С.

1

Волков Н.Н.

6

Демин А.В., Моисеева М.И.

9

Цуркан М.В., Собакинская Е.А., Смолянская О.А., Беспалов В.Г., Вакс В.Л., Балбекин Н.С.

15

Дырнаев А.В., Потапов А.С.

19

Егоров В.И., Латыпов И.З., Рунаков А.В.

23

Глейм А.В., Чивилихин С.А.

Арановский С.В., Бобцов А.А., Пыркин А.А.

28

Коротков А.В., Мухин А.И.

32

Рабин Е.Ю., Григорьев В.В., Быстров С.В.

36

Спорягин А.В.

Абдуллин А.А., Толмачев В.А.

41

Михеев К.Е., Томасов В.С.

46

Пантелеев М.Г., Кохтенко Н.В., Лебедев С.В.

53

Катериненко Р.С.

58

Шестаков В.К.

62

Попова С.В., Ходырев И.А.

67

Балакшин П.В., Петров Г.Ю.

71

Афанасьева А.С., Буздалов М.В.

77

Ахи А.А., Станкевич А.С., Шалыто А.А.

81

Баниных А.Г.

85

Степанов Д.В., Шалыто А.А.

90

Ульянцев В.И., Царев Ф.Н.

96

Тихомиров А.В., Шалыто А.А.

100

Родников С.А.

106

Кухтевич И.В., Букатин А.С., Мухин И.С.

111

Евстратов А.А.

Коробейников А.Г., Кувшинов С.С., Блинов С.Ю., Лейман А.В., Несторов С.И.

116

Волхонский В.В., Воробьев П.А.

120

Павленко М.Н., Шмидт А.Б.

124

Агаркова Н.В., Артемова Г.О., Гусарова Н.Ф.

128

Пирская А.С.

135

Воробьев А.С., Федосов Ю.В.

141

Новиков В.В., Генович В.В., Гриншпун Д.М.

145

Варжель С.В., Куликов А.В., Брунов В.С.

151

Асеев В.А.

Шаховат С.Н., Мельников Г.И.

152

Мельников В.Г., Кравчук Р.Ю., Мельников Г.И., Шаховат С.Н.

153

155



**Издание Санкт-Петербургского национального исследовательского университета
информационных технологий, механики и оптики**

ГЛАВНЫЙ РЕДАКТОР.....

доктор технических наук, профессор В.О. Никифоров

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

доктор технических наук, профессор А.А. Бобцов,

доктор технических наук А.В. Бухановский,

доктор технических наук, профессор В.А. Валетов,

доктор физико-математических наук, старший научный сотрудник Т.А. Вартанян,

доктор технических наук М.А. Ган,

доктор технических наук, профессор Ю.А. Гатчин,

кандидат технических наук, старший научный сотрудник Н.Ф. Гусарова,

доктор технических наук, профессор А.В. Демин,

кандидат технических наук, доцент Н.С. Кармановский (заместитель главного редактора),

доктор физико-математических наук, профессор Ю.Л. Колесников,

доктор физико-математических наук, профессор С.А. Козлов,

доктор технических наук, профессор А.Г. Коробейников,

доктор технических наук, профессор В.В. Курейчик,

доктор технических наук, доцент Л.С. Лисицына,

кандидат технических наук, доцент В.Г. Мельников,

доктор технических наук, профессор Ю.И. Нечаев,

доктор технических наук, профессор Н.В. Никоноров,

доктор технических наук, профессор А.А. Ожиганов,

доктор технических наук, профессор П.П. Парамонов,

доктор физико-математических наук, старший научный сотрудник Е.Ю. Перлин,

доктор технических наук, профессор И.Г. Сидоркина,

доктор технических наук О.А. Степанов,

доктор технических наук, профессор В.Л. Ткалич,

доктор технических наук, профессор А.А. Шалыто,

доктор технических наук, профессор Ю.Г. Якушенков

Ответственный секретарь Г.О. Артемова

Редактор, верстка Н.Ф. Гусарова

Редактор Л.Н. Точилина

Перевод Н.Г. Лещикова

Графика М.В. Герасимова

Адрес: 197101, Санкт-Петербург, Кронверкский пр., 49, НИУ ИТМО, комн. 330

Телефон / факс (812) 233 45 51

<http://books.ifmo.ru/ntv/>

E-mail:karmanov@mail.ifmo.ru



ПРАВИЛА ПРЕДСТАВЛЕНИЯ РУКОПИСЕЙ

ОБЩИЕ ПОЛОЖЕНИЯ

Редакция принимает к рассмотрению оригинальные (нигде ранее не опубликованные, за исключением материалов научных конференций) работы по тематике журнала, представляющие новые научные результаты, полученные лично авторами статьи. Публикуются также обзоры, характеризующие современное состояние актуальных направлений развития науки и техники, сообщения о научных конференциях, материалы научных дискуссий и рецензии на новые книги.

Тематика журнала включает разделы, отражающие достижения науки, техники и технологии в области современного приборостроения по направлениям: оптика, оптические системы и технологии, оптическое материаловедение и нанотехнологии, информационные и компьютерные системы и технологии, методы и системы защиты информации, системный анализ, системы автоматического управления, мехатроника, системы автоматизированного проектирования, микроэлектроника, биомедицинские технологии. В качестве дополнительных разделов журнал также публикует работы по экономике и финансам.

Материал статьи должен быть изложен в соответствии со следующей структурой.

В аннотации, рассчитанной на самый широкий круг читателей, необходимо кратко (объемом до 100 слов) изложить научное содержание статьи.

В введении необходимо представить содержательную постановку рассматриваемого вопроса, краткий анализ известных из научной литературы решений, критику их недостатков и преимущества (особенности) предлагаемого подхода.

В основном тексте статьи должна быть представлена строгая постановка решаемой задачи, изложены и обстоятельно разъяснены (доказаны) полученные утверждения и выводы, приведены результаты экспериментальных исследований или математического моделирования, иллюстрирующие сделанные утверждения. Основной текст статьи может быть разбит на отдельные разделы: «Постановка задачи», «Основной результат», «Результаты моделирования» и т.п.

В заключении необходимо кратко сформулировать основные результаты, прокомментировать их и, если возможно, указать направления дальнейших исследований.

Все статьи проходят процедуру рецензирования. После ее завершения автору направляется экземпляр рукописи с замечаниями рецензента и научного редактора для доработки. При опубликовании статьи указываются даты ее поступления и последующих переработок.

ОСНОВНЫЕ ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ

Объем статьи, включая иллюстрации, таблицы и список литературы, не должен превышать 6 страниц.

Оригинал-макет статьи выполняется в редакторе Word for Windows. Шрифт Times New Roman, размер шрифта 12 pt, межстрочный интервал одинарный, поля сверху 20 мм, остальные поля 25 мм. Абзацный отступ 10 мм.

Аннотация печатается шрифтом Times New Roman 10 обычный, межстрочный интервал одинарный, выравнивается по ширине страницы. В аннотации не допускаются ссылки на цитированную литературу и громоздкие формулы.

Название раздела печатается шрифтом Times New Roman 12 полужирный, межстрочный интервал одинарный, отделяется от предыдущего раздела полуторной пустой строкой, а от последующего текста одной пустой строкой.

Название подраздела печатается шрифтом Times New Roman 12 полужирный, межстрочный интервал одинарный, пустыми строками не отделяется.

Формулы набираются в редакторе Microsoft Equation Ed. 3.0. Параметры стиля: текста — Times New Roman, прямой; функции — Times New Roman, прямой; переменной — Times New Roman, наклонный; греческих букв — Symbol, прямой; символов — Symbol, прямой; матрицы-вектора — Times New Roman, прямой, полужирный; чисел — Times New Roman, прямой. Размеры: обычный — 12 pt, крупный индекс — 10 pt, мелкий индекс — 8 pt, крупный символ — 16 pt, мелкий символ — 12 pt. Отдельные переменные в тексте допускается набирать в текстовом режиме с соблюдением требований, указанных для формул. Разделительный знак десятичных дробей — запятая.

При оформлении рисунков и графиков используется шрифт Times New Roman. Стиль написания переменных должен соответствовать требованиям к основному тексту.

Используются только единицы измерения, соответствующие системе СИ.

Список использованной литературы оформляется в соответствии с ГОСТ 7.0.5-2008. Нумерация позиций — согласно очередности ссылок в тексте.

Плата с аспирантов за публикацию рукописей не взимается.

Более подробную информацию смотрите на сайте <http://books.ifmo.ru/ntv/>

Подписано к печати 15.12.2011 Тираж 350 экз. Заказ № 25(77)

Отпечатано в учреждении «Университетские телекоммуникации»
Адрес: 197101, Санкт-Петербург, Кронверкский пр., 49

Журнал зарегистрирован Федеральной службой по надзору в сфере связи,
информационных технологий и массовых коммуникаций,
свидетельство ПИ № ФС77-47243 от 10.11.2011 г.
Подписной индекс 47 197 по каталогу агентства Роспечать
(Газеты. Журналы)

SCIENTIFIC AND TECHNICAL JOURNAL

OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS

JANUARY-FEBRUARY 2012

NUMBER 1 (77)



ISSN 2226-1494

NAUČNO-TEHNIČESKIJ VESTNIK INFORMACIONNYH TEHNOLOGIJ, MEHANIKI I OPTIKI

OPTICAL AND OPTICAL ELECTRONIC SYSTEMS. OPTICAL TECHNOLOGIES

- Structure formation principles of interference beam splitters
Choice of multiwave aerosol lidar parameters for remote atmosphere sounding
Invariant model for estimation of the atmosphere transmitting efficiency at objects monitoring in the optical spectral range

L. Gubanova, E. Putilin	1
N. Volkov	6
A. Demin, M. Moiseeva	9

PHOTONICS AND OPTICAL INFORMATICS

- Spectrum research of DNA molecule in the terahertz frequency domain

M. Tsurkan, E. Sobakinskaya, O. Smolyanskaya, V. Bespalov, V. Vaks, N. Balbekin	15
A. Dyrnaev, A. Potapov	19
V. Egorov, I. Latypov, A. Rupasov, A. Gleim, S. Chivilikhin	23

ANALYSIS AND SYNTHESIS OF COMPLEX SYSTEMS

- Identification of the linearly varying frequency for the sinusoidal signal
Improved time-frequency analysis algorithm of barely visible radar stations signals, based on windowed Fourier transform
Application of qualitative exponential instability conditions for dynamic processes estimation
Servo actuator control system with non-rigid executive axle
Analysis of energy performance of multilevel semiconductor converters of electric drive systems

S. Aranovskiy, A. Bobtsov, A. Pyrkin	28
A. Korotkov, A. Mukhin	32
E. Rabysh, V. Grigoriev, S. Bystrakov, A. Sporyagin	36
A. Abdullin, V. Tolmachev	41
K. Mikheev, V. Tomasov	46

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES

- Simulation environment of real-time agent-based systems
Effective inference in production rules model with column-oriented databases
Ontology extraction from Wiki-systems
Key phrases extraction
Implementation aspects of speech recognition applications in customer-service call centers

M. Panteleev, N. Kokhtenko, S. Lebedev	53
R. Katerinenko	58
V. Shestakov	62
S. Popova, I. Khodyrev	67
P. Balakshin, G. Petrov	71

PROGRAMMING TECHNOLOGIES

- Fitness function choosing for genetic algorithms with reinforcement learning
Automatic solution modifications for testing solution checkers
Trees application for multi-dimensional array mass updates
Genetic algorithm approach to observer's optimal trajectory design
Scenario-based finite state machine induction using SAT-solvers
Genetic programming application for cell automata generation

A. Afanasyeva, M. Buzdalov	77
A. Akhi, A. Stankevich, A. Shalyto	81
A. Bannykh	85
D. Stepanov, A. Shalyto	90
V. Ulyantsev, F. Tsarev	96
A. Tikhomirov, A. Shalyto	100

MECHANICS AND MECHATRONICS

- Airfoil formation setting for gas bearings

S. Rodinkov	106
-------------	-----

MATERIALS TECHNOLOGY AND NANOTECHNOLOGIES

- Microfluid chips for biological objects research by methods of high-resolution microscopy

I. Kukhlevich, A. Bukatin, I. Mukhin, A. Evstrapov	111
----------------------------------------------------	-----

INFORMATION SECURITY METHODS AND SYSTEMS

- Steganoalgorithm creation based on format and spatial principles of the data hiding

A. Korobeynikov, S. Kuvshinov, S. Blinov, A. Leyman, S. Nesterov	116
V. Volkhonskiy, P. Vorobyov	120

COMPUTER-AIDED DESIGN SYSTEMS

- Automated design and research system of wooden structures with metal gear plates
Design decision support system for scientific and technical information documentation

M. Pavlenko, A. Shmidt	124
N. Agarkova, G. Artemova, N. Gusarova	128

NEW TECHNOLOGIES IN EDUCATION

- Methodology of graduate competences evaluation
Electronic device design and manufacturing in a unified information environment for the students' educational process
Sets of the visual and computational laboratory models on the basis of digital electronic information processing

A. Pierskaya	135
A. Vorobyev, Y. Fedosov	141

BRIEF REPORTS

- Reflection coefficient decreasing method of fiber Bragg gratings by the effect of photochromism
Parametric identification of bodies' inertia tensors on spherical motions with slow spin
Inertia tensor identification on reverse-symmetric motions in a finite angular range

V. Novikov, V. Genovich, D. Grinshpun	145
S. Varzhel, A. Kulikov, V. Brunov, V. Aseev	151
S. Shakhval, G. Melnikov	152
V. Melnikov, R. Kravchuk, G. Melnikov, S. Shakhval	153
	155

SUMMARY

УДК 535.8, 535.4

ПРИНЦИПЫ ФОРМИРОВАНИЯ СТРУКТУРЫ
ИНТЕРФЕРЕНЦИОННЫХ СПЕКТРОДЕЛИТЕЛЕЙ

Л.А. Губанова, Э.С. Путилин

Рассмотрена методика формирования интерференционных спектропрелителей, образованных симметричными ячейками диэлектрических слоев. Приведена методика формирования интерференционных коротковолновых и длинноволновых спектропрелителей. Представлен анализ влияния количества симметричных ячеек и их структуры на выходные параметры спектропрелителей.

Ключевые слова: интерференционные покрытия, спектропрелители, симметричные ячейки, диэлектрические слои, дихроические элементы.

Введение

Интерференционные спектропрелители, типичные спектральные кривые которых представлены на рис. 1, используются в оптических приборах для выделения заданной спектральной области. В частности, в приборах с дифракционными решетками они используются для устранения нерабочих порядков спектра, в киноаппаратуре – для устранения тепловых потоков в виде холодных зеркал и теплофильтров; в цветном телевидении – для разделения цветовых потоков в качестве дихроических элементов, а также для подавления побочных полос пропускания диэлектрических узкополосных интерференционных светофильтров. Спектропрелители являются диэлектрическими зеркалами в области спектра, где необходимо получить минимальное пропускание. Их можно разделить на два типа – коротковолновые и длинноволновые отрезающие интерференционные фильтры, отличающиеся тем, что в первом случае область высокого пропускания лежит в коротковолновой области спектра, а во втором – в длинноволновой.

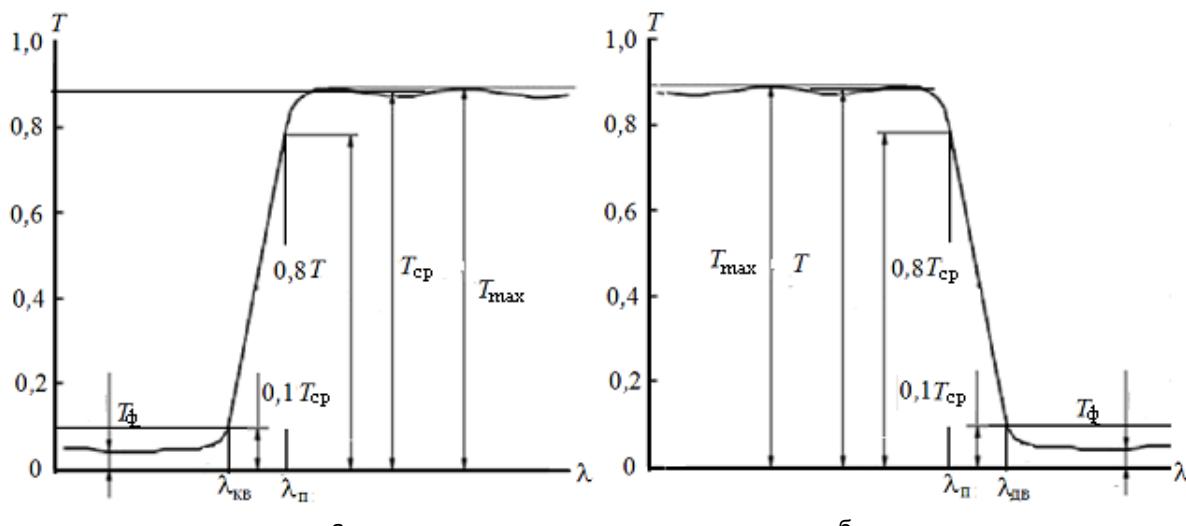


Рис. 1. Спектральная характеристика идеализированного спектропрелителя:
коротковолновый (а), длинноволновый (б)

Основными характеристиками спектропрелителей (рис. 1) являются: средний T_{cp} и максимальный T_{max} коэффициенты пропускания, ширина полосы пропускания $\Delta\lambda_{0,8}$ на уровне $T \geq 0,8T_{cp}$; ширина зоны подавления $\Delta\lambda_{0,1}$ $T \leq 0,1T_{cp}$; T_Φ – коэффициент пропускания в зоне подавления (области гашения); длины волн λ_{kb} и λ_{dv} , соответствующие пропусканию $T = 0,1T_{cp}$ на коротковолновой и длинноволновой границах зоны минимального пропускания. Крутизна коротковолновой K_k или длинноволновой K_d границ пропускания спектропрелителей определяется как отношение разности пропускания на уровнях $0,8T_{cp}$ и $0,1T_{cp}$ к разности соответствующих им длин волн, т.е. $K_k = 0,7T_{cp}/(\lambda_{kb} - \lambda_{ni})$, $K_d = 0,7T_{cp}/(\lambda_{dv} - \lambda_{ni})$, где λ_{ni} – длина волны, на которой коэффициент пропускания $T = 0,8T_{cp}$.

Понятие «симметричная система слоев»

Рассмотрим симметричные системы слоев диэлектриков, из которых возможно формирование диэлектрических зеркал, спектропрелителей, светофильтров и т.д., т.е. системы, которые имеют плоскость симметрии. Минимальное число слоев такой системы – три. В этом случае $\varphi_1 = \varphi_3$, а $n_1 = n_3$. Особен-

стью симметричных систем является то, что диагональные элементы матрицы интерференции, описывающей их свойства, одинаковы. Положим, что $\varphi_1 = \varphi_3 = \varphi$, $n_1 = n_3 = n$, и определим элементы матрицы интерференции [1]:

$$\begin{cases} m_{11} = \cos 2\varphi_1 \cos \varphi_2 - 0,5(n_1 \cdot n_2^{-1} + n_2 \cdot n_1^{-1}) \sin 2\varphi_1 \sin \varphi_2, \\ m_{12} = n_1^{-1} \sin 2\varphi_1 \cos \varphi_2 + (n_2^{-1} \cos^2 \varphi_1 - n_2 n_1^{-2} \sin^2 \varphi_1) \sin \varphi_2, \\ m_{21} = n_1 \sin 2\varphi_1 \cos \varphi_2 + (n_2 \cos^2 \varphi_1 - n_1^2 n_2^{-1} \sin^2 \varphi_1) \sin \varphi_2, \\ m_{22} = \cos 2\varphi_1 \cos \varphi_2 - 0,5(n_1 \cdot n_2^{-1} + n_2 \cdot n_1^{-1}) \sin 2\varphi_1 \sin \varphi_2. \end{cases} \quad (1)$$

Как видно из (1), $m_{11} = m_{22}$ и тогда в соответствии с [2] можно ввести понятие эффективных показателя преломления (ЭПП) N и фазовой толщины (ЭФТ) симметричной системы слоев $\varphi : N^2 = m_{21} \cdot (m_{12})^{-1}$, $\cos \varphi = m_{11} = m_{22}$, $\sin^2 \varphi = m_{12} \times m_{21}$.

Матрица интерференции симметричной системы слоев будет иметь вид, совпадающий с матрицей интерференции одного слоя.

Такая формальная запись удобна при рассмотрении многослойных систем, в которых симметричные системы многократно (ℓ раз) повторяются. При описании такой системы слоев матрица интерференции будет иметь вид

$$M = \begin{vmatrix} \cos \varphi & iN^{-1} \sin \varphi \\ iNs \in \varphi & \cos \varphi \end{vmatrix}^\ell = \begin{vmatrix} \cos \ell \varphi & iN^{-1} \sin \ell \varphi \\ iNs \in \ell \varphi & \cos \ell \varphi \end{vmatrix}.$$

Это значит, что ЭФТ интерференционной системы увеличивается в ℓ раз, а ЭПП остается прежним.

Задача сводится к тому, что необходимо вычислить эквивалентный показатель преломления такой системы и ее эквивалентную фазовую толщину. ЭФТ: $\cos \varphi = m_{11} = m_{22}$. В таком виде ее трудно анализировать, поскольку φ_1 и φ_2 – не всегда кратные числа. Рассмотрим для иллюстрации наиболее простую ситуацию, когда $\varphi_2 = 2\varphi_1$, т.е. $\varphi_1 = \varphi$, а $\varphi_2 = 2\varphi$, тогда

$$m_{11} = \cos^2 2\varphi - 0,5(n_1^2 + n_2^2) \cdot (n_1 n_2)^{-1} \sin^2 2\varphi = \cos \varphi.$$

Далее проанализируем, как ЭФТ такой системы φ и ЭПП меняются в зависимости от фазовой толщины входящих в эту трехслойную систему слоев:

$$\begin{aligned} \cos \varphi &= (2n_1 n_2)^{-1} (n_1 + n_2)^2 \left[\cos^2 \varphi - (n_1^2 + n_2^2)(n_1 + n_2)^2 \right] \\ N^2 &= n_1^2 \frac{\cos 2\varphi - (n_1 - n_2)/(n_1 + n_2)}{\cos 2\varphi + (n_1 - n_2)/(n_1 + n_2)}. \end{aligned} \quad (2)$$

Выражения (2) целесообразнее записать в виде дроби, потому что для оценки поведения этих функций удобнее иметь дело с величиной, меньшей единицы.

Проанализируем, как ведет себя ЭПП в зависимости от фазовой толщины и от величины показателей преломления слоев, образующих симметричную систему, вернее, даже не от величины показателей преломления, а от их относительной разности. Очевидно, что в случае, если фазовая толщина слоя равна нулю, эквивалентный показатель преломления $N^2 = n_1 n_2$. Фазовая толщина слоя равна нулю, когда слой отсутствует или когда длина волны равна бесконечности, т.е. при достаточно больших длинах волн или при очень малых толщинах слоя.

Далее проанализируем, что будет происходить по мере увеличения фазовой толщины слоев φ . Изобразим зависимость N^2 и $\cos \varphi$ от фазовой толщины слоев и количества симметричных ячеек многослойника.

На рис. 2 изображены зависимости N^2 , $\cos \varphi$, определяемые формулой (2), от фазовой толщины слоев для двух случаев, $n_1 < n_2$ и $n_1 > n_2$ и разного числа повторения симметричных ячеек. Как видно из этого рисунка, на графических зависимостях наблюдаются зоны разрыва функции, а также зоны с отрицательным и положительным значением величины N^2 , положения которых зависят от знака разности показателей преломления образующих систему слоев. Величина $\cos \varphi$ в зонах разрыва больше единицы, особенно это заметно при многократном повторении числа симметричных систем слоев. Очевидно, что в этой зоне величина $\cos \varphi$ должна быть заменена гиперболической функцией $\operatorname{ch} \varphi$, которая быстро растет по мере увеличения числа периодов. В этой ситуации можно легко записать и коэффициент отражения системы ячеек, осажденной на подложку, через ЭПП и ЭФТ симметричных ячеек слоев.

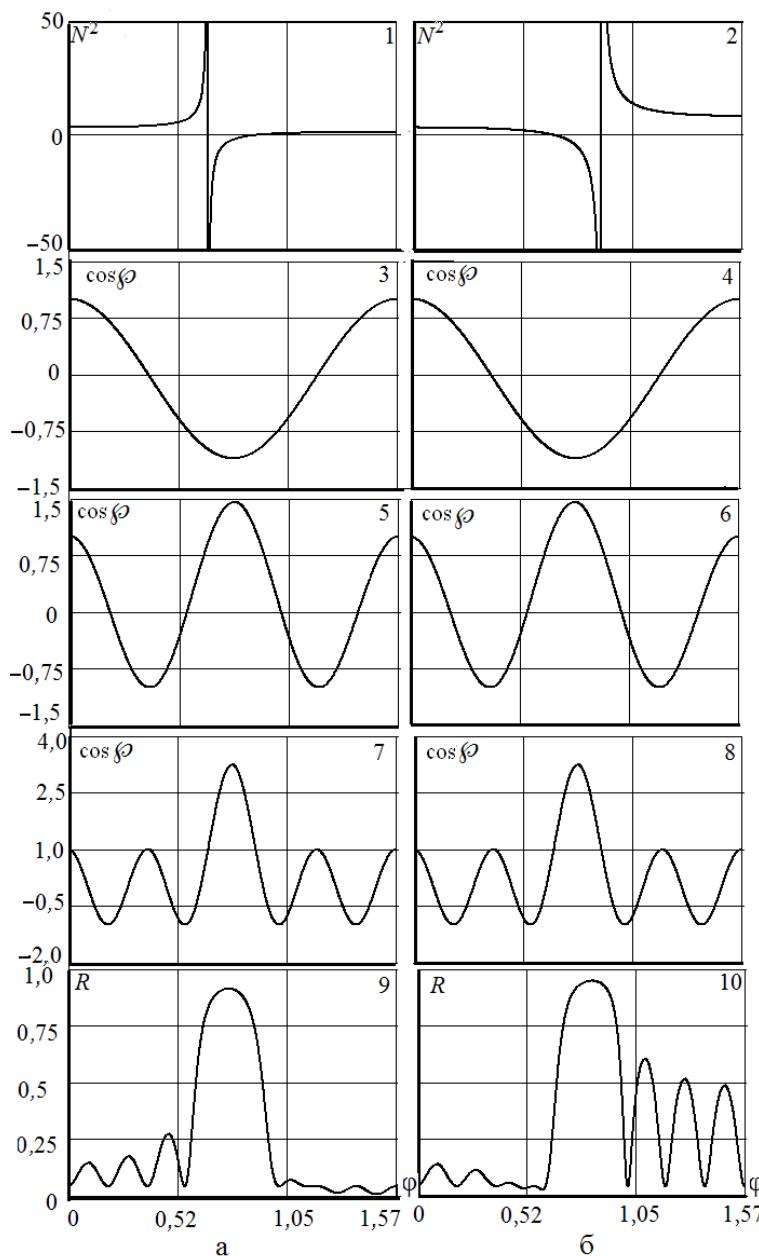


Рис. 2. Зависимости эквивалентных фазовой толщины (1, 2) и показателя преломления симметричных ячеек: (3, 4) – для одного, (5, 6) – для двух, (7, 8) – для четырех периодов, а также энергетического коэффициента отражения (9,10) для четырех периодов от фазовой толщины слоев покрытия, состоящего из симметричной системы слоев (H2BH – фрагменты (а); B2HB – (б)) при $n_H=1,45$, $n_B=2,30$

Для симметричной системы слоев коэффициент отражения будет иметь вид

$$R = \frac{N^2 (n_0 - n_m)^2 \cos^2 \varphi + (n_0 n_m - N^2)^2 \sin^2 \varphi}{N^2 (n_0 + n_m)^2 \cos^2 \varphi + (n_0 n_m + N^2)^2 \sin^2 \varphi}, \quad (3)$$

а при многократном (ℓ) повторении симметричных систем слоев:

$$R = \frac{N^2 (n_0 - n_m)^2 \cos^2 \ell \varphi + (n_0 n_m - N^2)^2 \sin^2 \ell \varphi}{N^2 (n_0 + n_m)^2 \cos^2 \ell \varphi + (n_0 n_m + N^2)^2 \sin^2 \ell \varphi}. \quad (4)$$

Как видно из фрагментов (9, 10) на рис. 2 в соответствии с зависимостями (3), (4), там, где $\cos \varphi > 1$, а N^2 имеет разрыв, энергетический коэффициент отражения имеет максимальное значение и стремится к единице по мере увеличения числа симметричных систем слоев. Эти зоны в литературе получили название зон подавления (в этих областях пропускание минимально), на самом деле это зоны высокого отражения. Зоны, где величина N^2 имеет малое значение, а $|\cos \varphi| < 1$ – это зоны прозрачности

диэлектрической системы. Если выполняется условие $N^2 = n_0 n_m$, то такая система будет иметь минимум отражения при той фазовой толщине, которая соответствует точке пересечения.

Сравнение энергетических коэффициентов отражения диэлектрических систем, образованных повторением ячеек симметричных систем на подложке, показывает существенное различие в их спектральных характеристиках. Системы, состоящие из симметричных ячеек В2НВ, имеют малое отражение при значениях фазовых толщин слоев, меньших 0,6, т.е. в длинноволновой области спектра. Системы, образованные многократным повторением симметричных ячеек Н2ВН имеют малое отражение при фазовых толщинах больших 1,05, т.е. в коротковолновой области спектра. Иначе говоря, многократное повторение симметричных ячеек Н2ВН позволяет получать длинноволновые, а ячеек В2НВ – коротковолновые спектроределители, отличающиеся крутизной и особенностями поведения в области прозрачности в зависимости от их числа.

Спектроределители на основе симметричных ячеек слоев

Рассмотрим два вида спектроределителей. Первым рассмотрим спектроределитель, образованный структурой $m\text{PB}2\text{H}2\text{B}...2\text{B}2\text{H}2$, идентичной структуре $\Pi(\text{B}2\text{H}2)^{0.5(m-1)}$. Здесь m – цифра перед буквой П обозначает количество слоев покрытия, цифра перед В и Н – кратность данного слоя относительно четвертьволнового.

Далее проанализируем, как количество симметричных ячеек (слоев в многослойной системе) влияет на крутизну и ширину зоны подавления. На рис. 3 представлены спектральные кривые пропускания для многослойных систем, состоящих из 13, 15, 17 слоев или 6, 7, 8-ми кратным повторением симметричных ячеек В2НВ. Это длинноволновые спектроределители. Как видно из рисунка, с увеличением числа слоев незначительно увеличивается зона гашения, существенно уменьшается пропускание в минимуме и меняется крутизна спектроределителя.

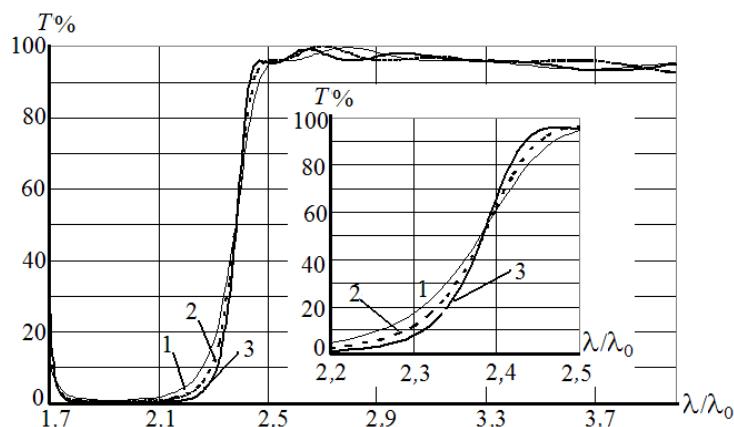


Рис. 3. Спектральные кривые пропускания спектроределителей, сформированных из симметричных ячеек (В2НВ) (количество ячеек): 6 – кривая 1; 7 – кривая 2; 8 – кривая 3, при $n_{\text{B}}=2,3$, $n_{\text{H}}=1,45$.

Вставка на графике показывает участок спектральной характеристики, который характеризует крутизну спектроределителя, в увеличенном масштабе

Далее рассмотрим аналогичную картину для коротковолновых спектроределителей. В этом случае в основе спектроределителя лежит ячейка вида Н2ВН. Как и выше, проанализируем покрытия, состоящие из 13, 15 и 17 слоев, что идентично структуре $\Pi(\text{H}2\text{B}N)^{0.5(m-1)}$, где, как и раньше, m – число слоев многослойной системы ($m = 13, 15, 17$). Спектральные характеристики таких систем представлены на рис. 4. Здесь, как и для длинноволновых спектроределителей, с увеличением числа слоев незначительно увеличивается зона гашения, существенно уменьшается пропускание в минимуме и увеличивается крутизна спектроределителя. Кроме того, по мере увеличения числа слоев величина ближайшего минимума пропускания, расположенного в коротковолновой части спектра, уменьшается, что существенно ухудшает эксплуатационные характеристики спектроределителя.

Для устранения этих недостатков, используя методы компьютерного синтеза структуры многослойных диэлектрических систем [3], можно оптимизировать спектральные характеристики спектроределителей за счет использования слоев с оптическими толщинами, не кратными $0,25\lambda_0$, что позволяет несколько увеличить крутизну при увеличении T_{cp} вблизи границы прозрачности. В качестве примера реализации можно рассмотреть следующие многослойные системы [3]:

13 П 1,145 В 1,096 НВН... В 0,995 Н 1,234 В

13 П 1,301 В 0,955 НВНВ... В 0,152 Н 0,151 В.

Спектральные характеристики таких систем представлены на рис. 5.

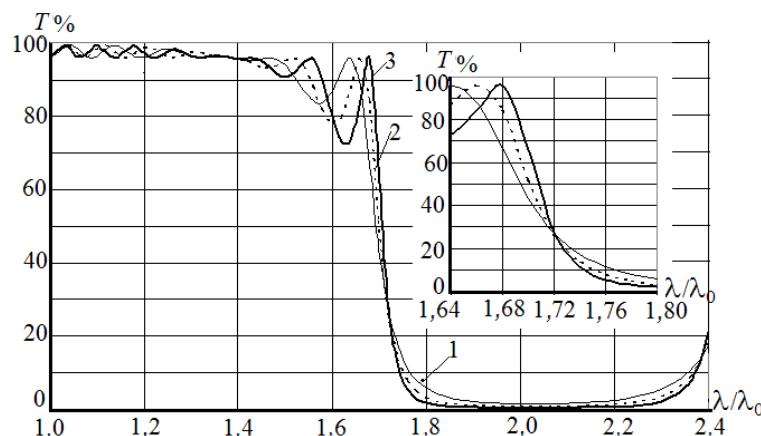


Рис. 4. Спектральные кривые пропускания коротковолновых спектрорадиоделителей, сформированных из симметричных ячеек (H2BH) (количество ячеек): 6 – кривая 1; 7 – кривая 2; 8 – кривая 3, при $n_B=2,3$, $n_H=1,45$. Вставка на графике показывает участок спектральной характеристики, который характеризует крутизну спектрорадиоделителя, в увеличенном масштабе

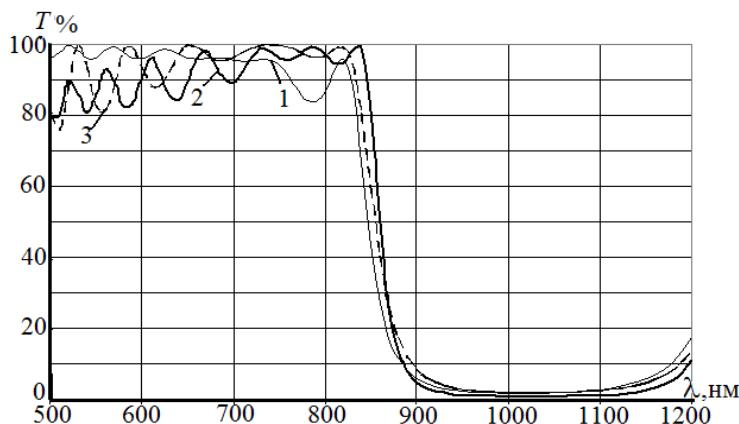


Рис. 5. Спектральные характеристики систем (H2BH)⁶ $\lambda_0=500$ нм – кривая 1; 13 П 1,145 В 1,096 НВН... В 0,995 Н 1,234 В $\lambda_0=1000$ нм – кривая 2; 13 П 1,301 В 0,955 НВНВН... В 0,152Н 0,151В – кривая 3, при $n_B=2,3$ $n_H=1,45$

На рис. 5 приведены спектральные характеристики диэлектрических систем, образованных шестикратным повторением симметричной ячейки (H2BH) и систем, состоящих из того же числа слоев, но имеющих оптические толщины слоев, прилегающих к подложке и граничащих с воздухом, отличающиеся от $0,25\lambda_0$. Необходимость наличия таких слоев обусловлена задачей максимального увеличения пропускания в диапазоне 700–850 нм.

Наличие в таких системах значительного числа слоев с оптической толщиной, не кратной $\lambda_0/4$, существенно затрудняет реализацию покрытий, а ошибки в слоях, неизбежно возникающие в процессе формирования системы, заметно ухудшают при реализации спектральную кривую. Кроме того, эти спектрорадиоделители имеют недостаточно широкую зону прозрачности. Современные методы синтеза позволяют найти конструкцию спектрорадиоделительного покрытия, лишенного вышеперечисленных недостатков, но состоящего из слоев с оптической толщиной, не равной $\lambda_0/4$ [4].

Заключение

В работе рассмотрены принципы конструирования спектрорадиоделителей, предназначенных для выделения участков спектра на основе симметричных ячеек. Показано, чем отличаются симметричные ячейки, формирующие коротковолновые спектрорадиоделители, от ячеек, входящих в состав длинноволновых спектрорадиоделителей. Проанализировано влияние расположения слоев в ячейке и количества ячеек на крутизну спектральной зависимости пропускания, величину зоны гашения и пропускание в минимуме. Показано, что с увеличением количества симметричных ячеек увеличивается крутизна спектральной зависимости и уменьшается пропускание в минимуме.

Литература

1. Путилин Э.С. Оптические покрытия: Учебное пособие. – СПб: СПбГУ ИТМО, 2010. – 227 с.
2. Фурман Ш.А. Тонкослойные оптические покрытия. – Л.: Машиностроение, 1977. – 264 с.
3. Телен А. Конструирование многослойных интерференционных светофильтров // Физика тонких пленок. – М.: Мир, 1972. – Т. 5. – С. 46–83.
4. Виногоров В.М., Исаев А.Г., Симонов В.П., Учайкин А.Г. Оптический интерференционный спектрорадиометр. Патент G02B5/28, 2006 [Электронный ресурс]. – Режим доступа: <http://bd.patent.su/2365000-2365999/pat/serv1/servlet3200.html>, св. Яз. рус. (дата обращения 01.12.2011).

Губанова Людмила Александровна	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, La7777@mail.ru
Путилин Эдуард Степанович	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, eputilin @yandex.ru

УДК 621.3.038.8.004.14

**ВЫБОР ПАРАМЕТРОВ МНОГОВОЛНОВОГО АЭРОЗОЛЬНОГО ЛИДАРА
ДЛЯ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ АТМОСФЕРЫ**

Н.Н. Волков

Описывается возможность дистанционного зондирования атмосферы при помощи лидара, позволяющего получать коэффициенты обратного рассеяния и экстинкции аэрозоля в диапазоне длин волн от 355 до 1064 нм. Оцениваются параметры многоволнового лидара для мониторинга аэрозоля в высотном диапазоне 1–10 км.

Ключевые слова: аэрозоль, лидар, рамановское рассеяние.

Введение

Увеличение количества антропогенных аэрозольных выбросов в атмосферу, повышение концентрации CO₂ и других парниковых газов оказывают воздействие на радиационный баланс планеты, и, как следствие, на ее климат. Следует отметить, что среди перечисленных факторов воздействие аэрозоля на радиационный баланс изучено в наименьшей степени. Воздействие атмосферного аэрозоля на радиационный баланс Земли двояко. Во-первых, аэрозоль влияет на радиационный баланс путем рассеяния и поглощения солнечной радиации (прямое воздействие). Во-вторых, это влияние происходит за счет модификации свойств облаков (косвенное воздействие). Для уменьшения соответствующих погрешностей в климатических моделях необходима глобальная информация об основных микрофизических параметрах аэрозоля, таких как радиус частиц, концентрация и комплексный показатель преломления.

Одним из наиболее перспективных инструментов, способных дистанционно проводить измерения параметров частиц, являются лидары. Малая длина волны излучения, сопоставимая с размером аэрозолей, и высокое пространственное разрешение делают лидары уникальным инструментом для исследования аэрозолей и облаков. Возможности метода значительно расширяются при проведении зондирования одновременно на нескольких длинах волн. Многоволновые системы позволяют получать информацию не только о спектральной зависимости коэффициентов рассеяния аэрозоля, но и о микрофизических параметрах частиц, таких как размер, концентрация и комплексный показатель преломления. Целью работы является оценка параметров многоволнового лидара, предназначенного для исследования аэрозолей различного типа: промышленного, пустынного и вулканического происхождения.

Существующие методы и средства для измерения параметров аэрозольных частиц атмосферы

Теоретические и экспериментальные исследования последних лет показывают, что измеренные коэффициенты обратного рассеяния и экстинкции аэрозоля могут быть инвертированы в микрофизическими параметры частиц путем решения обратной задачи зондирования [1]. Наиболее удобными с практической точки зрения являются лидары на основе Nd:YAG-лазера с генератором третьей гармоники. Такая система позволяет измерять три коэффициента обратного рассеяния (β) на длинах волн 355, 532, 1064 нм и два коэффициента экстинкции (α) на длинах волн 355, 532 нм. Вычисление коэффициентов экстинкции может быть проведено только с использованием рамановского рассеяния на молекулах азота или кислорода [1]. В измерениях, как правило, используется рассеяние на азоте, и длины волн соответствующих рамановских компонент для второй и третьей гармоники Nd:YAG составляют 608 нм и 387 нм. Излучаемая длина волны составляет 355 нм, в то время как после рассеяния длина волны составляет 387 нм, поэтому соответствующая экстинкция пересчитывается на 355 нм с использованием коэффициента Ангст-

рема. Соответствующая погрешность вычисления коэффициента экстинкции, связанная с неопределенностью выбора параметра Ангстрема, не велика и не превосходит 5%.

Сечение рамановского рассеяния почти на три порядка меньше сечения рэлеевского рассеяния, поэтому для надежной регистрации рамановского сигнала в нижней тропосфере необходимо использовать мощные лазерные источники излучения и приемные телескопы значительной апертуры. В работе проводится оценка параметров лазера и системы регистрации, требуемых для создания многоволнового аэрозольного лидара.

Процесс детектирования рассеянного излучения описывается статистикой Пуассона, и погрешность измерения может быть оценена как $\varepsilon = \frac{1}{\sqrt{N}}$, где N – количество зарегистрированных фотонов [2].

При измерении лидарного сигнала системой эта ошибка составлять не более 0,5%. Таким образом, из одного высотного слоя система должна регистрировать не менее $4 \cdot 10^4$ фотонов.

Количество детектируемых фотонов может быть оценено из лидарного уравнения [2]:

$$N = \frac{N_L \cdot S \cdot \Delta z \cdot T \cdot \eta \cdot \beta}{z^2} \cdot e^{-\int_0^z \alpha(z) dz}, \quad (*)$$

где N – количество фотонов, принимаемых системой; N_L – количество фотонов, излучаемых в атмосферу; S – площадь приемного зеркала телескопа; z – дистанция зондирования; T – коэффициент пропускания оптической системы; η – квантовая эффективность фотоприемника. Коэффициент обратного рассеяния молекул определяется как $\beta = \sigma \cdot n_{mol}$, где σ – дифференциальное сечение рассеяния (рэлеевского либо рамановского), n_{mol} – концентрация молекул. Коэффициент молекулярной экстинкции α связан с коэффициентом обратного рассеяния как $\alpha = \frac{8\pi}{3} \cdot \beta$.

Задача, описываемая уравнением (1), является неопределенной, поскольку это уравнение содержит произведение неизвестных $N_L \cdot S \cdot \Delta z \cdot T \cdot \eta$, а требуемое количество фотонов N может быть обеспечено как за счет увеличения энергии лазера, так и за счет увеличения апертуры телескопа, либо же за счет увеличения толщины слоя Δz .

Выбор и расчет параметров многоволнового аэрозольного лидара

При выборе параметров системы должны приниматься во внимание такие факторы, как стоимость, габариты, энергопотребление и т.д. Расчетные лидарные сигналы, обусловленные молекулярным рассеянием длин волн 355, 532, 1064 нм, показаны на рисунке. Результаты расчетов количества зарегистрированных фотонов за один лазерный импульс и соответствующие частоты счета фотонов на высотах 1 км и 10 км приведены в таблице. Частота счета v связана с количеством фотонов N как $v = \frac{N}{2\Delta z}$. В расчетах

апертура телескопа принималась равной 400 мм. Телескоп данной апертуры компактен, обеспечивает надежную регистрацию рамановского сигнала и не приводит к необоснованному удорожанию системы. Высотное разрешение $\Delta z = 100$ м позволяет профилировать аэрозольный состав атмосферы и является достаточным для большинства климатологических задач [3]. Максимальная высота зондирования предполагалась равной 10 км. Энергия излучения составляла 100 мДж на каждой длине волны. Результаты расчетов могут быть легко пересчитаны для любых других параметров системы.

Длина волны приемного канала λ , нм	Тип рассеяния	Скорость счета фотонов, принимаемых системой v , МГц		Количество фотонов, принимаемых системой N	
		1 км	10 км	1 км	10 км
355	Рэлеевское	$1,62 \cdot 10^6$	$3,05 \cdot 10^3$	$10,81 \cdot 10^5$	$2,04 \cdot 10^3$
532		$0,50 \cdot 10^6$	$1,71 \cdot 10^3$	$3,35 \cdot 10^5$	$1,41 \cdot 10^3$
1064		$0,062 \cdot 10^6$	$0,24 \cdot 10^3$	$0,41 \cdot 10^5$	$0,16 \cdot 10^3$
387	Рамановское на N_2	$1,61 \cdot 10^3$	3,38	$10,74 \cdot 10^2$	2,26
608		$0,43 \cdot 10^3$	1,51	$2,87 \cdot 10^2$	1,01

Таблица. Параметры многоволновой аэрозольной лидарной системы

При выборе параметров системы следует рассматривать сигнал рamanовского рассеяния как наиболее слабый. Из таблицы можно заключить, что на длине волны 608 нм регистрируется один фотон за импульс с высоты 10 км, т.е. для обеспечения требуемой точности измерения должны проводиться в течение 30 мин (при частоте повторения импульсов лазера 20 Гц). Такое временное разрешение достаточно для отслеживания временной эволюции вертикального распределения аэрозоля.

Помимо количества фотонов, определяемых по формуле (*), при выборе параметров системы необходимо учитывать динамический диапазон изменения лидарного сигнала (который для современных систем регистрации составляет четыре–пять порядков), а также максимальную и минимальную частоту счета фотонов фотодетектором. Лучшие фотоэлектронные умножители (ФЭУ), позволяющие проводить одновременное детектирование в аналоговом режиме и в режиме счета фотонов, надежно работают в диапазоне частот счета 10^3 – 10^2 МГц. На высоте 10 км скорость счета составляет 1,5 МГц для 608 нм, что достаточно для проведения надежных измерений (собственный шум ФЭУ составляет порядка 1 кГц). С уменьшением высоты скорость счета растет и на высоте 1 км достигает $1,6 \cdot 10^3$ МГц для 387 нм. Следует отметить, что в расчетах не рассматривалось влияние атмосферного аэрозоля. При типичных величинах экстинкции в пограничном слое атмосферы это может приводить к двукратному уменьшению сигнала. Таким образом, скорости счета фотонов в рamanовских каналах соответствуют характеристикам детектора. В то же время рэлеевский сигнал на три порядка превосходит рamanовский, поэтому для проведения измерений в каналы упругого рассеяния должны быть введены ослабляющие фильтры с коэффициентами ослабления от 100 до 1000.

В расчетах не рассматривался геометрический фактор перекрытия, характеризующий высоту, на которой лазерный пучок полностью входит в угловое поле телескопа. Для телескопа апертурой 400 мм с угловым полем 0,25 мрад эта высота составляет около 400 м при расходимости лазерного пучка 0,15 мрад. Таким образом, система позволяет производить измерения на высотах от 400 м, однако для этого часть апертуры телескопа должна быть заблокирована, чтобы обеспечить линейный режим работы ФЭУ. На рисунке показаны зависимости скорости счета фотодетектора от высоты зондирования для длин волн 355, 532, 1064 нм.

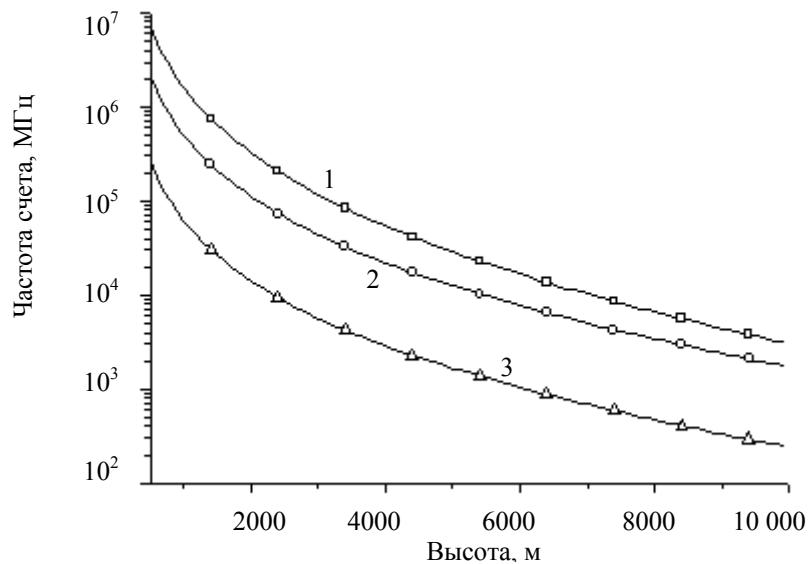


Рисунок. Зависимость частоты счета детектора от высоты зондирования для молекулярного рассеяния на длинах волн 355 нм (1), 532 нм (2) и 1064 нм (3)

Заключение

Проблема определения параметров лидарной системы является многопараметрической, и выбор конкретной конфигурации определяется спецификой поставленной задачи. В работе была проведена оценка параметров многоволнового рamanовского лидара для мониторинга аэрозоля в высотном диапазоне 1–10 км. Проведенные вычисления показывают, что для уверенного приема сигнала рamanовского рассеяния азота достаточной является система с апертурой телескопа 400 мм при энергии лазера не менее 100 мДж на длинах волн 355 и 532 нм. Для обеспечения требуемой точности на высоте 10 км измерение должно проводиться в течение 30 мин (при частоте повторения импульсов лазера 20 Гц). На основании проведенных расчетов спроектирована лидарная система, которая должна начать работу в 2012 г. в Центральной аэрологической обсерватории.

Литература

1. Ansmann A., Muller D. Lidar and atmospheric aerosol particles. – Springer Series in Optical Sciences, 2005 [Электронный ресурс]. – Режим доступа: <http://www.springerlink.com/content/j75258113x74r713/fulltext.pdf>, с. 1–510 с.
2. Межерис Р. Лазерное дистанционное зондирование: Пер. с англ. – М: Мир, 1987. – 510 с.
3. Althausen D., Muller D., Ansmann A., Wandinger U., Hube H., Clauder E., Zorner S. Scanning 6-wavelength 11-channel aerosol lidar // J. Atmos. and Oceanic Technol. – 2000. – V. 17. – P. 1469–1482.

Волков Николай Николаевич

– Московский государственный университет геодезии и картографии (МИИГАиК), аспирант, nn-volkov@yandex.ru

УДК 001.891.57

ИНВАРИАНТНАЯ МОДЕЛЬ ДЛЯ ОЦЕНКИ КОЭФФИЦИЕНТА ПРОПУСКАНИЯ АТМОСФЕРЫ ПРИ МОНИТОРИНГЕ ОБЪЕКТОВ В ОПТИЧЕСКОМ ДИАПАЗОНЕ СПЕКТРА

А.В. Демин, М.И. Моисеева

Рассматривается задача учета влияния атмосферы на пропускание излучения в среднем и дальнем диапазоне инфракрасного спектра с целью повышения достоверности результатов при мониторинге летательных аппаратов. Предложена аналитическая апостериорная модель пропускания излучения атмосферой. Показана общая структура такой модели. Приводятся результаты идентификации значения коэффициента пропускания атмосферы на основе экспериментальных данных.

Ключевые слова: аналитическая модель, идентификация, коэффициент пропускания атмосферы, мониторинг объектов, экспериментальные данные.

Введение

Проблема достоверности результатов при дистанционном зондировании среды в среднем и дальнем диапазоне инфракрасного (ИК) спектра излучения ($\text{ИК}_{\text{Сид}}$) является актуальной при решении задачи пеленгации высокоскоростных летательных аппаратов (ЛА). Зондирование среды представляет собой процесс определения состояния среды по измеренным параметрам с целью выявления в ней отклонений от некоего «эталона», характеризуемого совокупностью физико-химических и оптико-физических параметров. Отклонения среды от «эталона» возникают как при наличии в среде объектов, отличных от нее (например, летательный аппарат), так и вследствие техногенного вмешательства.

На достоверность определения пеленгационных характеристик ЛА при его мониторинге влияет атмосфера. Последнее особенно сказывается при пеленгации высокоскоростных ЛА, которая возможна только в $\text{ИК}_{\text{Сид}}$. Это объясняется их температурой, лежащей в пределах 270–1000 К.

Для учета состояния атмосферы, характеризуемого коэффициентом пропускания (входящим в уравнение пеленгации), необходимо наличие базы данных со сведениями об атмосфере при любых возможных условиях. Последнее практически невозможно на момент мониторинга без соответствующей аппаратуры. На рис. 1 изображена структурная схема процесса мониторинга объекта с помощью оптико-электронных систем. В этом случае общим уравнением для измеряемой яркости будет $\Omega_{\text{изм}} = I_{\text{атм}} \circ I_{\text{объекта}}$, где $I_{\text{атм}}$ и $I_{\text{объекта}}$ – функции, характеризующие яркость атмосферы и объекта мониторинга; \circ – композиция функций. Учет атмосферной компоненты $I_{\text{атм}}$ особенно важен тогда, когда необходима высокая вероятность принятия верного решения.

Известны работы [1, 2], посвященные измерениям коэффициента пропускания атмосферы в различных спектральных диапазонах. Следует заметить, что имеющиеся практические результаты получены при конкретных условиях, которые в общем случае могут отличаться от текущих. В настоящей работе предлагается получение апостериорной аналитической модели для учета коэффициента пропускания атмосферы $\tau_a(\lambda)$, инвариантной относительно условий наблюдения.

Аналитическая модель для оценки коэффициента пропускания атмосферы

При построении аналитической модели для оценки коэффициента пропускания атмосферы в $\text{ИК}_{\text{Сид}}$ необходимо получить расчетное соотношение для прохождения ИК излучения в атмосфере.

Известно, что аналитическая модель экспериментальных данных, полученных в результате исследования определенного физического явления или технического объекта, может быть представлена композицией функций вида

$$Z = F_1(r_1, r_2, \dots, r_k) \circ F_2(S_1, S_2, \dots, S_m),$$

где r_1, r_2, \dots, r_k – внутренние параметры, характеризующие явление, а S_1, S_2, \dots, S_m – внешние (рис. 2). Соответственно, компонента $F_1(r_1, r_2, \dots, r_k)$ описывает влияние внутренних параметров и представляет собой общую часть искомой модели. Вторая компонента – $F_2(S_1, S_2, \dots, S_m)$ – характеризует воздействие внешних параметров.

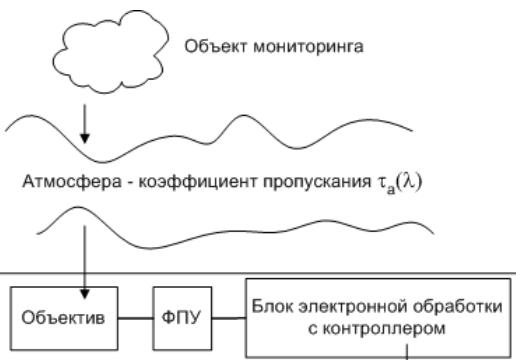


Рис. 1. Структурная схема процесса мониторинга объекта

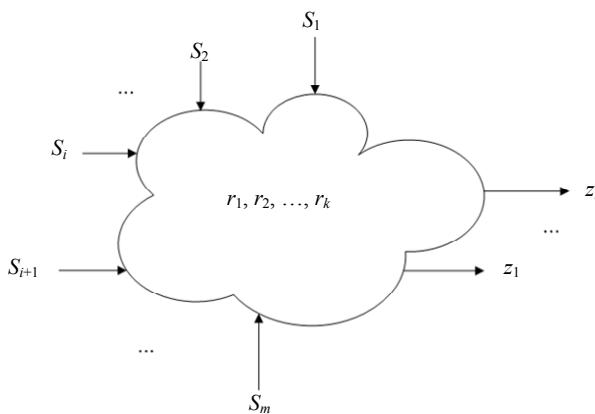


Рис. 2. Параметры, описывающие физическое явление

При получении аналитической модели не будем учитывать возможную зависимость состояния объекта от выходных параметров, т.е. предполагаем отсутствие обратной связи. Кроме того, будем считать, что все исходные эмпирические данные описывают одно и то же явление, которое с точки зрения физики должно подчиняться какому-то единому закону. Несмотря на то, что аналитическая модель строится апостериорно – на основе экспериментальных данных, полученных при различных условиях, – она должна содержать некоторые общие для всех ситуаций составляющие, учитывающие единый физический закон явления (см. рис. 1). Этим объясняется наличие компоненты F_1 . Компонента F_2 необходима для корректировки результатов с учетом различных внешних условий.

Известны [1, 2] экспериментальные данные о значении коэффициента пропускания атмосферы $\tau_a(\lambda)$ для двух спектральных диапазонов в инфракрасной области – [3; 5,2] и [8; 14] мкм. Результаты, представленные графически на рис. 3, получены для пяти климатических зон при различных значениях метеорологической дальности видимости d .

Различия форм графиков для разных участков диапазонов спектра, наличие множества локальных экстремумов обусловлены сложной физической природой явления пропускания атмосферы. Пропускание атмосферы определяется поглощением составляющих атмосферу газов и рассеянием на частицах, молекулах и аэрозолях. Для диапазона ИК излучения существенное влияние оказывает присутствующий в атмосфере водяной пар. На рис. 4 показаны спектры пропускания озона (O_3), углекислого газа (CO_2) и водяных паров (H_2O). Для сравнения показан солнечный спектр, наблюдаемый с земной поверхности [3].

Метод практического определения коэффициента пропускания атмосферы описан в [4] и применим при известных параметрах атмосферы: температуре, метеорологической дальности видимости и относительной влажности. Данный метод обременителен при мониторинге объектов, поскольку требует наличия дополнительных измерительных систем. Кроме того, для его применения необходимо наличие базы данных, содержащей сведения о поглощении газов в различных спектральных диапазонах, о коэффициенте рассеяния атмосферы и другие сведения. Для оценки коэффициента пропускания предлагается построение апостериорной аналитической модели в виде зависимости

$$\tau_a(\lambda) = f_1(\lambda) \circ f_2(\varphi, \xi),$$

где $f_1(\lambda)$ – функция пропускания в зависимости от длины волны излучения для среднестатистического физико-химического состава атмосферы; $f_2(\phi, \xi)$ – функция, учитывающая геофизический и климатический фактор (географическое место, его физические особенности и время года – например, Арктика и летний период).

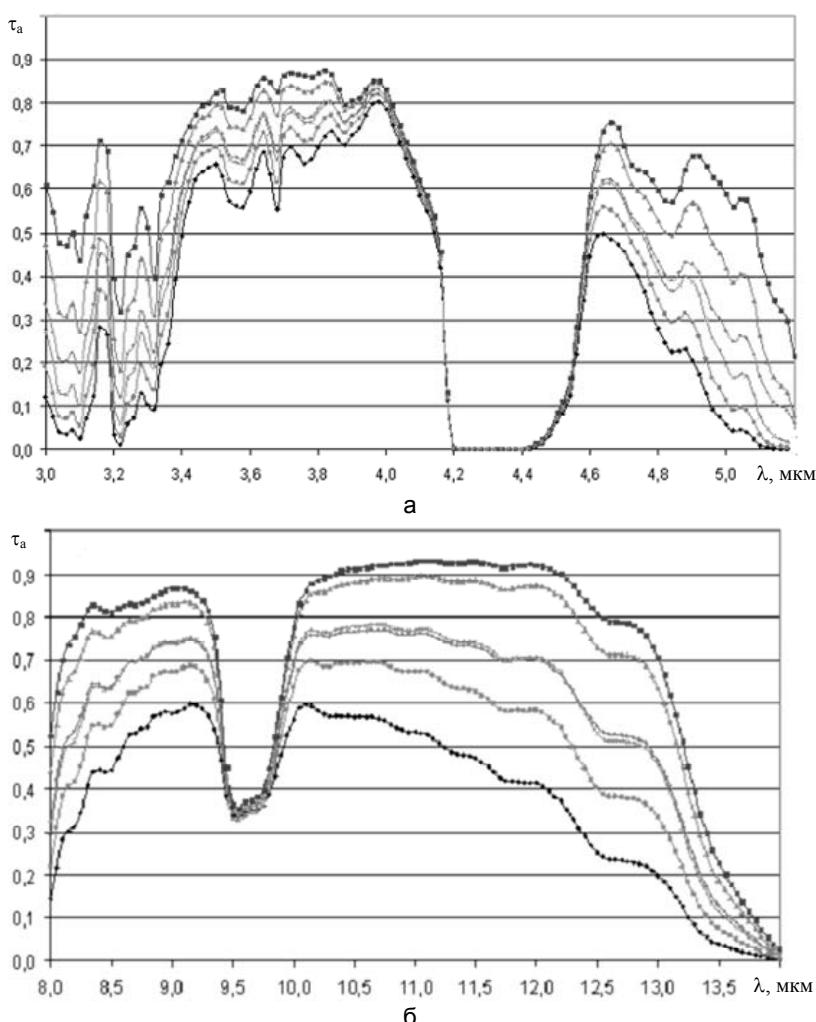


Рис. 3. Значения коэффициента пропускания атмосферы для $d = 5$ км в диапазоне длин волн [3; 5,2] мкм (а) и [8; 14] мкм (б). Сверху вниз показаны данные для климатических зон: Арктика зимой; средние широты зимой; усредненные данные; Арктика летом; средние широты летом; тропики

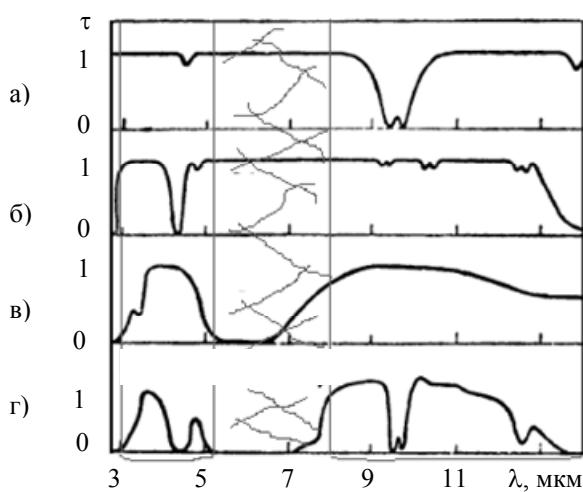


Рис. 4. Спектральное пропускание озона (а), углекислого газа (б), водяных паров (в) и атмосферы (г)

Исходные экспериментальные данные известны для разных метеорологических дальностей видимости и климатических зон. Для получения инвариантной относительно условий наблюдения зависимости усредним все исходные данные по величине дальности видимости d . Для удобства математического представления спектральной характеристики коэффициента пропускания $\tau_a(\lambda)$ и повышения точности вычислений выделим в спектре характерные участки. Подобный подход предложен авторами в [5]. Это позволяет использовать для каждого участка наименее приемлемую эмпирическую зависимость.

Вид эмпирической зависимости определим с помощью регрессионного анализа и перебора различных известных функций. Эмпирические коэффициенты для каждой климатической зоны получены на основе метода наименьших квадратов.

Критерием достоверности получаемой зависимости будем считать коэффициент детерминации R^2 , превышающий значение 0,9.

Результаты построения модели

Диапазон [3; 5,2] мкм был разделен на три участка (I, II, III) по точкам локальных минимумов (подробнее принцип деления представлен в [5]). Диапазон [8; 14] мкм был разделен на участки IV, V, VI по точкам локальных максимумов. Графически этот этап представлен на рис. 5.

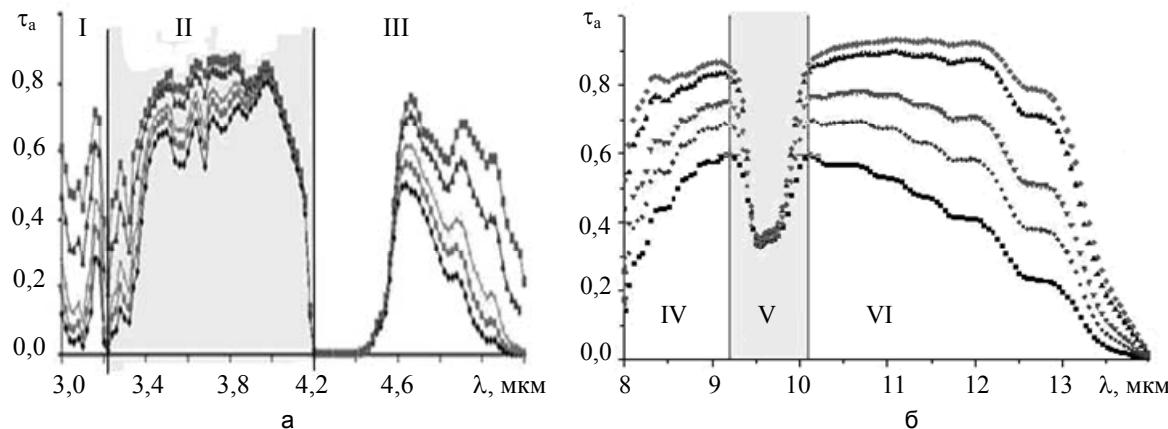


Рис. 5. Разбиение спектральных диапазонов на изотропные участки: диапазон длин волн [3; 5,2] мкм (участки I, II, III) (а); диапазон длин волн [8; 14] мкм (участки IV, V, VI) (б)

Для математического представления спектрального коэффициента пропускания на выделенных участках длин волн выбраны полином пятой степени

$$\tau(\lambda) = A_0 + A_1 \cdot \lambda + A_2 \cdot \lambda^2 + A_3 \cdot \lambda^3 + A_4 \cdot \lambda^4 + A_5 \cdot \lambda^5 \quad (1)$$

где $A_0, A_1, A_2, A_3, A_4, A_5$ – константы, и функция вида

$$\tau(\lambda) = \tau_0 + A \cdot \frac{1}{\frac{\lambda - \lambda_c + w_1/2}{w_2}} \cdot \left(1 - \frac{1}{1 + e^{-\frac{\lambda - \lambda_c - w_1/2}{w_3}}} \right), \quad (2)$$

иногда называемая asymmetric double sigmoid, где τ_0 – горизонтальная асимптота; A – амплитуда; λ_c – абсцисса центра «пика» функции; w_1, w_2 и w_3 – параметры, влияющие на ширину и «форму» пика кривой.

На каждом из шести выделенных участков спектрального диапазона применима только одна из представленных зависимостей. Зависимость (1) применена на первых двух (I, II) участках диапазона [3; 5,2] мкм, а зависимость (2) – для участка III диапазона [3; 5,2] мкм и на всех участках (IV, V, VI) диапазона длин волн [8; 14] мкм. Участки IV и VI аппроксимируются одной формулой. Вычисленные значения эмпирических коэффициентов в формулах приведены в табл. 1 и табл. 2. При этом величины коэффициента детерминации составили от 0,939 до 0,955 для полиномиальной аппроксимации (1) и от 0,975 до 0,997 для аппроксимации asymmetric double sigmoid (2), что показывает высокую достоверность полученных результатов.

Отметим, что математическое представление соотношения (1) – более простое, но коэффициенты аппроксимации должны иметь 15 значащих цифр и не несут физического смысла. Соотношение (2) больше соответствует реальным физическим законам пропускания атмосферы, эмпирические коэффициенты могут быть подсчитаны с точностью до четырех значащих цифр.

Коэффициенты	Арктика. Лето	Тропики	Арктика. Зима	Средние широты. Лето	Средние широты. Зима
	I: [3; 3,2] мкм				
A_0	19879555,8171565	19875345,1528570	19878738,8475666	19877752,8039432	19880602,3121815
A_1	-32259735,6738831	-32255664,8883141	-32258955,5542238	-32257992,0433420	-32260752,0135739
A_2	20937847,8100510	20936536,2672222	20937599,8373558	20937285,8598364	20938176,8571623
A_3	-6794064,3632233	-6793923,5474454	-6794038,1152467	-6794004,0082322	-6794099,8705824
A_4			1102181,0		
A_5			-71514,0		
R^2	0,976	0,935	0,940	0,972	0,969
II: [3,22; 4,22] мкм					
A_0	14168,7544688900	14169,8581257862	14173,0513145934	14168,8873745887	14170,5451200000
A_1	-19903,7802491497	-19904,5253558671	-19905,8000394155	-19903,9297913745	-19904,5681000000
A_2	11145,3869141245	11145,5017267942	11145,6249282324	11145,4149164221	11145,4731300000
A_3			-3110,5230		
A_4			432,7770		
A_5			-24,020		
R^2	0,952	0,939	0,947	0,946	0,955

Таблица 1. Эмпирические коэффициенты для соотношения (1).
Цифрами I–II обозначены соответствующие участки спектра

Коэффициенты	Арктика. Лето	Тропики	Арктика. Зима	Средние широты. Лето	Средние широты. Зима
	III: [4,24; 5,2] мкм				
y_0			-0,002		
x_c	4,722	4,653	4,862	4,687	4,803
A	0,708	0,801	0,718	0,741	0,693
w_1	0,313	0,159	0,594	0,234	0,476
w_2			0,024		
w_3			0,126		
R^2	0,982	0,990	0,976	0,987	0,975
IV: [8; 9,15], VI: [10,15; 14] мкм					
y_0	0,754	0,599	0,856	0,690	0,830
x_c			9,643		
A	-0,412	-0,259	-0,500	-0,360	-0,484
w_1			0,496		
w_2			0,036		
w_3			0,060		
R^2	0,996	0,997	0,997	0,996	0,997
V: [9,2; 10,1] мкм					
y_0			-0,198		
x_c			10,566		
A	0,970	0,771	1,113	0,889	1,078
w_1	5,320	4,980	5,675	5,124	5,580
w_2	0,292	0,203	0,307	0,248	0,326
w_3	0,506	0,803	0,339	0,623	0,372
R^2	0,989	0,979	0,987	0,989	0,987

Таблица 2. Эмпирические коэффициенты для соотношения (2).
Цифрами III–VI обозначены соответствующие участки спектра

Построенные по предложенным эмпирическим зависимостям спектральные характеристики пропускания атмосферы показаны на рис. 6.

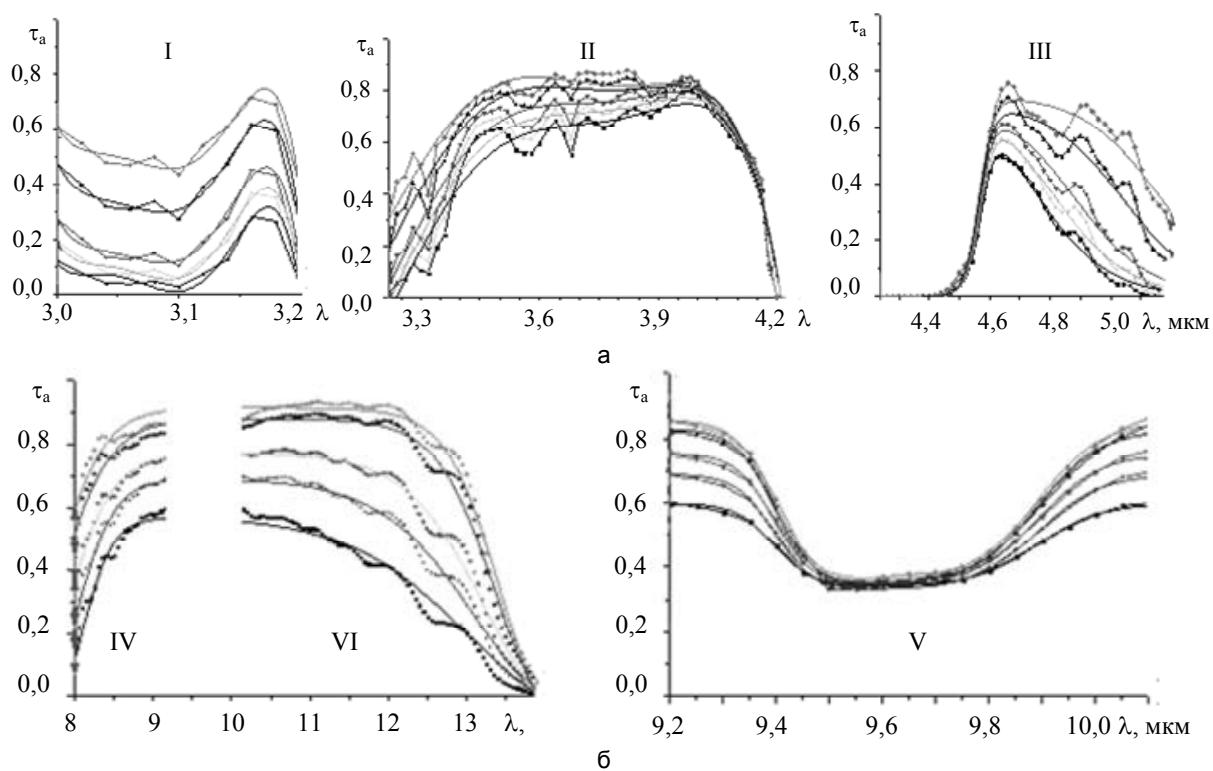


Рис. 6. Расчетные зависимости $\tau_a(\lambda)$ для диапазонов [3; 5,2] мкм (а) и [8; 14] мкм (б) на выделенных участках диапазонов длин волн. Сверху вниз показаны данные для климатических зон: Арктика зимой; средние широты зимой; усредненные данные; Арктика летом; средние широты летом; тропики (цифрами I–VI обозначены соответствующие участки спектра)

На сегодняшний день учет географических и климатических условий представлен в модели различными значениями коэффициентов в эмпирической формуле (табл. 1, 2). Задача поиска компоненты $f_2(\phi, \xi)$ модели, учитывающей внешние факторы, будет решена на последующих этапах работы.

Заключение

Рассмотрена проблема учета внешних условий при мониторинге среды на примере явления пропускания атмосферы в среднем и дальнем ИК диапазоне спектра излучения. Предложен алгоритм разбиения множества известных экспериментальных данных в диапазонах длин волн [3; 5,2] мкм и [8; 14] мкм на группы для представления их аналитическими зависимостями. Предложена инвариантная по отношению к дальности видимости аналитическая модель оценки коэффициента пропускания атмосферы при мониторинге объектов. Полученные результаты полезны для создателей тепловизионной техники.

Литература

- Ллойд Дж. Системы тепловидения: Пер. с англ. – М.: Мир, 1978. – 417 с.
- Справочник по инфракрасной технике / Под ред. У. Волф, Г. Цисис. Т.1. Физика инфракрасного излучения: Пер. с англ. – М.: Мир, 1995. – 606 с.
- Джемисон Дж.Э., Мак-Фи Р.Х., Пласс Дж.Н. и др. Физика и техника инфракрасного излучения. – М.: Советское радио, 1965. – 642 с.
- Госсорт Ж. Инфракрасная термография. Основы, техника, применение: Пер. с франц. – М.: Мир, 1988. – 416 с.
- Демин А.В., Моисеева М.И. Представление экспериментальных данных об оптических параметрах атмосферы аналитическими зависимостями // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 2 (72). – С. 90–95.

Демин Анатолий Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, dav_60@mail.ru

Моисеева Мария Игоревна

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, mim14@mail.ru

УДК 681.784.88, 535.36

ИССЛЕДОВАНИЕ СПЕКТРА МОЛЕКУЛЫ ДНК

В ТЕРАГЕРЦОВОЙ ОБЛАСТИ ЧАСТОТ

М.В. Цуркан, Е.А. Собакинская, О.А. Смолянская, В.Г. Беспалов, В.Л. Вакс, Н.С. Балбекин

Проведено исследование спектра деградированной ДНК в диапазоне частот 0,1–1,3 ТГц. Определены собственные линии поглощения ДНК, идентифицированы линии поглощения, отвечающие за взаимодействие с водой и конформационные изменения.

Ключевые слова: терагерцовое излучение, спектроскопия, спектр ДНК, линии поглощения.

Введение

Спектроскопический анализ биомолекул в терагерцовой (ТГц) области частот привлекает внимание исследователей во всем мире [1–5]. Связано это, прежде всего с тем, что диэлектрическая функция макромолекул в ТГц диапазоне сформирована низкочастотными колебаниями, представляющими собой коллективные движения больших групп атомов, образующих молекулу. Такие колебания очень чувствительны к внутримолекулярной структуре вещества и дают информацию о его конформационных состояниях. Эта информация крайне важна для биологии, так как конформационная гибкость биомолекул (ДНК, РНК, белки, витамины, и т.д.) напрямую связана со способностью молекулы изменять свою структуру, и поэтому играет важную роль в биохимических реакциях.

Методика, позволяющая определять возможные конформационные состояния, выявлять мутации, а также влиять на ход биохимических реакций, была бы крайне востребована, прежде всего, для анализа молекулы ДНК, являющейся основой многих химических процессов, происходящих в клетке. Спектр этой молекулы и ее компонентов в дальнем инфракрасном и ТГц диапазонах являлся предметом исследования нескольких работ. Существование специфических спектральных особенностей при изучении спектра нуклеотидов и нуклеозидов методами ТГц продемонстрировано в работах [1, 2]. Было показано, что для каждого из 4 оснований существуют специфические спектральные особенности как для коэффициента поглощения, так и для коэффициента отражения [2]. Найденные полосы поглощения были идентифицированы как резонансы водородных связей, ответственных за агрегацию молекул в микрокристаллическое состояние. Влияние степени гидратации ДНК на характеристики АТ и GC пар стало предметом исследования [3]. Регистрация спектра ДНК в растворе с водой была продемонстрирована в работах [4, 5].

В работе [6] получены линии поглощения и показатели преломления пленок ДНК сельди. В частности, было показано, что спектры гибридизованных и денатурированных молекул хорошо различаются между собой, что в принципе позволяет с помощью регистрации характерных полос поглощения различать состояния ДНК. Ключевым моментом такой методики является точное знание частот образца. Фактически информация о собственных частотах ДНК, а также о частотах связей этой молекулы с другими важными биовеществами (вода, белки) может дать новый импульс развитию других приложений: диагностика функционального состояния ДНК, выявление мутированных образцов, управление химическими реакциями.

В связи с этим целью данной работы стало подробное изучение спектра ДНК в диапазоне 0,1–1,3 ТГц, что включало в себя определение характерных собственных частот, а также линий поглощения, соответствующих связям молекулы с различными веществами.

Описание подготовки образцов и методики измерений

Для проведения исследований использовалась деградированная ДНК сельди (Канада) в виде сухого порошка. ДНК сельди – один из самых распространенных объектов исследования, что позволяет соопределять полученные данные с известными данными из литературы. Образцы ДНК приготавливались в виде пленок на двух различных подложках (полиэтиленовой и кварцевой), которые обеспечивали хорошее пропускание в ТГц диапазоне. В качестве растворителя использовались ректифицированный этиловый спирт и дистиллированная вода. Порошок ДНК смешивался в различных концентрациях K ($K_1 > K_2 > K_3$) с растворителем и помещался на подложку, где оставался высыхать в атмосферных условиях до появления твердой пленки.

Первый этап проведения измерений заключался в приготовлении образцов. Затем регистрировался спектр очищенной подложки. Для этого подложка очищалась ректифицированным спиртом, промывалась в дистиллированной воде и затем высушивалась. Далее проводилось измерение спектра образцов ДНК. Для проверки воспроизводимости результатов регистрация спектра для каждого образца проводилась 3 раза. Измерения проведены для различных концентраций ДНК.

Схема установки

В работе использовалась схема, представленная на рис. 1. Пучок накачки от фемтосекундного Yb:KYW-лазера Solar FL-1 ($\lambda = 800$ нм; $t_{\text{имп}} = 45$ фс; $f = 80$ МГц; $P = 1,5$ Вт), пройдя через моторизированную линию оптической задержки, попадает на полупроводниковый кристалл арсенида индия, помещенный в постоянное магнитное поле, который генерирует ТГц излучение. Получаемое ТГц излучение обладает следующими параметрами: спектральный диапазон 0,1–1,3 ТГц, средняя мощность 48 мкВт, длительность импульса 2,5 пс. Генерируемое излучение после прохождения фильтра из тефлона, отсекающего диапазон длин волн, меньших 50 мкм, попадает на ТГц поляризатор, выделяющий горизонтальную составляющую. Поляризованное излучение проходит сквозь образец, обладающий некоторым амплитудно-фазовым пропусканием. Образец фиксируется в вертикальном положении в фокусной плоскости, перпендикулярной лучу, с помощью двухкоординатного предметного столика, и при прохождении сквозь него происходит дифракция ТГц излучения. При одновременном попадании пробного пучка фемтосекундного излучения и пучка ТГц излучения на электрооптический кристалл CdTe ТГц импульс в кристалле наводит двулучепреломление для пробного пучка вследствие электрооптического эффекта. Величина двулучепреломления прямо пропорциональна напряженности электрического поля ТГц волны в данной временной точке $E(t)$. Схема измерения двулучепреломления состоит из четвертьвольновой пластины, призмы Волластона, балансного фотодетектора собственной разработки (диоды BPW34, микросхема усилителя OP177 фирмы Analog Devices) и синхронного усилителя LIA-MV-150 фирмы Femto, управляемого от оптико-механического модулятора MC-1000A фирмы Thorlabs, помещенного в пучок накачки. В отсутствие ТГц излучения пробный пучок не испытывает двулучепреломления и после прохождения четвертьвольновой пластины, превращающей горизонтальную поляризацию пучка в круговую, и призмы Волластона разделяется на два пучка с ортогональными поляризациями одинаковой интенсивности, при этом с балансного фотодетектора не регистрируется сигнал. При наведении ТГц импульсом двулучепреломления в пробном пучке пучки после призмы Волластона имеют уже не одинаковую интенсивность, и на балансном фотодетекторе регистрируется сигнал рассогласования. При изменении полярности ТГц импульса меняется и знак двулучепреломления, что соответственно вызывает изменение полярности сигнала с фотодетектора. С помощью линии оптической задержки изменяется время пересечения ТГц импульса и импульса пробного пучка в кристалле. Таким образом, измеряя при различных задержках сигнал рассогласования, можно измерить зависимость амплитуды ТГц излучения от времени $E(t)$. Для увеличения соотношения сигнал/шум в схеме используется синхронный усилитель. Отфильтрованный и усиленный сигнал передается в компьютер посредством цифрового вольтметра NI-4065 фирмы National Instruments. Схема регистрации установки обеспечивает частотную погрешность меньше 5,5 ГГц.

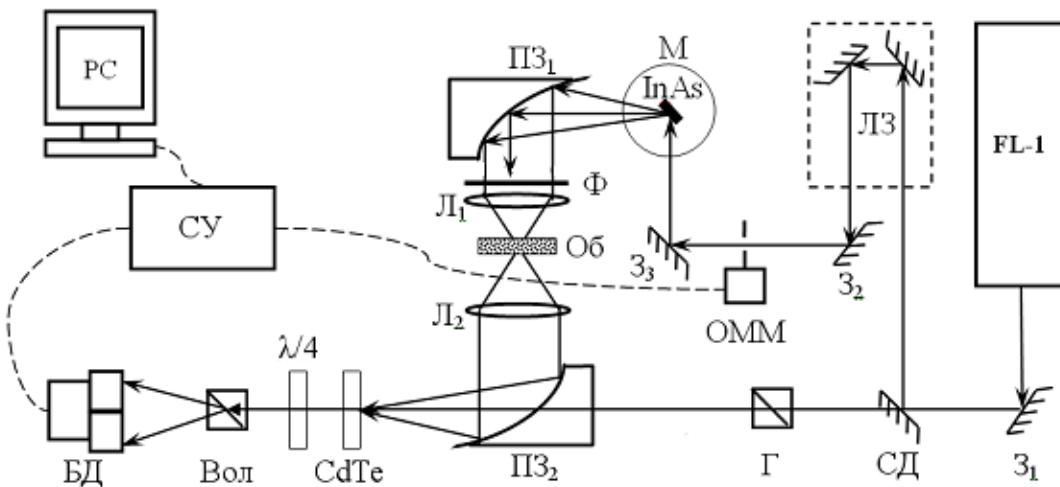


Рис. 1. Схема терагерцового спектрофотометра TS-5: FL-1 – Yb:KYW-лазер фемтосекундных световых импульсов; Z_1 , Z_2 , Z_3 – плоские зеркала; СД – светоделитель; ЛЗ – линия задержки; ОММ – оптико-механический модулятор; InAs – кристалл; М – постоянный магнит; Г – призма Грана; ПЗ₁ и ПЗ₂ – параболические зеркала; Ф – фильтр; Л₁ и Л₂ – линзы; Об – объект; CdTe – электрооптический кристалл; $\lambda/4$ – фазовая пластина; Вол – призма Волластона; БД – балансный детектор; СУ – синхронный усилитель; РС – компьютер

Программное обеспечение для спектрофотометра осуществляется через виртуальный прибор, созданный в среде разработки LabView, управляющий задержкой и процессом измерения и позволяющий осуществлять съемку временного профиля ТГц импульса. На дисплее компьютера отображается кривая

$E(t)$ ТГц излучения. Полученные профили полей проходят предварительную обработку в программе Origin 6.0, конечный вид спектра рассчитывается в программе Mathcad 13.

Анализ данных

Некоторые зарегистрированные спектры для разных подложек и растворителей показаны на рис. 2–4. На графиках по оси ординат отложена интенсивность сигнала в относительных единицах, а по оси абсцисс – частота. Прямоугольниками выделены линии поглощения образцов ДНК.

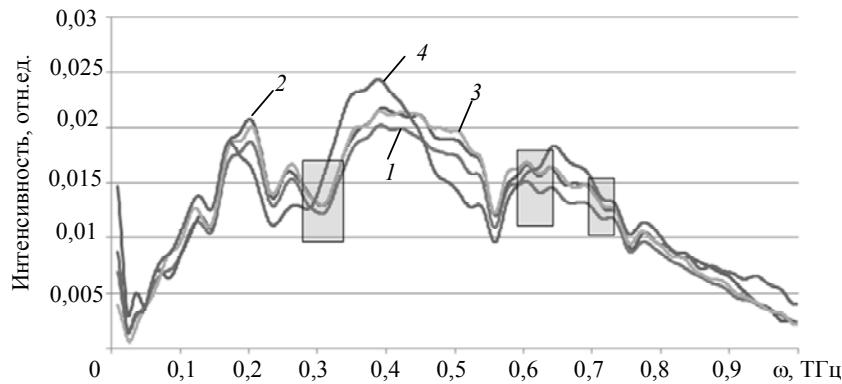


Рис. 2. Спектр образцов ДНК в растворе с водой, концентрация K_2 , на кварцевой подложке:
1 – первое измерение образца; 2 — второе измерение образца; 3 – третье измерение образца;
4 – спектр подложки. Прямоугольниками выделены полосы поглощения на 0,300; 0,625 и 0,712 ТГц

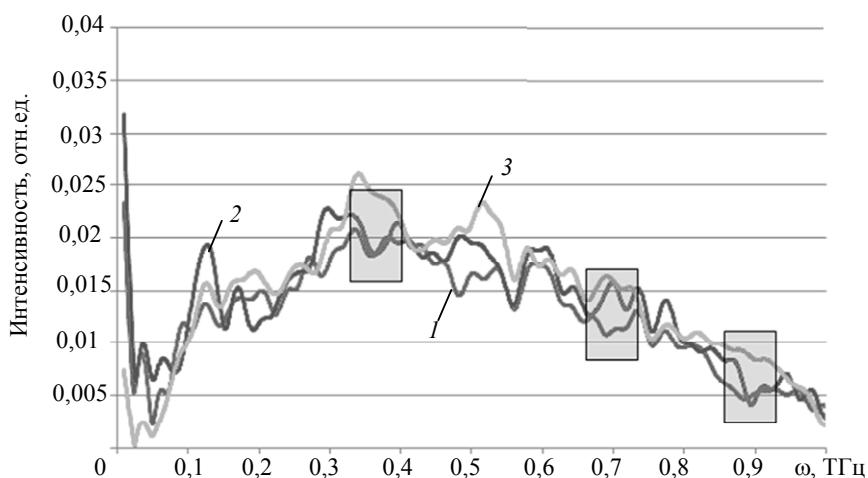


Рис. 3. Спектр образцов ДНК в растворе с водой, концентрация K_2 , на полиэтиленовой подложке:
1 – первое измерение образца; 2 – второе измерение образца; 3 – спектр подложки
Серыми прямоугольниками выделены полосы поглощения на 0,368; 0,711 и 0,900 ТГц

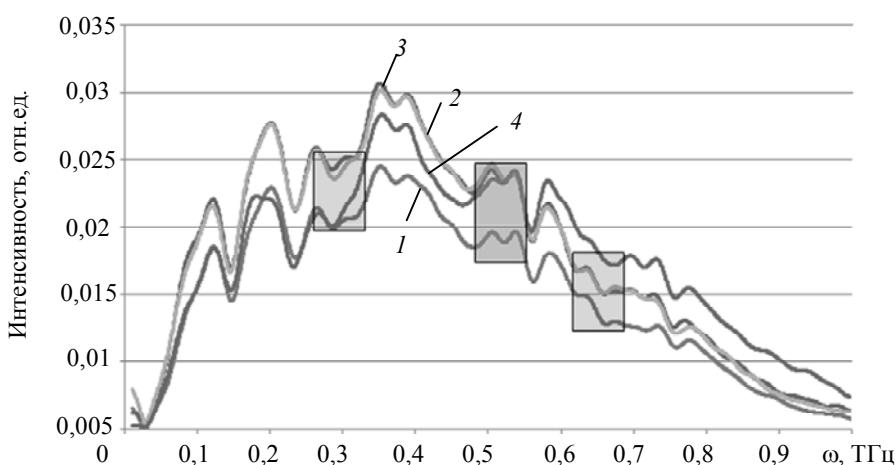


Рис. 4. Спектр образцов ДНК в растворе со спиртом, концентрация K_1 , на полиэтиленовой подложке:
1 – первое измерение образца; 2 – второе измерение образца; 3 – третье измерение образца; 4 – спектр подложки. Серыми прямоугольниками выделены полосы поглощения на 0,315; 0,519 и 0,662 ТГц

Определение характерных линий поглощения ДНК осуществлялось по зарегистрированным спектрам для разных подложек и растворителей. Критериями отбора служили отличия линий поглощения образца от спектральных линий подложки в центральной частоте, интенсивности и профиле. Все найденные линии поглощения образцов сведены в таблицу.

Подложка	Частоты линий поглощения образцов ДНК с водой, ТГц	Частоты линий поглощения образцов ДНК со спиртом, ТГц
Полиэтиленовая подложка	0,368; 0,623; 0,667; 0,711; 0,900	0,315; 0,519; 0,662
Кварцевая подложка	0,300; 0,481; 0,625; 0,712	0,415; 0,520; 0,665; 0,712

Таблица. Зарегистрированные линии поглощения образцов ДНК, усредненные по нескольким данным

Полученные данные были проанализированы путем сопоставления со спектрами бактериальной ДНК в растворе с водой [4], спектрами ДНК сельди в твердом состоянии [6] и спектрами воды [7]. Линии поглощения ДНК, зарегистрированные в работе [6] с помощью Фурье-спектрометра IFS-66 (спектральное разрешение порядка 6 ГГц), были получены без использования растворителей. Близость полученных авторами частот к результатам работы [6] считалась определяющей для собственных линий поглощения ДНК сельди. Таким образом, в настоящей работе следующие частоты были идентифицированы как собственные частоты ДНК: 0,368; 0,623 (0,625); 0,662 (0,665) ТГц.

Линии поглощения 0,315; 0,415; 0,519 (0,520); 0,711 (0,712) ТГц, близкие к линиям поглощения бактериальной ДНК, были отнесены к конформационным колебаниям в связи с тем, что в работе [4] они были получены для другого типа ДНК с разными растворителями. Линия поглощения на частоте 0,667 ТГц, полученная с использованием воды в качестве растворителя, возможно, отвечает за взаимодействие с водой, так как совпадает с линией поглощения бактериальной ДНК, также растворенной в воде [4]. Две линии поглощения на частотах 0,300 и 0,900 ТГц не были идентифицированы.

Заключение

С помощью спектрофотометра ТГц диапазона проведено исследование спектра деградированной ДНК сельди в диапазоне 0,1–1,3 ТГц. Исследования проведены для образцов в твердой фазе на подложках из полиэтилена и кварца. В результате исследования идентифицированы собственные линии поглощения ДНК. Впервые для ДНК сельди получены частоты, отвечающие за конформационные связи молекулы и за колебания связей ДНК с водой. Таким образом, в отличие от имеющихся в литературе работ по спектрам ДНК, в настоящей работе получены как собственные частоты образца, так и частоты конформационных колебаний и связей с растворителем.

Фактически такая информация является основой для создания методик диагностики состояния ДНК (структура, наличие мутаций) по собственным частотам и управления биохимическими реакциями через активацию/дезактивацию связей с окружающими веществами.

Однако для непосредственного применения в указанных приложениях полученные данные нуждаются в уточнении. В связи с этим следующим этапом работы станет исследование спектров ДНК вблизи полученных частот с более высоким спектральным разрешением, что даст возможность провести тщательный анализ результатов по профилю линии поглощения, сдвигу центральной частоты и изменению интенсивности. Такой анализ позволит уточнить центральные частоты и подтвердить идентификацию линий, что в конечном итоге обеспечит получение достоверных данных.

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (ГК № 16.513.11.3070 и ГК № 14.740.12.0841).

Литература

- Markelz A.G., Roitberg A., Heilweil E.J. Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz // Chemical Physics Letters. – 2000. – V. 320. – P. 42–48.
- Fischer B.M. Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy // Phys. Med. Biol. – 2002. – V. 47. – P. 3807–3814.
- Semenov M., Bolbukh T., Maleev V. Infrared study of the influence of water on DNA stability in the dependence on AT/GC composition // Journal of Mol. Structure. – 1997. – V. 408/409. – P. 213–217.
- Globus T., Khromova T., Gelmont B., Woolard D., Tamm L.K. Terahertz characterization of dilute solutions of DNA // Proc.of SPIE. – 2006. – V. 6093. – P. 609308-1–609308-12.
- Cherkasova O.P., Nazarov M.M., Shkurinov A.P. and V.I. Fedorov. Terahertz spectroscopy of biological molecules // Radiophysics and Quantum Electronics. – 2009. – V. 52. – № 7. – C. 518.
- Parthasarathy R., Globus T., Khromova T., Swami N., Woolard D. Dielectric properties of biological molecules in the Terahertz gap // Appl. Phys. Lett. – 2005. – V. 87. – P. 113901-1–113901-3.

7. Globus T., Anwar Mehdi, DeMaria Anthony J., Shur Michael S. Low-Terahertz spectroscopy of liquid water // Terahertz Physics, Devices and Systems II // Proc. of SPIE. – 2007. – V. 6772. – P. 67720S-1–67720S-11.

Цуркан Мария Валерьевна	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, мл. научный сотрудник, tsurkan.maria@yandex.ru
Собакинская Екатерина Александровна	– Институт физики микроструктур РАН, инженер, katja@ipm.sci-nnov.ru
Смолянская Ольга Алексеевна	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, доцент, o_smolyanskaya@mail.ru
Беспалов Виктор Георгиевич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, victorbespaloff@gmail.com
Вахс Владимир Лейбович	– Институт физики микроструктур РАН, кандидат физ.-мат. наук, ст. научный сотрудник, зав. отделом, vax@ipm.sci-nnov.ru
Балбекин Николай Сергеевич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, Godnikolay@yandex.ru

УДК 004.932

КОМБИНИРОВАННЫЙ МЕТОД ПОДСЧЕТА ЭРИТРОЦИТОВ НА ИЗОБРАЖЕНИЯХ МАЗКОВ КРОВИ

А.В. Дырнаев, А.С. Потапов

Рассмотрена задача подсчета эритроцитов на изображениях препаратов крови, полученных с помощью микроскопа. Предложен метод сегментации таких изображений, основанный на детектировании границ методом Канни с последующим выявлением окружностей на изображении с помощью преобразования Хафа. В методе используется уточнение результатов сегментации на основе бинарной маски объектов и фона. Метод устойчив к изменениям структуры, текстуры и цвета клеток, а также позволяет проводить подсчет клеток, в том числе перекрывающих друг друга, в условиях помех на изображении.

Ключевые слова: подсчет числа клеток, цифровая микроскопия, сегментация изображений.

Введение

В задачах анализа изображений, полученных с помощью микроскопа, в рамках цитологических исследований часто возникает необходимость подсчета числа клеток определенного типа. При исследовании препаратов крови важной задачей является подсчет числа эритроцитов, исходя из показателей которого можно диагностировать нарушения в кровообразовании или повреждения эритроцитов вследствие различных факторов.

Среди методов подсчета эритроцитов в крови можно выделить методы, использующие уже смоделированную базу изображений клеток крови и их характеристик [1–4], методы, основанные на пороговом разложении [5] или методы, использующие сегментацию методом управляемого водораздела [6]. Существуют также подходы, в которых для сегментации используются цветовые характеристики изображения [7, 8] или текстурные характеристики [8]. В [9] предложено использовать алгоритм активной контурной модели для выделения контуров клеток.

В работе [10] было предложено на изображении контуров клеток, полученных детектором границ Канни, определять центры кривизны контуров, которые, в свою очередь, являлись предполагаемыми центрами клеток.

В настоящей работе предлагается улучшение разработанного ранее метода путем использования преобразования Хафа [11] совместно с сегментированием изображения мазка крови на фон и объекты вместо определения кривизны контуров. Для выделения контуров применяется алгоритм Канни [12], для более успешной работы которого используется масштабирование изображения.

Метод выделения и подсчета эритроцитов

В предлагаемом методе на первом этапе выполняется масштабирование изображения. Коэффициент масштабирования определяется как отношение размера клетки, оптимального для проведения сегментации, в пикселях к реальному ее размеру. При масштабировании используется бикубическая интерполяция. Этот шаг необходим в связи с тем, что изображения могут быть получены при разных увеличениях, а реализация последующих шагов метода для произвольных размеров эритроцитов является затруднительной. В частности, используемый в предлагаемом методе детектор границ Канни содержит в себе этап фильтрации изображения гауссовым фильтром определенного размера, т.е. настроен на выделение контуров на фиксированном масштабном уровне.

После масштабирования выполняется алгоритм Канни [12] для выделения контуров на изображении. В нем сначала производится сглаживание изображения фильтром Гаусса [13] для удаления шума.

Следующим шагом метода Канни является вычисление градиента каждой точки изображения. Ядра фильтров для расчета градиента изображения с помощью оператора Собеля представляют собой массивы 3×3 :

$$MGy = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix}; \quad MGx = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}.$$

Границы выделяются там, где градиент изображения приобретает максимальное значение, при этом только локальные максимумы отмечаются как границы. Следующий этап алгоритма выделения границ Канни заключается в определении потенциальных границ двойной пороговой фильтрации. Итоговые границы определяются путем подавления всех краев, не связанных с определенными границами.

Так как эритроциты в большинстве имеют округлую форму, их сегментация может быть проведена с помощью методов выделения окружностей на изображении. Одним из таких методов является преобразование Хафа, которое позволяет найти на изображении объекты, принадлежащие определенному классу фигур [11]. Окружность может быть задана параметрическими уравнениями:

$$x = a + R \cos(\theta);$$

$$y = b + R \sin(\theta),$$

где x, y – координаты точек окружности; a, b – координаты центра окружности; R – радиус окружности.

Как известно, координаты центра окружности a, b и ее радиус R являются параметрами, с помощью которых можно описать любую окружность на плоскости. Преобразование Хафа использует специальный массив, называемый аккумулятором, размерность которого равна количеству неизвестных параметров объекта. Аккумулятор рассчитывается таким образом, что если достаточное количество контурных точек изображения удовлетворило параметрическим уравнениям, то содержимое соответствующей ячейки аккумулятора увеличивается на единицу. Таким образом, максимумы аккумулятора будут соответствовать параметрам объекта в данной точке.

Большая размерность аккумулятора (массива) сильно увеличивает сложность алгоритма, и поэтому для окружности лучше брать заранее известный радиус, сокращая, таким образом, размерность до двух. Не на всех участках изображения присутствуют клетки, поэтому целесообразно выделить отдельные участки контуров и для них рассчитывать центры окружностей с помощью преобразования Хафа.

Радиус клеток может варьироваться в пределах некоторых значений. Для расчета этих значений на изображении контуров выбираются наиболее удачно выделившиеся контуры, представляющие собой объекты, замкнутая поверхность которых обладает наименьшим эксцентризитетом и достаточно большим диаметром. Диапазон значений радиусов клеток используется в качестве параметра R в преобразовании Хафа.

Координаты предполагаемых центров окружностей образуют облака точек на изображении, имеющие наибольшую плотность в районе центров искомых клеток.

Некоторые плотные образования точек ошибочно располагаются не в районе центров клеток, а на фоне. Это происходит из-за несовершенного выделения границ и образования ложных контуров, имеющих форму окружности. Для устранения точек, лежащих на фоне, используется следующий метод. Учитывая тот факт, что фон на изображениях мазков крови по своей интенсивности всегда выше интенсивности объектов, можно построить маску фона. Она строится путем сильного контрастирования исходного полутонового изображения с последующей бинаризацией и морфологической обработкой. Морфологическая обработка используется для устранения пропусков в маске в центрах клеток, имеющих схожую с фоном интенсивность. После построения маски удаляются те точки из облака центров окружностей, которые принадлежат фону.

Для объединения плотных образований точек используется морфологическая операция закрытия, в результате чего в центрах клеток на бинарном изображении образуются сплошные участки. Слишком малые по площади участки удаляются, а количество оставшихся участков представляет собой количество клеток на изображении.

Результаты эксперимента

В ходе экспериментального исследования предложенного метода были взяты изображения образцов крови, полученные с помощью микроскопа. Пример такого изображения приведен на рис. 1.

В первую очередь изображения были преобразованы из цветных в полутоновые. Для масштабирования изображений был рассчитан коэффициент, предполагающий, что диаметр клеток после масштабирования станет равен 100 пикселям.

Для улучшения изображений после масштабирования также повышался их контраст. Затем был применен детектор границ Канни. Экспериментальным путем были выбраны пороги для двойной пороговой фильтрации – 0,12 и 0,3, а также значение среднеквадратического отклонения для фильтра Гаусса, равное 6.

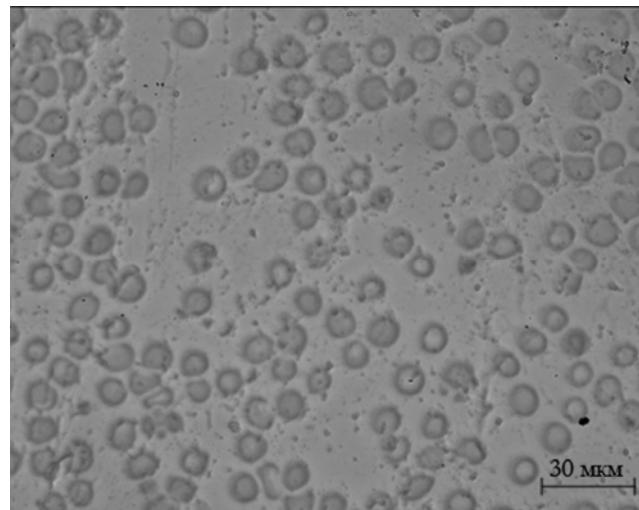


Рис. 1. Пример тестового изображения

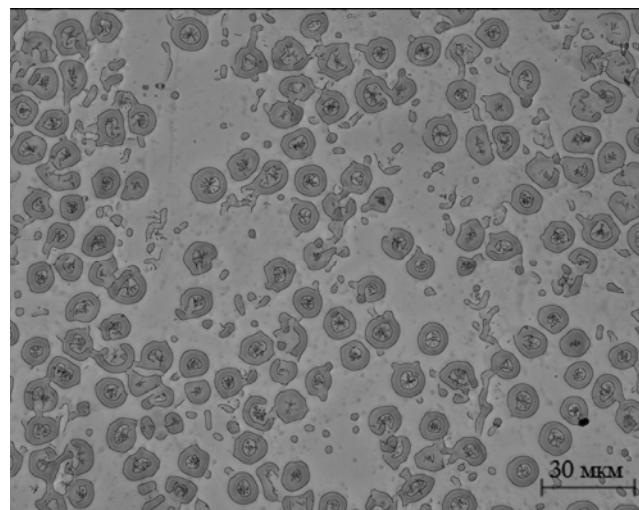


Рис. 2. Контуры и облако точек, наложенные на тестовое изображение

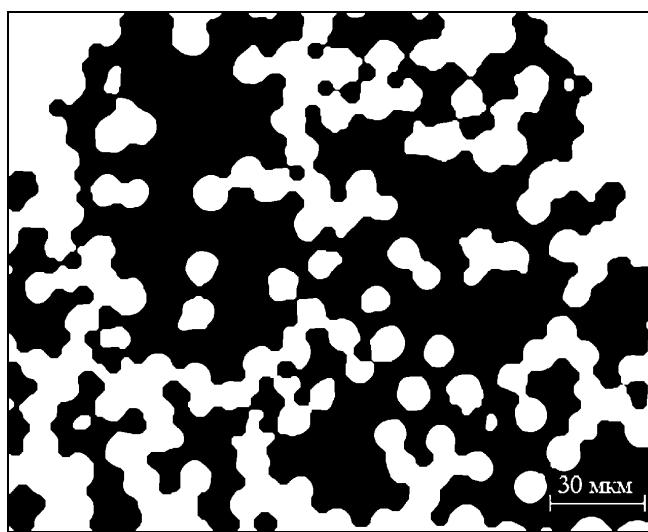


Рис. 3. Маска с фоном и объектами; объекты имеют белый цвет

Затем на изображениях были отмечены хорошо выделившиеся клетки с эксцентризитетом меньше 0,5 и размером больше некоторого порога, определенного с помощью метода Отсу [14]. Диапазон значений диаметров этих клеток использовался для преобразования Хафа. Тестовое изображение с нанесен-

ными на него границами и облаком точек представлено на рис. 2. Мaska, отделяющая объекты и фон, представлена на рис. 3.

Окончательная морфологическая обработка облака точек представляет собой закрытие изображений облаков точек элементом размером 10×10 , а также удаление отдельных элементов с площадью менее 100 пикселей. Результат окончательной морфологической обработки представлен на рис. 4.

Метод показал лучшие или сопоставимые с другими методами [1–9] результаты. Вероятность правильного определения количества клеток составила 96–98%.

Все тестовые изображения содержали в себе помехи, появившиеся в результате плохих условий съемки. Кроме этого, границы клеток были достаточно сильно размыты, что говорит о том, что метод может показать и лучшие результаты при условии, что изображения мазков крови будут более высокого качества.

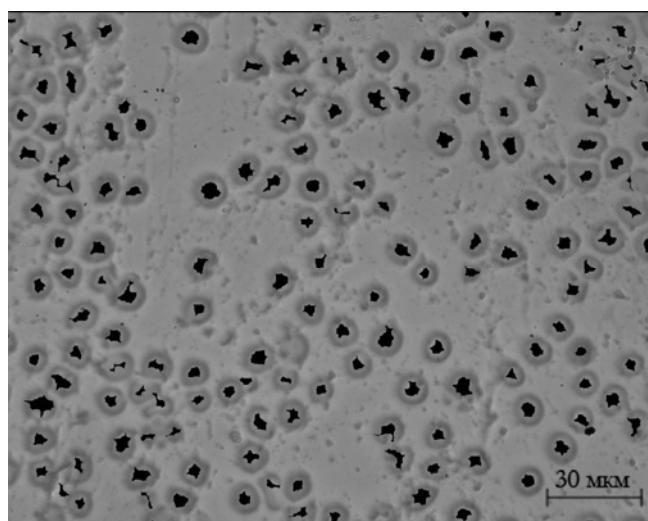


Рис. 4. Изображение центров клеток, наложенное на исходное

Заключение

Вероятность правильного подсчета количества эритроцитов разработанным методом составила 96–98%, что является хорошим результатом сегментации клеток в условиях невысокого качества изображений. Таким образом, метод пригоден для использования в задачах анализа образцов крови при различных лабораторных условиях.

Количество входных параметров метода может варьироваться. Параметр может быть один в случае, когда в алгоритме задан лишь средний размер клетки в пикселях или масштаб изображения. Для более тонкой подстройки метода под данное изображение в качестве входных параметров алгоритма могут выступать пороги для двойной пороговой фильтрации, значение среднеквадратического отклонения для фильтра Гаусса и минимальная площадь объектов на последнем этапе морфологической обработки изображения центров клеток. Автоматическое определение оптимальных параметров является целью дальнейших исследований.

Метод может быть использован не только для подсчета эритроцитов, но и для подсчета клеток другого размера либо любых объектов, имеющих округлую форму.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации.

Литература

1. Dahmen J., Hektor J., Perrey R., Ney H. Automatic Classification of Red Blood Cells Using Gaussian Mixture Densities // Proc. Bildverarbeitung für die Medizin. – 2000. – P. 331–335.
2. Costrarido L. Medical Image Analysis Methods: Evaluation Strategies for Medical-image Analysis. – Taylor & Francis. – United Stated of America, 2005. – P. 433–471.
3. Kumar B.R., Joseph D.K., Teager T.V.S. Energy Based Blood Cell Segmentation //14th International Conference on Digital Signal Processing. – DSP. 2002. – 1–3 July. – Santorini, Greece. – V. 2. – P. 619–622.
4. Bamford P. Empirical Comparison of Cell Segmentation Algorithms Using an Annotated Dataset // Proc. IEEE International Conference on Image Processing. – 2003. – V. 2. – P. 1073–1077.
5. Mukherjee D.P., Ray N., Acton S.T. Level Set Analysis for Leukocyte Detection and Tracking // IEEE Trans. Image Process. – 2004. – V. 13. – № 4. – P. 562–572.
6. Park J., Keller J.M. Snakes on the Watershed // IEEE Transactions on Pattern Analysis and Machine Intelligence. – PAMI, 2001. – V. 23. – № 10. – P. 1201–1205.

7. Sinha N., Ramakrishnan A.G. Blood Cell Segmentation Using EM Algorithm // Proc. Third Indian Conference on Computer Vision. Graphics Image Processing. – ICVGIP, 2002. – December 16–18. –Ahmadabad, India, 2002. – P. 376–382.
8. Kumar R.S., Verma A., Singh J. Color Image Segmentation and Multi-Level Thresholding by Maximization of Conditional Entropy // International Journal of Signal Processing. – 2006. – V. 3. – № 1. – P. 121–125.
9. Mcinerney T., Terzopoulos D. Deformable models in medical image analysis: A survey // Med Image Anal. – 1996. – P. 91–108.
10. Дырнаев А.В. Метод подсчета эритроцитов на изображениях мазков крови // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 6 (76). – С. 17–22.
11. Hough P.V.C. Method and means for recognizing complex patterns. – US Patent 3069654, December 18, 1962. – 8 с.
12. Canny J.F. A computational approach to edge detection // IEEE Trans. Pattern Analysis and Machine Intelligence. – 1986. – V. 8 – № 6. – P. 679–698.
13. Shapiro L.G., Stockman G.C. Computer Vision. – Prentence Hall, 2001. – 580 р.
14. Otsu N. A threshold selection method from gray-level histograms // IEEE Transactions on Systems, Man and Cybernetics. – 1979. – V. 9. – № 1. – P. 62–66.

Дырнаев Алексей Васильевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, dirnalex@yandex.ru
Потапов Алексей Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, pas.aicv@gmail.com

УДК 535.8

УСТАНОВКА КВАНТОВОЙ КРИПТОГРАФИИ С ИСТОЧНИКОМ ОДНОЧНЫХ ФОТОНОВ, ОСНОВАННЫМ НА ЯВЛЕНИИ СПОНТАННОГО ПАРАМЕТРИЧЕСКОГО РАССЕЯНИЯ СВЕТА

В.И. Егоров, И.З. Латыпов, А.В. Рупасов, А.В. Глейм, С.А. Чивилихин

Предложены схема однофотонного источника, основанного на явлении спонтанного параметрического рассеяния света, для приложений квантовой информатики и схема установки квантовой криптографии, содержащая такой источник. Проводится сравнение характеристик этой системы с альтернативными, оперирующими критически ослабленным классическим излучением.

Ключевые слова: квантовая криптография, спонтанное параметрическое рассеяние, бифотоны, поднесущие частоты.

Введение

Технология квантовой криптографии, опирающаяся на квантовые свойства света, позволяет передавать по незащищенному каналу связи случайную последовательность бит таким образом, что вмешательство злоумышленника (именуемого Евой) в процесс передачи неизбежно порождает дополнительный шум в канале и обнаруживается легитимными пользователями (именуемыми Алиса и Боб) [1].

Одним из важных элементов любой системы квантовой рассылки ключа является источник однофотонного излучения. В современных работах различают два основных подхода к пониманию квантового сигнала: одиночные фотоны и когерентные состояния [2, 3]. Известно несколько механизмов генерации одиночных фотонов, в частности, с использованием квантовых точек [4] и спонтанного параметрического рассеяния (СПР) [5]. К сожалению, на сегодняшний день не существует широкого выбора коммерческих продуктов, основывающихся на этих принципах. Используемой в экспериментах [2] и коммерческих образцах [6] альтернативой являются короткие лазерные импульсы, ослабленные до критического уровня так, чтобы среднее число фотонов в них было меньше единицы. В квантовой криптографии в связи с этим часто употребляется выражение «одна десятая фотона», подразумевающее, что одиночный фотон присутствует в среднем лишь в одном из десяти временных отсчетов.

Системы, использующие ослабленное классическое излучение («когерентные состояния») значительно более просты и доступны, а также имеют ряд других преимуществ, но обладают уязвимостью к определенным типам атак Евы. Кроме того, однофотонные источники необходимы для других приложений квантовой информатики (например, квантовой телепортации), поэтому их создание является важной задачей.

Оценка уязвимости систем квантовой криптографии с классическим источником

Безусловная безопасность систем квантовой рассылки ключа основывается на предположении, что в каждом рабочем импульсе содержится не более одного фотона [3]. Если это не так, у злоумышленника появляется возможность провести эффективную атаку, называемую beamsplitting (разделение пучка). Суть ее состоит в том, что Ева может отвести по одному фотону из каждого импульса, не тронув остальные, и сохранить их до того момента, как Алиса и Боб начнут обсуждение по открытому каналу, а затем провести

измерения в соответствии с полученной информацией. Подобная тактика не требует проведения измерения квантовых состояний непосредственно в процессе передачи, а потому никак не влияет ни на статистику получателя, ни на уровень шума в канале. В случае если Ева способна хранить фотоны в течение времени, затрачиваемого на генерацию ключа (для современных систем оно может не превышать нескольких секунд), эффективность этой атаки составляет 100% и не зависит от типа системы. Следует отметить, что, хотя число фотонов в пучке носит статистический характер, у Евы существует несколько возможностей для успешного определения импульсов, содержащих более одного фотона. Во-первых, теоретически перехватчик может обладать устройством, определяющим количество фотонов во временном интервале без их непосредственного измерения. Во-вторых, известно, что внутри каждого отсчета квантовые частицы идут с задержкой, а ворота счетчика в устройстве Боба, как правило, настроены таким образом, чтобы пропускать только первый фотон [2]. Этот подход значительно снижает уровень шума, но также позволяет Еве безнаказанно использовать запаздывающие компоненты импульса.

Для оценки эффективности атаки beamsplitting рассмотрим статистику излучения источника когерентных состояний [2]. Известно, что она описывается следующим выражением

$$P(n, \mu) = \frac{\mu^n}{n!} \cdot e^{-\mu},$$

где n – число фотонов в импульсе; μ – среднее число фотонов в импульсе. Доля ключа, известную Еве, можно получить, отнеся количество импульсов с числом фотонов больше единицы к количеству непустых импульсов, из которых будет формироваться ключ

$$P(n > 1 | n > 0, \mu) = \frac{1 - P(0, \mu) - P(1, \mu)}{1 - P(0, \mu)}.$$

На рис. 1 приведены графики зависимости количества импульсов, содержащих более одного фотона (I), и доли ключа, известной Еве (II), от величины μ в диапазоне 0,1–1.

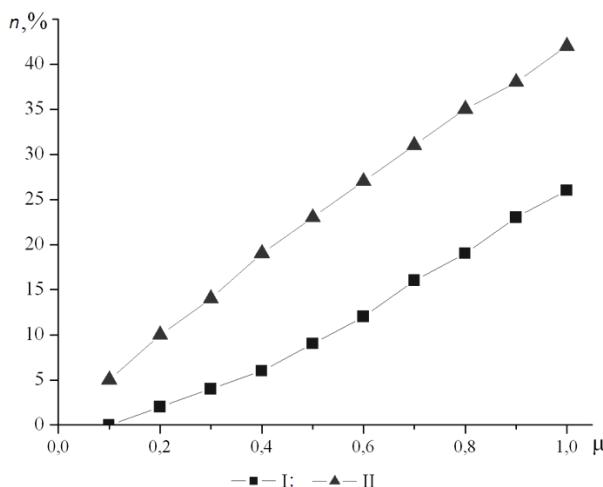


Рис. 1. Зависимость доли импульсов, содержащих более одного фотона (I) и доли ключа, известной Еве (II), от величины μ для источника когерентных состояний

Из графика видно, что хотя при значении μ порядка 0,1 число состояний с несколькими фотонами не превышает 1%, доля ключа, полученная Евой с применением атаки beamsplitting, составляет уже около 5%. Это связано с тем, что пустые отсчеты ($n = 0$), которых при низком значении μ большинство, не участвуют в процессе формирования ключа. Следует отметить, что разделение пучка может быть легко совмещено с другими типами атак, например, intercept-resend [2, 3]. В этом случае доли ключа, полученные Евой от атак разных типов, будут суммироваться. Хотя многие стратегии перехвата подразумевают повышение уровня шума в канале, Ева при определенных условиях может эффективно маскироваться под него, жертвуя информацией о ключе. Отметим также, что значение μ на практике ограничено снизу характеристиками приемника, так как отношение сигнал/шум линейно зависит от этого параметра [2].

Из этого следует, что использование источников когерентных состояний в реальных системах сопряжено с опасностью утечки секретных данных, которая может быть точно оценена в зависимости от характеристик применяемого пользователем оборудования.

Схема однофотонного источника для систем квантовой криптографии

Одним из возможных методов приготовления однофотонных состояний является СПР [5]. На основе СПР можно построить источник с вероятностным испусканием фотонов, обладающий высокой эффективностью, стабильностью и имеющий возможность перестройки в зависимости от задачи. Принципиальная схема источника, предлагаемого авторами, изображена на рис. 2.

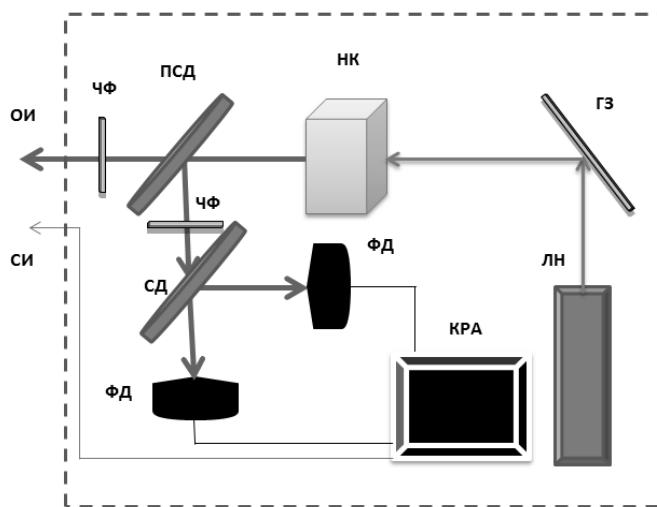


Рис. 2. Принципиальная схема однофотонного источника на основе СПР

Рассмотрим принцип работы такого источника. Излучение накачки генерируется непрерывным лазером (ЛН). Может быть использован как газовый лазер, так и диодный, с мощностью в диапазоне 10–100 мВт. Длина волны лазера определяет длины волн генерируемых фотонов. Фотон накачки, отражаясь от глухого зеркала (ГЗ) с определенной вероятностью (10^{-7} – 10^{-11}) рассеивается на нелинейном кристалле (НК), в результате чего рождается пара коррелированных фотонов (бифотон). Частоты пары фотонов могут быть одинаковыми (частотно-вырожденный режим СПР) и меняются при изменении ориентации нелинейного кристалла и ориентация его оптической оси относительно пучка накачки подбираются таким образом, что пара фотонов рождается с линейной и взаимно-ортогональной поляризацией (II тип синхронизма СПР). Далее бифотоны делятся по поляризации с помощью полупрозрачного светоделителя (ПСД). Один фотон (сигнальный) идет в оптический тракт (ОИ), а другой (холостой) проходит в боковой канал через частотный фильтр (ЧФ), используемый для отсечения излучения накачки и паразитных засветок, и светоделитель (СД). Затем холостой фотон регистрируется одним из лавинных фотодиодов (ФД), а сигнал о детектировании поступает на комплекс регистрирующей аппаратуры (КРА), включающий счетчики импульсов, электрическую линию задержки, схему совпадений, а также генератор синхроимпульсов (СИ). Наличие СД в холостом канале обусловлено необходимостью отсечения двухфотонных состояний света, представляющих наибольшую угрозу для безопасности квантового канала.

В процессе юстировки один из ФД ставится на выход ОИ, а схема совпадений соединяет прямой (сигнальный) и боковой каналы. При правильной настройке установки, когда выполняется условие орто-гональности поляризации, регистрируется высокий уровень совпадений детектирования фотонов в двух каналах. После этого схема совпадений перемещается в холостой канал, а регистрация совпадения в нем будет означать присутствие двухфотонного состояния света, и такой импульс не будет использоваться. Когда в холостом канале фотон детектируется только на одном из приемников, в сигнальном канале также присутствует ровно один фотон. В этом случае КРА посылает синхроимпульс на затвор однофотонного источника либо на затвор детектора на принимающей стороне криптографической схемы.

Частота генерации бифотонов настолько мала, что излучение даже при непрерывной накачке не высокой мощности носит импульсный характер и может регистрироваться счетчиками одиночных фотонов без предварительного ослабления излучения. Это позволяет получить надежный инструмент для отсеивания паразитных многофотонных состояний света, что делает данную схему «истинным» однофотонным источником.

Принципиальная схема системы квантовой криптографии с СПР-источником

На рис. 3 приведена предлагаемая схема системы квантовой рассылки ключа с рассмотренным выше однофотонным источником. Схема во многом повторяет классическую систему Plug-and-play [5], однако имеет и важные отличия. В частности, из-за заложенной в источнике возможности отсеивания холостых импульсов синхроимпульс может быть выведен из волокна ипущен по отдельной (не обязательно оптической) линии связи, что значительно снижает уровень шума засветки в системе. Это отличие выражается в отсутствии дополнительного источника лазерного излучения у Алисы, СД и детектора классического излучения в блоке Боба. Синхронизация может быть выполнена при передаче сигнала от холостого детектора в СПР-источнике Алисы на модулятор Боба.

Следует отметить, что жесткая привязка синхронизационного сигнала к детектированию фотона в холостом канале имеет и негативный эффект: при использовании легитимными пользователями протокола двух состояний (например, B92 [2]) Ева будет обладать выигрышной стратегией, так как всегда бу-

деть знать, был ли в импульсе фотон. Однако при использовании протоколов с большим числом состояний (например, BB84 [2]) с учетом понижения уровня шума такой возможностью перехватчик обладать не будет. Этот вопрос нуждается в дополнительном теоретическом исследовании.

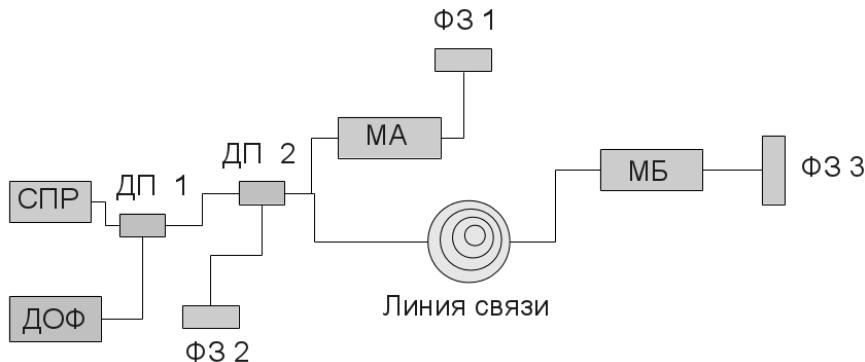


Рис. 3. Принципиальная схема Plug-and-Play системы квантовой криптографии с СПР-источником одиночных фотонов

Рассмотрим алгоритм работы этой схемы. Отправитель-Алиса запускает процесс генерации ключа, подавая на кристалл источника СПР излучение накачки. После рождения фотонной пары и детектирования «холостого» фотона импульс проходит через первый делитель (ДП 1) и разделяется на втором (ДП 2) на компоненты Р1 и Р2. Первая компонента, Р1, сразу идет по линии связи к получателю-Бобу, а вторая, Р2, предварительно задерживается в специальном («задерживающем») отрезке волокна между фарадеевскими зеркалами ФЗ 1 и ФЗ 2. Система из трех фарадеевских зеркал (ФЗ 1, ФЗ 2 и ФЗ 3) позволяет сохранить поляризацию импульсов Р1 и Р2 в момент их интерференции [7]. После прохождения света по линии связи Боб без изменений отражает обратно Р1 и вносит фазовый сдвиг в Р2 на модуляторе Боба (МБ), чтобы закодировать бит. При обратном проходе Алиса, в свою очередь, задерживает Р1 и модулирует его фазу на модуляторе Алисы (МА), после чего Р1 и Р2 интерферируют, одновременно оказываясь на выходе из делителя ДП 2. Результат интерференции Р1 и Р2 наблюдается с помощью счетчика одиночных фотонов ДОФ.

Преимущества и недостатки систем с однофотонным источником

Установка квантовой криптографии с СПР-источником обладает как преимуществами, так и недостатками по сравнению с другими современными типами систем квантовой рассылки ключа.

К недостаткам можно отнести высокую стоимость и сложность, низкую скорость работы (связанную с малой вероятностью генерации бифотонов) по сравнению с технологиями Plug-and-play [3] и квантовой рассылки криптографического ключа на поднесущей частоте модулированного света (КРКПЧ) [8], технические сложности с детектированием сигнального и холостого фотонов при определенных соотношениях их частот и частоты накачки. Для предложенной схемы установки актуальны проблемы, характерные для большинства Plug-and-play схем: увеличение потерь и снижение скорости за счет использования двунаправленной схемы распространения излучения, необходимость точного контроля фазы излучения. Кроме того, отсутствует возможность простой реализации мультиплексирования и интеграции в существующие линии оптической связи, присущая системам КРКПЧ [8].

Тем не менее, системы с СПР-источниками обладают фундаментально важными преимуществами: в них достигается значительно более высокая чистота состояний, требующаяся для достижения уровня секретности, описываемого в теории квантовой криптографии. Эти источники применимы для экспериментов в других областях квантовой информатики, демонстрируют низкий уровень шума и позволяют при необходимости легко отсекать пустые отсчеты.

Заключение

В работе продемонстрирована принципиальная возможность создания системы квантовой криптографии с бифотонным источником. Несмотря на то, что они позволяют достичь значительно более высокой чистоты квантовых состояний, целесообразность их применения должна оцениваться в зависимости от задачи. В частности, известные успешные эксперименты по взлому доступных на рынке систем квантовой криптографии базируются на уязвимостях аппаратуры принимающей стороны, а не атаке типа beamsplitting [6].

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы, ГК №16.513.11.3070.

Работа выполнена при поддержке ГК №14.740.12.08.41 «Использование излучения ультракороткой длительности для биомедицины, промышленности и защищенных коммуникаций», выполняемого в

рамках мероприятия «Проведение поисковых научно-исследовательских работ в целях развития обще-российской мобильности в области физики и астрономии».

Литература

1. Bennett C.H., Brassard G. Quantum Cryptography: Public Key Distribution and Coin Tossing // Proceedings of IEEE International Conference on Computers Systems and Signal Processing. – 1984. – P. 175–179.
2. Gisin N., Ribordy G., Tittel W., Zbinden H. Quantum cryptography // Rev. Mod. Phys. – 2002. – V. 74. – № 1. – P. 145–190.
3. Scarani V., Bechmann-Pasquinucci H., Cerf N.J. et al. The security of practical quantum key distribution // Rev. Mod. Phys. – 2009. – V. 81. – P. 1301–1350.
4. Unitt D.C., Bennett A.J., Atkinson P. et.al. Quantum dots as single-photon sources for quantum information processing // Journal of optic. – 2005. – V. 7. – № 7. – P. 129–134.
5. Калачев А.А., Калашников Д.А., Калинкин А.А., Митрофанова Т.Г., Самарцев В.В., Шкаликов А.В. Бифотонная спектроскопия кристалла рубина // Учен. зап. Казан. гос. ун-та. Сер. физ.-матем. науки. – 2008. – Т. 150. – Кн. 2. – С. 125–130.
6. Feihu Xu, Bing Qi, Hoi-Kwong Lo. Experimental demonstration of phase-remapping attack in a practical quantum key distribution system // New J. Phys. – 2010. – V. 12. – P. 113026.
7. Muller A., Herzog T., Huttner B., Tittel W., Zbinden H., Gisin N. «Plug and play» systems for quantum cryptography // Appl. Phys. Lett. – 1997. – V. 70. – P. 793–795.
8. Рупасов А.В., Глейм А.В., Егоров В.И., Мазуренко Ю.Т. Согласованная система квантовой рассылки криптографического ключа на поднесущей частоте модулированного света // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 2 (72). – С. 95–99.

Егоров Владимир Ильич

- Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, egorovvl@gmail.com

Латыпов Ильнур Зиннурович

- КФТИ КазНЦ РАН, мл. научный сотрудник, ilnur.latypov@gmail.com

Рупасов Андрей Игоревич

- Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, sadbender@yandex.ru

Глейм Артур Викторович

- Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, aglejm@yandex.ru

Чивилихин Сергей Анатольевич

- Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, ст. научный сотрудник, sergey.chivilikhin@gmail.com

3**АНАЛИЗ И СИНТЕЗ СЛОЖНЫХ СИСТЕМ**

УДК 681.51.015

**ИДЕНТИФИКАЦИЯ ЛИНЕЙНО МЕНЯЮЩЕЙСЯ ЧАСТОТЫ
СИНУСОИДАЛЬНОГО СИГНАЛА**
С.В. Арановский, А.А. Бобцов, А.А. Пыркин

Рассмотрена задача идентификации линейно меняющейся частоты синусоидального сигнала. Показана практическая значимость рассматриваемой задачи, получена динамическая модель сигнала и предложен алгоритм идентификации линейно меняющейся частоты. Приведен численный пример, иллюстрирующий работоспособность предложенного алгоритма.

Ключевые слова: идентификация, линейно меняющаяся частота, синусоидальный сигнал.

Введение

В работе рассматривается задача идентификации линейно меняющейся частоты синусоидального сигнала вида $y(t) = A \sin(\omega t^2/2 + \varphi)$. Несмотря на кажущуюся математическую экзотичность задачи (традиционно строится алгоритм идентификации для сигналов $A \sin(\omega t + \varphi)$), необходимо отметить, что проблема идентификации линейно меняющейся частоты синусоидального сигнала имеет реальное прикладное значение и может возникать, например, в задачах компенсации возмущений. Проблема идентификации постоянной частоты гармонического сигнала была хорошо изучена за последнее десятилетие, разработано большое число функционирующих в реальном времени алгоритмов [1–8]. В то же время задача идентификации переменной частоты синусоидального сигнала в литературе практически не рассматривается. На данный момент авторам не известны методы, позволяющие в реальном времени оценивать частоту такого сигнала. Существующие методы, в частности, алгоритм [8], позволяют оценивать частоту гармонического сигнала, имеющего трапециевидную форму (рис. 1), но не обеспечивают нулевую ошибку оценивания.

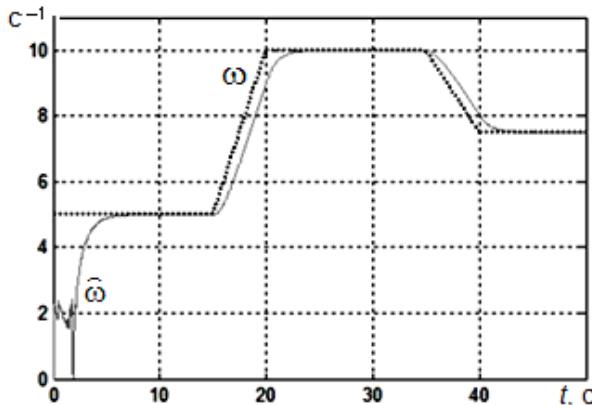


Рис. 1. Оценка частоты $\hat{\omega}$ при изменении частоты ω методом [8]

В работе ставится цель разработать алгоритм идентификации линейно меняющейся частоты гармонического сигнала.

Возмущения с линейно меняющейся частотой при управлении телескопом

Рассмотрим, как связана обсуждаемая в работе задача с прикладной проблемой компенсации возмущений при управлении телескопом. Широкое распространение в телескопостроении получили вентильные двигатели [9, 10]. Такие машины имеют много достоинств, но, к сожалению, не лишены недостатков. Одним из недостатков является наличие пульсирующего зубцового момента, который для трехфазных машин может быть описан как

$$M_p(\alpha) = M_a \sin\left(pn \frac{6}{2\pi} \alpha + \varphi_p\right), \quad (1)$$

где α – угол поворота двигателя; $M_p(\alpha)$ – значение зубцового момента в угле α ; M_a – амплитуда пульсаций момента; pn – число пар полюсов; φ_p – фаза зубцового момента. Очевидно, что если ось

телескопа вращается с постоянной скоростью $\dot{\alpha} = \text{const}$, то во временной области зубцовый момент описывается гармонической функцией $M_p(t) = M_a \sin(\omega_p t + \varphi_p)$ с частотой

$$\omega_p = pn \frac{6}{2\pi} \dot{\alpha} = \text{const}.$$

Проблема компенсации зубцового момента в приводах телескопов не нова. Она решалась, например, при построении систем управления группировкой телескопов VLT, Чили [11]. Однако при этом рассматривалась задача компенсации пульсаций момента только для той оси, приводами которой они вызваны. Естественно, что при этом значение угла α доступно прямому измерению, как и значение $\dot{\alpha}$, и задача компенсации сводится к определению только амплитуды и фазы сигнала при известной частоте.

Традиционно оптические телескопы включают в себя две оси – азимутальную и ось угла места, при этом в фотоприемном устройстве используются светочувствительные матрицы. Однако в ряде ситуаций оправдано использование вместо матрицы линейки светочувствительных элементов, размещенной на третьей оси телескопа – оси сканирования. В этом случае ось сканирования осуществляет периодические колебательные движения, а поступающие со светочувствительной линейки столбцы формируют полный кадр изображения. Недостатком такой конструкции является то, что, в отличие от азимутальной и угломестной осей, которые всегда ортогональны, угол между осью сканирования и азимутальной может меняться от 0° до 90° и совпадает с углом наклона оси угла места к горизонту. Соответственно моменты, создаваемые приводами одной оси, оказывают возмущающее воздействие на вторую ось с коэффициентом пропорциональности, равным косинусу угла между осями. Если угломестная ось ориентирована в зенит, то плоскость сканирования перпендикулярна азимутальной плоскости, и взаимное влияние осей отсутствует. Если же ось угла места ориентирована в горизонт, то взаимное влияние осей максимально, момент, создаваемый приводами азимутальной оси, полностью передается на ось сканирования, и наоборот.

Рассмотрим задачу компенсации пульсаций зубцового момента приводов азимутальной оси, передаваемых на сканирующую ось. Первый возможный подход заключается в передаче информации об угле поворота одной оси в систему управления другой осью. В этом случае задача компенсации становится тривиальной, но расплачиваться за это приходится увеличением информационного потока в каналах связи, появлением задержек и снижением максимальной частоты дискретизации систем управления. Второй путь – идентификация параметров возмущающего момента с его последующей компенсацией. Это потребует внедрения алгоритмов идентификации, но снизит объем передаваемой по каналам связи информации. Как было отмечено выше, при движении с постоянной скоростью зубцовый момент представляет собой гармоническую функцию времени с постоянной частотой. В этом случае для идентификации параметров возмущения (частоты, амплитуды и фазы) разработано большое число функционирующих в реальном времени алгоритмов [1–8], позволяющих решить задачу компенсации возмущения.

Более сложная задача идентификации возникает, если предположить, что траектория движения оси телескопа состоит из участков движения с постоянной скоростью (сопровождение медленных объектов) и участков равноускоренного движения (переход между различными объектами, сопровождение быстро движущихся объектов, проекция траекторий которых на азимутальную ось описывается параболой). В этом случае развитие зубцового момента (1) во времени будет описываться гармонической функцией с линейно меняющейся частотой, и для компенсации такого возмущения потребуется решить задачу идентификации линейно меняющейся частоты.

Математическая постановка задачи

Рассматривается синусоидальный сигнал вида

$$y(t) = A \sin(\omega t^2 / 2 + \varphi), \quad (2)$$

где A , ω и φ – неизвестные параметры. Ставится задача синтеза алгоритма идентификации неизвестной частоты ω .

Основной результат

Продифференцируем уравнение (2) трижды:

$$\ddot{y} = \theta(3ty + t^2 \dot{y})$$

или

$$p^3 y = \theta(3ty + t^2 py), \quad (3)$$

где $p = d / dt$ – оператор дифференцирования; $\theta = -\omega^2$ – неизвестный параметр. Для оценки неизвестного параметра θ произведем параметризацию модели (3) следующим образом:

$$p(p+1)^2 y = (2p^2 + p)y + \theta(3ty + t^2 py),$$

причем $t^2 py = pt^2 y - 2ty$. Запишем:

$$py = \frac{2p^2 + p}{(p+1)^2}y + \theta \frac{3}{(p+1)^2}ty + \theta \frac{p}{(p+1)^2}t^2y - \theta \frac{2}{(p+1)^2}ty. \quad (4)$$

Введем следующие обозначения

$$\xi_1 = \frac{1}{(p+1)^2}y, \quad \xi_2 = \frac{1}{(p+1)^2}ty, \quad \xi_3 = \frac{1}{(p+1)^2}t^2y,$$

при подстановке которых, выражение (4) примет вид

$$py = (2p^2 + p)\xi_1 + \theta\xi_2 + \theta\xi_3.$$

Тогда можно записать параметризованную модель в виде

$$\dot{y} = 2\ddot{\xi}_1 + \dot{\xi}_1 + \theta\xi_2 + \theta\dot{\xi}_3 = 2\ddot{\xi}_1 + \dot{\xi}_1 + \theta(\xi_2 + \dot{\xi}_3) = \xi_1 + \theta\xi_2, \quad (5)$$

где $\xi_1 = 2\ddot{\xi}_1 + \dot{\xi}_1$ и $\xi_2 = \xi_2 + \dot{\xi}_3$ – измеряемые сигналы.

Теперь построим алгоритм идентификации параметра θ . Временно предположим, что параметр θ известен. Запишем алгоритм идентификации следующим образом:

$$\dot{\hat{\theta}} = -k\xi_2^2\hat{\theta} + k\xi_2^2\theta, \quad (6)$$

где коэффициент $k > 0$. Для доказательства работоспособности алгоритма (6) рассмотрим ошибку $\tilde{\theta} = \theta - \hat{\theta}$. Дифференцируя $\tilde{\theta}$, получаем

$$\dot{\tilde{\theta}} = \dot{\theta} - \dot{\hat{\theta}} = -\dot{\hat{\theta}} = -k\xi_2^2\tilde{\theta}.$$

Рассмотрим функцию Ляпунова вида

$$V = \tilde{\theta}^2 / 2. \quad (7)$$

Дифференцируя (7), получаем

$$\dot{V} = -k\xi_2^2\tilde{\theta}^2 = -k\xi_2^2V \leq 0. \quad (8)$$

Из (8) следует устойчивость положения равновесия $\tilde{\theta} = 0$. Интегрируя (8), получаем

$$V(t) = V(t_0)e^{-ka(t)},$$

где t_0 – начальный момент времени; $a(t) = \int_{t_0}^t \xi_2^2 d\tau$. Легко видеть, что $a(t) \rightarrow \infty$ при $t \rightarrow \infty$, а, следовательно, $\lim_{t \rightarrow \infty} \tilde{\theta}(t) = 0$.

Таким образом, алгоритм идентификации (6) обеспечивает сходимость настраиваемого параметра $\hat{\theta}(t)$ к θ . Теперь построим реализуемый алгоритм, в котором параметр θ не измеряется. Из уравнения (5) находим

$$\dot{y} - \xi_1 = \theta\xi_2. \quad (9)$$

Подставляя (9) в (6), имеем

$$\dot{\hat{\theta}} = -k\xi_2^2\hat{\theta} + k\xi_2\dot{y} - k\xi_2\xi_1. \quad (10)$$

Поскольку переменная \dot{y} неизвестна, то введем в рассмотрение сигнал

$$\chi = \hat{\theta} - k\xi_2y. \quad (11)$$

Дифференцируя (11), получаем реализацию алгоритма идентификации вида

$$\dot{\chi} = -k\xi_2^2\hat{\theta} - k\xi_2y - k\xi_2\xi_1, \quad (12)$$

$$\hat{\theta} = \chi + k\xi_2y. \quad (13)$$

Очевидно, что при дифференцировании (13) будет получено выражение (10). Алгоритм идентификации (12), (13) обеспечивает сходимость $\hat{\theta}(t)$ к θ . Для поиска параметра ω достаточно провести следующую несложную процедуру $\hat{\omega} = \sqrt{|\hat{\theta}(t)|}$, где $\hat{\omega}$ – оценка частоты ω .

Пример

Рассмотрим сигнал вида (2) с параметрами $\omega = 0,5$; $A = 2$; $\varphi = 1$. На рис. 2 представлена временная диаграмма сигнала $y(t)$. На рис. 3, а, б, представлены результаты моделирования алгоритма идентификации параметра ω с коэффициентами $k = 1$ и $k = 0,01$ соответственно. На рис. 4, а, б, представлены графики переходных процессов оценки $\hat{\omega}$ при коэффициентах алгоритма идентификации $k = 0,001$ и $k = 0,2$ для сигнала $y(t)$ с параметрами $\omega = 2$; $A = 5$; $\varphi = -2$.

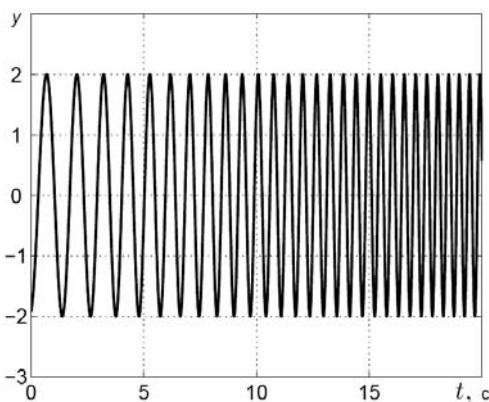


Рис. 2. Временная диаграмма сигнала $y(t)$

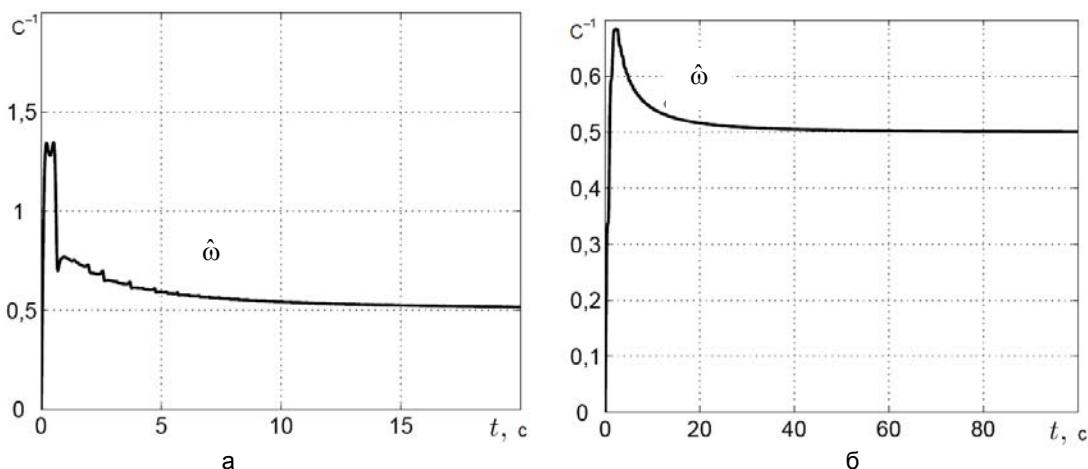


Рис. 3. Переходные процессы для переменной $\hat{\omega}$ при $\omega = 0,5 : k = 1$ (а); $k = 0,01$ (б)

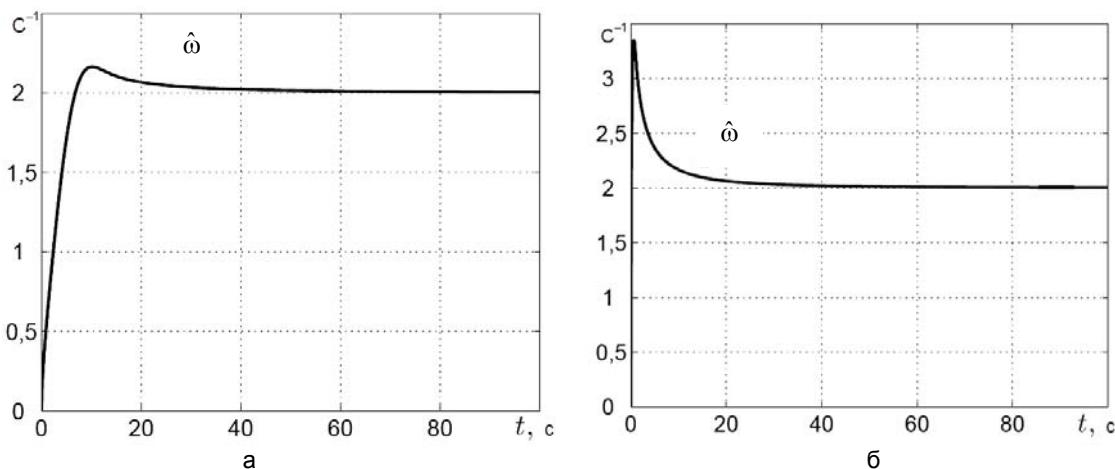


Рис. 4. Переходные процессы для переменной $\hat{\omega}$ при $\omega = 2 : k = 0,001$ (а); $k = 0,2$ (б)

Заключение

Рассмотрена задача идентификации частоты синусоидального сигнала с линейно меняющейся частотой. Получена динамическая модель генератора сигнала вида $y(t) = A \sin(\omega t^2 / 2 + \varphi)$, предложен алгоритм идентификации вида (12), (13), обеспечивающий сходимость настраиваемого параметра $\hat{\omega}$ к истинному значению ω . Предложенный алгоритм имеет настроочный коэффициент k , варьируя который можно изменять быстродействие алгоритма.

Работа выполнена при финансовой поддержке ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы» (контракт № 11.519.11.4007).

Литература

1. Арановский С.В., Бобцов А.А., Кремлев А.С., Лукьянова Г.В., Николаев Н.А. Идентификация частоты смещенного синусоидального сигнала // Автоматика и телемеханика. – 2008. – № 9. – С. 3–9.
2. Арановский С.В., Бобцов А.А., Пыркин А.А. Адаптивный наблюдатель неизвестного синусоидально-го выходного возмущения для линейного объекта // Автоматика и телемеханика. – 2009. – № 11. – С. 108–116.
3. Арановский С.В., Бобцов А.А., Кремлев А.С. Компенсация конечномерного квазигармонического возмущения для нелинейного объекта // Известия РАН. Теория и системы управления. – 2006. – № 4. – С. 14–21.
4. Арановский С.В., Бобцов А.А., Никифоров В.О. Синтез наблюдателя для нелинейного объекта в ус-ловиях гармонического возмущения, приложенного к выходной переменной // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 3. – С. 32–39.
5. Marino R., Tomei R. Global Estimation of Unknown Frequencies // IEEE Transactions on Automatic Control. – 2002. – № 47. – P. 1324–1328.
6. Hou M. Amplitude and frequency estimator of a sinusoid // IEEE Transactions on Automatic Control. – 2005. – V. 50. – P. 855–858.
7. Marino R., Santosuosso G.L. Output regulation for a class of nonlinear systems with unknown sinusoidal disturbances // Proceedings of the 44th IEEE Conference on Decision and Control. – 2005. – C. 3321–3326.
8. Stanislav Aranovskiy, Alexey Bobtsov, Artem Kremlev, Nikolay Nikolaev, Olga Slita. Identification of fre- quency of biased harmonic signal // European Journal of Control. – 2010. – № 4. – P. 129–139.
9. Васильев В.Н., Томасов В.С., Шаргородский В.Д., Садовников М.А. Состояние и перспективы разви-тия прецизионных электроприводов комплексов высокоточных наблюдений // Изв. вузов. Приboro-строение. – 2008. – № 6. – С. 5–12.
10. Аракелян А.К., Афанасьев А.А. Вентильтные электрические машины. – М.: Энергоатомиздат, 1997. – 236 с.
11. Erm T., Sandrock S. Adaptive correction of periodic errors improves telescope performance // Proceedings of the American Control Conference. – 2005. – V. 6. – P. 3776–3777.

**Арановский Станислав
Владимирович**

- Санкт-Петербургский национальный исследовательский университет ин-формационных технологий, механики и оптики, кандидат технических наук, ст. научный сотрудник, s.aranovskiy@gmail.com

Бобцов Алексей Алексеевич

- Санкт-Петербургский национальный исследовательский университет ин-формационных технологий, механики и оптики, доктор технических наук, профессор, декан, bobtsov@mail.ifmo.ru

Пыркин Антон Александрович

- Санкт-Петербургский национальный исследовательский университет ин-формационных технологий, механики и оптики, кандидат технических наук, мл. научный сотрудник, a.pyrkin@gmail.com

УДК 621.37

**УСОВЕРШЕНСТВОВАННЫЙ АЛГОРИТМ АНАЛИЗА СИГНАЛОВ
МАЛОЗАМЕТНЫХ РАДИОЛОКАЦИОННЫХ СТАНЦИЙ НА ОСНОВЕ
ОКОННОГО ПРЕОБРАЗОВАНИЯ ФУРЬЕ**

А.В. Коротков, А.И. Мухин

Рассмотрена задача анализа сигналов малозаметных радиолокационных станций. Предложен усовершенствованный алгоритм, основанный на некогерентной интеграции выходных отсчетов дискретного преобразования Фурье. При-менение такого алгоритма позволяет учесть особенности данного класса сигналов, в первую очередь низкое соотно-шение сигнал/шум.

Ключевые слова: спектральный анализ, интеграция, малозаметные радиолокационные станции.

Введение

В настоящее время широко используются радиолокационные станции с низкой вероятностью пе-рехвата. Сигналы таких станций излучаются на уровне шума в широкой полосе, что затрудняет их обна-ружение. Такие радиолокационные станции получили название малозаметных радиолокационных стан-ций (МРЛС).

Основными видами внутриимпульсной модуляции, применяемой в МРЛС, являются частотная модуляция (ЧМ), частотная и фазовая манипуляция, а также их комбинации. Наиболее распространенной является ЧМ с линейно изменяющейся частотой (ЛЧМ). Для кодирования фазы зондирующего сигнала широко используются коды Баркера, многофазные коды Франка Р1, Р2, Р3 и Р4, а также многопозиционные коды Т1, Т2, Т3 и Т4 [1].

Для таких сигналов, в первую очередь из-за низкого соотношения сигнал/шум, классический метод частотно-временного преобразования на основе дискретного преобразования Фурье (ДПФ) оказывается недостаточно эффективным, поэтому возникает необходимость дополнительной обработки полученных спектрограмм.

Оконное преобразование Фурье

Пусть $x(n)$ – последовательность известных отсчетов исследуемого сигнала длиной M , M -точечное ДПФ которой имеет вид [2]

$$X(k) = \sum_{n=0}^{M-1} x(n)e^{-j2\pi nk/M}.$$

Далее рассмотрим разбиение $x(n)$ на P субпоследовательностей длиной N отсчетов каждая. Следовательно, $PN=M$. Поэлементное суммирование этих субпоследовательностей дает новую последовательность $y(n)$ длиной N отсчетов. ДПФ данной последовательности, $Y(m)$, представляет собой версию $X(m)$, прореженную в P раз, причем $|Y(m)| = |X(Pm)|$. Таким образом, разделяем взвешенную окном последовательность $x(n)$, содержащую N отсчетов, на P субпоследовательностей, далее суммируем субпоследовательности поэлементно для получения последовательности с наложениями длиной M отсчетов. Затем вычисляем модули спектральных отсчетов с помощью ДПФ. Для получения требуемой частотной характеристики рассчитывается M -точечное окно, которое представляет собой фильтр нижних частот с заданными параметрами. Для реализации такого анализатора длина исходной последовательности должна быть больше N (количество точек ДПФ) в целое число раз.

Ширина полос пропускания (каналов) анализатора спектра определяется главным образом шириной главного лепестка окна. Центральные частоты полос пропускания определяются как f_s / N , где f_s – частота дискретизации. Это значит, что перекрытие полос пропускания анализатора зависит как от ширины главного лепестка окна, так и от f_s и P . Динамический диапазон анализатора можно расширить, увеличив P , что приводит к увеличению M и длины последовательности $x(m)$. С увеличением M более длинные окна будут давать полосы пропускания анализатора, приближающиеся к прямоугольной форме, с более низким уровнем боковых лепестков и неравномерностью амплитудно-частотной характеристики в полосе пропускания.

Усовершенствованный алгоритм анализа сигналов

ДПФ имеет фиксированное частотно-временное разрешение, которое зависит от окна. Более того, разрешение по времени и частоте подчиняется принципу неопределенности, а именно, хорошее временное разрешение дает плохое частотное, и наоборот. Этот недостаток оказывает сильное влияние на характеристику обнаружения: если сигнал имеет полосу шире частотного разрешения, то энергия сигнала распределяется по нескольким составляющим (каналам) на частотной оси, так что вероятность обнаружения, зависящая от этих составляющих, уменьшается. Аналогично, сигналы с большей, чем разрешение по времени, длительностью распределяют свою энергию по нескольким составляющим на временной оси. Предлагаемый усовершенствованный анализатор спектра осуществляет более гибкое частотно-временное представление. Он основывается на некогерентной интеграции выходных отсчетов ДПФ с разными длинами, которую можно представить в виде следующей формулы:

$$I(m) = \sum_1^{l_i} |X_w(m)|,$$

где l_i – длина интеграции; $X_w(m)$ – отсчеты ДПФ, взвешенной окном входной последовательности. Величину $I(m)$ назовем модифицированной спектрограммой. Множество таких спектрограмм с разными значениями l_i может быть применено для анализа сигналов с низким соотношением сигнал/шум.

Что касается количества интеграторов, то их должно быть, по крайней мере, три. Первый, длиной l_1 , сохраняет наилучшее разрешение по времени, второй (l_2), с наибольшей длиной, позволяет получить максимальное превышение уровня сигнала над шумом, третий (l_3), с промежуточным значением, призван снизить потери (частотного и временного разрешения) при анализе вторым интегратором. Максимальная возможная длина интегратора равна n/N , где n – длина выборки, N – число каналов ДПФ. На практике целесообразно выбирать данный параметр от $n/2N$ до $n/4N$. При реализации непрерывной обработки выбор длин интеграторов зависит от требуемого частотно-временного разрешения и типа сигнала. Экспериментально выявлено, что такой подход (интеграция) позволяет получить увеличение амплитуды сигнала относительно шума (в Δ раз). Величина Δ зависит от числа каналов ДПФ, длины интегратора и длины выборки и может быть выражена следующей формулой [3]:

$$\Delta = 5 \lg \frac{A(l_{i+1})}{A(l_i)},$$

где $A(l_i)$ – среднее значение амплитуды при длине i -го интегратора. Следует отметить, что для априорно неизвестных сигналов необходим совместный анализ полученных спектрограмм.

Применение усовершенствованного алгоритма анализа сигналов

Для примера взяты три широкополосных сигнала со следующими видами модуляции: ЛЧМ, фазовая манипуляция (ФМн), фазо-частотная манипуляция (ФЧМн). Их параметры приведены в табл. 1.

Параметр	Сигнал 1	Сигнал 2	Сигнал 3
Полоса сигнала, МГц	40	10	20
Длительность, мкс	200	168	100
Частота дискретизации, МГц	128	128	128
Длина выборки (число отсчетов)	26500	21504	12800
Вид модуляции	ЛЧМ	ФМн	ФЧМн
Соотношение сигнал/шум, дБ	-6	-6	-6

Таблица 1. Параметры сигналов

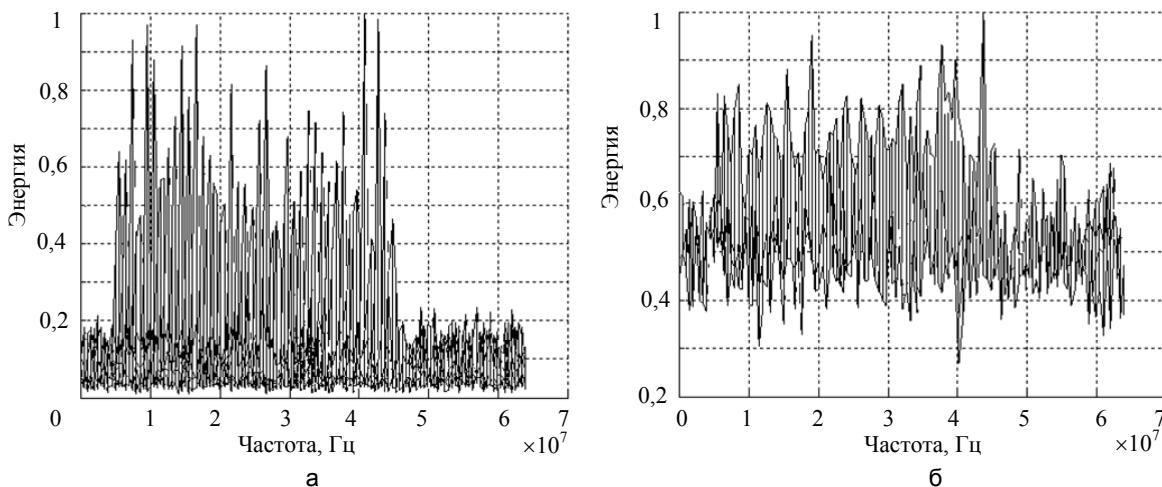


Рис. 1. Спектр ЛЧМ сигнала: длина интегратора $l_1 = 1$ (а); длина интегратора $l_3 = 25$ (б)

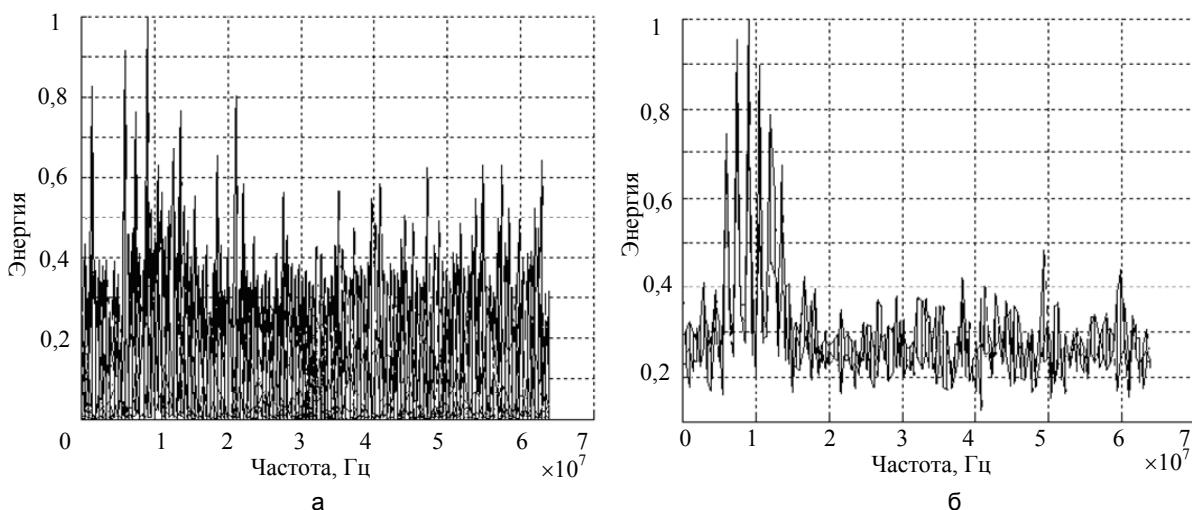
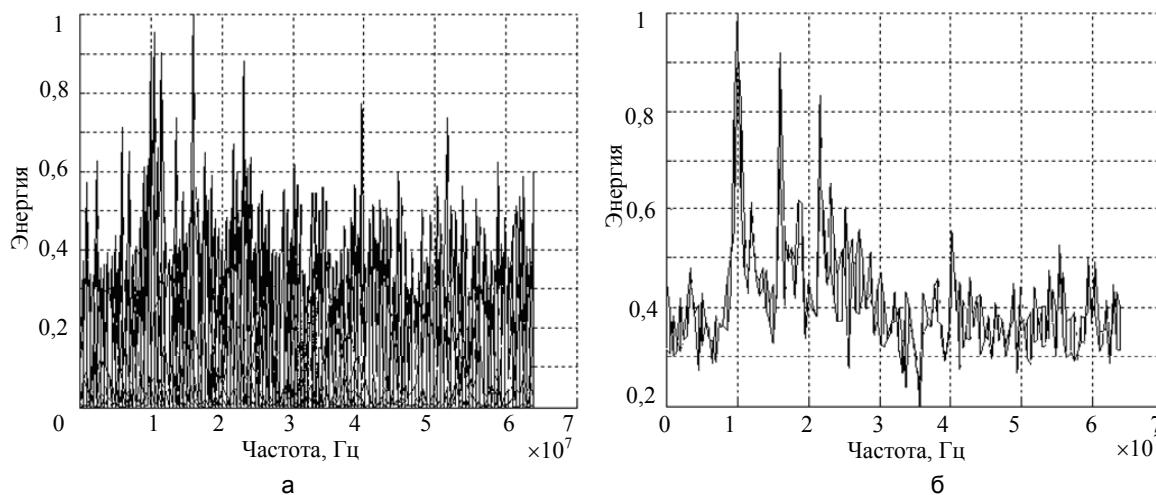


Рис. 2. Спектр ФМн сигнала: длина интегратора $l_1=1$ (а); длина интегратора $l_3=24$ (б)

Рис. 3. Спектр ФЧМн сигнала: длина интегратора $l_1 = 1$ (а); длина интегратора $l_3 = 50$ (б)

Вид модуляции	Длины интеграторов $l_1; l_2; l_3$	Средние значения амплитуд $A(l_1); A(l_2); A(l_3)$	Значение Δ (дБ) для длин интеграторов		
			l_2 и l_1	l_3 и l_2	l_3 и l_1
ЛЧМ	1; 5; 25	29; 146; 733	3,4	3,4	6,8
ФМн	1; 4; 24	30; 119; 710	3,0	3,9	6,9
ФЧМн	1; 25; 50	29; 741; 1400	6,9	1,6	8,4

Таблица 2. Результаты моделирования

Рассмотрены три варианта частотно-временного анализатора с различными длинами интеграции. Результаты, полученные при помощи имитационного моделирования, представлены в табл. 2 и на рис. 1–3. В качестве ФМн сигнала рассмотрен сигнал, содержащий код Баркера (длина кода 7) [4], ФЧМн – код Франка (длина кода 16).

Проведенный анализ позволяет сделать следующие выводы.

1. Применение усовершенствованного алгоритма позволяет получить выигрыш в соотношении сигнал/шум до 8 дБ (табл. 2), причем, чем больше длина интеграции (l), тем больше значение Δ .
2. Предложенный алгоритм наиболее эффективен для выделения из шума ФМн и ФЧМн сигналов.
3. Для ЛЧМ сигналов целесообразно применение интеграции с параметром $l=1$.
4. Усовершенствованный алгоритм анализа позволяет обнаруживать и производить предварительную оценку параметров ФМн и ФЧМн сигналов при соотношениях сигнал/шум вплоть до – 6 дБ, что является важным преимуществом при анализе излучений МРЛС.
5. С увеличением длины интеграции ухудшается частотно-временное разрешение, поэтому необходим совместный анализ полученных спектрограмм с разными длинами интеграции.

Заключение

Рассмотрена задача анализа сигналов малозаметных радиолокационных станций, с помощью усовершенствованного алгоритма на основе оконного дискретного преобразования Фурье. Продемонстрировано, что применение данного алгоритма позволяет получить выигрыш в соотношении сигнал/шум до 8 дБ, а также эффективно выделять из шума и анализировать ФМн и ФЧМн сигналы малозаметных радиолокационных станций. Для получения более полной картины об исследуемом сигнале необходим совместный анализ полученных данных с разными длинами интеграции.

Литература

1. Pace Phillip E. Detecting and Classifying Low Probability of Intercept Radar. – Artech House, 2009. – 857 р.
2. Айфичер Эммануил С., Джервис Барри У. Цифровая обработка сигналов: Пер. с англ. – М.: Вильямс, 2004. – 992 с.
3. Лайонс Ричард. Цифровая обработка сигналов: Пер. с англ. – 2-е изд. – М.: Бином пресс, 2006. – 656 с.
4. Коротков А.В. Частотно-временной анализ сигналов малозаметных радиолокационных станций с использованием вейвлет-преобразования // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 5. – С. 44–48.

Коротков Андрей Владимирович

– ОАО «НИИ Вектор», инженер, Andrey.k0206516@mail.ru

Мухин Анатолий Иванович

– Военно-космическая академия им. А.Ф. Можайского, адъюнкт, tolik516@yandex.ru

УДК 681.5

ИСПОЛЬЗОВАНИЕ УСЛОВИЙ КАЧЕСТВЕННОЙ ЭКСПОНЕНЦИАЛЬНОЙ НЕУСТОЙЧИВОСТИ ДЛЯ ОЦЕНКИ ДИНАМИЧЕСКИХ ПРОЦЕССОВ

Е.Ю. Рабыш, В.В. Григорьев, С.В. Быстров, А.В. Спорягин

На основе прямого метода Ляпунова и условий качественной экспоненциальной неустойчивости найдены оценки динамических показателей качества переходных процессов, позволившие создавать эффективные процедуры аналитического анализа многомерных неустойчивых непрерывных и дискретных систем управления.

Ключевые слова: качественная экспоненциальная неустойчивость, оценки качества, непрерывные и дискретные системы, анализ поведения неустойчивых систем, потеря управления.

Введение

Одной из актуальных проблем теории управления является анализ поведения неустойчивых систем управления (систем с параметрическими нарушениями). Результаты этого анализа являются ценными для принятия решений при выходе из строя автоматической системы управления, когда неустойчивая система управления может представлять собой существенную угрозу, опасность и для человека, и для окружающей среды. При проектировании такой опасной системы управления необходимо позаботиться о том, чтобы при потере управления, вызванной той или иной причиной, срабатывала система защиты и сигнализации, основанная на динамических свойствах самой системы управления и обеспечивающая минимизацию потерь, связанных с таким инцидентом. Для этого используется понятие качественной экспоненциальной неустойчивости, тесно связанной с качественными показателями процессов неустойчивых систем управления благодаря введению условий, ограничивающих фактически значения скорости изменения нормы вектора состояния системы, что непосредственно связано со степенью расходности переходных процессов [1–4].

Постановка задачи

Пусть поведение непрерывной динамической системы описывается дифференциальным уравнением

$$\dot{x}(t) = f(x(t)), \quad (1)$$

где $x \in R^n$ – вектор состояния динамической системы; $x(0) = x_0 \in R^n$ – вектор начальных состояний; $t \geq 0$ – время; $f(x)$ – n -мерная нелинейная вектор-функция векторного аргумента, такая, что при любых $x_0 \in R^n$ решение $x \in R^n$ уравнения (1) существует и единствено.

Непрерывная система (1) с положением равновесия $x = 0$ называется качественно экспоненциально (β, r) неустойчивой, если существуют такие параметры r ($r > 0$) и β ($\beta > r$), что для любых траекторий движения системы, исходящих из произвольных начальных условий $x_0 \in R^n$, в любой момент времени $t \geq 0$ выполняется условие

$$\|x(t) - e^{\beta \cdot t} \cdot x_0\| \leq \rho \cdot (e^{[\beta+r] \cdot t} - e^{\beta \cdot t}) \cdot \|x_0\|, \quad (2)$$

где $\rho \geq 1$. Здесь норма вектора задается соотношением

$$\|x\| = \left[\sum_{i=1}^n x_i^2 \right]^{1/2},$$

где x_i – i -ая компонента вектора состояния x .

Пусть поведение дискретной динамической системы описывается разностным уравнением

$$x(m+1) = f(x(m)), \quad (3)$$

где $x \in R^n$ – вектор состояния динамической системы; $x(0) = x_0 \in R^n$ – вектор начальных состояний; $m = 0, 1, 2, \dots$ – номер интервала дискретности; $f(x)$ – n -мерная нелинейная вектор-функция векторного аргумента, такая, что при любых $x_0 \in R^n$ решение $x \in R^n$ уравнения (1) существует и единствено. Дискретная система (3) с положением равновесия $x = 0$ называется качественно экспоненциально (β, r) неустойчивой, если существуют такие параметры r ($r > 0$) и β ($\beta > 1+r$), что для любых траекторий движения системы, исходящих из произвольных начальных условий $x_0 \in R^n$, при которых для любого номера интервала дискретности $m \geq 0$ выполняется условие

$$\|x(m) - \beta^m \cdot x_0\| \leq \rho \cdot ((\beta + r)^m - \beta^m) \cdot \|x_0\|, \quad (4)$$

где $\rho \geq 1$. Параметр β подобен коэффициенту сноса и для неустойчивых систем определяет среднюю скорость расходимости траекторий движения от начального состояния. Параметр r подобен коэффициенту диффузии и определяет отклонения траекторий движения от усредненной траектории.

Под критическим временем переходного процесса в непрерывных и дискретных динамических системах соответственно будем понимать значение $t = t_c$, такое, что

$$\|x(t)\| = \delta_c \cdot \|x_0\|, \quad (5)$$

$$\|x(m)\| = \delta_c \cdot \|x_0\|, \quad (6)$$

т.е. момент времени, в который переходной процесс выходит за заданную критическую δ_c -окрестность начального положения ($\delta_c > 1$). Выбор относительной величины окрестности δ_c определяется требованиями конкретной задачи и зависит от технологических параметров объекта управления. При этом критическое время переходного процесса для неустойчивых систем характеризует среднюю степень расходимости переходных процессов.

Под выбросом в непрерывных и дискретных динамических системах будем понимать величины σ_0 ($\sigma_0 \geq 1$), определяемые соответственно уравнениями

$$\sigma_0 = \frac{\max_{t \in [0, \infty)} x_m(t)}{\|x_0\|}, \quad (7)$$

$$\sigma_0 = \frac{\max_{m \in [0, \infty)} x_m(m)}{\|x_0\|}, \quad (8)$$

где x_m – миноранта $\|x\|$, т.е. функция, ограничивающая снизу текущие значения нормы вектора состояния, так, что $x_m \leq \|x\|$ для любого момента времени. Выброс косвенно характеризует колебательность в неустойчивой динамической системе, т.е. разброс от средней степени расходимости. При значении σ_0 , стремящимся к бесконечности, процесс носит монотонный характер.

Ставится задача на основе достаточных условий качественной экспоненциальной неустойчивости (2) и (4) для непрерывных и дискретных динамических систем, задаваемых уравнениями (1) и (3) соответственно, отыскать оценки динамических показателей качества в виде критического времени переходного процесса и выброса, которые совместно с достаточными условиями качественной экспоненциальной неустойчивости позволяли бы создать эффективные численные процедуры анализа неустойчивых динамических систем.

Основные результаты

В дальнейшем для оценки процессов будем использовать квадратичную функцию Ляпунова вида:

$$V(x) = x^T \cdot P \cdot x,$$

где P – симметрическая положительно определенная $n \times n$ матрица. Будем говорить, что функция Ляпунова квадратичная, если эта функция является выпуклой положительно однородной степени 2 и выполняется соотношение Релея:

$$\tilde{\eta}_1 \cdot \|x\|^2 \leq V(x) \leq \tilde{\eta}_2 \cdot \|x\|^2,$$

где значения $\tilde{\eta}_1$ и $\tilde{\eta}_2$ являются минимальным и максимальным собственными числами матрицы P соответственно. Выпуклая положительно однородная функция степени 2 обладает следующими свойствами:

$$V(0) = 0, V(\gamma x) = \gamma^2 V(x),$$

при любых $\gamma > 0$ и при любых $x_0 \in R^n$.

Непрерывная система (1) с положением равновесия $x = 0$ качественно экспоненциально (β, r) неустойчива, если существуют такая квадратичная функция Ляпунова и такие параметры r ($r > 0$) и β ($\beta > r$), что для любых траекторий движения системы, исходящих из произвольных начальных условий $x_0 \in R^n$, в любой момент времени $t \geq 0$ выполняется условие

$$V\left(\frac{d}{dt}x(t) - \beta \cdot x(t)\right) \leq r^2 \cdot V(x(t)).$$

Дискретная система (3) с положением равновесия $x = 0$ качественно экспоненциально (β, r) неустойчива, если существуют такая квадратичная функция Ляпунова и такие параметры r ($r > 0$) и β

$(\beta > 1 + r)$, что для любых траекторий движения системы, исходящих из произвольных начальных условий $x_0 \in R^n$, для любого номера интервала дискретности $m \geq 0$ выполняется условие

$$V(x(m+1) - \beta \cdot x(m)) \leq r^2 \cdot V(x(m)).$$

Утверждение 1. Оценки критического времени переходного процесса и выброса для непрерывных динамических систем имеют вид

$$t_c = \frac{1}{\beta} \cdot \ln(\delta_c), \quad (9)$$

$$\sigma_0 = (\rho + 1) \cdot e^{\frac{\beta \cdot \ln((\rho + 1) \cdot \beta)}{r}} - \rho \cdot e^{\frac{(\beta + r) \cdot \ln((\rho + 1) \cdot \beta)}{r}}. \quad (10)$$

Утверждение 2. Оценки критического времени переходного процесса и выброса для дискретных динамических систем имеют вид

$$t_c = T \cdot \log_{\beta}(\delta_c), \quad (11)$$

$$\sigma_0 = (\rho + 1) \cdot \beta^{\frac{\log(\frac{(\rho + 1) \cdot \ln \beta}{\rho \cdot \ln(\beta + r)})}{r}} - \rho \cdot (\beta + r)^{\frac{\log(\frac{(\rho + 1) \cdot \ln \beta}{\rho \cdot \ln(\beta + r)})}{r}}. \quad (12)$$

Здесь T – интервал квантования.

Доказательства утверждений приведены в Приложении.

Приведем алгоритм аналитического анализа динамических свойств неустойчивых непрерывных и дискретных систем с исходными данными – матрицей описания замкнутой системы F_u .

1. По заданным показателям качества t_c и σ_0 при $\rho = 1$ определить значения параметров β и r .

3. Как для непрерывных, так и для дискретных систем проверить выполнение условия:

$$\max_i \lambda_i \leq 0, i = 1, 2, \dots, n, \quad (13)$$

где λ_i определяется из характеристического уравнения

$$\det \left[\left[(F_u - \beta I)^T (F_u - \beta I) - r^2 I \right] - \lambda_i \right] = 0,$$

где I – единичная $n \times n$ матрица. Если условие (13) выполняется, то выполняются и заданные оценки качества переходных процессов.

Для демонстрации эффективности предлагаемого алгоритма представим результаты математического моделирования системы, динамика которой описывается уравнением

$$x(m+1) = F_u x(m), \quad (14)$$

где матрица описания F_u имеет вид

$$F_u = \begin{bmatrix} 1,078 & 0 & 0 & 0,013 \\ 0 & 1,077 & 0,011 & 0 \\ 0,012 & 0 & 1,081 & 0 \\ 0 & 0,014 & 0 & 1,082 \end{bmatrix}$$

с интервалом квантования $T=0,1$ с.

Проанализируем исходную неустойчивую систему управления, при этом возьмем параметры качества

$$t_c = 3, \delta_c = 10, \sigma_0 = 5, \quad (15)$$

используя которые, находим:

$$\beta = 1,08, r = 0,0158.$$

Проверим выполнение условия (13):

$$\max_i \lambda_i = -0,00003 \leq 0,$$

т.е. проверяемое условие выполняется, таким образом, и заданные показатели качества тоже должны выполняться. Теперь проверим удовлетворение другим показателям качества:

$$t_c = 3, \delta_c = 10, \sigma_0 = 20, \quad (16)$$

откуда находим параметры:

$$\beta = 1,08, r = 0,0106.$$

Проверим выполнение условия (13):

$$\max_i \lambda_i = 0,0001 > 0,$$

т.е. проверяемое условие не выполняется, таким образом, и заданные показатели качества тоже не должны выполняться. Желаемые оценочные трубы и реакция системы управления (14) на начальные отклонения

$$x_0^T = [0, 50, 50, 50, 5]$$

представлены на рисунке. Рисунок подтверждает справедливость полученных заключений.

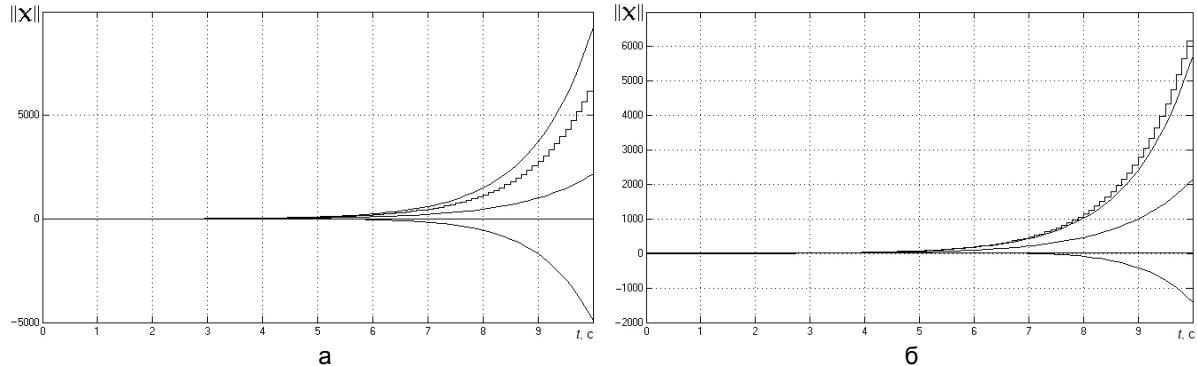


Рисунок. Оценочная трубка из условия качественной экспоненциальной неустойчивости построенная:
(а) – по параметрам качества (15); (б) – по параметрам качества (16)

Заключение

Полученные оценки динамических показателей качества в виде критического времени переходного процесса и выброса совместно с достаточными условиями качественной экспоненциальной неустойчивости позволили создать эффективные численные процедуры анализа неустойчивых непрерывных и дискретных динамических систем.

Работа выполнена при финансовой поддержке РФФИ (грант 09-08-00857-а «Методология применения теории качественной устойчивости при проектировании систем управления адаптивной оптикой»).

Приложение. Доказательства утверждений

Доказательство утверждения I

Из свойств нормы:

$$\|x(t)\| - \|e^{\beta \cdot t} x_0\| \leq \|x(t) - e^{\beta \cdot t} x_0\|,$$

$$\|e^{\beta \cdot t} x_0\| = e^{\beta \cdot t} \|x_0\|,$$

откуда, учитывая (2), получим:

$$\|x(t)\| - \|e^{\beta \cdot t} x_0\| \leq \rho \cdot (e^{[\beta+r]t} - e^{\beta \cdot t}) \cdot \|x_0\|, \quad (\text{П.1})$$

из которого при $r=0$ получим:

$$\|x(t)\| = e^{\beta \cdot t} \|x_0\|,$$

разрешив которое относительно t , с учетом (5) получим:

$$t_c = \frac{1}{\beta} \cdot \ln(\delta_c). \quad (\text{П.2})$$

Рассмотрим миноранту из неравенства (П.1):

$$\max_t x_m(t) = (\rho+1) \cdot e^{\beta \cdot t_\sigma} \|x_0\| - \rho \cdot e^{[\beta+r]t_\sigma} \cdot \|x_0\|. \quad (3)$$

Чтобы найти t_σ , возьмем производную по времени и, приравняв к нулю, разрешим относительно t_σ :

$$t_\sigma = \frac{1}{r} \ln \left(\frac{(\rho+1)\beta}{\rho(\beta+r)} \right). \quad (\text{П.44})$$

Подставив (П.3) и (П.4) в (7), получим

$$\sigma_0 = (\rho+1) \cdot e^{\frac{\beta \cdot \ln((\rho+1)\beta)}{r(\beta+r)}} - \rho \cdot e^{\frac{(\beta+r) \cdot \ln((\rho+1)\beta)}{r(\beta+r)}}. \quad (\text{П.55})$$

Равенства (П.2) и (П.5) соответствуют равенствам (9) и (10), что и требовалось доказать.

Доказательство утверждения 2

Из свойств нормы

$$\|x(m)\| - \|\beta^m x_0\| \leq \|x(m) - \beta^m x_0\|,$$

$$\|\beta^m x_0\| = \beta^m \|x_0\| \text{ при } \beta \geq 0,$$

откуда, учитывая (4), получим

$$\|x(m)\| - \beta^m \|x_0\| \leq \rho \cdot ((\beta + r)^m - \beta^m) \cdot \|x_0\|, \quad (\Pi.6)$$

из которого при $r = 0$ получим:

$$\|x(m)\| = \beta^m \|x_0\|,$$

разрешив которое относительно m , с учетом (6) и $t = mT$ получим

$$t_c = T \cdot \log_{\beta}(\delta_c). \quad (\Pi.7)$$

Рассмотрим монотонность неравенства (П.6):

$$\max_m x_m(m) = (\rho + 1) \cdot \beta^{m_\sigma} \|x_0\| - \rho \cdot (\beta + r)^{m_\sigma} \cdot \|x_0\|. \quad (\Pi.8)$$

Чтобы найти m_σ , возьмем производную по времени и, приравняв к нулю, разрешим относительно m_σ :

$$m_\sigma = \log_{\left(\frac{\beta+r}{r}\right)} \left(\frac{(\rho+1) \ln \beta}{\rho \ln(\beta+r)} \right). \quad (\Pi.9)$$

Подставив (П.8) и (П.9) в (8), получим

$$\sigma_0 = (\rho + 1) \cdot \beta^{\log_{\left(\frac{\beta+r}{r}\right)} \left(\frac{(\rho+1) \ln \beta}{\rho \ln(\beta+r)} \right)} - \rho \cdot (\beta + r)^{\log_{\left(\frac{\beta+r}{r}\right)} \left(\frac{(\rho+1) \ln \beta}{\rho \ln(\beta+r)} \right)}. \quad (\Pi.10)$$

Равенства (П.7) и (П.5) соответствуют равенствам (11) и (12), что и требовалось доказать.

Литература

1. Бобцов А.А., Быстров С.В., Григорьев В.В., Мансурова О.К., Мотылькова М.М. Качественная устойчивость и неустойчивость непрерывных и дискретных динамических систем // Труды 2-й Российской мультиконференции по проблемам управления. – СПб: ФГУП ЦНИИ «Электроприбор», 2008. – С. 41–43.
2. Рабыш Е.Ю., Григорьев В.В., Быстров С.В. Анализ поведения неустойчивых непрерывных и дискретных динамических систем // Сборник статей I международной заочной научно-технической конференции «Информационные технологии. Радиоэлектроника. Телекоммуникации». – Тольятти: Изд-во ПВГУС, 2011. – С. 263–270.
3. Grigoryev V.V., Mansurova O.K. Qualitative Exponential Stability and Instability of Dynamical System // Preprints of 5-th IFAK Symposium on Nonlinear Control Systems (NOLCOS'01). – St.Petersburg: IPME RAS, 2001. – P. 899–902.
4. Grigoryev V.V., Michailov S.V. Analysis and Synthesis Methods Based on Lyapunov's Method // Abstracts the Second Int. Conf. D. Eq. and Appl. – St. Petersburg: SPBSPU, 1998. – P. 37–38.

Рабыш Евгений Юрьевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, Rabysh@yandex.ru

Григорьев Валерий Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, grigvv@yandex.ru

Быстров Сергей Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, sbystrov@mail.ru

Спорягин Анатолий Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, Avsporyagin@yandex.ru

СИСТЕМА УПРАВЛЕНИЯ СЛЕДЯЩЕГО ЭЛЕКТРОПРИВОДА С НЕЖЕСТКОЙ ИСПОЛНИТЕЛЬНОЙ ОСЬЮ

А.А. Абдуллин, В.А. Толмачев

Приведены результаты математического моделирования процессов слежения электропривода азимутальной оси большого телескопа ТИ-3.12м Алтайского оптико-лазерного центра. Предложены два структурных решения системы управления. Показана эффективность применения дополнительной обратной связи по скорости второй массы исполнительной оси. Применение обратной связи позволило расширить диапазон скоростей плавного слежения и уменьшить среднеквадратичные ошибки слежения.

Ключевые слова: следящий электропривод, исполнительная ось, двухмассовый механизм, регулирование положения, скоростная подсистема.

Введение

Объектом исследования в настоящей работе является следящий электропривод азимутальной оси опорно-поворотного устройства большого телескопа ТИ-3.12м. Угломестная и азимутальная оси опорно-поворотного устройства (ОПУ) являются нежесткими [1].

Следящие электроприводы осей, как правило, строятся по структуре с подчиненным регулированием параметров. При этом полоса пропускания частот, определяющая динамические ошибки слежения при наличии лишь датчиков угла и скорости только двигателя (первой массы), ограничивается наименьшей из частот механических резонансов оси и коэффициентом соотношения масс [2].

Расширение полосы пропускания системы в условиях нежесткости осей ОПУ возможно при введении дополнительных обратных связей по скорости и углу поворота второй массы. Последнее может быть достигнуто либо установкой соответствующих датчиков на обеих массах, либо использованием наблюдателей неизмеряемых датчиками координат.

Целью работы является исследование динамических характеристик безредукторного следящего электропривода двухмассовой азимутальной оси с обратными связями по измеряемым скорости и углу поворота двигателя и скорости вилки с угломестной осью (второй массы).

Описание системы

Структурная схема системы управления следящего электропривода азимутальной оси представлена на рис. 1. Структурно система содержит контур регулирования угла и подчиненную ему трехконтурную скоростную подсистему, при этом внутренний контур регулирования момента представлен апериодическим звеном первого порядка. Настройка скоростной подсистемы приведена в работе [3].

По данным разработчика ОПУ телескопа ТИ-3.12м расчетные параметры двухмассовой модели механизма азимутальной оси следующие: момент инерции первой массы $J_1=6600 \text{ кг}\cdot\text{м}^2$; момент инерции второй массы $J_2=197300 \text{ кг}\cdot\text{м}^2$; коэффициент жесткости $C_{12}=8,62\times 10^8$; момент нагрузки на валу двигателя $M_{c1}=800 \text{ Н}\cdot\text{м}$ типа «сухое трение»; момент нагрузки на второй массе – ветровой с заданной спектральной плотностью и максимальным значением до $4400 \text{ Н}\cdot\text{м}$. Примем коэффициенты передачи датчиков момента и скорости равными $K_m=1,34 \text{ В}/(\text{Н}\cdot\text{м})$, $K_\Omega=38,1 \text{ В}\cdot\text{с}$ соответственно и $T_{m1}=4\times 10^{-4} \text{ с}$. Тогда, используя методику синтеза регуляторов контуров скоростной подсистемы с обратной связью только по скорости первой массы, изложенную в работе [4], получим $K_{n1}=201$, $T_{n1}=0,0713 \text{ с}$. Полоса пропускания внешнего контура регулирования скорости составляет величину $f_n=4,5 \text{ Гц}$ при переходном процессе, соответствующем настройке на «технический оптимум» со временем $t_{n1}=0,214 \text{ с}$ и перерегулированием 4,3%.

Попытка поднять полосу пропускания частот до значения $f_n=17,8 \text{ Гц}$ с сохранением эталонного качества переходного процесса, пересчитав параметры регуляторов контуров ($K_{n1}=804$, $T_{n1}=0,0178 \text{ с}$) в системе с обратной связью только по скорости двигателя, не удается, поскольку при этом возрастает колебательность переходного процесса относительно эталонного.

Введение дополнительных обратных связей по сигналу идеального датчика скорости второй массы на вход контуров с $K_1=0,1$ и $K_2=0,9$ позволяет реализовать требуемую полосу пропускания частот скоростной подсистемы, при этом время переходного процесса уменьшается до $t_{n1}=0,054 \text{ с}$.

Контур регулирования угла содержит пропорционально-интегральный регулятор (ПИ-регулятор) угла, на входе которого сравнивается напряжение $U_{задa}$, пропорциональное заданному значению угла поворота вала двигателя, с напряжением, поставляемым датчиком угла с коэффициентом передачи K_a . Выходной сигнал регулятора угла с коэффициентом передачи K_{n2} и постоянной времени интегрирования T_{n2} поставляет напряжение $U_{зад\Omega}$ для скоростной подсистемы.

При синтезе контура регулирования угла подчиненный контур регулирования скорости заменяется апериодическим звеном первого порядка. Контур настраивается на «симметричный оптимум» с параметрами регулятора $K_{n1}=804$ и $T_{n1}=0,0713 \text{ с}$, при этом расчетное время реакции углового контура на задающие и возмущающие воздействия определяется величиной $t_{n2}=0,214 \text{ с}$.

Система с выбранной структурой обладает астатизмом второго порядка относительно задающих воздействий, т.е. при идеальном датчике угла она обеспечивает нулевые установившиеся значения ошибки при работе системы с постоянным и с линейно возрастающим сигналом задания.

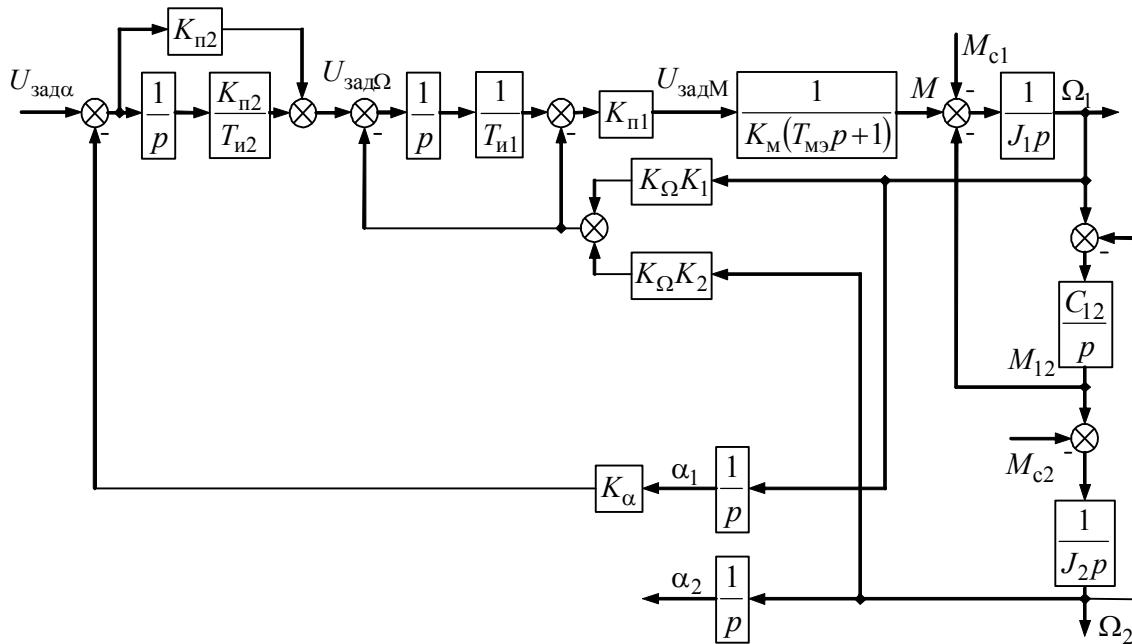


Рис. 1. Структурная схема системы управления следящего электропривода азимутальной оси

При слежении за сигналом, меняющимся во времени с постоянным ускорением, системе свойственна постоянная по величине ошибка, пропорциональная ускорению

$$E_{\text{уск}} = \frac{\alpha_0''}{D_\varepsilon}.$$

Коэффициент пропорциональности между ошибкой по ускорению и ускорением α_0'' задания называют добротностью системы (D_ε) по ускорению. Этот коэффициент численно равен коэффициенту передачи разомкнутой системы, т.е. в данном случае

$$D_\varepsilon = \frac{K_{\text{n}2}}{T_{\text{i}2}} \cdot \frac{K_\alpha}{K_\Omega}.$$

Уменьшение ошибки следящей системы, построенной на основе подчиненного регулирования, в основном может быть достигнуто двумя путями: повышением порядка астатизма основного контура (положения) и использованием принципа комбинированного управления. Первый путь практически нереализуем, поскольку введение еще одного интегрирующего звена в контур регулирования угла приведет к потере устойчивости.

Комбинированное управление предусматривает создание двух каналов – разомкнутого и замкнутого позиционного. Разомкнутый канал с целью достижения минимальных ошибок по ускорению при использовании структуры подчиненного регулирования целесообразно организовать на основе блока вычисления второй производной задающего сигнала, подавая его выходной сигнал непосредственно на вход контура регулирования положения. Выбором коэффициента $K_{\text{уск}}$ можно свести ошибку по ускорению к нулю, не влияя на динамические качества собственно контура регулирования положения. Для полной компенсации скоростной ошибки достаточно выполнить условие

$$K_{\text{уск}} = \frac{1}{D_\varepsilon}.$$

Моделирование и анализ

Исследованию подлежали две системы с параметрами двухмассового механизма исполнительной оси, приведенными выше.

- «Система 1» – система регулирования угла с датчиками скорости и угла только первой массы; полоса пропускания системы ограничивается резонансной частотой двухмассового механизма и коэффициентом соотношения масс. Расчетные параметры звеньев системы с $f_{\text{n}}=4,5$ Гц; $K_m=0,0013$ В/(Н·м); $K_\Omega=38,1$ В·с; $K_\alpha=1$; $K_{\text{n}1}=201$; $T_{\text{i}1}=0,0713$ с; $K_{\text{n}2}=267$; $T_{\text{i}2}=0,285$ с; $K_1=1$; $K_2=0$.

- «Система 2» – система регулирования угла с датчиками скорости обеих масс и угла первой массы; полоса пропускания системы расширена за счет введения обратной связи по скорости второй массы в скоростной подсистеме. Расчетные параметры звеньев системы $f_n=17,8$ Гц; $K_m=0,0013$ В/(Н·м); $K_\Omega=38,1$ В·с; $K_a=1$; $K_{n1}=804$; $T_{u1}=0,0178$ с; $K_{n2}=1068$; $T_{u2}=0,0713$ с; $K_1=0,1$, $K_2=0,9$.

Исследовались режимы слежения за сигналами задания, меняющимися с постоянной скоростью и постоянным ускорением. Моделирование процессов слежения проводилось в среде MatLab/Simulink.

Время моделирования при слежении за сигналом, меняющимся с постоянной скоростью, составляло 10 с при различных скоростях задания. На первую массу системы действовал момент сопротивления типа «сухое трение» $M_{c1}=800$ Н·м, а на вторую – «ветровая нагрузка» с максимальным моментом $M_{c2}=4400$ Н·м. Результат моделирования ветрового момента нагрузки представлен на рис. 2.

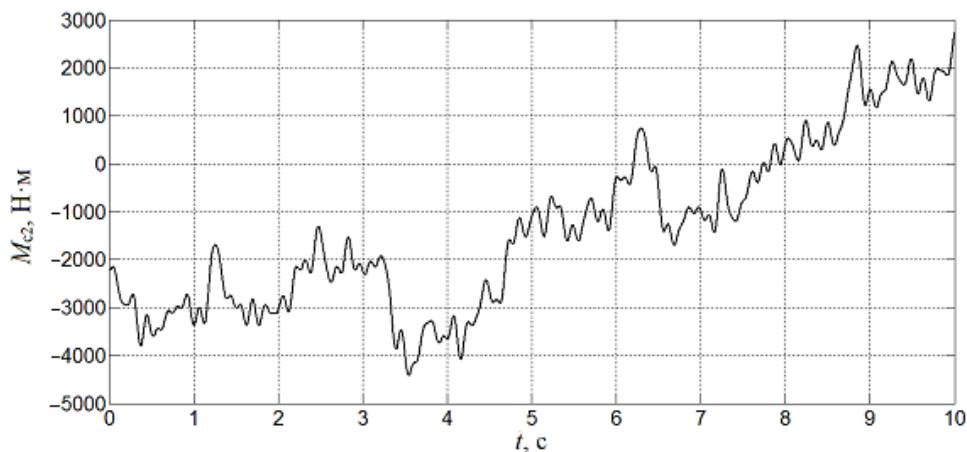


Рис. 2. Результат моделирования ветрового момента нагрузки

Результаты моделирования процессов в системе при слежении за сигналом задания, изменяющемся с постоянной скоростью, после выхода на установившийся режим представлены в табл. 1. Приняты следующие обозначения: Δ сп. кв. – среднеквадратичное отклонение угла от его заданного значения в установившемся режиме; $\Delta \alpha_m$ – максимальная амплитуда ошибки. На рис. 3–6 представлены: α_z – заданное значение угла; α_1 – угол первой массы; α_2 – угол второй массы.

Скорость, "/с	Вид нагрузки	«Система 1»		«Система 2»	
		Δ сп. кв., "	$\Delta \alpha_m$, "	Δ сп. кв., "	$\Delta \alpha_m$, "
0,375	Трение	автоколебания	–	$5,987 \times 10^{-10}$	0,052
1,75		$2,7 \times 10^{-6}$	0,879	$4,2 \times 10^{-10}$	0,131
5		$1,34 \times 10^{-5}$	1,1	$4,64 \times 10^{-10}$	0,27
0,375	Трение и ветер	автоколебания	–	0,51	1,11
1,75		1,8	4,103	0,49	1,06
5		1,65	4,717	0,52	1,11

Таблица 1. Результаты моделирования при слежении за сигналом, меняющимся с постоянной скоростью

Из приведенных в табл. 1 результатов видно, что при наличии лишь момента нагрузки M_{c1} и при изменении задания со скоростью 0,375 "/с, в «Системе 1» на всем промежутке времени моделирования слежение системы за сигналом сопровождается автоколебаниями.

На рис. 3 представлены переходные процессы в «Системе 1» при скорости задания 1,5 "/с и наличии момента нагрузки на первой массе. С уменьшением скорости задания амплитуда колебаний увеличивается, а период уменьшается. При скорости задания 1,75 "/с спустя 4,5 с от начала движения система выходит на установившийся режим слежения. Этую скорость будем называть нижней границей скоростей плавного слежения в «Системе 1».

В «Системе 2» нижняя граница скорости плавного слежения, при которой в системе не возникают автоколебания, снизилась до 0,375 "/с (рис. 4), что в 4 раза ниже, чем в «Системе 1».

При моделировании на тех же скоростях, но с нагрузкой на обеих массах в установившемся режиме слежения за сигналом задания, значение среднеквадратичной ошибки Δ сп. кв. не превышает 1,8" для «Системы 1». При этом максимальная амплитуда ошибки $\Delta \alpha_m$ в среднем составляет около 4,4".

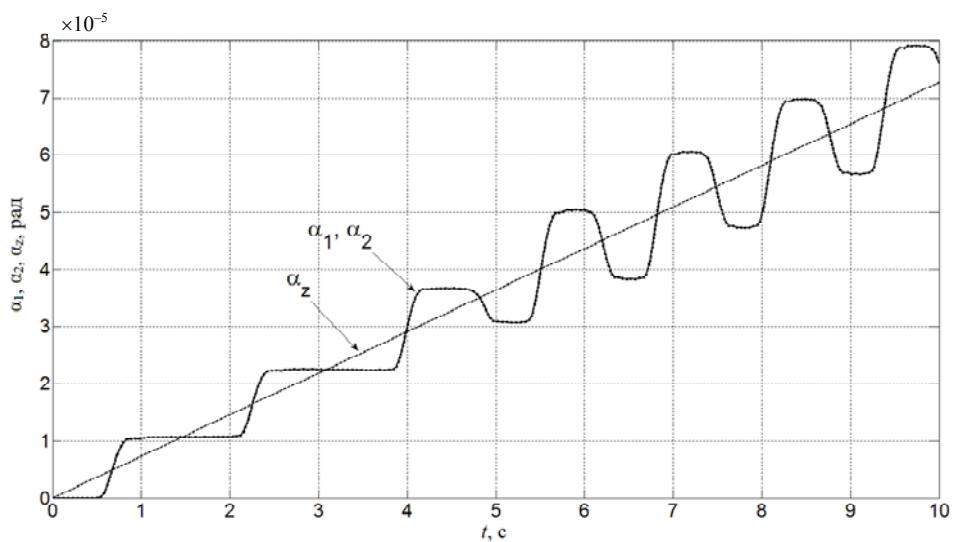


Рис. 3. Переходные процессы в «Системе 1» при скорости задания – 1,5 °/с и наличии момента нагрузки на первой массе

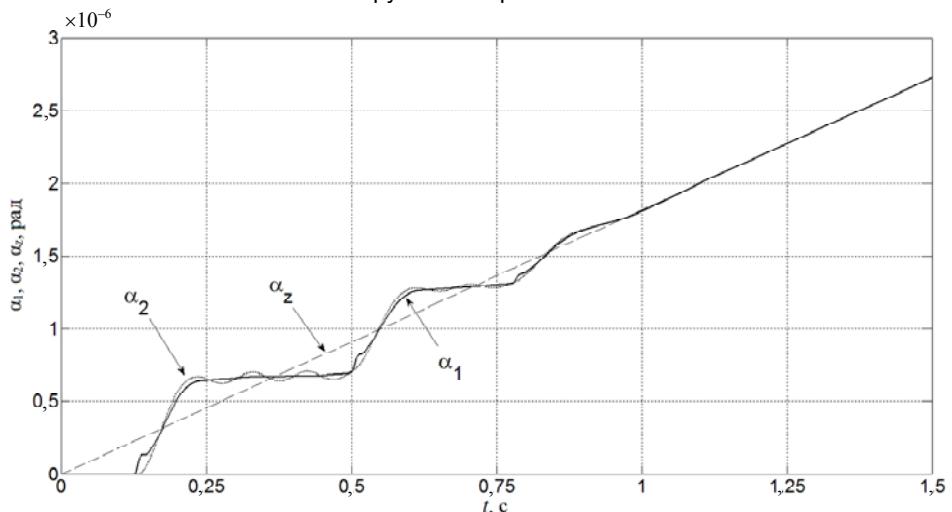


Рис. 4. Процессы выхода на режим слежения в «Системе 2» при скорости задания – 0,375 °/с и наличии момента нагрузки на первой массе

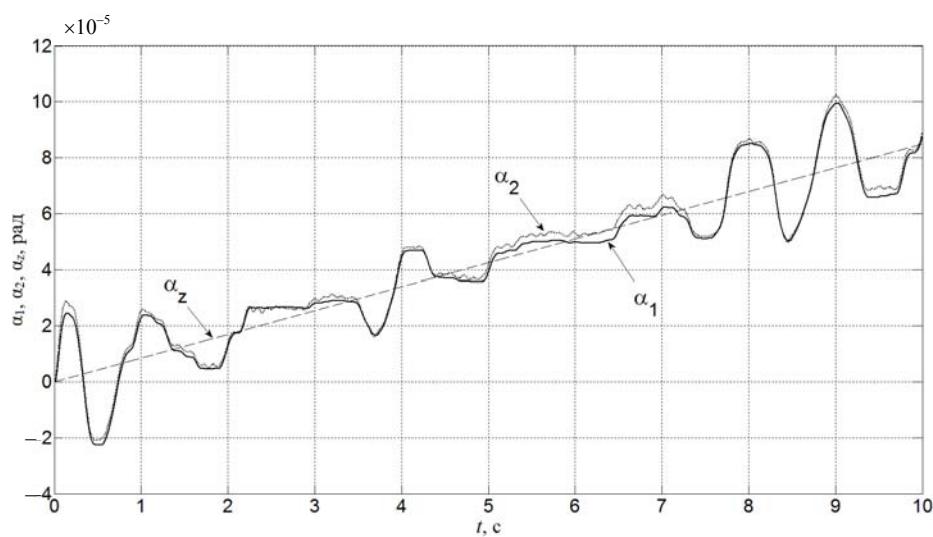


Рис. 5. Переходные процессы в «Системе 1» при скорости задания – 1,75 °/с и наличии момента нагрузки на обеих массах

На рис. 5 представлены результаты моделирования при сигнале задания, изменяющемся со скоростью 1,75 "с. В случае «Системы 2» Δ ср. кв. не превышает 0,52", что в 3–3,5 раза меньше, чем в «Системе 1». На рис. 6 представлены результаты моделирования системы с сигналом задания, изменяющимся со скоростью 1,75 "с, и наличием нагрузки на обеих массах.

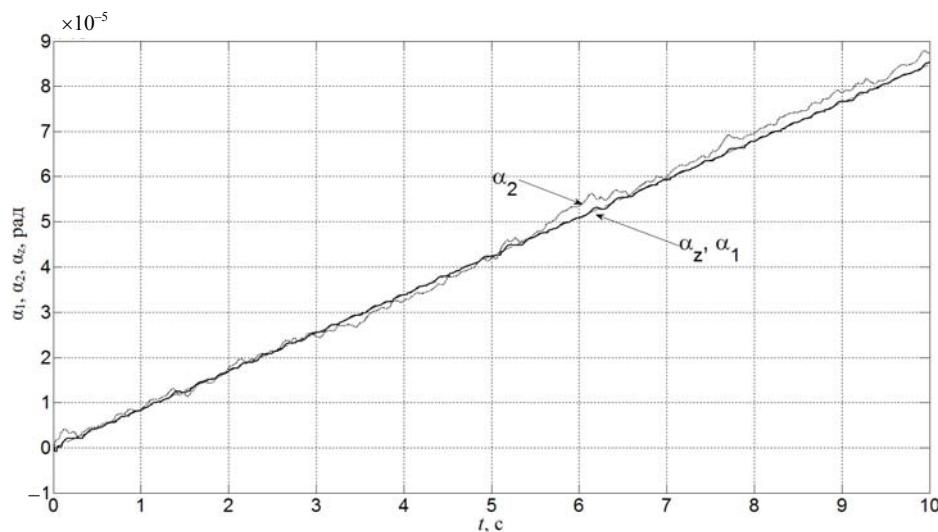


Рис. 6. Переходные процессы в «Системе 2» при скорости задания – 1,75 °/с и наличии момента нагрузки на обеих массах

Результаты моделирования процессов в системе при слежении за сигналом задания, изменяющимся с постоянным ускорением, после выхода на установившийся режим представлены в табл. 2.

Ускорение, "/с²	Вид нагрузки	«Система 1»		«Система 2»	
		Δ ср. кв., "	$\Delta \alpha_m$, "	Δ ср. кв., "	$\Delta \alpha_m$, "
0,5	Трение	0,0205	0,943	0,0014	0,0395
1		0,0409	0,636	0,0028	0,0379
5		0,2047	1,046	0,0139	0,0665
0,5	Трение и ветер	1,8	4,331	0,62	1,038
1		1,99	5,166	0,65	1,031
5		1,47	4,748	0,64	1,094

Таблица 2. Результаты моделирования при слежении за сигналом, меняющимся с постоянным ускорением

По результатам, приведенным в табл. 2, видно, что при наличии лишь момента нагрузки M_{cl} , среднеквадратичная ошибка Δ ср. кв. в обеих системах увеличивается практически пропорционально ускорению, но не превышает допустимого значения 1", при этом в «Системе 2» ошибка меньше практически на порядок. При наличии нагрузки на обеих массах значение Δ ср. кв. в «Системе 1» не превышает 2", а в «Системе 2» – 0,65", что практически в 4 раза меньше.

Заключение

Предложена структура системы управления с дополнительной обратной связью по скорости второй массы. Сравнительный анализ результатов моделирования позволяет сделать заключение о возможности снижения динамических ошибок и снижения нижней границы скоростей плавного слежения практически пропорционально расширению полосы пропускания частот скоростной подсистемы за счет введения дополнительной обратной связи по скорости второй массы. Применение в следящем приводе трехконтурной скоростной подсистемы с дополнительной обратной связью по скорости второй массы позволяет снизить нижнюю границу диапазона скоростей плавного слежения практически в 4 раза и во столько же раз уменьшить среднеквадратичные ошибки в типовых режимах слежения.

Результаты работы планируется использовать при разработке и изготовлении электроприводов систем наведения телескопов измерений.

Литература

- Исупов А.Н., Исупов К.С., Храмов С.Н. Определение амплитудно-частотных характеристик альт-азимутального опорно-поворотного устройства крупногабаритного наземного телескопа // Изв. вузов. Приборостроение. – 2008. – Т. 51. – № 6. – С. 38–44.

2. Борцов Ю.А., Соколовский Ю.Г. Автоматизированный электропривод с упругими связями. – СПб: Энергоатомиздат, 1992. – 288 с.
3. Абдуллин А.А., Толмачев В.А. Система регулирования скорости двухмассового механизма с использованием наблюдателя // Изв. вузов. Приборостроение. – 2011. – Т. 54. – № 5. – С. 66–71.
4. Толмачев В.А. Синтез следящего электропривода оси опорно-поворотного устройства // Изв. вузов. Приборостроение. – 2008. – Т. 51. – № 6. – С. 68–72.

Абдуллин Артур Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, artur.abdullin@gmail.com

Толмачев Валерий Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, tolmachev@ets.ifmo.ru

УДК 621.314.333

**АНАЛИЗ ЭНЕРГЕТИЧЕСКИХ ПОКАЗАТЕЛЕЙ МНОГОУРОВНЕННЫХ
ПОЛУПРОВОДНИКОВЫХ ПРЕОБРАЗОВАТЕЛЕЙ
СИСТЕМ ЭЛЕКТРОПРИВОДА**

К.Е. Михеев, В.С. Томасов

Проведен анализ энергетических показателей энергоподсистем электроприводов переменного тока, построенных на базе многоуровневых преобразователей напряжения различной топологии. В пакете Matlab/Simulink с использованием библиотек Simulink и SimPowerSystems разработан специализированный блок вычислений «Measurements», позволяющий проводить анализ и сравнение энергетических показателей многоуровневых преобразователей.

Ключевые слова: энергоподсистема, энергетические показатели, многоуровневые преобразователи, электропривод, синусоидальная широтно-импульсная модуляция, ШИМ, моделирование.

Введение

Диапазон регулирования угловой скорости в современных системах прецизионного электропривода может определяться значениями от нескольких угловых секунд в секунду до десятков градусов в секунду. Такой диапазон скорости требует соответствующего диапазона изменения входного напряжения, что зачастую не удается реализовать при построении энергоподсистем на основе традиционных мостовых схем с широтно-импульсной модуляцией (ШИМ). По этой причине в настоящее время практикуется применение многоуровневых преобразователей напряжения, позволяющих значительно расширить диапазон регулирования напряжения. Одним из важнейших преимуществ многоуровневых преобразователей, помимо большого диапазона напряжения, является их способность генерировать на выходе напряжение, более качественное по энергетическим показателям и, в частности, по гармоническому составу, чем у одноуровневых преобразователей. Методы определения, формулы расчета энергетических показателей качества электроэнергии хорошо известны и широко применяются при расчетах цепей синусоидального тока [1–3]. Оценка качества электроэнергии существенно усложняется при анализе многоуровневых преобразователей, так как на выходе последних имеется квазисинусоида сложной формы, разложение которой в ряд Фурье без помощи ЭВМ практически невозможно. Вычислительные мощности современных пакетов программ позволяют без особого труда смоделировать работу любых преобразователей и провести анализ их работы. Целью работы является анализ энергетических показателей энергоподсистем прецизионных электроприводов с широким диапазоном регулирования скорости, построенных на базе многоуровневых преобразователей напряжения, с применением пакета Matlab/Simulink.

Обзор основных схем многоуровневых преобразователей

В настоящее время в основном распространены две основные топологии построения многоуровневых инверторов [3, 4].

Мостовая схема. Строится по принципу традиционной трехфазной мостовой схемы инвертора, но с той лишь разницей, что на каждое плечо моста приходится не два полностью управляемых ключевых элемента, а 4, 6 и т.д., с включенными параллельно им обратными диодами. Кроме того, для построения многоуровневых инверторов используются дополнительные элементы, такие как блокирующие диоды или конденсаторы с нефиксированным уровнем напряжения на них (возможны две разновидности топологии), необходимые для баланса нейтральной точки и корректной работы схемы. Общий вид одного плеча такой схемы для двух разновидностей топологии представлен на рис. 1. Принципиальные схемы, изображенные на рис. 1, а также обозначения элементов заимствованы из [5].

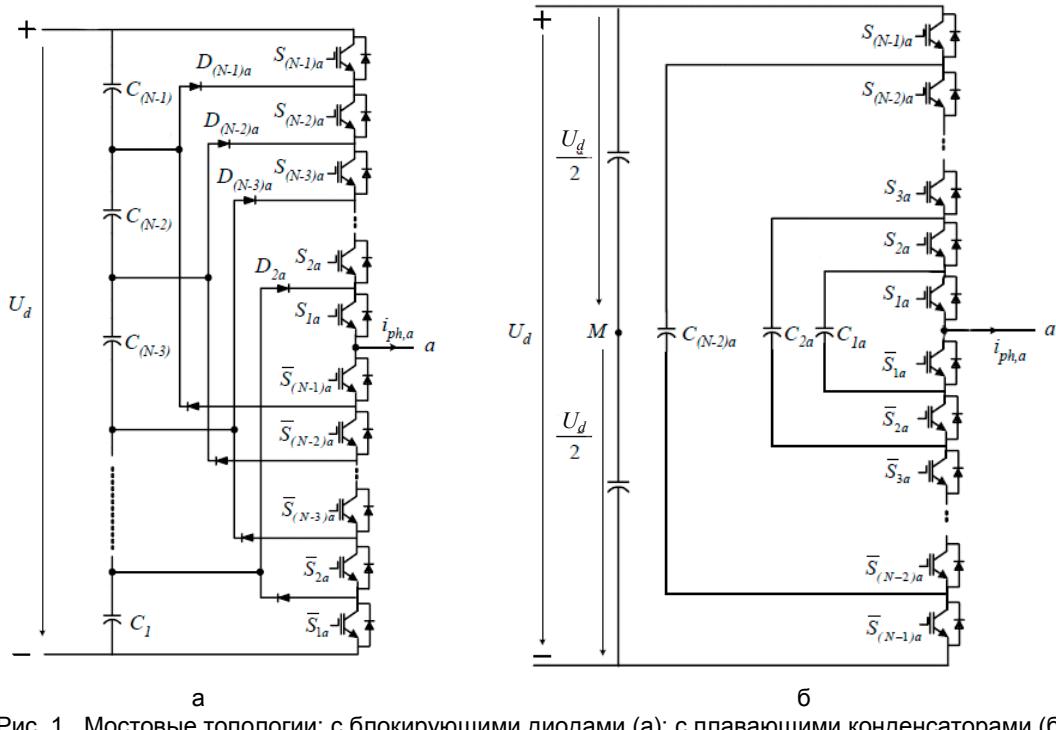


Рис. 1. Мостовые топологии: с блокирующими диодами (а); с плавающими конденсаторами (б)

Схема инвертора на базе однофазных Н-мостовых схем. Простейший вариант такой схемы строится на базе трех однофазных мостов, положительный вывод каждого из которых подключается к соответствующей фазе нагрузки, а отрицательные выводы всех трех соединены в общую точку. Большее число уровней в кривой выходного напряжения достигается последовательным включением нескольких Н-мостовых схем для каждой фазы. Общий вид данной топологии представлен на рис. 2. Обозначения и сама схема также заимствованы из источника [5]. Данная топология имеет, в силу ряда причин, большее распространение. К ним относятся простота схемы по сравнению с мостовой, отсутствие делительных или плавающих конденсаторов, что исключает сложные процессы зарядки-разрядки, более простое управление, возможность работы даже при выходе из строя одного или нескольких Н-мостов. Недостатком Н-мостовых схем является необходимость создания множества источников постоянного напряжения, что обычно достигается путем применения сложного многообмоточного трансформатора с последующим выпрямлением напряжения.

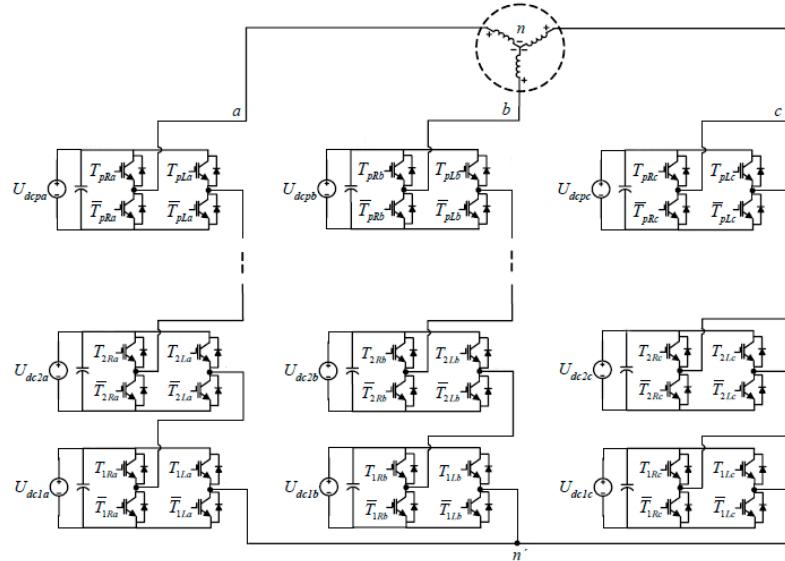


Рис. 2. Н-мостовая топология

Параметры моделирования

В работе принятые следующие параметры управления и нагрузки: амплитуда первой гармоники фазного напряжения на нагрузке $U_{\phi,1\max}=100$ В; индекс модуляции $m=1$; частота коммутации ключей инвертора $f_k=1$ кГц; трехфазная активно-индуктивная нагрузка $L_\phi=10$ мГн; $R_\phi=1$ Ом; способ модуляции

кривой напряжения – синусоидальная ШИМ. Моделирование реального источника постоянного напряжения (ИПН) не проводилось, так как это усложнило бы модель, и не является обязательным для анализа работы непосредственно самого инвертора. Реальный ИПН при моделировании заменен идеализированным источником – блоком DC Voltage Source (библиотека SimPowerSystems/Electrical Sources). Значение постоянного напряжения выбирается исходя из топологии схемы [3, 4]:

- для мостовых схем при применении синусоидальной ШИМ оно равно удвоенной амплитуде первой гармоники;
- для схем на базе Н-мостовых ячеек напряжение на входе каждой ячейки равно амплитуде первой гармоники выходного напряжения, деленной на число последовательно включенных Н-мостов.

Параметры ключевых элементов были выбраны одинаковыми для всех схем.

Для определения энергетических показателей качества собрана подсистема «Measurements» на основе стандартных блоков библиотек Simulink и SimPowerSystems (рис. 3).

Определение энергетических показателей

Известно [1–3], что полная, или кажущаяся, мощность S определяется произведением действующих значений напряжения и тока на входе системы. Она всегда превышает фактически передаваемую в нагрузку активную мощность P из-за появления неактивных составляющих полной мощности обменного характера – мощности сдвига Q (реактивной), мощности искажения T , мощности несимметрии H – и определяется выражением

$$S = UI, \quad (1)$$

где U, I – действующие значения напряжения и тока. В свою очередь, ее можно выразить через составляющие следующим образом:

$$S = \sqrt{P^2 + Q^2 + T^2 + H^2}. \quad (2)$$

Мощность несимметрии H в данном случае равняется нулю. Активная мощность P равна среднему значению мгновенной мощности за период питающего напряжения и определяет количество электромагнитной энергии, необратимо преобразующейся в другие формы энергии:

$$P = \frac{1}{T} \int_0^T u(t)i(t)dt, \quad (3)$$

где $u(t), i(t)$ – мгновенные значения напряжения и тока; T – период кривой напряжения (тока).

Полная и активная мощности связаны через коэффициент мощности:

$$K_{\text{мощн}} = \frac{P}{S}. \quad (4)$$

Реактивная мощность, или мощность сдвига Q , обусловлена сдвигом по фазе основной гармоники тока относительно напряжения питающей сети и связана с активной мощностью через коэффициент сдвига

$$K_{\text{сдв}} = \frac{P}{\sqrt{P^2 + Q^2}}. \quad (5)$$

Для линейных систем с синусоидальным напряжением коэффициент мощности и коэффициент сдвига одинаковы и равны косинусу угла сдвига фазы φ между током и напряжением:

$$K_{\text{мощн}} = K_{\text{сдв}} = \cos \varphi. \quad (6)$$

Для определения активной и реактивной мощности на основе измерений тока и напряжения цепи в пакете Matlab используется блок Active & reactive Power библиотеки SimPowerSystems.

При появлении несинусоидальных составляющих тока необходимо учитывать мощность искажения T

$$T = \sqrt{S^2 - (P^2 + Q^2)}. \quad (7)$$

Коэффициент искажения характеризует обмен энергией между источником и приемником, обусловленный высшими гармоническими составляющими тока:

$$K_{\text{иск}} = \frac{\sqrt{P^2 + Q^2}}{\sqrt{P^2 + Q^2 + T^2}}. \quad (8)$$

Коэффициент гармоник, или интегральный показатель гармонического состава тока (Total Harmonic Distortion), характеризует соотношение между энергией, обусловленной высшими гармониками тока, и энергией обусловленной основной (первой) гармоникой:

$$K_{\text{гарм}} = \frac{T}{\sqrt{P^2 + Q^2}} = \sqrt{\left(\frac{1}{K_{\text{иск}}}\right)^2 - 1}. \quad (9)$$

Коэффициент полезного действия энергоподсистемы определяется как отношение активной мощности, фактически передаваемой в нагрузку P_d , к потребляемой из сети активной мощности P :

$$\eta = \frac{P_d}{P} . \quad (10)$$

На рис. 3 представлена разработанная подсистема «Measurements», позволяющая на основе (1)–(10) определить активную (P), реактивную (Q), кажущуюся (S) фазные мощности; фазную мощность искажения (T), коэффициент искажения (K_{isk}), коэффициент мощности (K_m), коэффициент сдвига (K_{sdv}), коэффициент гармоник (K_garm), коэффициент гармоник тока (K_garm_I), коэффициент полезного действия инвертора η (KPD), а также среднее значение тока питания (I_{in}) посредством применения стандартных блоков библиотек Simulink и SimPowerSystems. В скобках приведены обозначения параметров, используемые в пакете Matlab при построении подсистемы и ее моделировании.

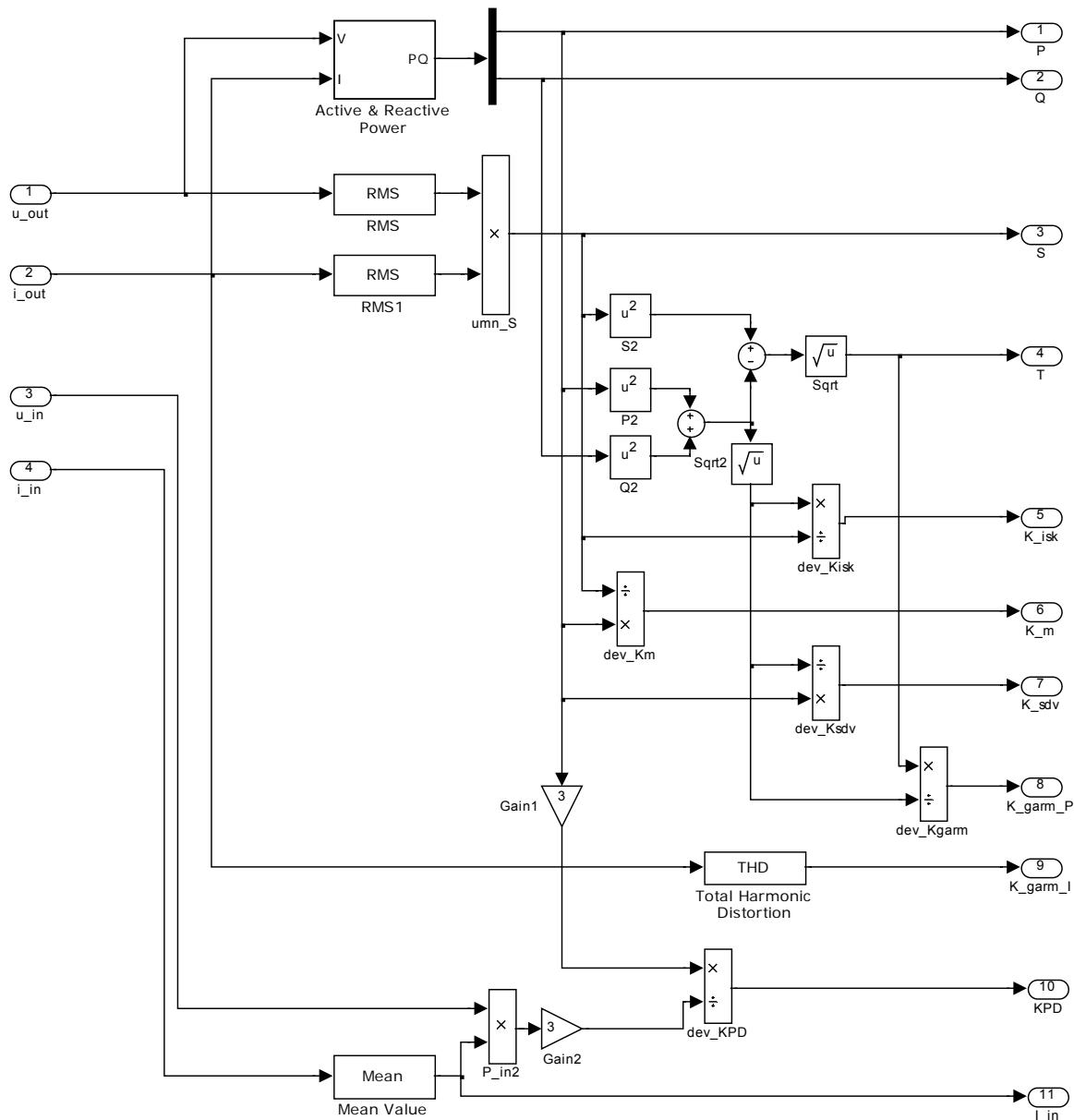


Рис. 3. Подсистема «Measurements»

Моделирование

Блоки управления транзисторами для каждой схемы состоят из стандартных блоков и представлены на моделях в виде подсистем. При синусоидальной ШИМ происходит сравнение опорного синусоидального напряжения, величина амплитуды которого определяется индексом модуляции m , с пилообразным напряжением постоянной амплитуды. В момент равенства этих сигналов компаратор выдает импульсы управления ключами соответствующей длительности [3, 4].

На рис. 4 приведены: схема трехуровневого мостового инвертора (мост 3×4) – вариант топологии рис. 1, б, с четырьмя транзисторами в каждом плече моста (рис. 4, а); схема пятиуровневого Н-мостового инвертора (Н-мост $3 \times 2H$) – вариант топологии рис. 2, с двумя Н-мостовыми ячейками на каждую фазу нагрузки (рис. 4, б). Изображения элементов и обозначения на рисунках соответствуют обозначениям, принятым в пакете Matlab. Для сравнения энергетических показателей авторами смоделированы мостовая двухуровневая схема (2 транзистора в плече – мост 3×2), мостовая четырехуровневая схема (6 транзисторов в плече – мост 3×6) и Н-мостовая двухуровневая схема (одна Н-мостовая ячейка на фазу, Н-мост $3 \times 1H$).

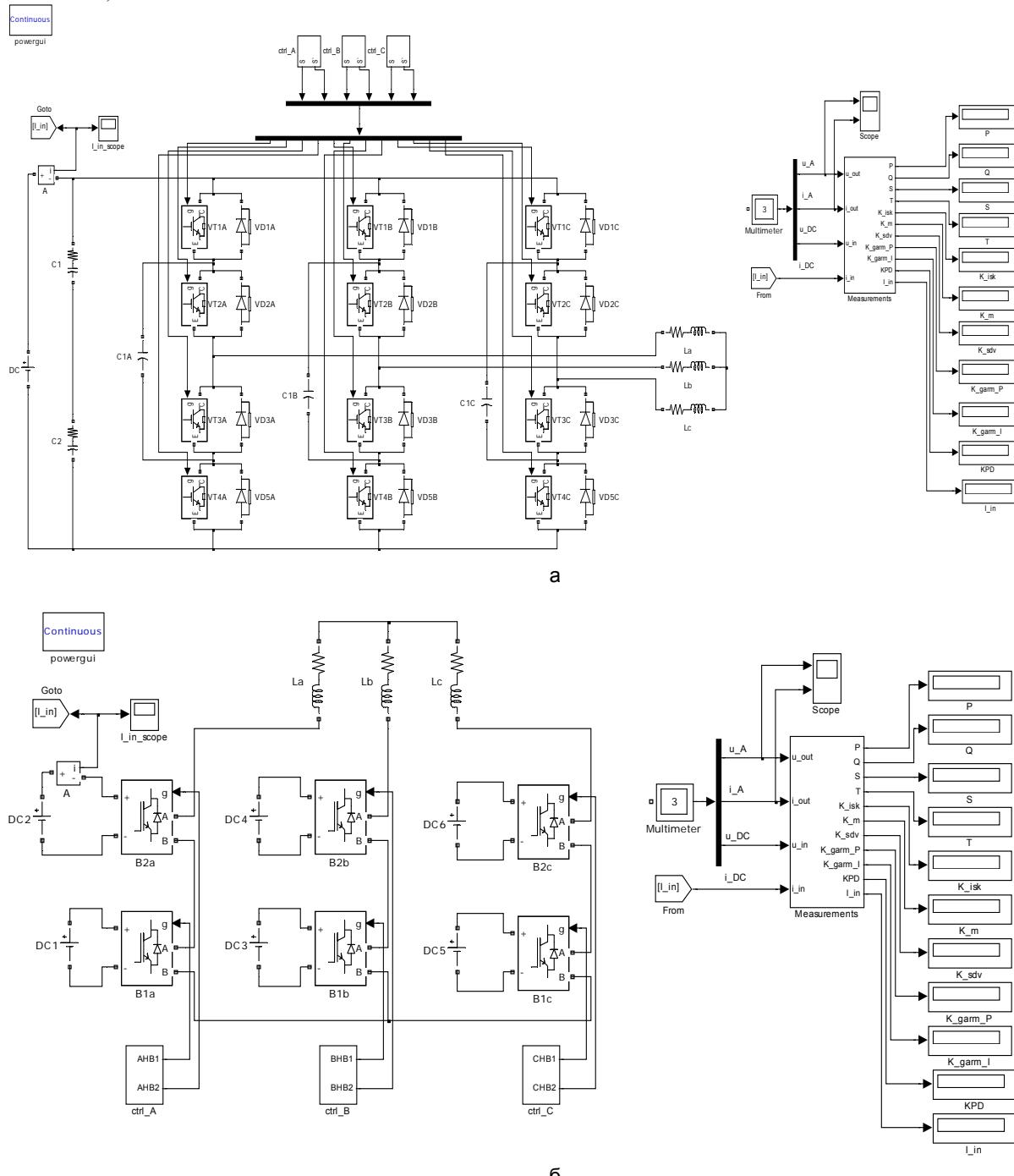


Рис. 4. Схемы моделирования: трехуровневый мостовой инвертор (а); пятиуровневый Н-мостовой (б)

Для анализа энергетических показателей схем получены кривые напряжения и тока на выходе инвертора. Для этого применен блок осциллографа Scope (библиотека Simulink/Sinks). Для сравнения проведено моделирование пяти схем многоуровневых инверторов.

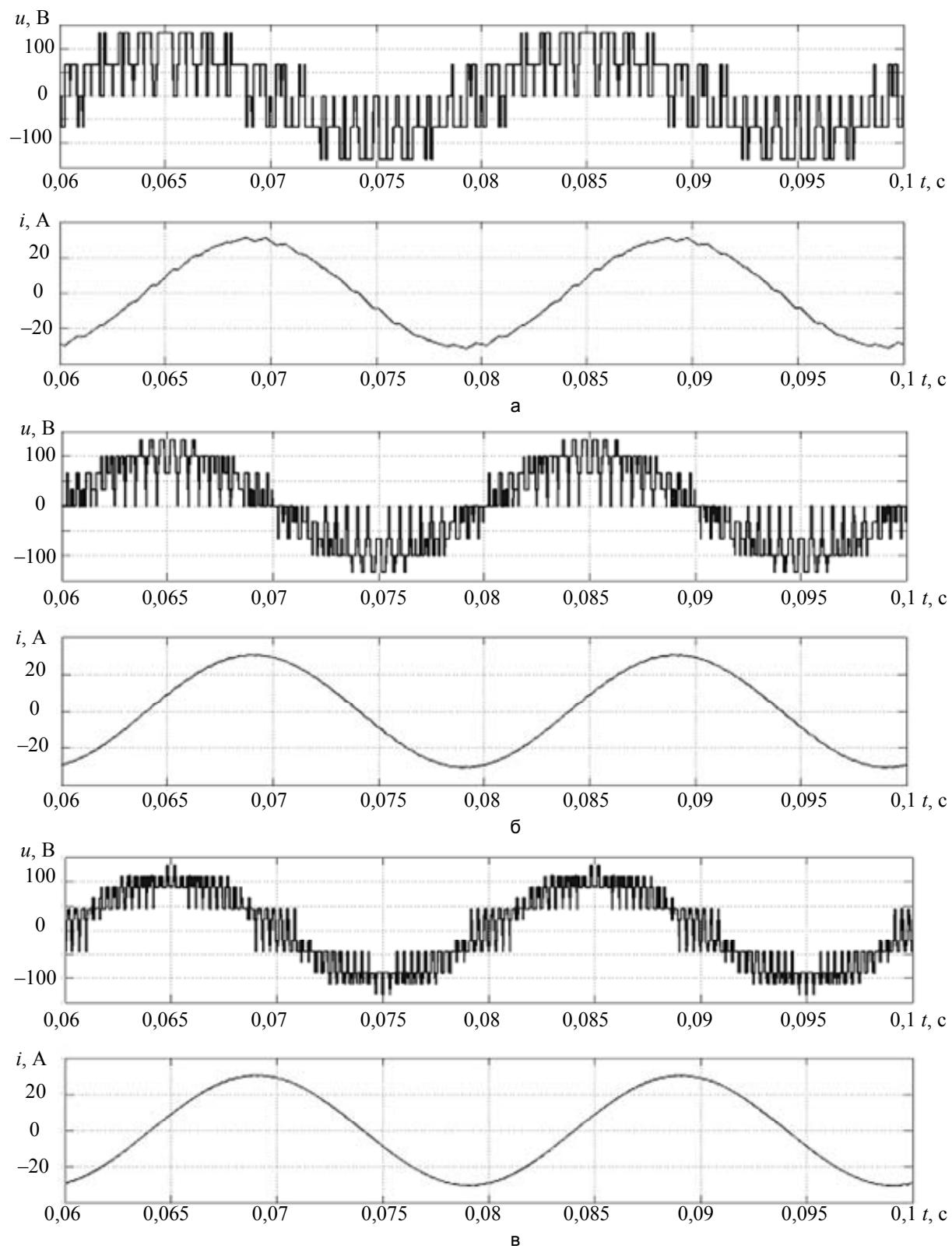


Рис. 5. Диаграммы выходных напряжений (u) и токов (i): мостовая двухуровневая (а); мостовая трехуровневая (б); мостовая четырехуровневая (в)

Сравнение результатов показало, что усложнение схемы приводит к увеличению числа ступеней в кривой фазного напряжения и, как следствие, уменьшению пульсаций в кривой фазного тока, что наблюдается на диаграммах напряжений и токов и выражается в улучшении энергетических показателей, в первую очередь, за счет значительного уменьшения мощности искажения. На рис. 5 приведены диаграммы токов и напряжений для трех мостовых схем. Для моделирования изображения взят временной участок от 0,06 с до 0,1 с, на котором уже закончились переходные процессы, имеющие место в начале ра-

боты схемы. Форма токов и напряжений для Н-мостовых схем почти идентична форме токов и напряжений для трех- и четырехуровневых мостовых схем, что подтверждается значениями показателей качества (таблица).

В таблице приведены наиболее значимые показатели качества электроэнергии на выходе преобразователя, полученные в результате моделирования: фазная активная мощность (P_ϕ), фазная реактивная мощность (Q_ϕ), фазная мощность искажения (T_ϕ), полная фазная мощность (S_ϕ), коэффициент искажения ($K_{иск}$), коэффициент мощности ($K_{мощн}$), коэффициент гармоник по напряжению ($K_{гарм}$), коэффициент гармоник тока ($K_{гарм тока}$). Как было отмечено выше, более сложные схемы инверторов дают лучшее качество электроэнергии, что выражается в уменьшении мощности искажения и соответствующем улучшении показателей качества.

Схема	P_ϕ , Вт	Q_ϕ , вар	T_ϕ , Вт	S_ϕ , В·А	$K_{иск}$	$K_{мощн}$	$K_{гарм}$	$K_{гарм тока}$
мост 3×2	454	1437	1040	1831	0,82	0,25	0,69	0,026
мост 3×4	460	1442	606	1631	0,93	0,28	0,4	0,008
мост 3×6	459	1442	458	1582	0,96	0,29	0,3	0,005
Н-мост $3 \times 1H$	456	1436	612	1626	0,93	0,28	0,41	0,008
Н-мост $3 \times 2H$	456	1435	389	1555	0,97	0,29	0,26	0,003

Таблица. Показатели качества электроэнергии, полученные в результате моделирования

Заключение

Разработаны математические модели многоуровневых полупроводниковых преобразователей, которые позволили осуществить анализ сложных по своей топологии и по математическому описанию схем силовой электроники, находящих применение в прецизионных быстродействующих и широкодиапазонных электроприводах. Анализ напряжений в фазах нагрузки подтвердил, что с увеличением числа уровней преобразуемого напряжения уменьшается амплитуда пульсаций тока нагрузки и улучшается его гармонический состав. Это позволяет увеличить срок работы электрической машины, снять ограничения на длину кабеля подключения двигателя, исключить установку дополнительных входных и выходных фильтров. Определены численные значения показателей качества электроэнергии для энергоподсистем систем электропривода с различной топологией и структурой выходных каскадов полупроводниковых преобразователей. Показано, что применение сложных схем инверторов позволяет получить лучшее качество электроэнергии, что выражается в уменьшении мощности искажения и соответствующем улучшении показателей качества.

Результаты работы нашли практическое применение при проектировании и разработке энергоподсистем прецизионных электроприводов телескопов траекторных измерений и оптико-электронных комплексов регистрации космических объектов и аппаратов, проводимых на кафедре электротехники и прецизионных электромеханических систем Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики по заказам Роскосмоса РФ.

Литература

1. Маевский О.А. Энергетические показатели вентильных преобразователей. – М.: Энергия, 1978. – 320 с.
2. Борисов П.А., Томасов В.С. Определение составляющих полной мощности энергоподсистем электротехнических комплексов в пакете Matlab // Exponenta Pro. Математика в применениях. – 2004. – № 1. – С. 40–44.
3. Зиновьев Г.С. Основы силовой электроники: Учебник. – Новосибирск: Изд-во НГТУ, 1999. – 199 с.
4. Пронин М.В., Воронцов А.Г. Силовые полностью управляемые полупроводниковые преобразователи (моделирование и расчет). – СПб: Электросила, 2003. – 172 с.
5. C. Boit. Investigation and Comparison of Multi-Level Converters for Medium Voltage Applications. – Berlin: Technische Universität, 2007. – 160 p.

Михеев Кирилл Евгеньевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, mikhkirill@yandex.ru
Томасов Валентин Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, зав. кафедрой, tomasov@ets.ifmo.ru

УДК 004.94:519.876.5

СРЕДА ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ АГЕНТНЫХ СИСТЕМ
РЕАЛЬНОГО ВРЕМЕНИ

М.Г. Пантелейев, Н.В. Кохтенко, С.В. Лебедев

Рассмотрена проблема создания среды имитационного моделирования многоагентных систем для экспериментального исследования и отработки различных моделей поведения интеллектуальных агентов реального времени в условиях командного противодействия. Сформулированы требования, предъявляемые к такой среде. Предложена архитектура среды, показаны особенности ее функционирования.

Ключевые слова: интеллектуальные агенты, многоагентные системы, среда имитационного моделирования, реальное время.

Введение

Концепция интеллектуальных агентов (ИА), реализующих автономное целенаправленное поведение в открытой, динамической, частично наблюдаемой среде, в настоящее время используется как при создании систем, функционирующих в физических мирах (автономные роботы, беспилотные летательные аппараты и др.), так и в виртуальных средах (программные агенты управления знаниями в Интернет и корпоративных сетях) [1, 2]. В связи с расширением практических приложений ИА и основанных на них многоагентных систем (МАС) особую актуальность приобретает создание технологических средств (инструментов и платформ), поддерживающих их разработку и исследование поведения [3]. Натурное исследование моделей автономного и группового поведения агентных систем в различных ситуациях, как правило, экономически нецелесообразно, поэтому актуальным является создание сред имитационного моделирования таких систем [4]. Существующие в настоящее время среды имитационного моделирования агентных систем, как показал анализ, обладают рядом недостатков, в частности: имеют высокую стоимость (например, VR-Forces), не поддерживают возможность моделирования ИА реального времени со сложной делиберативной архитектурой (например, AnyLogic), ориентированы на узкий класс агентных приложений и не имеют развитой поддержки проектирования агентов (например, Robosoccer).

Таким образом, важнейшей проблемой, содержащей широкое практическое внедрение агентных систем, является отсутствие платформы, позволяющей исследовать поведение ИА и МАС различной архитектуры средствами имитационного моделирования. Целью настоящей работы является создание программной среды имитационного моделирования (СИМ) агентных систем, функционирующих в открытых динамических мирах, с учетом требований реального времени.

Требования к СИМ

Общими требованиями к СИМ МАС являются:

- универсальность – поддержка моделирования различных типов ИА в части как физической подсистемы агента (ФПА), так и архитектуры интеллектуальной подсистемы агента (ИПА);
- открытость – возможность добавления новых компонентов, используемых при построении имитационных моделей ИА (например, новых типов сенсоров), включая его интеллектуальную подсистему;
- масштабируемость – возможность наращивания числа моделируемых объектов (в том числе ИА) без снижения общей производительности системы.

Важнейшим требованием к СИМ МАС является обеспечение возможности моделирования агентных систем в реальном масштабе времени, когда длительность такта моделирования соответствует динамике процессов в моделируемой физической среде. При этом ИА, функционирующие в открытых динамических мирах, необходимо строить и моделировать как системы ограниченной рациональности. Такие системы, в общем случае, не могут за отведенное время оптимально решать стоящие перед ними задачи, а должны рационально использовать имеющееся в конкретной ситуации время обдумывания решения. В высокодинамичных средах решения должны приниматься агентом в темпе обновления данных о состоянии среды. Вследствие этого ИПА должна выдавать команды управления физической подсистеме агента, а СИМ МАС – обеспечивать обработку этих команд (т.е. пересчет физического состояния всех моделируемых агентов) в пределах такта моделирования.

Другое важное требование к создаваемой СИМ МАС – возможность мониторинга ментальной деятельности агента, т.е. фиксации в каждом такте моделирования промежуточных результатов процесса обдумывания решений агентом на каждом значимом для используемой архитектуры и модели шаге принятия решений. Эта информация должна сохраняться для последующего воспроизведения и анализа сеанса моделирования вне режима реального времени. Данное требование обусловлено необходимостью

сравнительного исследования эффективности различных архитектур ИПА как систем ограниченной рациональности.

Еще одним требованием к СИМ МАС является возможность моделирования различных стратегий координации действий в группах агентов. В настоящее время в теории МАС разработано множество подходов и моделей координации поведения агентов, ориентированных на различные классы агентных систем и решаемые ими задачи. Выбор конкретных стратегий и моделей координации выполняется разработчиком моделируемой МАС, а СИМ не должна накладывать в этой части никаких ограничений. Задачей среды является предоставление системных механизмов, позволяющих моделировать различные стратегии координации как при наличии непосредственной коммуникации между агентами, так и без нее. В первом случае СИМ МАС должна поддерживать моделирование каналов связи между агентами с учетом их ограниченной пропускной способности и дальности действия, наличия шумов и ненадежной доставки сообщений. Координация в условиях отсутствия непосредственной коммуникации предполагает наблюдение за поведением других агентов посредством собственных сенсоров и (или) получение информации о других агентах от единого центра управления (выделенного агента-координатора).

Архитектура и особенности реализации СИМ МАС

На рис. 1 представлена архитектура СИМ МАС, удовлетворяющая сформулированным выше требованиям, в составе которой выделены среда разработки и среда исполнения.

Среда разработки позволяет создавать новых агентов, задавать начальные условия сценариев моделирования и конфигурировать среду исполнения. При создании ФПА используются библиотеки классов, реализующих различные модели динамики физического носителя агента (например, управляемого полета беспилотного летательного аппарата).

Разрабатываемая СИМ МАС ориентирована, в первую очередь, на исследование агентов с делиберативной архитектурой ИПА [1, 2]. В общем случае делиберативные агенты используют символическую модель окружающей среды и основанные на этой модели методы прогнозирования событий и планирования действий. Класс таких архитектур достаточно широк. СИМ МАС не накладывает никаких ограничений на выбор конкретной архитектуры ИПА, этот выбор целиком определяется разработчиком агентной системы. Вместе с тем для ряда широко известных типовых архитектур процесс проектирования ИПА поддерживается в СИМ МАС соответствующими шаблонами и библиотеками классов.

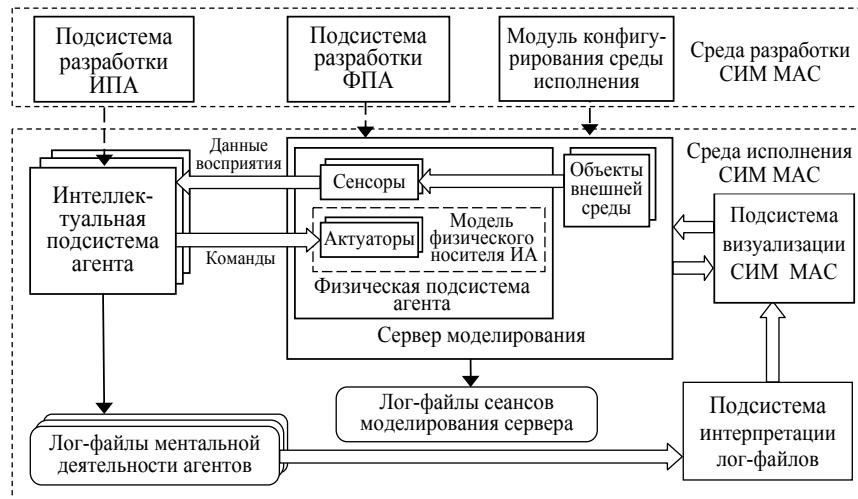


Рис. 1. Архитектура СИМ МАС

На первом шаге проектирования выбирается базовая архитектура, тем самым определяется состав подсистем ИПА и базовый ментальный цикл. Например, архитектура, основанная на полезности [1], включает подсистемы оценки обстановки, прогнозирования, планирования. На втором этапе выполняется разработка отдельных подсистем ИПА. При построении разных подсистем используются различные модели представления и обработки знаний – семантические сети, продукционные системы, фреймы, а также различные модели обработки неопределенности. Программная архитектура допускает использование существующих инструментов построения систем, основанных на знаниях (CLIPS, JESS и др.). Среда разработки ИПА поддерживает редактирование баз знаний агентов и конфигурирование методов их обработки.

Среда исполнения обеспечивает реализацию сеансов моделирования и включает следующие основные компоненты: сервер моделирования, подсистему визуализации и подсистему интерпретации лог-файлов сеансов моделирования. Модуль конфигурирования среды исполнения позволяет задавать длительность такта моделирования, режим работы подсистемы визуализации и логгирования процесса моделирования и т.п.

Среда исполнения СИМ MAC имеет клиент-серверную архитектуру. В соответствии с требованиями масштабирования и обеспечения реального времени ИПА различных агентов MAC реализуются в отдельных процессах и могут выполняться на отдельных компьютерах локальной сети. Подсистема визуализации также поддерживает клиент-серверное взаимодействие с сервером моделирования и может быть реализована на выделенном сервере для отображения моделируемых процессов в 3D-графике, что требует значительных вычислительных ресурсов.

Подсистема интерпретации лог-файлов предназначена для воспроизведения сеансов моделирования вне режима реального времени с возможностью детального анализа динамики ментальной деятельности агентов. Поскольку моделирование состояния многоагентного физического мира и ментальной деятельности отдельных агентов в общем случае выполняется на разных компьютерах, лог-файлы этих процессов также сохраняются на разных компьютерах. Состав сохраняемой информации в общем случае зависит от выбранной архитектуры агента и моделей принятия решений. В частности, для архитектуры ИА, основанной на полезности, эта информация может включать варианты действий, текущие значения их полезности, прогнозируемые ситуации и др. В режиме последующего воспроизведения сеанса моделирования подсистема интерпретации лог-файлов работает с распределенной файловой системой и обеспечивает отображение физического процесса и состояния ментальной деятельности агентов в каждом такте. Базовый алгоритм работы сервера моделирования представлен на рис. 2.

В каждом такте моделирования через заданный интервал времени Δt сервер опрашивает порты, к которым подключены ИПА моделируемых агентов, и считывает поступившие от них команды. Такт работы сервера задает интервал, в рамках которого ИПА должны принимать сенсорные данные, обрабатывать их, вырабатывать решения и выдавать команды управления. На основе этих команд сервер вычисляет новые физические состояния всех агентов (состояния ФПА). При этом реализуемые модели агентов как систем ограниченной rationalности должны гарантировать выполнение цикла работы ИПА в рамках заданного такта моделирования.

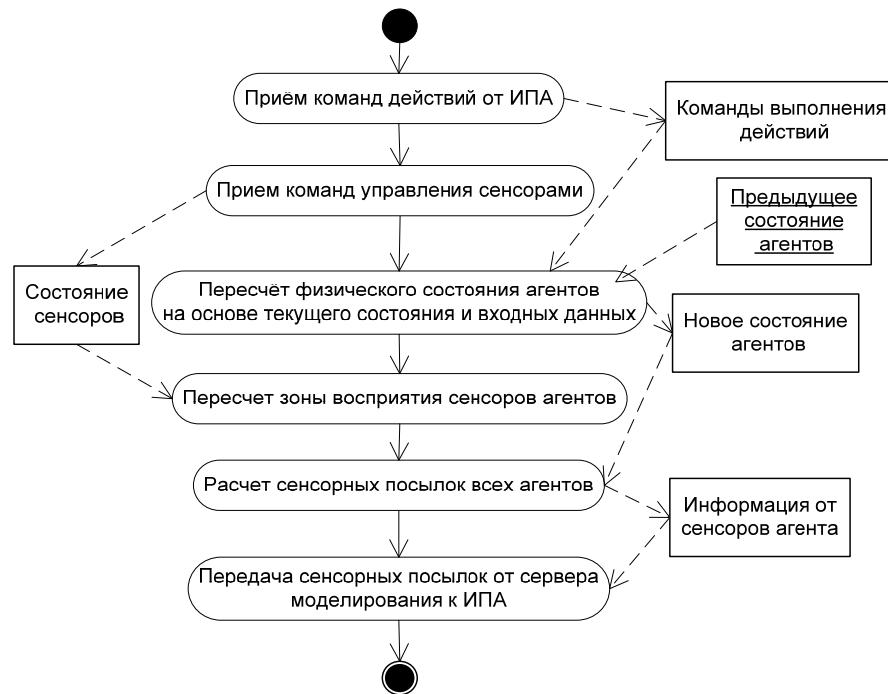


Рис. 2. Базовый алгоритм работы сервера моделирования

Новые зоны сенсорного покрытия определяются для всех агентов с учетом нового пространственного положения агентов и режимов работы их сенсоров. При этом для каждого агента определяются объекты внешней среды, находящиеся в пределах зоны его восприятия, и вычисляются сенсорные данные с учетом заданной модели ошибок. СИМ MAC поддерживает возможность моделирования различных типов сенсоров, входящих в состав подсистемы восприятия (ПВсп) агента. Существенными особенностями имитационного моделирования сенсоров агентных систем являются:

- ограниченность зон восприятия;
- возможность управления режимами работы сенсора;
- наличие ошибок (шумов) в воспринимаемой информации.

Среда разработки СИМ MAC поддерживает возможность создания новых типов сенсоров, а также использования при построении конкретного ИА существующих типов сенсоров с возможностью конфи-

турирования их параметров. При создании имитационной модели сенсора задаются параметры зоны восприятия, характеристики объектов, воспринимаемые данным типом сенсора, и модель ошибок измеряемых параметров. При наличии у сенсора разных режимов работы для каждого из них определяются собственные характеристики.

В общем случае модель внесения погрешностей в получаемые сенсором данные имеет вид

$$\Delta = f(d, R, P),$$

где Δ – относительная погрешность, вносимая в измеряемые данные; d – расстояние до наблюдаемого объекта; R – режим работы сенсора; P – текущая характеристика помеховой обстановки; f – функция, описывающая зависимость величины погрешности от указанных выше параметров. При моделировании помеховой обстановки учитывается, что она может быть обусловлена как естественными шумами, так и помехами, целенаправленно создаваемыми другими агентами.

Моделируемая МАС может включать произвольное число агентов, при этом подсистема восприятия каждого агента может содержать любое число сенсоров различных типов. С учетом этого алгоритм работы сервера, реализующий моделирование восприятия агентов, представлен на рис. 3.

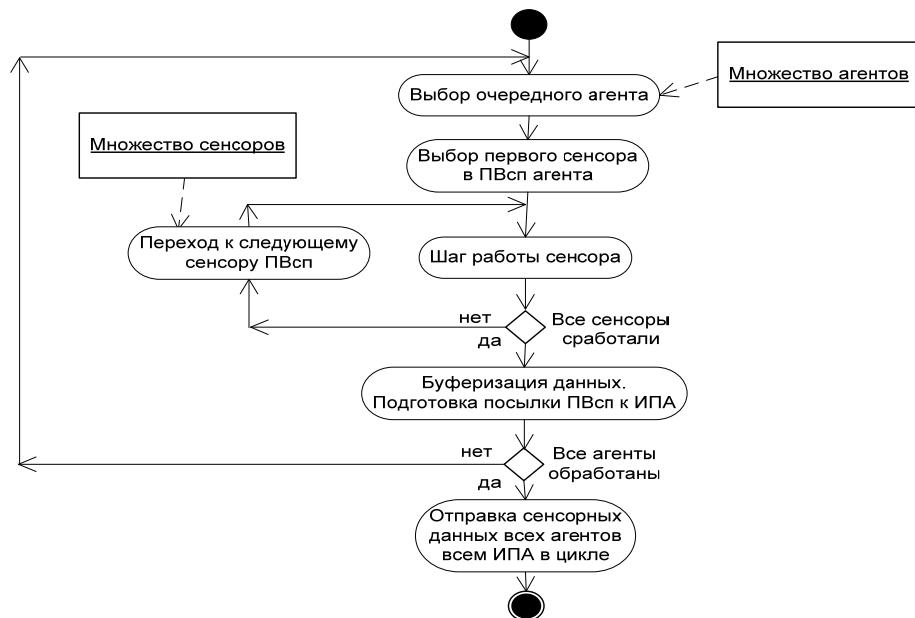


Рис. 3. Алгоритм работы сервера в части моделирования восприятия агентов

Собранные в текущем такте сенсорные данные сериализуются и отсылаются ИПА соответствующих агентов через сокетные соединения. На основе полученных сообщений ИПА обновляет внутреннюю модель мира агента текущими значениями параметров наблюдаемых объектов, на основе которых вычисляются значения высоконивневых параметров ситуации и выполняется ее оценка. Обобщенный алгоритм работы сенсора (рис. 4) выполняет обход графа сцены и выбор объектов, попавших в зону восприятия сенсора. Он реализован в составе шаблонного класса, на основе которого создаются модели конкретных сенсоров путем задания их параметров.

Для отработки моделей кооперативного поведения агентов, основанных на коммуникации, в СИМ МАС реализована возможность моделирования ненадежных каналов связи с ограниченной пропускной способностью. Поддерживается возможность моделирования, как широковещательных каналов связи, так и выделенных между парой агентов («точка–точка»). При создании модели канала связи в среде разработки задаются его характеристики, в частности: пропускная способность, вероятность безошибочной доставки сообщения, максимальная дальность связи (при моделировании радиоканалов) и др. Указанные характеристики могут описываться случайными величинами с задаваемыми параметрами распределения. Поддерживается возможность приема нескольких сообщений от различных агентов в одном такте моделирования.

На уровне среды исполнения модели каналов связи реализуются основным сервером моделирования. Для отправки сообщения агент посылает серверу соответствующую команду, содержащую передаваемое сообщение и идентификатор получателя. В следующем такте принимающий агент получает от сервера команду, содержащую переданное сообщение. При моделировании радиоканалов сервер добавляет во входное сообщение направление на источник и расстояние до него с внесением погрешностей относительно истинных значений. Таким образом, агенты имеют возможность с определенной степенью точности идентифицировать источник сообщения по его положению в пространстве.

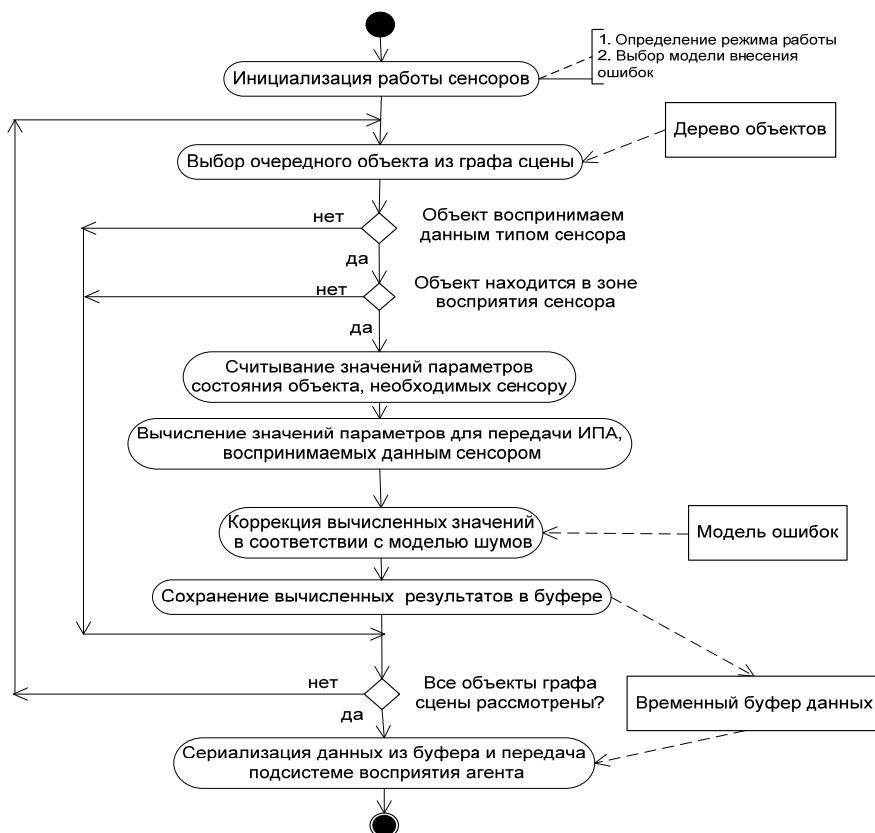


Рис. 4. Базовый цикл работы модели сенсора

Заключение

Разрабатываемая среда имитационного моделирования позволяет исследовать модели автономного и группового поведения широкого класса интеллектуальных агентов с учетом реальных ограничений их сенсоров и каналов связи, поддерживает режим реального времени и возможность журналирования и анализа ментальной деятельности агентов в процессе принятия ими решений в динамических мирах. Реализованный прототип среды имитационного моделирования многоагентных систем был успешно апробирован при моделировании координируемых действий группы автономных беспилотных летательных аппаратов в условиях противодействия.

Литература

1. Рассел С., Норвиг П. Искусственный интеллект: Современный подход. – 2-е изд. Пер. с англ. – М.: Вильямс, 2006. – 1408 с.
2. Пузанков Д.В., Мирошников В.И., Пантелейев М.Г., Серегин А.В. Интеллектуальные агенты, многоагентные системы и семантический Web: концепции, технологии, приложения. – СПб: Технолит, 2008. – 292 с.
3. Uhrmacher Adelinde M., Weyns Danny. Multi-Agent Systems: Simulation and Applications // Computational Analysis, Synthesis and Design of Dynamic Modes. – Boca Raton: CRC Press, 2009. – 566 p.
4. Кельтон В., Лоу А. Имитационное моделирование. Классика CS. – 3-е изд. – СПб: Питер; Киев: BHV, 2004. – 847 с.

Пантелейев Михаил Георгиевич

– Санкт-Петербургский государственный электротехнический университет «ЛЭТИ», кандидат технических наук, доцент, mpanteleyev@gmail.com

Кохтенко Николай Владимирович

– Санкт-Петербургский государственный электротехнический университет «ЛЭТИ», аспирант, nukokhtenko@gmail.com

Лебедев Сергей Вячеславович

– Санкт-Петербургский государственный электротехнический университет «ЛЭТИ», аспирант, lebedev.sv.etu@gmail.com

УДК 004.89

ЭФФЕКТИВНЫЙ ЛОГИЧЕСКИЙ ВЫВОД В ПРОДУКЦИОННОЙ МОДЕЛИ С ПОМОЩЬЮ БАЗ ДАННЫХ С ВЕРТИКАЛЬНОЙ АРХИТЕКТУРОЙ

Р.С. Катериненко

Предлагается идея применения вертикальных (column-oriented) баз данных для логического вывода в продукционной модели знаний. Рассматриваются потенциальные преимущества использования указанного типа архитектуры по сравнению с традиционными реляционными базами данных для специфичных SQL-запросов, возникающих в продукционной модели. Приводятся результаты сравнительного эксперимента с вертикальной базой данных KDB+ и реляционной OracleXE.

Ключевые слова: продукционная модель, логический вывод, вертикальная база данных, векторные языки программирования.

Введение

По своей природе логический вывод имеет большую алгоритмическую трудоемкость. Но в некоторых моделях, таких как дескрипционная логика и продукционные системы, трудоемкость ниже. Благодаря этому возможно практическое применение этих моделей. С целью повышения скорости логического вывода предлагается использовать базы данных (БД) с вертикальной архитектурой. В последнее время сообщество профессионалов все чаще обращает свое внимание на нереляционные БД [1, 2], но неизвестно ни одного применения этого типа для логического вывода. Предлагаемый подход является развитием идеи использования реляционных систем управления базами данных (СУБД) для логического вывода в продукционной модели знаний, предложенной в работе [3].

Трансляция в SQL

В работе [3] предлагается алгоритм трансляции процедуры логического вывода в SQL-запросы к реляционной БД. В продукционной модели знания описываются с помощью фактов и правил (продукций). Факты представляют собой атомарные формулы или, другими словами, реализацию какого-либо предиката. Например, факты SubClass(Fruct, Lemon) или Parent(John, Bob), являются реализацией предикатов SubClass(X, Y) и Parent(X, Y) соответственно. Основной объем базы знаний заключен в фактах. Их количество, как правило, превосходит количество правил. Предикаты могут быть произвольной арности. В предлагаемом подходе предикаты хранятся в таблицах БД. В столбцах таблицы, соответствующей предикату, находятся термы, при которых он принимает истинное значение. Логический вывод осуществляется путем последовательного применения продукции к фактам и порождения новых фактов до тех пор, пока продолжают выводиться ранее неизвестные факты (Fixpoint iteration).

С помощью SQL-запросов осуществляется поиск фактов, при которых искомый предикат является истинным. В контексте концепции закрытого мира (Closed World Assumption) это означает, что истинны те и только те факты, которые занесены в БД или могут быть получены логическим выводом из них. Например, если запрос пользователя имеет вид « $P(x)$ », то должны быть выданы все факты с предикатом P , подтверждающие его истинность. При этом должны быть учтены факты, порожденные применением продукции.

При загрузке продукции анализируется их зависимость между собой по предикатам. Две продукции $P1$ и $P2$ называются зависимыми, если выполняется хотя бы одно условие:

- предикат из условия продукции $P1$ совпадает с предикатом из следствия продукции $P2$;
- предикат из следствия продукции $P1$ совпадает с предикатом из условия продукции $P2$.

По критерию зависимости строится граф (Program graph). Это позволяет найти для искомого предиката все необходимые продукции (цепочку продукции или дерево), выполнение которых может повлиять на значение искомого предиката. Найденное подмножество, как правило, содержит значительно меньшее количество продукции, чем вся база знаний. Далее по найденным продукции генерируется SQL-запрос, возвращающий множество фактов для искомого предиката. Каждый предикат вида $b(x)$ определяет множество фактов:

$S = \{\text{SELECT * FROM } \text{Predicate_B}\};$

Продукция, содержащая предикаты $a1$ и $a2$, соединенные конъюнкцией по общей переменной, например, $a(y) :- a1(y), a2(y)$, задает множество вида

$S = \{\text{SELECT * FROM } S1 \text{ JOIN } S2\}.$

Две и более продукции с общим выводимым предикатом, например, $a1(t) :- a2(t)$, $a1(t) :- a3(t)$, задают дизъюнкцию множеств фактов $S1$ и $S2$ и объединяются в одно множество с предикатом $a1$:

$S = \{\text{SELECT * FROM } S1 \text{ UNION } S2\}.$

Благодаря эффективности современных СУБД SQL-запросы выполняются достаточно быстро, и достигается ускорение логического вывода.

Постановка эксперимента

После небольшого введения в трансляцию продукции в SQL рассмотрим особенности генерируемого SQL. Наличие предикатов с общей переменной, связанных конъюнкцией, трансформируется в операцию соединения (join). Кроме того, как минимум одна конъюнкция имеется почти в каждой продукции. Таким образом, для логического вывода очень критична скорость выполнения соединения. В реляционной алгебре эта операция принимает на вход предикат. Путем вычисления истинности предиката устанавливается, должны ли обрабатываемые строки входить в результатирующее множество. При трансляции продукции предикат всегда вырожденный – это равенство соответствующих термов.

Еще одной особенностью продукции модели является то, что чаще всего предикаты имеют небольшую арность, но большое количество реализаций. При этом не все термы предиката связаны переменными, т.е. не все столбцы соответствующей предикату таблицы участвуют в вычислении запроса. Это наводит на мысль об использовании такой СУБД, в которой бы имелась возможность проводить операции не над строками, а над отдельными столбцами. Иначе говоря, целесообразно применение вместо реляционной БД вертикальной, ключевым свойством которой является возможность эффективно запросить значение отдельного столбца таблицы, а не всей строки, и тем самым избежать считывания лишней, нерелевантной информации и ограничить множество используемых в запросе столбцов [1]. Рассмотрим основные априорные свойства систем этого класса.

Особенности архитектуры вертикальных СУБД позволяют функционировать, занимая меньший объем хранилища по сравнению с традиционными строко-ориентированными аналогами. Это достигается благодаря более эффективной работе с архивированными столбцами в вертикальных БД. Некоторые системы этого класса реализуют свои операторы для работы с архивированными столбцами и используют их напрямую в плане вычисления запроса [2]. Например, рассматриваемая в эксперименте система KDB+ [4] позволяет хранить каждый столбец таблицы в отдельном файле и считывать его в память (выгружать из памяти) по мере необходимости. Кэширование осуществляется по столбцам, а не по строкам. Чем больше обращений к столбцу, тем выше вероятность нахождения его в кэше. Благодаря постолбцовому хранению значительно сокращается количество модифицируемых при удалении/добавлении структур данных [5].

Как уже упоминалось ранее, увеличив скорость выполнения соединения по равенству, можно добиться существенного прироста производительности. Вертикальная архитектура СУБД позволяет сделать подобное соединение очень эффективным. В некоторых системах применяется ВАТ-алгебра (BAT-Algebra, Binary Associated Table) [4].

Вертикальные БД ориентированы на высокопроизводительное чтение данных, а не на запись. Это позволяет поддерживать сложные структуры хранилища с трудоемкой операцией вставки, но быстрым чтением. При работе с продукции базой знаний возникает схожая ситуация: факты редко меняются, но при этом часто требуется логический вывод, порождающий громоздкие запросы к базе. Типичный пример – экспертная система.

Учитывая вышеописанные предпосылки и априорные свойства вертикальных БД, было решено провести сравнительный эксперимент по скорости выполнения запроса. В качестве представителя систем указанного класса была взята система KDB+. Поскольку данная работа является первой работой автора в этом направлении, то не ставилась цель сделать исчерпывающий обзор вертикальных БД и выявить лучшую. KDB+ представляет собой СУБД, базирующуюся на векторном языке программирования Q. Данная технология нашла широкое применение в биржевой торговле для анализа больших массивов котировок биржевых инструментов. Синтаксис и семантика языка Q очень сильно отличается от SQL, от процедурных и объектных языков программирования. Базовые объекты языка – это функция и список, т.е. упорядоченный набор объектов. Соответственно, в контексте языка Q таблица является совокупностью списков, по одному для каждого столбца.

Для сравнительного эксперимента мы выбрали продукцию вида

$$Q(x) :- A(q, X) \wedge C(m, X, Y, n) \wedge B(Y, p),$$

в которой четырехместный предикат С связан переменной X с двухместным предикатом А и переменной Y с двухместным предикатом В. Каждый определяется соответствующей таблице, содержащей все его реализации:

```
create table tableA (
    t0 char(20),
    t1 char(20),
    primary key (t0)
);
create table tableB (
    t0 char(20),
    t1 char(20),
    primary key (t0)
);
```

```
create table tableC (
    t0 char(20),
    t1 char(20),
    t2 char(20),
    t3 char(20),
    primary key (t0)
);
```

SQL-запрос для Oracle, соответствующий продукции, выглядит следующим образом:

```
select count(ta.t1)
from tableA ta, tableB tb, tableC tc
where ta.t1 = tc.t1 and tb.t1 = tc.t2;
```

При этом выполняется только подсчет выводимых фактов, чтобы исключить задержки, связанные с выводом на терминал. Как уже говорилось, синтаксис и семантика языка Q сильно отличается от SQL, поэтому аналогичный Q-запрос выглядит по-другому:

```
\t select cnt:count(i)
from tableC
where (tableC.t1 in tableA.t1) and (tableC.t2 in tableB.t1).
```

Анализ результатов эксперимента

Ниже приведены результаты эксперимента. Графики описывают время выполнения запроса в зависимости от количества фактов в каждой таблице. Шкала по оси Y выбрана логарифмической, так как наибольший интерес представляет время исполнения запроса, нежели шаг по оси X. Графикам Oracle соответствуют линии с нечетными номерами, графикам KDB+ – линии с четными номерами.

На рис. 1 приведены графики максимальных и минимальных времен для обеих СУБД. Хорошо заметно, что минимальное время KDB+ никогда не превосходит минимальное время Oracle. Максимальные времена (выбросы) чаще всего соответствуют первому запуску запроса. Это связано с отсутствием информации в кэше. На больших объемах (больше 350 тысяч фактов) у Oracle наблюдается значительный рост времени первого исполнения запроса.

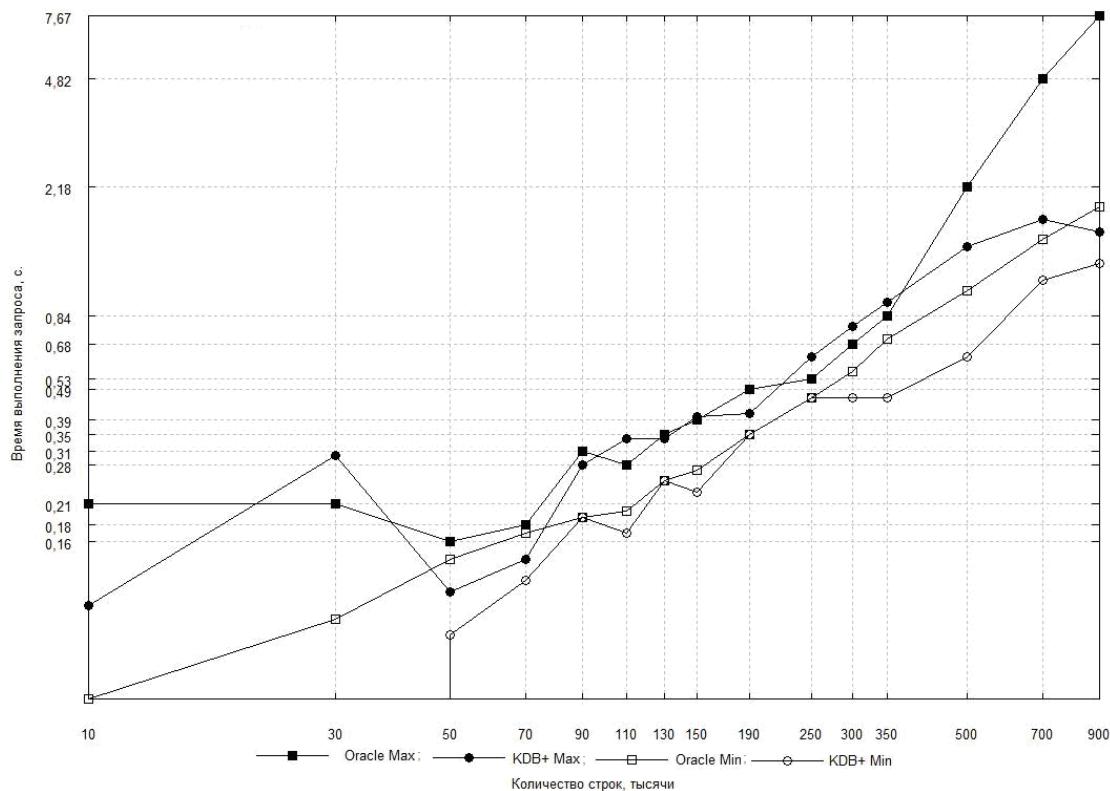


Рис. 1. Максимальное и минимальное время выполнения запроса

На рис. 2 приведены графики усредненных времен с исключенными выбросами. В среднем KDB+ показывает стабильный результат, в отличие от Oracle: при числе фактов до 90 000 время выполнения запроса намного превосходит KDB+, затем, при числе фактов до 300 000, приближается к нему и иногда оказывается меньшим.

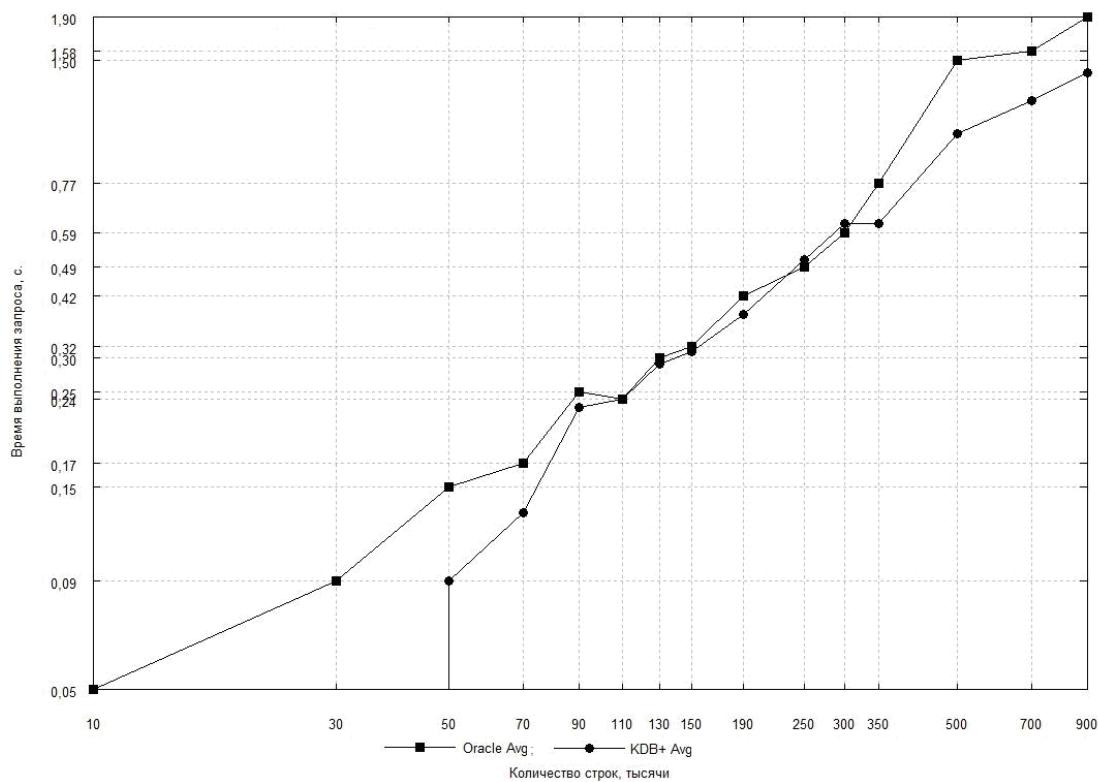


Рис. 2. Среднее время выполнения запроса

```

SQL> select count(ta.t1) from tableA ta, tableB tb, tableC tc where ta.t1 = tc.t1 and tb.t1 = tc.t2
COUNT(TA.T1)
-----
175000
Elapsed: 00:00:00.84
Execution Plan
Plan hash value: 2254415514

| Id  | Operation          | Name   | Rows  | Bytes | TempSpc | Cost <CPU> |
|-----|
| 01:10 | SELECT STATEMENT |        | 1    | 88   |          | 5780  <2>  |
| 1   | SORT AGGREGATE   |        | 1    | 88   |          |          1      |
| *2  | HASH JOIN         |        | 301K | 25M  | 13M    | 5780  <2>  |
| 00:08 | TABLE ACCESS FULL | TABLEB | 425K | 9145K|          | 646   <3>  |
| *4  | HASH JOIN         |        | 301K | 18M  | 13M    | 3315  <2>  |
| 00:08 | TABLE ACCESS FULL | TABLEA | 408K | 8780K|          | 645   <3>  |
| 00:15 | TABLE ACCESS FULL | TABLEC | 301K | 12M  |          | 1192  <2>  |

Predicate Information (identified by operation id):
2 - access('TB'."T1"='TC'."T2")
4 - access('TA'."T1"='TC'."T1")

```

Рис. 3. План выполнения запроса

После 350 000 фактов время запроса Oracle опять начинает расти. Представляется, что это связано с расходами на создание дополнительных структур при операции соединения, как видно из плана исполнения SQL-запроса на рис. 3. Соединение реализовано алгоритмом «Hash join». Для одной из таблиц, участвующей в соединении, создается хэш-таблица. Далее при последовательном проходе по другой таблице, применяя хэш-функцию, находятся совпадения.

Таким образом, на маленьких объемах построение хэш-таблицы не окупалось быстрым поиском по хэш-функции, но дало хороший результат на средних объемах. На больших объемах данных (приближающихся к 1000000) хэш-таблица увеличивается в размере настолько, что затраченное на ее создание время впоследствии не компенсируется быстрым поиском. К сожалению, пока не удалось найти информацию, какой алгоритм используется для операции соединения в KDB+. На рис. 3 также можно заметить, что реляционная БД создает временные структуры размером 100–200% от начального размера таблиц. Для вертикальной БД эта цифра составляет около 10%.

Заключение

При использовании СУБД для логического вывода в производственной модели знаний происходит генерация специфичных SQL-запросов. Специфика заключается в потенциально большом количестве операций соединения по равенству и операций объединения. Чаще всего в запросах используется обращение только к небольшому количеству столбцов таблицы. Для эффективного использования этих особенностей может быть целесообразным использование вертикальной базы данных вместо традиционной строково-ориентированной базы данных. Преимущество заключается в эффективной реализации требуемых операций, исключающей считывание нерелевантных столбцов и оптимизации занимаемого объема хранилища.

На сравнительных экспериментах вертикальная база данных продемонстрировала лучшее время исполнения запроса при меньшем потреблении памяти. СУБД Oracle также предоставляет большие возможности по оптимизации: сбор статистики с целью изменения плана запроса, построение индексов по столбцам и т. д. Этими средствами можно улучшить производительность, но при этом будет затрачена память на дополнительные структуры. На данный момент автор не располагает информацией о способах настройки производительности KDB+. Таким образом, можно говорить о преимуществе вертикальной базы данных для логического вывода в производственной модели, только если обе взяты с настройками по умолчанию, без оптимизации структуры хранения и запросов.

Литература

1. Stonebraker M., Bear Chuck, Cetintemel Ugur, Cherniack Mitch, Hackem Tingjian Ge Nabil, Harizopoulos Stavros, Lifter John, Rogers Jennie, Zdonik Stan. One Size Fits All? Part 2: Benchmarking Results // Proceedings of Conference on Innovative Database Research (CIDR). – 2007. – P. 173–184.
2. Stonebraker M., Abadi Daniel J., Batkin Adam, Chen Xuedong, Cherniak Mitch, Ferreira Miguel, Lau Edmond, Lin Amerson, Madden Sam, O’Neil Elizabeth, O’Neil Pat, Rasin Alex, Tran Nga, Zdonik Stan. CStore: A Column-Oriented DBMS // Proceedings of the 31st International Conference on Very Large Data Bases. – 2005. – P. 553–564.
3. Bessmertny I., Katerinenko R. Inference acceleration in production model of knowledge // Programming and Computer Software. – 2011. – V. 42. – P. 197–199.
4. Kx Systems. Система управления базой данных KDB+. Документация [Электронный ресурс]. – Режим доступа: <http://www.kx.com/products/database.php>, св. Яз. англ. (дата обращения 10.10 2011).
5. Svensson Per, Boncz Peter, Ivanova Milena, Kersten Martin, Nes Niels. Emerging database systems in support of scientific data // Scientific data management: challenges technology and deployment. – 2010. – P. 235–281.

Катериненко Роман Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, innuWorks@gmail.com

УДК 004.738.52

ИЗВЛЕЧЕНИЕ ОНТОЛОГИЙ ИЗ WIKI-СИСТЕМ

В.К. Шестаков

Рассматривается подход к извлечению онтологий из Wiki-систем, а также его использование в разработке и сопровождении Wiki-систем, заполняемых содержимым на основе онтологий, и другие возможные применения. Описывается реализация данного подхода в виде клиентского приложения.

Ключевые слова: Wiki, онтологии, извлечение данных, Pywikibot, Semantic MediaWiki.

Введение

Для удовлетворения все возрастающих информационных потребностей пользователей разработаны разнообразные средства построения информационных систем. Одним из таких удобных и простых в использовании средств сбора и хранения информации являются Wiki-системы [1]. Они позволяют работать не только с текстовым, но и с мультимедийным контентом, имеют удобный и интуитивно понятный интерфейс, просты в освоении. Однако их большим недостатком является то, что они позволяют отслеживать в создаваемых информационных системах только структурную целостность ссылок, не обеспечи-

вая при этом логической непротиворечивости и семантической согласованности (semantic consistency) [2] используемых в них понятий.

Общая идея предлагаемого подхода состоит в создании инструментария, который бы обеспечивал заполнение Wiki-систем содержимым с согласованной системой понятий (семантически согласованных Wiki). Wiki с такими свойствами можно получить, если строить ее на основе логически непротиворечивой онтологии, описывающей предметную область будущей системы. В этот инструментарий должны также входить средства контроля и отслеживания изменений в онтологии.

Первая составляющая данного подхода, заключающаяся в заполнении Wiki-систем содержимым на основе онтологий, была описана в предыдущей работе автора [3]. В настоящей работе рассматривается его вторая составляющая, обратная первой, посвященная извлечению онтологий из уже существующих Wiki-систем. Как уже говорилось, она необходима для контроля, трансформации, сопровождения и развития Wiki, которые уже наполнены содержимым. Кроме того, она может иметь и другие применения, например, построение предварительного, чернового варианта онтологии предметной области по уже существующей Wiki-системе или объединение нескольких Wiki по близким предметным областям.

Существует много подходов к извлечению знаний из Wiki-систем в виде онтологий. Чаще всего используется Википедия. На ее основе осуществляют извлечение структурированной информации и предоставление доступа к ней [4], построение онтологии верхнего уровня (general-purpose ontology) [5] и крупномасштабной онтологии людей [6], ее используют в качестве среды для разработки онтологий [7] и как источник корпуса текстов для построения онтологии конкретной предметной области [8], а также для автоматического построения крупномасштабной мультимодальной онтологии для классификации веб-изображений [9]. Все упомянутые подходы и системы используют Википедию только в качестве источника информации, а извлеченную информацию они затем используют в своих целях, например, для построения своей собственной онтологии, базы данных или классификации чего-либо.

В настоящей работе предлагается несколько другой подход. Он заключается в том, что на основе информации, извлекаемой из выбранной Wiki-системы, строится не онтология вообще (например, какой-то предметной области), а онтология именно данной конкретной Wiki. Это дает возможность получить для дальнейшего анализа и использования ее структуру и содержание.

Метод извлечения онтологий

В основе любой Wiki-системы лежит так называемый Wiki-движок – комплекс программных средств для преобразования Wiki-разметки в код, предназначенный для отображения в браузере. Одним из самых распространенных движков является MediaWiki [10] (на его основе работает широко известная Википедия). Для него существуют специальные дополнения, называемые расширениями, позволяющие получить определенную функциональность. Так, расширение Semantic MediaWiki [11, 12] дает возможность добавлять семантическую информацию за счет расширения разметки, а также предлагает средства для работы с этой информацией.

Для проведения работ технического характера в Wiki-системах используют боты – специальные программы для выполнения заданного набора операций. Они являются клиентскими приложениями, поэтому не требуют внесения изменений на стороне сервера (т.е. для их работы не нужно модифицировать движок или ставить какое-либо расширение). Например, в Википедии боты используются для таких задач, как переименование категорий и статей, расстановка интервики-ссылок (ссылок на родственные проекты), исправление ссылок, удаление спама и т.п. Для реализации ботов используются различные языки программирования, а также существуют разные библиотеки для облегчения их написания. Одной из наиболее развитых библиотек является Python WikipediaBot Framework [13]. Она использует MediaWiki API (специальный интерфейс прикладного программирования) для взаимодействия с MediaWiki-системой – авторизации, получения данных и внесения изменений.

Пояснив основные понятия, можно перейти к рассмотрению общей схемы работы инструментальной системы, которая представлена на рисунке.

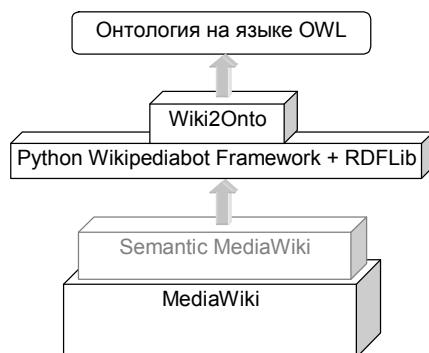


Рисунок. Общая схема работы инструментальной системы

Сначала программный модуль Wiki2Onto, разработанный в рамках данного проекта, при помощи Python WikipediaBot Framework извлекает онтологию из Wiki-системы, работающей на базе MediaWiki, возможно, с расширением Semantic MediaWiki, а затем с использованием библиотеки RDFLib [14] сохраняет в файл на языке OWL [15]. В табл. 1 представлено соответствие конструкций Semantic MediaWiki и языка OWL.

Semantic MediaWiki	Конструкция в онтологии
Категория	owl:Class
Подкатегория	rdfs:subClassOf
Страница	owl:NamedIndividual
Обычная ссылка	owl:ObjectProperty «Ссылается на»
Семантическая ссылка	Зависит от типа

Таблица 1. Соответствие конструкций Semantic MediaWiki и онтологии

Рассмотрим этот процесс более подробно. Онтология из Wiki-системы извлекается в следующем порядке. Сначала извлекаются все классы, при этом каждому классу соответствует одна категория Wiki, а структура вложенности категорий Wiki определяет иерархию классов. Затем извлекаются все страницы как экземпляры соответствующих классов. Для пустых страниц, на которые в Wiki имеются ссылки, заводится специальный служебный класс «Несуществующие страницы». После этого просматриваются все ссылки на каждой странице. Для начала определяется, является ли ссылка обычной или семантической. Если ссылка обычная, то для соответствующего экземпляра класса в OWL-онтологии заводится объектное свойство «Ссылается на» (так как ссылка обычная, а не семантическая, то у нее нет собственного имени, и данное имя выбрано для всех таких ссылок) со значением в виде экземпляра, имя которого совпадает с именем страницы, на которую указывает ссылка. Если ссылка семантическая, то она имеет структуру <название свойства, значение свойства>, и для нее сначала определяется тип ее свойства. Если свойство имеет тип «Страница» или его тип не указан, то в OWL-онтологии заводится объектное свойство с соответствующими именем (название свойства) и значением (значение свойства). (Заметим, что по умолчанию свойство ссылки имеет тип «Страница».) Если же свойство имеет какой-то другой стандартный тип, то тип свойства данных в OWL-онтологии определяется согласно табл. 2 (в качестве значения «owl:Annotation property» [16] используется соответствующий URI: «tel:» для телефонного номера [17], «mailto:» для адреса электронной почты, «http:» для URL и URI аннотации). Для пользовательских типов свойств создается собственный тип свойства данных.

Так как реализация модуля Wiki2Onto еще не доведена до финальной стадии, то пока поддерживаются не все стандартные типы свойств Semantic MediaWiki, а только те, что представлены в табл. 2. На данный момент этот модуль работает только с Wiki-системами на базе MediaWiki.

Извлекаемая онтология сохраняется в формате OWL, а не ограничивается RDFS по нескольким причинам. Во-первых, используется разделение свойств на два класса – объектные (owl:ObjectProperty) и типов данных (owl:DatatypeProperty). Во-вторых, в Semantic MediaWiki существуют ограничения на значения свойств с помощью специального свойства «Allows value» [18] и это нужно учитывать. В-третьих, один из способов проверки получаемой онтологии заключается в добавлении в нее аксиом.

Следует отметить, что Wiki-система, из которой извлекается онтология, не обязательно должна функционировать с расширением Semantic MediaWiki. Однако в случае использования Wiki без этого расширения извлекаемая онтология будет гораздо беднее, так как в ней не будет присутствовать специальная семантическая информация. В частности, нельзя будет извлечь атрибуты, разнообразие отношений также будет невелико. Правда, путем индивидуальной настройки на конкретную Wiki-систему объем извлекаемой из нее информации можно увеличить (например, если некоторые данные в ней приведены в однотипном формате или для их представления используются шаблоны [19], то для их извлечения можно реализовать специальную функцию).

Тип свойства	Тип в онтологии
Строка	string
Число	double
Булево	boolean
Дата	dateTime
Текст	string
Код	string
Телефонный номер	owl:Annotation property
URL	owl:Annotation property
Почта	owl:Annotation property
URI аннотации	owl:Annotation property

Таблица 2. Соответствие при извлечении стандартных типов

В Semantic MediaWiki существует свой собственный встроенный инструмент для экспорта в RDF, позволяющий извлекать размеченную семантическую информацию из списка страниц. Кроме того, для извлечения сразу всех семантических данных из Wiki, оборудованной данным расширением, существует специальный внешний инструмент [20]. Но, как можно заметить, таким способом можно извлечь только явно размеченную семантическую информацию, а вся остальная останется неохваченной. Подход, описанный в настоящей работе, позволяет извлекать дополнительную часть этой информации, причем методы ее извлечения планируется развивать (в том числе за счет лингвистических средств и возможности индивидуальной подстройки под конкретную систему).

Также существует возможность оборудовать Semantic MediaWiki хранилищем RDF и хранить всю семантическую информацию в нем [21]. Это упрощает ее экспорт, но не избавляет от вышеописанного недостатка. Кроме того, этот способ сужает область применения за счет дополнительного требования на наличие хранилища. Если же на Wiki не установлено расширение Semantic MediaWiki, то в ней нет подобных встроенных средств экспорта информации в RDF.

В то же время предлагаемый подход, как указывалось выше, не требует наличия этого расширения в обязательном порядке. В качестве способа реализации было выбрано именно написание собственного модуля, работающего по принципу MediaWiki-бота, а не разработка расширения для MediaWiki для увеличения области применения. Ведь не всегда к Wiki-системе, из которой нужно извлечь онтологию, есть доступ для того, чтобы установить на ней свое расширение.

Итак, существуют подходы для извлечения таксономии из MediaWiki-систем без семантических расширений (например, в проектах YAGO и DBpedia, упомянутых во введении), есть стандартные средства для экспорта семантической информации в расширении Semantic MediaWiki, но нет систем, объединяющих в себе и то, и другое. И именно в этом заключается новизна данной работы.

Варианты применения метода извлечения онтологий

В первую очередь, метод позволяет проследить, как изменилась онтология после того, как эксперты поработали над содержимым Wiki-сайта, созданного на базе первоначальной онтологии. Это может понадобиться не только ради обычного любопытства, но и для вполне серьезных целей – например, отслеживание развития проекта, верификация получаемой онтологии (ontology verification) и ее реинжиниринг (ontology reengineering) [22], проверка качества и сбалансированности получаемой структуры данных [23], координация в развитии отдельных частей предметной области. К полученной онтологии также можно применить одну из существующих машин вывода для получения неявных знаний.

Кроме того, извлекать онтологию можно не только из той Wiki-системы, которая была ранее заполнена на основании некоторой онтологии, но и из уже существующей Wiki, заполненной обычным образом. Например, хотим построить онтологию некоторой предметной области и обнаружили Wiki, содержащую нужные нам сведения из этой области. Вместо того чтобы вручную строить нужную онтологию «с нуля», можно извлечь из этой системы ее предварительный, черновой вариант, а затем уже дорабатывать его, что гораздо проще.

Еще одно применение метода извлечения онтологий заключается в объединении нескольких Wiki-систем по близким предметным областям. Непосредственное объединение систем может быть довольно сложным и потребовать много ручной работы, так как крайне затруднительно отследить все связи и пересечения между двумя системами. Используя предлагаемый подход, можно поступить гораздо проще: извлечь онтологию из каждой системы и провести их слияние (ontology merging) [22], а после этого по объединенной онтологии заполнить содержимым требуемую Wiki-систему.

Заключение

В работе рассмотрен подход и предложен метод извлечения онтологий из Wiki-систем, разработан прототип инструментальной системы, реализующий данный метод в указанном объеме.

Этот метод позволяет не только сопровождать, контролировать и развивать Wiki-системы, наполнение информацией которых производилось на основе онтологии, но и имеет некоторые другие применения, такие как построение предварительного, чернового варианта, онтологии предметной области по уже существующей Wiki или объединение нескольких Wiki по близким предметным областям.

В дальнейшем планируется завершить реализацию модуля Wiki2Onto, а также расширить его возможности, в частности, применить методы компьютерной лингвистики для извлечения информации со страниц Wiki-систем.

Литература

1. Leuf B., Cunningham W. The Wiki Way: Quick Collaboration on the Web. – Addison-Wesley Professional, 2001. – 464 p.
2. Baader F., Nutt W. Basic Description Logics: The Description Logic Handbook. – Cambridge University Press. – 2002. – P. 47–100.

3. Шестаков В.К. Инструменты построения информационных систем на основе Wiki-технологии и онтологий предметных областей // Сборник трудов конференции «Управление знаниями и технологиями семантического веба – 2010». – СПб: СПбГУ ИТМО, 2010. – С. 150–159.
4. Bizer C., Lehmann J., Kobilarov G., Auer S., Becker C., Cyganiak R., Hellmann S. DBpedia – A Crystallization Point for the Web of Data // Journal of Web Semantics: Science, Services and Agents on the World Wide Web. – 2009. – Is. 7. – P. 154–165.
5. Suchanek F.M., Kasneci G., Weikum G. YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia // Proceedings of the 16th International Conference on World Wide Web (Banff, Alberta, Canada, May 8–12, 2007). WWW '07. – NY: ACM Press, 2007. – P. 697–706.
6. Shibaki Y., Nagata M., Yamamoto K. Constructing Large-Scale Person Ontology from Wikipedia // Proceedings of the 2nd Workshop on «Collaboratively Constructed Semantic Resources». – Coling, 2010. – P. 1–9.
7. Hepp M., Bachlechner D., Siorpaes K. Harvesting Wiki Consensus – Using Wikipedia Entries as Ontology Elements // Proceedings of the First Workshop on Semantic Wikis – From Wiki to Semantics, Annual European Semantic Web Conference (ESWC 2006). – 2006. – P. 124–138.
8. Cui G.Y., Lu Q., Li W.J., Chen Y.R. Corpus Exploitation from Wikipedia for Ontology Construction // Proceedings of the Sixth International Language Resources and Evaluation (LREC 2008). – Marrakech, 2008. – P. 2125–2132.
9. Wang H., Jiang X., Chia L.-T., Tan A.-H. Wikipedia2Onto --- Adding Wikipedia Semantics to Web Image Retrieval // Proceedings of the WebSci'09: Society On-Line, 18-20 March 2009. – Greece: Athens, 2009. – P. 297–306.
10. Сайт проекта MediaWiki [Электронный ресурс]. – Режим доступа: <http://mediawiki.org>, св. Яз. англ. (дата обращения 30.11.2011).
11. Сайт проекта Semantic MediaWiki [Электронный ресурс]. – Режим доступа: <http://semantic-mediawiki.org>, св. Яз. англ. (дата обращения 30.11.2011).
12. Krötzsch M., Vrandecic D., Völkel M., Haller H., Studer R. Semantic Wikipedia // Journal of Web Semantics. – Elsevier, 2007. – № 5. – P. 251–261.
13. Сайт проекта Python WikipediaBot Framework [Электронный ресурс]. – Режим доступа: <http://pywikipediabot.sourceforge.net>, свободный. Яз. англ. (дата обращения 30.11.2011).
14. Сайт проекта RDFLib [Электронный ресурс]. – Режим доступа: <http://rdflib.net>, св. Яз. англ. (дата обращения 30.11.2011).
15. Motik B., Patel-Schneider P.F., Parsia B., eds. OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. W3C Recommendation, 27 October, 2009 [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/>, св. Яз. англ. (дата обращения 30.11.2011).
16. OWL Web Ontology Language Reference [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/owl-ref/>, св. Яз. англ. (дата обращения 30.11.2011).
17. The tel URI for Telephone Numbers [Электронный ресурс]. – Режим доступа: <http://www.ietf.org/rfc/rfc3966.txt>, св. Яз. англ. (дата обращения 30.11.2011).
18. Semantic MediaWiki Property:Allows value [Электронный ресурс]. – Режим доступа: http://semantic-mediawiki.org/wiki/Property:_Allows_value, св. Яз. англ. (дата обращения 30.11.2011).
19. Википедия: Механизм шаблонов [Электронный ресурс]. – Режим доступа: http://ru.wikipedia.org/wiki/Википедия:_Механизм_шаблонов, св. Яз. англ. (дата обращения 30.11.2011).
20. Semantic MediaWiki: RDF export [Электронный ресурс]. – Режим доступа: http://semantic-mediawiki.org/wiki/Help:RDF_export, св. Яз. англ. (дата обращения 30.11.2011).
21. Semantic MediaWiki: Using SPARQL and RDF stores [Электронный ресурс]. – Режим доступа: http://semantic-mediawiki.org/wiki/Help:Using_SPARQL_and_RDF_stores, св. Яз. англ. (дата обращения 30.11.2011).
22. Suarez-Figueroa M.C., Gomez-Perez A. Towards a Glossary of Activities in the Ontology Engineering Field // Proceedings of 6th International Conference on Language Resources and Evaluation (LREC'08). – Marrakech, 2008. – P. 870–873.
23. Гаврилова Т.А., Горовой В.А., Болотникова Е.С., Горелов В.В. Субъективные метрики оценки онтологий // Материалы Всероссийской конф. с между. участием «Знания-Онтологии-Теории» (ЗОНТ-2009). Т. 1. – Новосибирск: Институт математики СО РАН, 2009. – С. 178–186.

Шестаков Владимир Константинович – Новосибирский государственный университет, аспирант, zfc@ngs.ru

ИЗВЛЕЧЕНИЕ КЛЮЧЕВЫХ СЛОВОСОЧЕТАНИЙ

С.В. Попова, И.А. Ходырев

Исследованы задачи аннотирования ключевыми словами и словосочетаниями семантически близких групп текстов в маленьких коллекциях узкотематических документов короткой длины. Показана возможность извлечения ключевых слов с одновременной фильтрацией общеупотребительных слов. Предложена формула для оценки взаимной зависимости слов и алгоритм извлечения на ее основе ключевых словосочетаний. Представлены результаты тестирования используемых алгоритмов.

Ключевые слова: ключевые слова, ключевые словосочетания, аннотирование, кластеризация, анализ коротких текстов, информационный поиск.

Введение

Задача извлечения ключевых слов из коллекций документов и составление коротких аннотаций к коллекциям, их частям, отдельным текстам или частям текстов (topic detection, topic interpretation, summarization, TextTiling) [1, 2] получила широкое распространение вследствие высокой применимости на практике. Извлечение ключевых слов и словосочетаний из коллекций и текстов позволяет пользователю понять, насколько полезен для него некоторый документ, не требуя просмотра всего документа. Используя ключевые словосочетания, пользователь может найти документы, релевантные заинтересовавшему его документу. Появление новых словосочетаний в подборках научных статей, ранжированных по времени, может свидетельствовать о появлении нового направления или новой тенденции в некоторой научной области. В научной области важной является задача извлечения новой терминологии развивающейся области. Данная задача напрямую связана с выделением устойчивых словосочетаний [3]. Определение основных тем и ключевых слов, представленных в коллекциях новостей, является подзадачей направления, получившего широкое распространение под названием «Topic detection and tracking» (TDT) [4–7] и связанного с отслеживанием во времени изменений, модификаций, группировкой и возникновением «нового» в новостях. Решение задачи извлечения ключевых словосочетаний полезно также в задачах определения различных контекстов слова, например, при построении словарей или разработке систем автоматического машинного перевода. Сложность поставленной задачи связана с тремя основными проблемами, присущими узкотематическим коллекциям текстов короткой длины: низкая частота встречи термов (слов) в текстах, большое перекрытие по общим словам, недостаток данных для накопления статистической информации. Из-за низкой частоты встречи термов, маленького размера коллекций и небольшой частоты совместной встречаемости слов существующие методы извлечения словосочетаний, например, основанные на вычислении MI (меры взаимной информации), могут оказываться неэффективными, так как, в первую очередь, будут выделяться словосочетания, в состав которых входят слова, редко встречающиеся в коллекции. Целью данной работы является разработка алгоритмов для извлечения семантически близких групп текстов коллекции и аннотирования полученных групп ключевыми словами и словосочетаниями.

Постановка задачи и этапы решения

В работе рассматривается задача автоматического аннотирования коллекции документов ключевыми словами и словосочетаниями, характеризующими группы семантически близких документов данной коллекции. Извлечение из коллекции, помимо самих ключевых слов, контекстных словосочетаний с этими словами позволяет лучше отследить семантику использования выделенных слов. Задача извлечения ключевых словосочетаний и слов была разделена на два этапа: этап кластеризации и этап аннотирования ключевыми словосочетаниями полученных кластеров. В области кластеризации коротких текстов интересны работы [8–10]. Разработка собственного алгоритма потребовалась по следующим причинам. Если требуется ориентироваться на пользователя, результат кластеризации должен быть всегда достаточно высоким и желательно стабильным (не меняться при различных прогонах алгоритма). Стабильная работа алгоритма кластеризации необходима при использовании результатов кластеризации для аннотирования. В противном случае будут получаться различные аннотации для разных результатов кластеризации. Современные алгоритмы [8–10], дающие достаточно высокие результаты кластеризации для рассматриваемого типа коллекций, не дают этот результат стабильно (в случае фиксированного числа кластеров): при одном прогоне алгоритма результат может быть высоким, при другом относительно низким. В экспериментах авторов использованы те же коллекции, что и в работах [8–10]: CICling_2002, SEPLN_CICling, EasyAbstracts и Micro4News. Первые три коллекции содержат в себе аннотации научных статей и являются узкотематическими; последняя коллекция содержит короткие новости и относится к широкотематическим коллекциям. Названия внутренних тем коллекций (названия кластеров, которые должны быть получены на типе кластеризации) представлены в табл. 1. Данные коллекции, включая «Золотые стандарты» (Golden Standard, результат классификации коллекции экспертом, когда человек определяет, какие семантически близкие группы документов содержатся в коллекции), находятся в открытом доступе [11], где также можно найти информацию с описанием коллекций.

Название Коллекции	Названия внутренних тем коллекции
CICling_2002	Linguistic, Ambiguity, Lexicon and Text, Processing
SEPLN_CICling	Morphological – syntactic, analysis, Categorization of Documents, Corpus linguistics, Machine translation
Easy Abstracts	Machine Learning, Heuristics in Optimization, Automated reasoning and Autonomous intelligent agents
Micro4News	Sci.med, soc.religion.christian, rec.autos, comp.os.ms-windows

Таблица 1. Названия внутренних тем коллекций

Решение задачи кластеризации

Для решения задачи кластеризации были рассмотрены алгоритмы иерархической кластеризации: Single Linkage и Complete Linkage [12], Between Groups Linkage (UPGMA [13]). Исследовалось влияние на качество кластеризации данными алгоритмами сужения пространства кластеризации (пространства признаков, задаваемого словарем коллекции). Рассматривалось сужение пространства кластеризации, основанное на удалении термов с низкими и высокими значениями document frequency (где значением document frequency для терма является число документов, в которых данный терм встретился). Для оценки расстояния/подобия между двумя текстами были рассмотрены: расстояние Эвклида, Jaccard index, косинус угла между векторами, корреляция Пирсона. Описание данных оценок расстояния/подобия и результаты использования их для некоторых коллекций можно найти в работе [14].

Требовалось выбрать метод, дающий самые высокие результаты в среднем, при условии, что изначально точное число кластеров неизвестно и может быть задано в интервале от 3 до 8. В результате проведенных экспериментов с использованием описанных выше тестовых коллекций были выбраны метод Between Groups Linkage и коэффициент корреляции Пирсона. Данные алгоритм и оценка подобия легли в основу алгоритма кластеризации, предложенного в работе [15] и используемого в настоящей работе. Этот алгоритм использует векторную модель представления текстов.

Удалось добиться достаточно хорошей работы алгоритма за счет специального подбора техники отбора терминов (terms selection), которая позволяет сузить пространство кластеризации. Идея принципа отбора термов в том, что нужно из всего словаря коллекции оставить только такие термины, которые вносят «положительный» вклад при вычислении корреляции между двумя текстами. К таким терминам не относятся термины, встречающиеся в большом числе документов, так как эти термины усиливают корреляцию внутри группы текстов, заведомо превышающей размер наибольшего из кластеров. Для рассматриваемого в работе типа коллекций таких слов немного (обычно 2–4 слова). С другой стороны, слова, встречающиеся в коллекции редко, не несут в себе информации о корреляции. Именно таких слов оказывается большинство. Например, удаление слов, встречающихся менее чем в 3–4 текстах, может приводить к тому, что от всего словаря коллекции остается порядка 10% слов. Из-за большого числа таких слов в векторах, представляющих тексты, оказывается большое число не взаимосвязанных ненулевых значений, что оказывает на результат вычисления корреляции эффект шума. Исходя из этих наблюдений, можно считать, что целесообразно сужать пространство кластеризации с помощью удаления из словаря коллекции 90% слов с самыми низкими значениями document frequency и 3–4 слова с самыми высокими значениями document frequency.

Выбор иерархической кластеризации оправдан тем, что с ее помощью можно получить стабильный результат кластеризации, зависящий только от определения числа кластеров. Так как точное определение числа кластеров часто затруднительно, было введено предположение, что для используемых в настоящей работе коллекций число кластеров может быть определено в интервале от 3 до 8. Использование описанных выше средств (алгоритма кластеризации, меры подобия между текстами и техники отбора терминов) позволило получить результаты, сравнимые с результатами кластеризации алгоритмами, опубликованными в работах [8–10]. Оценка результатов кластеризации проводилась с помощью той же меры

$$\text{measure}(F) = \sum_i \frac{G_i}{|D|} \max F_{ij}, \quad F_{ij} = \frac{2 \cdot P_{ij} \cdot R_{ij}}{P_{ij} + R_{ij}}, \quad P_{ij} = \frac{|G_i \cap C_j|}{G_i}, \quad R_{ij} = \frac{|G_i \cap C_j|}{C_j}.$$

Здесь G_i задает кластеры, полученные в результате автоматической обработки; C_j задает кластеры, выделенные экспертами; $|D|$ – размер коллекции. В табл. 2 приведена оценка результатов кластеризации описанным выше алгоритмом (*) в лучшем (Max), в худшем случае (Min), и в среднем (Avg) в зависимости от того, какое число кластеров порождалось (от 3 до 8 кластеров). Для сравнения в табл. 2 приводятся данные, опубликованные в работе [8] для алгоритма AntSA-CLU и в работе [9] для алгоритма CLUDISPO.

Название Коллекции		CICling_2002	SEPLN_CICling	Easy Abstracts	Micro4News
*	Max	0,73	0,84	0,82	0,96
	Avg	0,65	0,72	0,79	0,87
	Min	0,59	0,65	0,72	0,79
CLUDISPO	Max	0,73	0,85	0,98	1
	Avg	0,6	0,72	0,92	0,93
	Min	0,47	0,58	0,85	0,85
AntSA-CLU	Max	0,75	0,85	0,98	1
	Avg	0,61	0,75	0,96	0,96
	Min	0,47	0,63	0,92	0,88

Таблица 2. Оценка результата автоматической кластеризации для тестовых коллекций

Результаты работы алгоритмов K-Means [12], MajorClust [16], DBSCAN [17] в таблице не приводятся, так как качество кластеризации с помощью этих алгоритмов хуже, чем с помощью AntSA-CLU и CLUDISPO. Для последних алгоритмов результаты могут быть найдены в работе [8].

Решение задачи выделения ключевых слов и словосочетаний

Второй частью работы является задача выделения ключевых слов для кластеров и определение контекста использования выделенных слов (определение ключевых словосочетаний для найденных слов, биграмм, коллокаций). Отбор происходит только среди слов, полученных в результате сужения пространства кластеризации. В основе алгоритма лежит простая идея: словами, характеризующими тематическую направленность кластера, являются слова, встречающиеся в большом числе документов данного кластера и в малом числе документов за его пределами. Слово отбирается как ключевое для кластера, если число текстов, в которых частота слова меньше α (мы выбирали $\alpha = 4$), не превышает размер кластера, и если данное слово не встречается только в β текстах данного кластера. В работе [15] показано, что эти два условия позволяют выделить ключевые для кластера слова, которые встречаются во многих документах кластера и появляются в ряде документов за пределами кластера (первая группа слов). Эти же два условия позволяют отсеять слова, частые в нескольких кластерах и типичные для данной коллекции (вторая группа слов). Связано это с тем, что практически нет текстов, в которых слова из второй группы встречались бы часто, в отличие от слов первой группы.

Параметр β является подвижным, его увеличение приводит к выделению большого числа ключевых слов, однако качество выделяемых слов при этом падает. В работе данный параметр автоматически изменялся в зависимости от того, какое число ключевых слов требуется выделить. В настоящей работе на этапе выделения ключевых слов для каждого кластера отбирались ключевые слова до тех пор, пока не было отобрано как минимум 10 слов. Из каждого отобранных слов на следующем этапе оставлялось всего 3 слова, встречающихся в наибольшем числе документов кластера.

В работе для решения задачи выделения словосочетаний предлагается оценка взаимосвязанности слов в виде

$$c = \frac{t_1 + t_2}{2 \cdot f(t^1 t^2)},$$

где t_1 и t_2 отражает число различных пар с первым или вторым словом из рассматриваемого словосочетания, $f(t^1 t^2)$ отражает число появлений словосочетания типа «первое–второе слово вместе». Чем меньше значение c , тем лучшей считается пара слов (t^1, t^2) . Словосочетанием, поясняющим ключевое слово, является устойчивое словосочетание, где одно из слов является ключевым. Пара слов (t_1, t_2) является устойчивым словосочетанием, если для слов t_1 и t_2 величина c меньше некоторого фиксированного порога, а сами значения t_1 , t_2 , $f(t^1 t^2)$ больше 3. Максимальное пороговое значение, используемое в работе, равно 20. Хорошим является также порог, равный 15, менее хорошим – 30, при пороге, равном 10, очень мало словосочетаний, определяемых как устойчивые. Идея,ложенная в основание предложенной формулы, состоит в следующем: устойчивые словосочетания часто встречаются вместе и редко порознь.

В настоящей работе приводятся результаты, полученные для коллекции Easy Abstracts. В табл. 3 приведены результаты для случая восьми кластеров (при числе кластеров в диапазоне от 3 до 8 результаты похожи). Пустые графы в табл. 3 говорят о том, что кластеры содержат всего 1–2 документа, т.е. такого числа документов недостаточно для выделения ключевых слов кластера.

Ключевые слова	Ключевые словосочетания
search, objective, function	objective+single, objective+genetic objective+multi, search+local search+space, search+tabu function+approximation
proof, theorem, based	theorem+proving
agent, models, agents	agent+oriented, agent+communication, agent+patterns, agents+esrl
learning, machine, function	machine+learning, machine+boltzmann machine+tabu, function+approximation, learning+machine, learning+classifier earning+reinforcement
selection, large, probabilistic	large+data, selection+feature, probabilistic+svm
---	---
---	---
---	---

Таблица 3. Результат аннотирования ключевыми словосочетаниями коллекции Easy Abstracts в случае восьми кластеров

Просмотрев документы коллекции Easy Abstracts, можно убедиться, что найденные словосочетания являются типичными для конкретных кластеров этой коллекции. К сожалению, у нас сейчас нет иной меры оценки качества найденных словосочетаний, чем оценка человеком-экспертом.

Заключение

В работе предложен алгоритм для автоматического аннотирования узкотематических маленьких коллекций коротких текстов, заключающийся в извлечении из коллекций ключевых словосочетаний и ключевых слов на основе предварительной кластеризации. Предложены алгоритмы кластеризации, аннотирования полученных кластеров ключевыми словами и алгоритм выделения ключевых словосочетаний. Предложенный алгоритм кластеризации дает достаточно высокие результаты для указанных коллекций. С помощью алгоритма выделения ключевых словосочетаний удается выделить словосочетания, отражающие специфику каждого кластера коллекции. Авторам неизвестны попытки выделения ключевых слов для рассматриваемых в этой работе коллекций, в частности, для коллекции Easy Abstracts. Это затрудняет оценку качества представленного алгоритма. В дальнейшем планируется модификация предложенного в работе алгоритма выделения ключевых словосочетаний: для вычисления $f(t^1t^2)$ планируется использовать информацию об отдельных кластерах, а не информацию обо всей коллекции в целом, как это сделано в настоящей работе.

Литература

1. Lloret E. Topic Detection and Segmentation in Automatic Text Summarization [Электронный ресурс]. – Режим доступа: <http://www.dlsi.ua.es/~ellobre/publications/SumTopics.pdf>, св. Яз. англ. (дата обращения 01.10.2011).
2. Teufel S., Moens M. Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status [Электронный ресурс]. – Режим доступа: <http://acl.ldc.upenn.edu/J/J02/J02-4002.pdf>, св. Яз. англ. (дата обращения 01.10.2011).
3. Ягунова Е.В., Пивоварова Л.М. Извлечение и классификация коллокаций на материале научных текстов. Предварительные наблюдения // V Международная научно-практическая конференция «Прикладная лингвистика в науке и образовании» памяти Р.Г. Пиотровского (1922–2009): Материалы. – СПб: 2010. – С. 356–364.
4. Makkonen J. Semantic Classes in Topic Detection and Tracking [Электронный ресурс]. – Режим доступа: <http://www.doria.fi/bitstream/handle/10024/48180/semantic.pdf>, св. Яз. англ. (дата обращения 01.10.2011).
5. Smith S.C., Rodríguez M.A. Clustering-based Searching and Navigation in an Online News Source [Электронный ресурс]. – Режим доступа: http://captura.uchile.cl/jspui/bitstream/2250/6257/1/Smith_Simon.pdf, св. Яз. англ. (дата обращения 01.10.2011).
6. Shih C., Peng T. Building Topic / Trend Detection System based on Slow Intelligence [Электронный ресурс]. – Режим доступа: <http://www.cs.pitt.edu/~chang/265/proj10/57shih.pdf>, св. Яз. англ. (дата обращения 01.10.2011).
7. He Q., Chang K., Lim E., Banerjee A. Keep It Simple with Time: A Re-examination of Probabilistic Topic Detection Models [Электронный ресурс]. – Режим доступа: <http://www-users.cs.umn.edu/~banerjee/papers/09/pami-tdt.pdf>, св. Яз. англ. (дата обращения 01.10.2011).
8. Errecalde M., Ingaramo D., Rosso P. A new AntTree-based Algorithm for Clustering Short-text Corpora // Journal of Computer Science and Technology. – 2010. – V. 10. – № 1. – P. 1–7.

9. Ingaramo D., Cagnina L., Errecalde M., Rosso P. A Particle Swarm Optimizer to cluster short-text corpora: a performance study // Proc. Workshop on Natural Language Processing and Web-based Technologies, 12th edition of the Ibero-American Conference on Artificial Intelligence. IBERAMIA. – 2010. – Р. 71–79.
10. Pinto D. Analysis of narrow-domain short texts clustering. Research report for Diploma de Estudios Avanzados. DEA // Department of Information Systems and Computation. UPV. – 2007 – [Электронный ресурс]. – Режим доступа: <http://users.dsic.upv.es/~prossoresources/PintoDEA.pdf>, св. Яз. англ. (дата обращения 01.10.2011).
11. PLN Resources // Data Sets for Short-texts Experimental Works [Электронный ресурс]. – Режим доступа: <https://sites.google.com/site/merrecalde/resources>, св. Яз. англ. (дата обращения 01.10.2011).
12. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze Introduction to Information Retrieval. – Cambridge University Press. – 2008. – С. 377–402.
13. Local methods – UPGMA (Unweighted Pair Group Method) // Phylogenetics workshop 09: Methods in Bioinformatics and Molecular Evolution [Электронный ресурс]. – Режим доступа: http://www.adelaide.edu.au/acad/events/workshop/LockhartUPGMA&NJ_calculation.pdf, св. Яз. англ. (дата обращения 01.10.2011).
14. Huang A. Similarity Measures for Text Document Clustering. Department of Computer Science The University of Waikato, Hamilton, New Zealand [Электронный ресурс]. – Режим доступа: http://nzcsrc08.canterbury.ac.nz/site/proceedings/Individual_Papers/pg049_Similarity_Measures_for_Text_Document_Clustering.pdf, св. Яз. англ. (дата обращения 01.10.2011).
15. Popova S.V., Khodyrev I.A. Local theme detection and annotation with key words for narrow and wide domain short text collections // The Fifth International Conference on Advances in Semantic Processing. SEMAPRO. 2011. – Lisbon: Portugal, 2011. – Р. 49–55.
16. Stein B., Niggemann O. On the Nature of Structure and its Identification // In Proc. of the 25th International Workshop on Graph Theoretic Concepts in Computer Science. LNCS. – Springer-Verlag, 1999. – V. 1665. – Р. 122–134.
17. Ester M., Kriegel H., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise // Proc. of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96). – 1996. – Р. 226–231.

Попова Светлана Владимировна
Ходырев Иван Александрович

– Санкт-Петербургский государственный университет, spbu@bk.ru
– ОЛИМП, программист, kivan.mih@gmail.com

УДК 004.934

НЕКОТОРЫЕ АСПЕКТЫ ИССЛЕДОВАНИЯ СИСТЕМ РАСПОЗНАВАНИЯ РЕЧИ В ТЕЛЕФОННЫХ СЛУЖБАХ ПОДДЕРЖКИ

П.В. Балакшин, Г.Ю. Петров

Представлены проблемы телефонных служб поддержки клиентов (call-центров) крупных компаний. Определены преимущества и недостатки применения систем распознавания речи в работе таких центров. Показана важность метрики FCR (First Call Resolution) и особенности ее вычисления в системах автоматического распознавания речи.

Ключевые слова: распознавание речи, телефонная служба поддержки, call-центр, FCR.

Введение

В настоящее время активно ведутся исследования в области речевого взаимодействия «человек – компьютер», идет поиск новых параметров исходных данных, новых методов, возможных ограничений и объединений, позволяющих повысить точность и другие показатели распознавания речи компьютерными системами. Актуальность данных исследований подтверждается большим количеством международных научно-технических конференций, посвященных этим проблемам (Interspeech (1993–2011 г.г.), SpeCom (1996–2011 г.г.)), крупными программно-техническими разработками последних лет (Dragon Naturally Speaking, встроенное речевое управление в операционной системе Vista, речевой поиск Google).

Автоматическое распознавание речи находит все более широкое применение в работе call-центров (call center) – телефонных служб поддержки клиентов крупных компаний, деятельность которых сводится к оперативному решению телефонных запросов от клиентов или потенциальных потребителей. В небольших компаниях с этим могут справиться несколько человек, но для крупных организаций (например, страховых или транспортных компаний, финансовых учреждений) такое положение дел неприемлемо (табл. 1).

В настоящее время ряд таких центров рассматривают возможность частичной замены операторов на компьютерные устройства, способные распознавать речь, обрабатывать, анализировать ее и выдавать осмысленный и необходимый ответ клиенту [2].

Вероятность услышать сигнал «Занято», %	Количество линий
90	3
80	6
70	8
....
10	28
5	31
4	32
3	33
2	34
1	36

Таблица 1. Необходимое количество линий при 500 входящих вызовах в час [1]

Целью настоящей работы являлось исследование и оптимизация систем автоматического распознавания речи в работе call-центров.

Построение модели системы распознавания

Следует подчеркнуть, что общение оператора телефонной службы поддержки и клиента происходит в виде живой беседы. Следовательно, распознавание речи должно происходить если не в реальном, то в квазиреальном масштабе времени. Преимуществами использования систем распознавания речи в телефонных службах поддержки являются [3–6]:

- существенное снижение времени ожидания (handle time), что позволит снизить затраты на работу;
- сокращение времени звонка в 1,5–2 раза за счет сокращения времени ввода информации оператором;
- возможность автоматического ответа на простые вопросы, освобождая время работы операторов для работы со сложными звонками;
- возможность работы с клиентами в круглосуточном режиме (включая праздники);
- быстрая верификация голоса клиента во время ответа на один–два несложных вопроса, что особенно актуально в банковском секторе для защиты от краж персональных карт и секретных документов;
- возможность работы с большим числом коротких звонков (телефонная служба поддержки букмекерских компаний);
- возможность замены сложной и допускающей частые ошибки IVR-системы (Interactive Voice Response), работающей в тональном режиме;
- возможность использовать распознавание речи в качестве дополнительного источника информации не только при разговоре, но и при дальнейшем анализе работы call-центра. В частности, данный анализ помогает повысить ключевую метрику FCR – разрешение проблемы (вопроса) за один звонок. Это приводит к снижению числа повторных звонков и повышению удовлетворенности клиентов одновременно, что в свою очередь приводит к снижению операционных расходов.

К основным недостаткам использования систем распознавания речи следует отнести:

- невозможность распознавания с точностью 100% из-за большого количества различных особенностей произношения, таких как диалект, дефекты речи, эмоции, и ассоциативного восприятия речи человеком;
- очень большая начальная стоимость внедрения системы распознавания речи.

Предлагаемая авторами модель системы распознавания речи включает в себя блок обратной связи на изменение размера речевого словаря после обработки полученной информации (рис. 1). С одной стороны, меньший словарь обеспечит более высокую скорость распознавания. С другой стороны, существует вероятность распознать редкий вопрос с очень плохой точностью. По этой причине данный параметр, а именно, размер речевого словаря системы распознавания, очень важен: от него зависит скорость вовлечения распознавания речи в работу телефонной службы поддержки.

В рамках проведенных исследований по предлагаемой модели [7] было определено, что для русского языка словарь размером 2500 слов является достаточным для покрытия более 95% информации усредненной предметной области. Исходя из этого, не менее 95% информации будет распознаваться с точностью, присущей используемой модели.

В итоге, возможность модификации размера речевого словаря внутри самой системы в режиме реального (квазиреального) времени становится преимуществом использования распознавания речи. Особенно очевидно данное преимущество в работе телефонных служб поддержки, так как каждая из них имеет строго конкретизированную предметную область.

Данные результаты для русского языка получены впервые. Однако они находятся в соответствии с аналогичными исследованиями по английскому языку. Исследователь Ч. Огден, сформировавший в 1930 г. BASIC English, показал, что при строгой грамматике необходимо 850 слов. В дальнейшем он

увеличил словарь до 1350 слов, а «стандартный» английский язык, использующийся в упрощенной версии Википедии, состоит из 2000 слов.

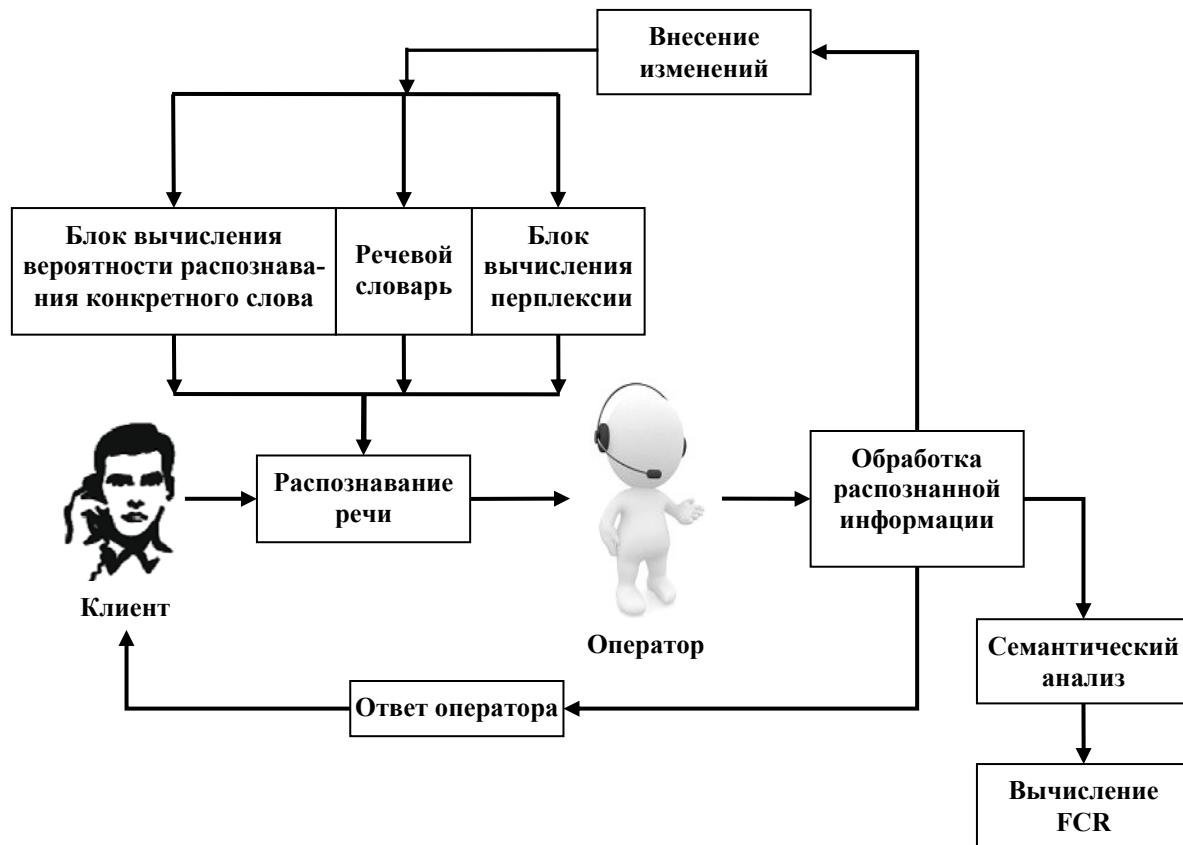


Рис. 1. Модель системы распознавания речи

Дополнительной особенностью модели является блок обратной связи на изменение вероятности распознавания конкретного слова после обработки полученной информации (рис. 1). Таким образом, блок вычисления вероятности распознавания конкретного слова работает в паре с речевым словарем. Благодаря этому появляется возможность увеличивать вероятность распознавания как заранее определенных, так и новых слов.

Выделим еще одну важную особенность предложенной модели. Авторами предложено учитывать в модели связность слов, или так называемую перплексию. Связность учитывает количество слов, которые логически могут встретиться после данного слова. Именно связность повышает точность распознавания речи, поскольку вопросы, задаваемые пользователями, имеют достаточно четкий синтаксис, что особенно ярко выражено в английском языке. Для этого выделен специальный блок вычисления перплексии (рис. 1), используемый во время распознавания, в который с помощью блока обратной связи вносятся изменения по результатам обработки распознаваемой информации.

Перплексия вычислялась по формуле [8]

$$PPL = 2^{\sum_{i=1}^N \frac{1}{N} \log_2 q(x_i)},$$

где N – количество состояний модели q ; x_i – состояние модели. Чем меньше значение перплексии, тем лучше модель предсказывает следующее слово.

Для проведения исследований за основу были взята языковая модель, использованная при реализации системы Decipher [9]. В данной системе при размере словаря в 998 слов в модели, использующей пары слов, перплексия равна 60, а в модели, использующей отдельные слова, перплексия равна 998, т.е. размеру всего словаря. В качестве исходных данных для исследования была составлена языковая модель на основе речевого словаря телефонной службы поддержки системного интегратора «Системные решения» (г. Москва). Словарь, составленный по результатам прослушивания звонков, содержал 1867 слов. Результаты исследования аналогичны системе Decipher (табл. 2).

Таким образом, лучшие результаты были получены при использовании групп из трех слов. Дополнительным подтверждением корректности результатов могут служить исследования о величине перплексии в зависимости от предметной области (табл. 3).

Количество связанных слов в языковой модели	1	2	3	4	5	6
Перплексия	1867	243	96	180	491	702

Таблица 2. Соотношение связанных слов и перплексии в языковой модели

Предметная область	Перплексия
Радиология	20
Неотложная медицинская помощь	60
Журналистика	105
Общий английский	247

Таблица 3. Перплексия триграммной языковой модели для различных предметных областей [10]

Как было отмечено ранее, преимуществом использования распознавания речи в телефонных службах поддержки является как само распознавание, так и дальнейший анализ полученной информации. В соответствии с результатами последних исследований call-центров [6, 8, 11], проведенных для различных отраслей бизнеса, 33% от общего числа звонков являются повторными (рис. 2).

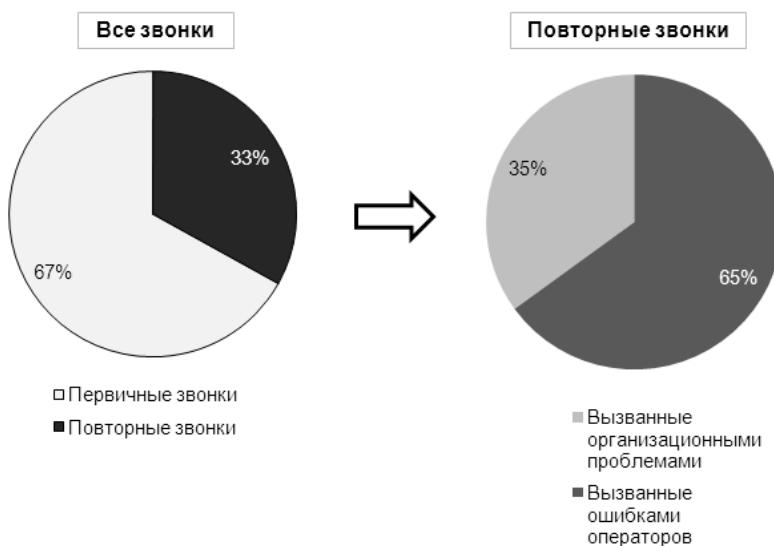


Рис. 2. Исследование причин возникновения повторных звонков

Особенности вычисления и использования метрики FCR

Наши исследования показали, что предлагаемая модель может с успехом использоваться для снижения количества повторных звонков. Это приводит к снижению общего количества звонков и снижению операционных расходов телефонной службы поддержки. Указанные изменения неразрывно связаны с улучшением основной метрики FCR, определяющей работу всей телефонной службы поддержки. Данная метрика оценивает процент разрешения проблемы (вопроса) за один звонок и вычисляется следующим образом [12]:

$$FCR_k = \frac{E_k}{N_k}, \quad FCR_K = \frac{\sum_{k=1}^K E_k}{\sum_{k=1}^K N_k},$$

где K – общее количество причин звонка; k – элемент из множества причин K ; N_k – количество цепочек звонков по k -ой причине; E_k – количество цепочек звонков по k -ой причине, состоящих из одного звонка. Тогда общее значение FCR можно определить как среднее из всех FCR_k .

Более высокое значение метрики является показателем более успешной работы. По итогам опроса руководителей 54 телефонных служб поддержки Великобритании 81% из них считает, что FCR – это именно тот показатель, который может объективно оценить работу как всей телефонной службы поддержки, так и каждого оператора (или группу операторов), в частности [3].

Существует несколько методов определения метрики FCR, среди которых выделяют [13]:

- текущий контроль за соединениями;
- опрос с помощью интерактивной IVR-системы по окончании звонка;
- телефонный опрос спустя один-три дня после первоначального звонка (survey – анкетирование, оценка, экспертиза).

В предлагаемом методе по итогам распознавания причины звонка и клиента данная информация передается в блок вычисления FCR (рис. 1), где на основании подсчета повторных звонков от каждого клиента по каждой причине вычисляется значение метрики FCR для каждой причины (FCR_k), а затем и итоговая FCR. Текущее значение FCR составляет 65–75%, что влечет довольно большие расходы на работу технической службы поддержки.

Рассмотрим более подробно, почему именно распознавание речи помогает повысить метрику FCR. Одной из главных проблем call-центров, которую возможно решить с применением автоматического распознавания речи, является определение причины звонка. Оператор либо выбирает эту информацию из списка уже известных причин, либо кратко записывает в текстовое поле информацию, услышанную от клиента. Но для части звонков определить причину звонка все равно невозможно, во многом из-за некорректной информации, введенной в специальное поле вручную. Так, в call-центре одного американского оператора связи процент звонков без причины составляет 10–12, а в call-центре одного довольно крупного американского банка – 27–30.

Показано, что в системе, в которой осуществлена предлагаемая авторами модель, количество звонков с конкретной причиной увеличивается, а за счет дальнейшего анализа повышается и значение метрики FCR до 72–79%. Исследованные звонки были распознаны со скоростью около 95 слов в минуту. Данный показатель является неплохим результатом, так как средняя скорость русской речи составляет 105–125 слов. Однако часто клиенты телефонной службы поддержки произносят свою речь с большим количеством эмоций, заметно ускоряя скорость речи и приближая ее к показателям английской речи – 140–150 слов в минуту.

Отметим еще один положительный момент. Введение в модель системы распознавания речи блока семантического анализа (рис. 1) позволяет повысить именно семантическую точность распознавания и улучшить значение метрики FCR.

Рассмотрим, например, следующие звонки:

- «Банкомат съел мою карту»;
- «Банкомат не вернул мне карту»;
- «Банкомат не принимает карту»;
- «Карта не выдается банкоматом обратно».

В данном случае важно то, что банкомат имеет проблемы при работе с картой. Таким образом, как причина звонка будет регистрироваться именно эта информация, а не ее семантически менее важные детали. За счет этого для определения метрики FCR будет использоваться уникальная причина звонка вместо различных схожих, т.е. семантический анализ выполняет некоторую предобработку данных.

Благодаря использованию семантического анализа в данных экспериментах метрика FCR увеличивается до 78–86%.

Заключение

В работе показано, что применение компьютерных систем распознавания речи, основанных на модели с модификацией словаря, и дальнейший анализ полученной информации позволяют повысить производительность телефонных служб поддержки и снизить затраты на их работу. Предложенная разработка помогает повысить ключевую метрику FCR. Это приводит к снижению числа повторных звонков, значительно снижает операционные расходы, повышает удовлетворенность клиентов.

Работа выполнена при поддержке гранта Правительства Петербурга № 3.11/04-06/50.

Литература

1. Официальный сайт компании Град Телеком [Электронный ресурс]. – Режим доступа: <http://www.gradtelecom.ru/ru/papers/callcenter/>, св. Яз. рус. (дата обращения 07.04.2010).
2. Хижинский Д. Распознавание речи взорвет рынок call-центров // cnews.com: сетевой журн. 22.05.2007. [Электронный ресурс]. – Режим доступа: <http://www.cnews.ru/reviews/index.shtml?2007/05/22/251191>, св. Яз. рус. (дата обращения: 11.05.2010).
3. Contact centres rate call efficiency over customer service // callcentrehelper.com: сетевой журн. 07.04.2010. [Электронный ресурс]. – Режим доступа: <http://www.callcentrehelper.com/contact-centres-rate-call-efficiency-over-customer-service-9921.htm#more-9921>, св. Яз. англ. (дата обращения: 28.04.2010).
4. How is speech recognition applied in call centers? [Электронный ресурс]. – Режим доступа: <http://www.gloccal.com/voice-recognition/speech-recognition-in-call-centers.html>, св. Яз. англ. (дата обращения 19.12.2010).

5. Feldon B. The top five uses of speech recognition technology [Электронный ресурс]. – Режим доступа: <http://www.callcentrehelper.com/the-top-five-uses-of-speech-recognition-technology-1536.htm>, св. Яз. англ. (дата обращения 19.12.2010).
6. Neustein A. Advanced in Speech Recognition. Mobile Environments, Call Centers and Clinics. – Springer Science. – 2010. – 370 p.
7. Балакшин П.В., Тропченко А.Ю. Выбор размера словаря при реализации тестовой системы распознавания речи // Интеллектуальные и информационные системы. Материалы Всероссийской научно-технической конференции. – Тула: Изд-во ТулГУ, 2009. – С. 51–54.
8. Call Centre Monitoring // callcentrehelper.com: сетевой журн. 2010 [Электронный ресурс]. – Режим доступа: <http://www.callcentrehelper.com/call-centre-monitoring-16.htm>, св. Яз. англ. (дата обращения: 22.04.2010).
9. Cohen M., Murveit H., Bernstein J., Price P., Weintraub M. The Decipher speech recognition system // IEEE ICASSP. – Albuquerque, 1990. – P. 77–80.
10. Roukos S., Battista Varile Giovanni, Antonio Zampolli. Language Representation. Survey of the State of the Art in Human Language Technology. – 1995 [Электронный ресурс]. – Режим доступа: <http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>, св. Яз. англ. (дата обращения: 05.09.2011).
11. Increase First Call Resolution: Eliminate the Top 10 Agent Errors Causing Repeat Calls // crmxchange.com: сетевая библ.: 2008 [Электронный ресурс]. – Режим доступа: <http://www.crmxchange.com/offer/enkatadec08.asp>, св. Яз. англ. (дата обращения: 28.12.2009).
12. Fluss D. FCR Improves Contact Center Performance // g-cem.org: сетевой журн. 2010 [Электронный ресурс]. – Режим доступа: http://www.gccrm.com/eng/content_details.jsp?contentid=2245&subjectid=107, св. Яз. англ. (дата обращения: 13.04.2011).
13. Стандарт EN 15838:2009 [Электронный ресурс]. – Режим доступа: http://www.iccici.ru/iccici/call_centers/en_15838/, св. Яз. рус. (дата обращения: 13.04.2011).

Балакшин Павел Валерьевич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, pvbalakshin@gmail.com
Петров Григорий Юрьевич	– Энката Технолоджис, Инк., кандидат экономических наук, директор отдела внедрения, gpetrov@enkata.com

5

ТЕХНОЛОГИИ ПРОГРАММИРОВАНИЯ

УДК 004.85

ВЫБОР ФУНКЦИИ ПРИСПОСОБЛЕННОСТИ ОСОБЕЙ ГЕНЕТИЧЕСКОГО АЛГОРИТМА С ПОМОЩЬЮ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

А.С. Афанасьева, М.В. Буздалов

Предлагается метод, позволяющий динамически выбирать вспомогательную функцию приспособленности, наиболее выгодную для использования в генетическом алгоритме. Метод основан на применении обучения с подкреплением. Приведены результаты его использования для решения модельной задачи.

Ключевые слова: генетические алгоритмы, машинное обучение, обучение с подкреплением.

Введение

В теории оптимизации известны задачи скалярной и многокритериальной оптимизации [1]. На практике возникает необходимость решать модифицированную скалярную задачу оптимизации, предполагающую наличие дополнительных критериев [2]. Она отличается от многокритериальной задачи тем, что целью ее решения является максимизация единственной целевой функции, а дополнительные критерии имеют лишь вспомогательное значение. В таких задачах целевая функция может некоторым образом зависеть от дополнительных критериев, поэтому в ряде случаев вместо максимизации целевой функции оказывается выгодным оптимизировать дополнительные критерии.

Важным классом алгоритмов оптимизации являются генетические алгоритмы (ГА) [3], где в качестве критерия выступает функция приспособленности (ФП). При наличии нескольких вспомогательных ФП выбор наиболее выгодной из них приходится производить вручную [4]. Подобный подход не вполне эффективен, так как предполагает многократный перезапуск ГА. В представляющей работе предлагается метод, позволяющий автоматизировать этот процесс. Более того, метод позволяет осуществлять динамический выбор в случае, когда на разных этапах выполнения алгоритма выгодны разные ФП.

Введем некоторые понятия. Пусть имеется набор функций приспособленности. Целью применения ГА является получение особи, имеющей как можно лучшее (например, наибольшее) значение одной из этих функций. Будем называть такую ФП целевой ФП (ЦФП), остальные ФП будем называть вспомогательными. Функцию приспособленности, используемую в ГА для отбора особей в данный момент времени, будем называть текущей ФП.

Будем рассматривать случай, в котором применение одной из вспомогательных ФП в качестве текущей приводит к более быстрому получению особей с более высокими значениями ЦФП, нежели применение самой ЦФП. Целью настоящей работы является демонстрация на модельной задаче того, что применение обучения с подкреплением [5] позволяет выбирать оптимальные текущие ФП во время работы ГА.

Модельная задача

Рассмотрим следующую модельную задачу. Пусть особь представляется битовой строкой фиксированной длины n . Будем обозначать число единиц в такой строке как x . Функции приспособленности особи будем задавать как функции от x . ЦФП задается формулой $g(x) = \left\lfloor \frac{x}{k} \right\rfloor$. Вспомогательные ФП

имеют следующий вид: $h_1(x) = \begin{cases} x, & x < p \\ p, & x \geq p \end{cases}$ и $h_2(x) = \begin{cases} p, & x < p \\ x, & x \geq p \end{cases}$, где p – положительное целое число. Будем называть p точкой переключения.

Можно видеть, что для особей с числом единиц, меньшим значения точки переключения, выгоднее использовать функцию h_1 в качестве текущей ФП. Для особей, число единиц в представлении которых превышает значение точки переключения, выгодно использовать h_2 . Применение этих функций будет быстрее приводить к более высоким значениям целевой функции g .

Задачей разрабатываемого метода является динамическое переключение на текущую ФП, наиболее выгодную для особей, составляющих текущее поколение. Другими словами, метод должен сначала выбирать функцию h_1 в качестве текущей, а после достижения точки переключения выбирать функцию h_2 .

Обучение с подкреплением

Большинство алгоритмов обучения с подкреплением относится к методам машинного обучения, не требующим наличия заранее подготовленного набора тестовых примеров [5, 7]. Инкрементальный характер этих алгоритмов позволяет обеспечить динамический характер переключения функций приспо-

собленности и освобождает от необходимости подготовки набора тестовых примеров, требующей выполнения многократных запусков ГА.

Кратко опишем идеи обучения с подкреплением. Агент обучения применяет действия к среде, которая отвечает на каждое действие поощрением. Целью агента является максимизация суммарного поощрения. Для эффективного применения использованных в данной работе алгоритмов необходимо, чтобы задача могла быть представлена в виде марковского процесса принятия решений.

Марковский процесс принятия решений состоит из:

- множества состояний S ;
- множества действий A ;
- функции поощрения $W : S \times A \rightarrow R$;
- функции перехода между состояниями $T : S \times A \times S \rightarrow R$, заданной таким образом, что вероятность перехода среды из состояния s в состояние s' после действия агента a равна $T(s, a, s')$.

В следующем разделе будет показано, как описанная ранее модельная задача была представлена в виде марковского процесса принятия решений.

Представление модельной задачи в терминах обучения с подкреплением

Множество действий агента задается с помощью функций приспособленности: $A = \{g, h_1, h_2\}$.

Применение того или иного действия означает выбор соответствующей ФП в качестве текущей и последующее формирование нового поколения ГА.

Состояние среды соответствует состоянию ГА. Оно однозначно определяется упорядоченным вектором значений производных всех применяемых ФП в точке x , задаваемой числом единиц в лучшем индивидууме поколения. Например, если выполняется неравенство $h_1'(x) \leq g'(x) \leq h_2'(x)$, то состояние имеет вид $\langle h_1, g, h_2 \rangle$.

Значение функции поощрения зависит от изменения значений функций приспособленности при переходе из одного состояния в другое, вызванном действием агента. Введем вспомогательную функцию D_f , такую, что

$$D_f(x_1, x_2) = \begin{cases} 0 : f(x_2) - f(x_1) < 0 \\ 0,5 : f(x_2) - f(x_1) = 0, \\ 1 : f(x_2) - f(x_1) > 0 \end{cases}$$

где f – некоторая фиксированная ФП. Функция поощрения имеет вид $W(s, a) = D_g(x_s, x_{s'}) + c(D_{h_1}(x_s, x_{s'}) + D_{h_2}(x_s, x_{s'}))$, где $c \in [0, 1]$ – вещественный коэффициент, позволяющий варьировать вклад вспомогательных функций в величину поощрения; s' – состояние, в которое переходит среда из состояния s после действия a ; x_s и $x_{s'}$ – значения, соответствующие лучшим особям в состояниях s и s' .

Наряду с приведенными определениями рассматривались и альтернативные: состояние среды определялось упорядоченным вектором значений функций приспособленности, значение функции поощрения – изменением значения ЦФП. Однако такие определения в ходе предварительного эксперимента не позволяли эффективно бороться с зашумленностью среды и приводили к непостоянству значений функции поощрения, иными словами, к переменному характеру среды. Для борьбы с подобным эффектом важно, чтобы после точки переключения среда переходила в другое состояние, в котором функция поощрения может быть определена по-новому без ущерба для постоянства среды. Предложенные в конечном итоге определения удовлетворяют этому требованию.

Описание реализации алгоритма

Для решения поставленной задачи было применено Q -обучение, которое является разновидностью инкрементальных алгоритмов обучения с подкреплением, не строящих модель среды. Особенностью таких алгоритмов является малая требовательность к вычислительным ресурсам на каждом шаге [7]. Эта особенность важна для того, чтобы не замедлять работу ГА.

Использовалось две разновидности Q -обучения: Q -обучение с ϵ -жадным правилом [5, 8] и отложенное Q -обучение, предложенное в работе [9]. Был реализован инкрементальный алгоритм. Каждый шаг этого алгоритма состоит из шага обучения, задачей которого является выбор главной ФП, и шага ГА, состоящего в генерации следующего поколения.

Разработаны независимые интерфейсы среды и агента обучения с подкреплением, что должно позволить данному методу работать не только с ГА, но и с другими алгоритмами оптимизации. Для добавления нового типа алгоритма следует реализовать соответствующий интерфейс среды, что предполагает

разработку состояний среды и функции поощрения. После этого к новой среде можно будет применять ранее реализованные алгоритмы обучения с подкреплением.

Описание и результаты эксперимента

В качестве эксперимента модельная задача решалась с помощью двух алгоритмов, один из которых основан на отложенном Q -обучении, другой – на ε -жадном. Была собрана статистика, описывающая запуски ГА с применением этих алгоритмов при различных комбинациях значений параметров как самой задачи, так и использовавшихся алгоритмов обучения. Вероятности кроссовера и мутации, применявшиеся в ГА, составляли 70% и 3% соответственно.

В качестве параметров задачи выступали длина строки l , представляющей особь ГА, значение точки переключения p , а также делитель k , входящий в определение целевой функции $g(x)$. Характер параметров обучения зависел от конкретного алгоритма. В случае отложенного обучения настраивались такие параметры, как период обновления значений Q -функции m , дополнительное поощрение ε_1 и величина дисконтного фактора γ [9]. Для ε -жадного обучения задавались вероятность выбора случайного действия ε , скорость обучения α и также величина дисконтного фактора γ [8].

Во время предварительного эксперимента было выявлено, что в качестве значения константы c , входящей в определение функции поощрения W , разумно брать значение $c = 0,5$. Подобный выбор позволяет детальнее учитывать состояние ГА, чем выбор, при котором значения c близки к нулю, и сохранить преимущество целевой функции перед вспомогательными, что не достигается при значениях c , близких к единице.

В ходе основного эксперимента было перебрано более тысячи комбинаций различных значений параметров, каждая конфигурация запускалась 100 раз с целью усреднения. Рассмотрим результаты для параметров задачи $l = 400$; $p = 266$; $k = 10$.

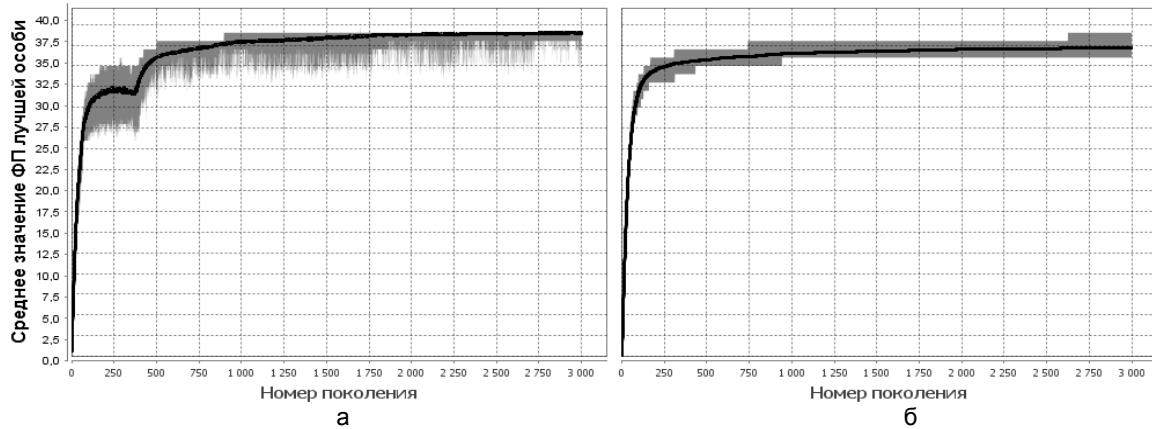


Рис. 1. Графики работы ГА, к которому применен алгоритм, основанный на отложенном Q -обучении (а), и обычного ГА (б)

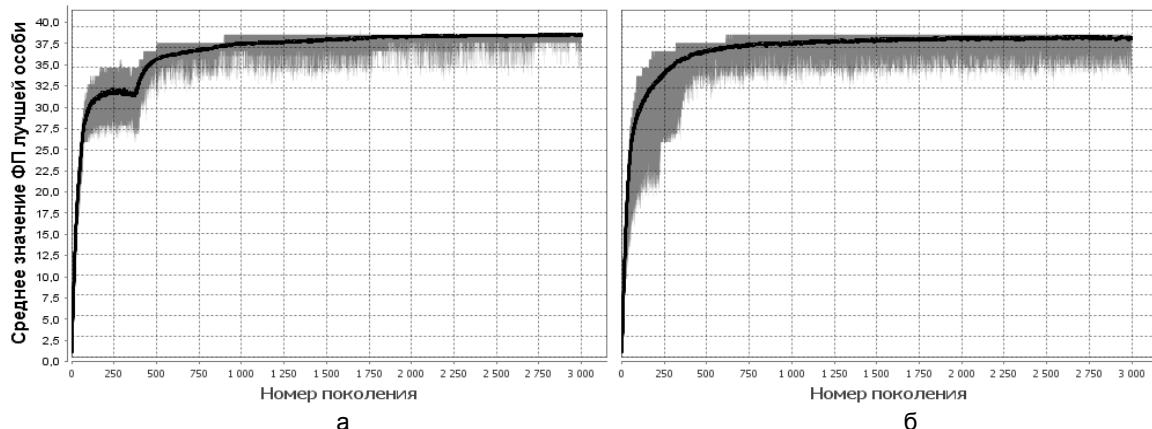


Рис. 2. Графики работы ГА, к которому применены алгоритмы, основанные на отложенном (а) и на ε -жадном (б) Q -обучении

На рис. 1 показан сравнительный график работы ГА без обучения (рис. 1, б) и работы алгоритма с отложенным Q -обучением (рис. 1, а). Для построения графика использованы оптимальные параметры алгоритма отложенного обучения, выявленные в ходе эксперимента: $m = 100$; $\varepsilon_1 = 0,2$; $\gamma = 0,01$. Можно видеть, что скорость работы ГА с обучением превосходит скорость работы обычного ГА. Это объясняет-

ся тем, что обучение справляется с задачей выбора оптимальной ФП. Отставание отложенного обучения на промежутке до 375-го поколения соответствует периоду накопления опыта агентом.

На рис. 2 представлен сравнительный график работы алгоритмов, основанных на отложенном и ε -жадном Q -обучении. Параметры отложенного обучения – те же, что и в предыдущем примере. Для жадного обучения были подобраны параметры $\varepsilon = 0,1$; $\alpha = 0,1$; $\gamma = 0,01$. Жадное обучение быстрее реагирует на изменение состояния среды, комбинируя накопление опыта с его применением. В целях исследования оно выбирает случайное действие с вероятностью ε . Этим объясняется большой разброс значений ФП особей, и, как следствие, отставание среднего значения ФП от значения, обеспечиваемого другим алгоритмом. Алгоритм, применяющий отложенное Q -обучение, характеризуется меньшим разбросом значений ФП особей. После прохождения периода накопления опыта он показывает лучшие результаты, чем алгоритм с жадным обучением.

На рис. 3 представлены графики, показывающие число переключений на ту или иную ФП. ФП 1 на рисунке обозначает $h_1(x) = \begin{cases} x, & x < p \\ p, & x \geq p \end{cases}$; ФП 2 соответствует $h_2(x) = \begin{cases} p, & x < p \\ x, & x \geq p \end{cases}$; ФП 3 представляет собой целевую функцию $g(x) = \left\lfloor \frac{x}{k} \right\rfloor$, где $p = 266$; $k = 10$. Можно видеть, что ε -жадное обучение быстрее и стабильнее выбирает ФП 1 на первом промежутке значений ЦФП особей, расположенному перед точкой переключения. Отложенное обучение на этом промежутке находится в стадии накопления опыта, что отражается на графике в виде примерно равного числа переключений на все три ФП.

Отложенное обучение после накопления опыта выбирает $g(x)$ в качестве главной ФП, чем объясняется отсутствие роста графика, соответствующего отложенному обучению, на промежутке от 125-го до 375-го поколений на рис. 1, 2, но затем достаточно быстро переключается на оптимальную $h_2(x)$. Таким образом, оба алгоритма обеспечивают стабильный выбор наиболее эффективной ФП 2 на втором участке значений ЦФП особей, следующем после точки переключения.

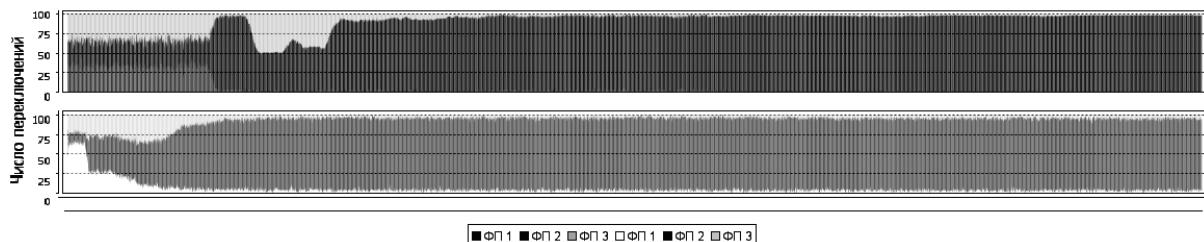


Рис. 3. Переключение ФП с применением отложенного (верхняя диаграмма) и ε -жадного (нижняя диаграмма) Q -обучения

Можно сделать вывод о том, что оба алгоритма способны переключаться на оптимальную ФП, что приводит к повышению производительности ГА. Использование ε -жадного обучения позволяет переключаться на оптимальную ФП практически сразу. Алгоритм, использующий отложенное Q -обучение, переключается при условии предоставления достаточного времени для накопления опыта, однако он показывает более стабильные и высокие результаты на длительном временном промежутке.

Заключение

В работе описан метод, позволяющий динамически настраивать генетический алгоритм с целью ускорения поиска наиболее приспособленных особей. Метод основан на применении обучения с подкреплением для выявления наиболее выгодной вспомогательной функции приспособленности. В предлагаемый метод заложены возможности расширения для совместного использования с произвольным алгоритмом оптимизации. Проведенный эксперимент показал, что метод позволяет эффективно решить поставленную модельную задачу. Это позволяет судить о перспективности применения методов машинного обучения к выбору вспомогательных оптимизируемых величин. Ниже перечислены возможные направления дальнейшей работы.

- Реализация затухающего в рамках фиксированного состояния Q -обучения с ε -жадным правилом, которое могло бы совместить в себе стабильность отложенного Q -обучения и быстроту реакции ε -жадного Q -обучения.
- Исследование целесообразности применения обучения с подкреплением, строящего модель среды. Этот подход позволяет более детально учитывать особенности среды, однако обычно характеризуется большими требованиями к вычислительной производительности [7].
- Применение обучения, ориентированного на меняющуюся среду [10].

- Применение предложенного подхода к решению практических задач, в частности, применение метода к генетическому выращиванию тестов, описанных в работе [4].

Литература

1. Лотов А.В., Поспелова И.И. Многокритериальные задачи принятия решений: Учебное пособие. – М.: МАКС Пресс, 2008. – 197 с.
2. Буздалов М.В. Генерация тестов для олимпиадных задач по теории графов с использованием эволюционных алгоритмов. Магистерская диссертация. СПбГУ ИТМО, 2011 [Электронный ресурс]. – Режим доступа: <http://is.ifmo.ru/papers/2011-master-buzdalov/>, св. Яз. рус. (дата обращения 25.11.2011).
3. Mitchell M. An Introduction to Genetic Algorithms. – MA: MIT Press, 1999. – 221 p.
4. Буздалов М.В. Генерация тестов для олимпиадных задач по программированию с использованием генетических алгоритмов // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 2 (72). – С. 72–77.
5. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. – MA: MIT Press, 1998. – 322 p.
6. Mitchell T.M. Machine Learning. – McGraw Hill, 1997. – 432 p.
7. Kaelbling L.P., Littman M.L., Moore A.W. Reinforcement Learning: A Survey // Journal of Artificial Intelligence Research. – 1996. – V. 4. – P. 237–285.
8. От моделей поведения к искусственному интеллекту / Под ред. В.Г. Редько. – 2-е. изд. – М.: КомКнига, 2010. – 456 с.
9. Strehl A.L., Li L., Wiewora E., Langford J., Littman M.L. PAC model-free reinforcement learning // Proceedings of the 23rd international conference on Machine learning. ICML'06. – 2006. – P. 881–888.
10. Da Silva B.C., Basso E.W., Bazzan A.L.C., Engel P.M. Dealing with non-stationary environments using context detection // Proceedings of the 23rd international conference on Machine learning. ICML'06. – 2006. – P. 217–224.

Афанасьева Арина Сергеевна

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, afanasyevrina@gmail.com

Буздалов Максим Викторович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, mbuzdalov@gmail.com

УДК 004.415.53

АВТОМАТИЧЕСКИЕ МЕТОДЫ МОДИФИКАЦИИ РЕШЕНИЙ ДЛЯ ТЕСТИРОВАНИЯ ПРОВЕРЯЮЩИХ ПРОГРАММ

А.А. Ахи, А.С. Станкевич, А.А. Шалыто

Предлагается метод автоматизированного тестирования проверяющих программ для олимпиадных задач по программированию, предназначенный для выявления ошибок в тестируемых программах. Предлагаемый метод основан на использовании модификации решений, проверяемых проверяющей программой. Описывается использование метода для тестирования олимпиадных задач полуфинала чемпионата мира ACM ICPC NEERC 2011 и соревнований Russian Code Cup 2011, при этом в ряде проверяющих программ были обнаружены ошибки.

Ключевые слова: олимпиады по программированию, олимпиадные задачи, проверяющие программы, тестирование.

Введение

Олимпиады по программированию проводятся в большом числе по всему миру. Они способствуют выявлению талантливых программистов среди школьников и студентов. Одной из наиболее значимых является международная студенческая олимпиада по программированию International Collegiate Programming Contest [1], которая проводится Association for Computing Machinery (далее олимпиада будет упоминаться как ACM ICPC). Данная олимпиада обладает обширной сетью отборочных соревнований, в число которых входят полуфинальные соревнования North East European Regional Contest [2] (NEERC), проходящие на территории России. Значимой олимпиадой на территории России также является соревнование Russian Code Cup [3], проводимое корпорацией Mail.ru Group.

На олимпиадах по программированию предлагается решить несколько задач. Формулировка задачи предполагает чтение входных данных, формат которых описан в условии задачи, и получение требуемых выходных данных, формат которых также описан в условии задачи. Решением задачи является программа, написанная на одном из языков программирования (например, в соревновании ACM ICPC используются языки C, C++ и Java [4]) и преобразующая входные данные в выходные.

Для проверки программа запускается на наборе заранее подготовленных тестов, не известных участникам. Решение считается прошедшим определенный тест, если оно, при работе с ним, не нарушило ограничений, указанных в условии задачи, завершилось корректно (без ошибок времени выполнения), а

его ответ признан правильным. Для проверки правильности выходных данных служат проверяющие программы (ПП), которым на вход подаются входные данные теста, выходные данные, полученные запуском решения участника, а также выходные данные, полученные запуском решения жюри олимпиады.

Соревнования на высоком уровне требуют высокого качества подготовленных задач. Для этого собираются интересные идеи для будущих задач, пишутся условия, решения и ПП, а также составляются тестовые наборы. В настоящей работе предлагается новый метод автоматизированного тестирования ПП, позволяющий выявлять ошибки и странности поведения тестируемых программ.

Описание предлагаемого подхода

В работе рассматривается автоматическое тестирование ПП, написанных на языке Java. Тестом для ПП является тройка <входные данные, выходные данные участника, выходные данные жюри>. В качестве входных данных в работе используется подготовленный жюри набор тестов. Выходные данные жюри получаются в результате запуска решения жюри на имеющемся наборе тестов. Предлагается метод автоматической генерации выходных данных участника.

Генерация выходных данных участника

Для генерации большого числа выходных данных участника в работе производится модификация имеющихся решений жюри на языке Java и последующий запуск модифицированных решений на подготовленных тестах. Для изменения решений жюри используется метод внесения ошибки, применяемый при мутационном тестировании программного обеспечения [5]. Ошибками могут являться использование неверной переменной или константы, отсутствие в коде одной из строк, использование неверного оператора и другие ошибки, часто допускаемые при написании программ. Более подробный список возможных изменений приведен в [6]. Модифицированные решения производят обширный и разнообразный набор выходных данных участника. Полученные данные используются для тестирования ПП.

Запуск модифицированных решений

Запуск модифицированных решений представляет собой отдельную проблему. Полученные программы могут работать продолжительное время, не завершаться или же иметь ошибки времени выполнения. В связи с этим измененные решения запускаются в отдельных виртуальных машинах Java Virtual Machine (JVM). При запуске новой машины указывается время, отведенное на исполнение одной программы. Это время немного превышает указанное в условии задачи ограничение на время работы решения на одном teste. По истечении отведенного времени JVM прекращает выполнение программы. Производится параллельный запуск модифицированного решения на всех тестах, что позволяет сократить временные затраты.

Анализ поведения ПП

Для проверки правильности исполнения ПП на полученном тестовом наборе производится анализ покрытия кода [7] и результата работы тестируемой программы.

Для анализа покрытия кода в настоящей работе используются две метрики: покрытия строк и покрытия решений. Первая проверяет, что каждая строка исходного кода была выполнена хотя бы раз, вторая – каждое условие в операторах ветвления хоть раз приняло как значение true, так и false. Для сбора статистических данных о покрытии в процессе проверки в исходный код ПП вносятся изменения, не меняющие поведения ПП и отвечающие лишь за сбор информации.

В случае отсутствия полного покрытия кода ПП в одной из метрик стоит уделить больше внимания непокрытым частям. Может оказаться, что данные части исходного кода тестируемой программы излишни или же условия в операторах ветвления составлены ошибочно. Исследования показали, что ряд строк ПП не будут покрыты никогда. Например, строки и условия, отвечающие за корректность выходных данных жюри, будут покрыты не полностью, так как предоставляемые ответы жюри всегда верны и корректны. Для покрытия этих строк производится запуск ПП с предоставленными во входной тройке выходными данными жюри и участника.

Результат работы ПП также представляет интерес. Возможны следующие результаты:

- OK – выходные данные участника правильны;
- WA (Wrong Answer) – выходные данные участника неправильны;
- PE (Presentation Error) – выходные данные участника некорректны с точки зрения формата, описанного в условии;
- FAIL – произошла ошибка выполнения ПП или была найдена некорректность во входных данных или выходных данных жюри.

Для ряда случаев некоторые результаты являются заведомо ошибочными. Прежде всего, результат проверки никогда не должен быть FAIL. При проверке решения жюри следует ожидать только результатов OK. При запуске ПП с измененными местами выходными данными участника и жюри нужно исключить

чить возможность результата РЕ, так как ответ жюри корректен с точки зрения условия. В ряде задач ПП производит проверку корректности ответов участника и жюри, после чего сравнивает их, определяя лучший. Для таких задач при «обратном» запуске ПП результат должен быть OK или FAIL. В случае обнаружения неверного результата тестирование прекращается. При этом пользователю становится доступна тройка <входные данные, выходные данные участника, выходные данные жюри>, на которой и наблюдается ошибочное поведение ПП.

Применение предлагаемого подхода

Для апробации описываемого подхода в условиях реальных олимпиадных задач были выбраны задачи полуфинальных соревнований ACM ICPC NEERC 2010, задачи соревнования Russian Code Cup 2011 и задача quest, использовавшаяся на петрозаводских сборах [8].

Результаты тестирования ПП полуфинала чемпионата мира по программированию NEERC 2010 (табл. 1) показывают, что ПП выполнены на качественном уровне. Разработанный метод достигает полного покрытия исходного кода ПП и не выявляет ошибок в процессе исполнения ПП, а также анализа результата их работы. Заметим, что для некоторых задач тестирование занимает значительное время. Это связано с большим временем исполнения правильного решения, которое составляет для этих задач порядка секунды, а также большим размером тестового набора, составляющего более 70 тестов.

Задача	Число строк в решении	Число единичных модификаций	Число компилирующихся решений	Число правильных решений	Время работы, мин	Число непокрытых строк проверяющей программы
alignment	79	1994	1143 (57%)	335 (17%)	9	0 (0) из 12 (4)
binary	208	5597	3349 (60%)	629 (11%)	8	0 (0) из 7 (2)
cactus	280	7811	4738 (61%)	1340 (17%)	300	0 (0) из 58 (19)
dome	191	3482	2237 (64%)	513 (15%)	75	0 (0) из 25 (7)
evacuation	133	6446	4009 (62%)	1750 (27%)	400	0 (0) из 36 (12)
factorial	156	5770	3623 (63%)	331 (6%)	400	0 (0) из 14 (4)
hands	308	7919	4200 (53%)	757 (10%)	145	0 (0) из 7 (2)
ideal	201	6502	3919 (60%)	2214 (34%)	700	0 (0) из 10 (3)
jungle	161	2991	1610 (54%)	357 (12%)	100	0 (0) из 7 (2)
kgraph	206	2941	1840 (63%)	556 (19%)	200	0 (0) из 27 (7)

Таблица 1. Результаты тестирования ПП соревнования NEERC 2010. В последнем столбце без скобок приведена метрика покрытия строк, в скобках – метрика покрытия решений

Выявленная проблема может иметь несколько решений. Первое состоит в улучшении используемого решения жюри с целью уменьшения времени его работы. Другим решением может являться исключение некоторых тестов из тестового набора задачи, так как обычно при подготовке задачи тестовый набор дополняется случайно сгенерированными тестами, не проверяющими каких-либо крайних случаев поставленной задачи. От части таких тестов часто можно отказаться.

Задача	Число строк в решении	Число единичных модификаций	Число компилирующихся решений	Число правильных решений	Время работы, мин	Число непокрытых строк проверяющей программы
guess	108	4100	2213 (54%)	926 (23%)	10	0 (0) из 53 (15)
birds	93	3791	2679 (71%)	903 (24%)	3	0 (0) из 40 (7)
queuesort	113	4518	2922 (65%)	1114 (25%)	5	0 (0) из 45 (15)
square	105	2192	1214 (55%)	316 (14%)	3	0 (0) из 37 (13)

Таблица 2. Результаты тестирования ПП соревнования Russian Code Cup 2011. В последнем столбце без скобок приведена метрика покрытия строк, в скобках – метрика покрытия решений

Как видно из табл. 2, при тестировании задач соревнования по программированию Russian Code Cup 2011 проблем со временем работы выявлено не было. Тем не менее, были выявлены ошибки в ПП к задачам guess и birds. Выявленные ошибки были устранены до проведения указанного соревнования. Заметим, что не все ошибки были найдены и устраниены с первого раза. Повторное тестирование выявляло новые ошибки в ПП задачи birds, которые также были устраниены.

Рассмотрим более подробно ошибку, выявленную в ПП задачи birds. Ошибка проявилась при подаче пустого выходного файла на вход ПП. Это приводило к ошибке исполнения из-за попытки преобразовать пустую строку в число. Данная ошибка была устранена с помощью окружения этой строки блоком

ком try-catch, сообщающим о некорректном формате выходного файла в случае возникновения ошибки. Исправленный код ПП приведен на сайте соревнования.

Интересными представляются результаты запуска разработанной программы для автоматизированного тестирования ПП задачи quest (табл. 3). В результате тестирования ПП в ней был выявлен ряд ошибок. Одна из ошибок возникала при генерации сообщения о неправильности ответа участника. Данная ошибка не была ранее обнаружена при использовании задачи на соревнованиях. Ошибка была успешно устранена, и ПП была перетестирована.

Задача	Число строк в решении	Число единичных модификаций	Число компилирующихся решений	Число правильных решений	Время работы, мин	Число непокрытых строк проверяющей программы
quest	337	9652	5911 (61%)	1410 (15%)	50	9 (15) из 280 (92)

Таблица 3. Результаты тестирования ПП задачи quest. В последнем столбце без скобок приведена метрика покрытия строк, в скобках – метрика покрытия решений

Не все строки ПП были покрыты. Интересно отметить, что первые три непокрытых строки соответствуют созданию сообщения об ошибке в случае некорректного входного файла. Так как разработанный метод не производит каких-либо изменений входных файлов, то данные строки не будут покрыты никогда. Проверка корректности входного файла и частей кода, ее проверяющих, остается на совести программиста, занимающегося подготовкой задачи к соревнованию. Остальные непокрытые строки соответствуют нарушению формата выходного файла. Такие строки наиболее сложны для покрытия, так как иногда требуют от программы вывода чрезвычайно странных выходных данных.

Также в ПП задачи quest наблюдается большое число непокрытых решений. Девять из них являются операторами if, находящимися непосредственно перед непокрытыми строками. Именно тот факт, что в процессе работы условия операторов всегда не выполнялись, и привел к неполноте покрытия строк исходного кода ПП. Оставшиеся непокрытые решения соответствуют имеющимся в коде конструкциям while-true, условие которых очевидным образом всегда выполняется.

Помимо перечисленных задач, на этапе разработки тестированию были подвергнуты ПП задач из цикла интернет-олимпиад сезона 2010–2011 г.г. [9]. Стоит отметить, что в одной из задач была выявлена часть кода, которая никогда не могла быть исполнена. Данная часть кода была устранена, тем самым был уменьшен объем ПП.

Выводы

В процессе тестирования было обнаружено, что наиболее сложными для покрытия строками исходного кода ПП являются строки, проверяющие некоторые специфические ограничения по формату или содержанию выходного файла решения участника, а также строки, соответствующие генерации сообщения в случае невыполнения этих ограничений. Например, блок кода, проверяющий, что выходной файл не содержит более никакой информации, чаще всего будет являться последней покрытой частью ПП. Стоит отметить, что необходимость именно в таком блоке чаще всего отсутствует, так как при запуске ПП данное условие проверяется системой автоматически.

Результаты показывают, что описанный в работе метод позволяет находить ошибки, не найденные даже во время использования задачи при проведении соревнований. Это свидетельствует об эффективности разработанного метода. Отметим, что тестирование ПП может требовать длительного времени, в связи с этим рекомендуется резервировать значительное время после подготовки олимпиады и до проведения соревнования для тестирования ПП и исправления выявленных ошибок.

Заключение

В работе описан метод, позволяющий автоматически тестировать проверяющие программы олимпиадных задач по программированию с помощью модификации правильных решений. Данный метод был применен для тестирования проверяющих программ реальных олимпиадных задач, где успешно выявил ряд ошибок, которые могли бы существенно повлиять на результаты соревнования. Полученные результаты позволяют утверждать, что описанный метод является достаточно перспективным в плане его применения при подготовке задач по олимпиадному программированию в целях повышения качества олимпиад.

Литература

1. ACM International Collegiate Programming Contest [Электронный ресурс]. – Режим доступа: http://en.wikipedia.org/wiki/ACM_ICPC, св. Яз. англ. (дата обращения 13.10.2011).

2. North East European Regional Contest [Электронный ресурс]. – Режим доступа: <http://neerc.ifmo.ru>, св. Яз. англ. (дата обращения 13.10.2011).
3. Russian Code Cup [Электронный ресурс]. – Режим доступа: <http://russiancodecup.ru>, св. Яз. англ. (дата обращения 13.10.2011).
4. Правила проведения полуфинала NEERC [Электронный ресурс]. – Режим доступа: <http://neerc.ifmo.ru/information/contest-rules.html>, св. Яз. рус., англ. (дата обращения 13.10.2011).
5. Мутационное тестирование [Электронный ресурс]. – Режим доступа: http://en.wikipedia.org/wiki/Mutation_testing, св. Яз. англ. (дата обращения 13.10.2011).
6. Ахи А.А. Автоматические методы модификации решений для тестирования проверяющих программ [Электронный ресурс]. – Режим доступа: <http://is.ifmo.ru/papers/2011-bachelor-akhi/>, св. Яз. рус. (дата обращения 13.10.2011).
7. Покрытие кода [Электронный ресурс]. – Режим доступа: http://en.wikipedia.org/wiki/Code_coverage, св. Яз. англ. (дата обращения 13.10.2011).
8. Сборы команд вузов-участников чемпионата мира по программированию в Петрозаводске [Электронный ресурс]. – Режим доступа: <http://karelia.snarknews.info/>, св. Яз. рус. (дата обращения 13.10.2011).
9. Интернет-олимпиады по информатике [Электронный ресурс]. – Режим доступа: <http://neerc.ifmo.ru/school/io/>, св. Яз. рус. (дата обращения 13.10.2011).

Ахи Антон Андреевич

- Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, anton.akhi@gmail.com

Станкевич Андрей Сергеевич

- Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доцент, stankev@rain.ifmo.ru

Шалыто Анатолий Абрамович

- Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, shalyto@mail.ifmo.ru

УДК 004.021

ПРИМЕНЕНИЕ ДЕРЕВЬЕВ ДЛЯ РЕАЛИЗАЦИИ МАССОВЫХ ОПЕРАЦИЙ НА МНОГОМЕРНЫХ МАССИВАХ ДАННЫХ

А.Г. Банных

Предлагается метод построения структур данных для выполнения массовых операций на многомерных структурах данных для узкого класса задач. Предлагаемый метод применим в том случае, если элементы многомерной структуры данных принадлежат абелевой группе, и позволяет эффективно выполнять вычисление суммы и прибавление значения к многомерной области.

Ключевые слова: структуры данных, массовые операции, многомерные массивы данных, дерево отрезков, дерево Фенвика.

Введение

Для эффективной работы с информацией были разработаны разнообразные структуры данных. Одна из самых распространенных из них – дерево. Существует огромное множество различных типов деревьев. Эти структуры позволяют собирать статистику на отрезках и изменять отдельные элементы. Представление об этих структурах данных можно получить в классических трудах [1–4].

Практически все эти деревья предназначены для данных, на которых можно ввести линейный порядок. В данной работе рассматриваются многомерные массивы данных. Введение нескольких измерений может быть как естественным (растровые изображения и видео), так и искусственным (базы данных). В последнем случае дополнительные измерения позволяют точнее указывать интересующую выборку информации. Широко известны квадродеревья [5], кд-деревья [6], обобщения дерева отрезков [7] и дерева Фенвика [8]. Существующие структуры данных для работы с многомерными структурами данных позволяют эффективно получать статистическую информацию и изменять отдельные элементы.

Цель настоящей работы заключается в том, чтобы исследовать возможность выполнения массовых обновлений – единообразного изменения целых областей данных. Ни одна вышеперечисленная структура данных в чистом виде не предоставляет подобную функциональность. Приведем простой пример. Пусть дан двумерный массив целых чисел размера $n \times n$. Требуется эффективно выполнять две операции – вычислять сумму на прямоугольнике и прибавлять число к прямоугольнику. Оказывается, что добиться асимптотики $O(\log^2 n)$ для выполнения этих операций – вовсе не тривиальная задача, как может показаться на первый взгляд. Подробнее о возникающих проблемах можно узнать в работе [9].

Постановка задачи

Рассмотрим d -мерное векторное пространство $U_d = \mathbb{Z}^d$. Элементы этого пространства будем обозначать жирным шрифтом. Компоненты векторов будем обозначать нижними индексами. Например, $\mathbf{x} = (x_1, x_2, \dots, x_d)$. Введем частичный порядок на элементах этого пространства. Скажем, что $\mathbf{x} \leq \mathbf{y} \Leftrightarrow \forall k \in \{1, 2..d\} x_k \leq y_k$. Кроме того, будем считать, что $\mathbf{1}_d = (1, 1, \dots, 1) \in U_d$.

Для любых $x, y \in U_d$ $x \leq y$ областью будем называть множество $P_{x,y} = [x_1..y_1] \times [x_2..y_2] \times \dots \times [x_d..y_d]$. Для повышения читаемости формул будем опускать нижние индексы в тех случаях, когда это не приводит к неоднозначностям. Будем считать, что мы работаем с областью $W = P_{\mathbf{1}, \mathbf{N}}$, где $\mathbf{N} \in U_d, \mathbf{1} \leq \mathbf{N}$. Эта область – не что иное, как множество индексов некоторого d -мерного массива.

Сопоставим индексам значения. Рассмотрим множество G и два бинарных оператора: $+$ и \circ ($(+, \circ : G \times G \rightarrow G)$). Далее оператор $+$ будем называть сложением, а \circ – умножением. Сложение используется для выполнения массовых запросов, таких как суммирование на прямоугольнике. Умножение используется при выполнении массовых обновлений, таких как умножение всех элементов подпрямоугольника на число.

Рассмотрим функцию $A : W \rightarrow G$. Можно рассматривать A как многомерный массив с множеством индексов W и множеством допустимых значений ячеек G .

Массовый запрос определим следующим образом:

$$get(A, P) = \sum_{x \in P} A(x).$$

Массовое обновление определим следующим образом:

$$update(A, P, v) = A' : W \rightarrow G,$$

где

$$A'(x) = \begin{cases} A(x) \circ v & x \in P \\ A(x) & x \notin P \end{cases}$$

Задача состоит в том, чтобы реализовать структуру данных, которая позволила бы эффективно выполнять эти массовые операции.

Чтобы массовые операции имели смысл, на них нужно наложить некоторые ограничения. Обычно полагают, что $\langle G, + \rangle$ – коммутативная полугруппа, а $\langle G, \circ \rangle$ – полугруппа. Кроме того, потребуем выполнения дистрибутивного закона умножения относительно сложения.

Подобная модель описывает не все возможные массовые запросы. Тем не менее, она проста, многие естественные запросы легко описываются в терминах этой модели, и на операции наложены требования, которые учитывают особенности реализации массовых операций в одномерном случае. Это позволяет надеяться на то, что задача имеет эффективное решение.

Настоящая работа посвящена узкому случаю: $+ \equiv \circ$, $\langle G, + \rangle$ – абелева группа. Стоит подчеркнуть, что у всех элементов группы существует обратный, т.е. $\forall a \in G \exists a^{-1} \in G : a + a^{-1} = 0$. Это свойство лежит в основе всех дальнейших рассуждений.

Основные утверждения

Чтобы упростить дальнейшее изложение, введем несколько операторов. Под ними скрываются преобразования массивов, которые напоминают дифференцирование и интегрирование функций. Для случая одномерной последовательности подобные преобразования хорошо описаны в работе [10].

Префиксом будем называть область вида $P_{\mathbf{1}, \mathbf{R}}$, $\mathbf{R} \in W$. Сuffixком будем называть область вида $P_{\mathbf{L}, \mathbf{N}}$, $\mathbf{L} \in W$. Обозначим замену i -й компоненты вектора следующим образом: $\mathbf{x}_{x_i \leftarrow x'_i}$.

Определение 1. Дифференциальный оператор по i -му индексу

$$D_i A(\mathbf{x}) = A(\mathbf{x}) - A(\mathbf{x}_{x_i \leftarrow x_i - 1}).$$

Определение 2. Интегральный оператор по i -му индексу

$$S_i A(\mathbf{x}) = \sum_{j=1}^{x_i} A(\mathbf{x}_{x_i \leftarrow j}).$$

Приведем ключевые свойства этих операторов. Для экономии места доказательства тривиальных фактов опущены. Ознакомиться с ними можно в работе [10].

Утверждение 1. Операторы D_a и D_b коммутируют.

Утверждение 2. Операторы S_a и S_b коммутируют.

Утверждение 3. Операторы S_i и D_i взаимообратны. $S_i D_i = D_i S_i = E$, где E – тождественный оператор.

Утверждение 4. Операторы S_a и D_b коммутируют.

Определение 3. Дифференциальный оператор $D = D_1 D_2 \cdots D_d$.

Определение 4. Интегральный оператор $S = S_1 S_2 \cdots S_d$.

Введение этих операторов позволяет вплотную подобраться к основному результату.

Лемма 1. $SD = DS = E$.

Лемма 2. $SA(\mathbf{x}) = \text{get}(A, P_{1,\mathbf{x}})$.

Доказательство.

$$SA(\mathbf{x}) = S_1 S_2 \cdots S_d A(\mathbf{x}) = \sum_{j_1=1}^{x_1} \sum_{j_2=1}^{x_2} \cdots \sum_{j_d=1}^{x_d} A(\mathbf{x}_{x_1 \leftarrow j_1, x_2 \leftarrow j_2, \dots, x_d \leftarrow j_d}) = \sum_{\mathbf{j} \leq \mathbf{x}} A(\mathbf{j}) = \text{get}(A, P_{1,\mathbf{x}}).$$

Таким образом, запись $SA(\mathbf{x})$ можно воспринимать как обычный запрос на сумму на префикссе.

Обратимся теперь к оператору D . Это вполне определенное преобразование, которое можно совершить с любым массивом. В работе массив $DA = B$ будет храниться в явном виде. Сформулируем некоторые его свойства.

Лемма 3. Сумма на префикссе в преобразованном оператором D массиве данных A есть элемент исходного массива.

Доказательство. Используя лемму 1, получаем:

$$B = DA \Rightarrow SB = SDA = EA = A \Rightarrow \forall \mathbf{x} SB(\mathbf{x}) = A(\mathbf{x}).$$

Полученный результат показывает связь между преобразованным массивом (B) и тем, что был изначально (A). По сути, эта лемма доказывает эквивалентность этих двух массивов, что позволяет хранить лишь один из них. Осталось понять, чем же новый массив лучше старого.

Лемма 4. Изменение одного элемента B влечет изменения целого суффикса массива $A = SB$.

Доказательство. Действительно, выберем произвольные $\mathbf{x} \in W$ и $v \in G$ и предположим, что

$$B'(\mathbf{y}) = \begin{cases} B(\mathbf{y}) + v & \mathbf{y} = \mathbf{x} \\ B(\mathbf{y}) & \mathbf{y} \neq \mathbf{x} \end{cases}.$$

$$\text{Положим } A = SB \text{ и } A' = SB'. \text{ Тогда } A'(\mathbf{y}) = \begin{cases} A(\mathbf{y}) + v & \mathbf{x} \leq \mathbf{y} \\ A(\mathbf{y}) & \mathbf{x} \not\leq \mathbf{y} \end{cases}.$$

Преобразование данных и запросов

Построим массив $B = DA$. Один из самых простых способов сделать это – используя определение 3, по очереди применить дифференциальные операторы к исходному массиву.

Преобразование запроса на сумму. Действие оператора S тесно связано с запросом на сумму на префикссе. Например, по леммам 2 и 3 $\text{get}(A, P_{1,\mathbf{x}}) = SA(\mathbf{x}) = SSB(\mathbf{x})$.

Последнее можно записать и иначе:

$$\text{get}(A, P_{1,\mathbf{x}}) = SSB(\mathbf{x}) = \sum_{1 \leq \mathbf{y} \leq \mathbf{x}} \sum_{1 \leq \mathbf{z} \leq \mathbf{y}} B(\mathbf{z}).$$

Значение этой записи интуитивно понятно: сумма превратилась в «сумму сумм».

Преобразование запроса на прибавление. Как было показано в лемме 4, изменение одной ячейки в B эквивалентно массивовому прибавлению к суффиксу массива A . Таким образом, массивовое обновление сводится к изменению единственной ячейки массива B .

Эквивалентность. На данный момент будет не совсем корректно говорить о решении задачи. Действительно, было показано, чему соответствуют запросы на префиксах или суффиксах, однако исходные запросы формулируются в терминах произвольных областей. В дальнейшем запросы на префикссе для суммы и на суффиксе для обновления будем называть ослабленными. Покажем, что любой запрос можно выразить не более чем через 2^d ослабленных.

Чтобы решить эту проблему, нужно воспользоваться формулой включения-исключения. Предположим, необходимо ответить на запрос $\text{get}(A, P_{\mathbf{x}, \mathbf{y}})$. Предположим, что $x_1 \neq 1$. Тогда выразим исходный запрос следующим образом:

$$get(A, P_{x,y}) = get\left(A, P_{x_{x_1 \leftarrow -1}, y}\right) - get\left(A, P_{x_{x_1 \leftarrow -1}, y_{y_1 \leftarrow -1}}\right).$$

Повторяя подобное преобразование для остальных индексов, выразим исходный запрос только через запросы на префиксах. Для этого потребуется не более d шагов, потому итоговое число запросов не превзойдет 2^d . Аналогично можно выполнить массовые обновления. Здесь снова используется существование обратных элементов, что позволяет одним массовым обновлениям компенсировать другие.

Эффективность. Поскольку большинство алгоритмов работы с многомерными массивами данных содержат $O(\log^d N)$ в своей асимптотике, то умножение на 2^d не изменит асимптотическую оценку.

Решение задачи

Сосредоточимся на выполнении запроса на «сумму сумм». В дальнейшем будет удобно воспользоваться нотацией, предложенной в работе [11]:

$$[\text{утверждение}] = \begin{cases} 1, & \text{если утверждение истинно;} \\ 0, & \text{в противном случае.} \end{cases}$$

Лемма 5. Для любого $\mathbf{x} \leq \mathbf{z}$ значение $B(\mathbf{x})$ входит в $SSB(\mathbf{z})$ ровно $\prod_{i=1}^d (z_i - x_i)$ раз.

Доказательство. Действительно $SSB(\mathbf{z}) = \sum_{1 \leq y \leq z} \sum_{1 \leq y' \leq y} B(y')$. Количество вхождений значения $B(\mathbf{x})$ в

эту сумму равно $\sum_{1 \leq y \leq z} \sum_{1 \leq y' \leq y} [\mathbf{y}' = \mathbf{x}]$.

$$\sum_{1 \leq y' \leq y} [\mathbf{y}' = \mathbf{x}] = \begin{cases} 1 & \mathbf{x} \leq \mathbf{y} \\ 0 & \mathbf{x} \not\leq \mathbf{y} \end{cases} = [\mathbf{x} \leq \mathbf{y}] \Rightarrow \sum_{1 \leq y \leq z} \sum_{1 \leq y' \leq y} [\mathbf{y}' = \mathbf{x}] = \sum_{1 \leq y \leq z} [\mathbf{x} \leq \mathbf{y}] = \sum_{\mathbf{x} \leq \mathbf{z}} 1 = \prod_{i=1}^d (z_i - x_i).$$

Будем в каждой ячейке хранить не просто значение B_x , а многочлен $B_x \prod_{i=1}^d (z_i - x_i)$, где z_i заранее не известны. Отметим, что после раскрытия всех скобок получится 2^d слагаемых.

Необходимо пояснить, что подразумевается под умножением элемента абелевой группы $\langle G, + \rangle$ на число. Можно было бы решить, что мы имеем дело с кольцом, но это не так. Подчеркнем, что все коэффициенты целые, поэтому умножение $g \in G$ на число $a \geq 0$ есть всего лишь $\sum_{i=1}^a g$. Аналогично определяется эта операция для случая $a < 0$. В итоге приходим к тому, что в каждой ячейке хранится 2^d элементов G . Поскольку это коэффициенты многочлена, то их можно складывать как векторы. Массив, содержащий эти многочлены, обозначим как C .

Рассмотрим $\sum_{1 \leq x \leq z} C_x$. Получим многочлен, представляющий собой число вхождений соответствующих элементов в запрос на сумму $sum(A, P_{1,z})$ (по леммам 2, 3, 5). Это означает, что

$$sum(A, P_{1,z}) = \left[\sum_{1 \leq x \leq z} C_x \right] (\mathbf{z}).$$

Чтобы ответить на запрос на сумму на префикссе, нужно вычислить сумму на префикссе в массиве C , после чего в получившийся многочлен подставить параметры запроса. При выполнении ослабленного запроса на массовое обновление изменяется одна ячейка, т.е. прибавление на суффиксе было сведено к прибавлению нового многочлена к одной ячейке. Таким образом, для решения финальной задачи достаточно поддержки двух запросов: сумма на префикссе и изменение одного элемента. Для решения этой задачи существует множество структур данных, позволяющих выполнять указанные запросы за $O(\log^d N)$.

Оценка потребляемых ресурсов

Время работы. Характерное время обработки запросов популярными структурами данных составляет $O(\log^d N)$. Сведение исходных запросов к ослабленным приводит к не более чем 2^d запросам к структуре. Таким образом, асимптотическая оценка времени обработки одного запроса составляет $O(2^d \log^d N) = O(\log^d N)$ при фиксированной размерности пространства d .

Для завершения оценки времени работы осталось определить, какое время занимают элементарные операции над многочленами. Будем считать операцию $+$ в G элементарной, т.е. на ее выполнение уходит $O(1)$ времени. Тогда после преобразования каждый элемент заменяется на многочлен сте-

пени 2^d . Значит, сложение двух многочленов выполняется за $O(2^d)$. Таким образом, асимптотической оценкой времени выполнения запроса осталось $O(\log^d N)$.

Обратимся теперь к массовым обновлениям. По аналогии с предыдущими рассуждениями можно считать, что ослабление запросов не ведет к ухудшению асимптотики. Единственное, с чем осталось разобраться, – построение нового многочлена.

Ключевой вопрос – за какое время можно выполнять умножение на целое число. Если мы имеем дело с примитивными типами данных, то эту операцию можно выполнить за $O(1)$ встроенными командами процессора. В иных случаях можно представить умножение на число через $O(\log(L))$ операций $+$, где L – абсолютное значение множителя [12]. Коэффициенты многочлена не превышают N^d , поэтому общая оценка времени, затраченного на построение нового многочлена, есть $O(2^d \log(N^d)) = O(\log^d N)$.

Память. Потребление памяти зависит от используемой структуры данных, поэтому сложно привести какие-либо конкретные оценки. Тем не менее, можно уверенно утверждать, что после всех преобразований потребление памяти возрастет не менее чем в 2^d раз, что связано с хранением многочленов. Кроме того, в случае разреженных данных количество ненулевых ячеек может также возрасти, но не более чем в 2^d раз.

Для большинства структур данных применение разработанного метода не ухудшает асимптотическую оценку времени их работы. Тем не менее, при возрастании размерности массива данных константа, скрытая в асимптотической оценке, возрастает экспоненциально.

Заключение

В работе был выбран частный случай массовых операций: $+ \equiv \circ$, $\langle G, + \rangle$ – абелева группа. Для него был разработан общий метод сведения задачи с массовыми обновлениями к хорошо изученной задаче без них. Разработанный метод можно использовать с любыми существующими структурами данных. Это свойство позволяет делать выбор оптимальной структуры данных для каждой задачи отдельно. Например, если используются все ячейки массива, то стоит сделать выбор в пользу дерева Фенвика. С другой стороны, в случае сильно разреженных данных можно использовать разреженное дерево отрезков. Подобная гибкость позволяет применять метод в различных областях.

Таким образом, метод получился универсальным настолько, насколько это возможно для данной задачи. К сожалению, вопрос об обобщении метода на более широкие классы задач остается открытым.

Литература

1. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. – 2-е изд. – М.: МЦНМО, 2010. – 1296 с.
2. Шень А. Программирование: теоремы и задачи. – 2-е изд. – М.: МЦНМО, 2004. – 296 с.
3. Tarjan R.E. Data Structures and Network Algorithms. – PA.: Society for Industrial Mathematics, 1983. – 140 p.
4. Кнут Д.Э. Искусство программирования. Т. 3. Сортировка и поиск. – М.: Вильямс, 2007. – 824 с.
5. Романовский И.В. Дискретный анализ. – 4-е изд. – СПб: Невский Диалект; БХВ-Петербург, 2008. – 336 с.
6. Bentley J.L. Multidimensional binary search trees used for associative searching // Communications of the ACM. – 1975. – V. 18. – № 9. – P. 509–517.
7. Препарата Ф., Шеймос М. Вычислительная геометрия: Введение. – М.: Мир, 1989. – 478 с.
8. Fenwick P.M. A New Data Structure for Cumulative Frequency Tables // Software: Practice and experience. – 1994. – V. 24 – № 3. – P. 327–336.
9. Кнут Д., Грэхем Р., Паташник О. Конкретная математика. Основание информатики. – М.: Вильямс, 2005. – 784 с.
10. Банных А.Г. Применение деревьев для реализации массовых операций на многомерных массивах данных [Электронный ресурс]. – Режим доступа: <http://is.ifmo.ru/papers/2011-bachelor-bannykh/>, св. Яз. рус. (дата обращения 25.11.2011).
11. Кнут Д.Э. Искусство программирования. Т. 1. Основные алгоритмы. – М.: Вильямс, 2007. – 720 с.
12. Brauer A. On addition chains // Bulletin of the American Mathematical Society. – 1939. – V. 45. – № 10. – P. 736–739.

Банных Антон Геннадьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, anton.bannykh@gmail.com

ИСПОЛЬЗОВАНИЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ ПОИСКА ОПТИМАЛЬНОЙ ТРАЕКТОРИИ НАБЛЮДАТЕЛЯ

Д.В. Степанов, А.А. Шалыто

Предложена модификация генетического алгоритма, предназначенная для поиска оптимальной траектории наблюдателя в задаче оценивания параметров движущейся цели по угловым измерениям (Bearing-Only Target Motion Analysis, BOTMA). Возникающие проблемы сходимости и попадания в локальные экстремумы, характерные для генетических алгоритмов, решены с помощью идей, заложенных в оптимизационные методы случайного восхождения и имитации отжига. Для ускорения работы алгоритма отбор особей производился на уровне хорошо приспособленных схем.

Ключевые слова: генетические алгоритмы, метод случайного восхождения, метод имитации отжига, теорема схем Холланда, гипотеза строительных блоков Гольдберга, расстояние Хэмминга, задача BOTMA.

Введение

В задаче определения параметров движения цели по угловым измерениям BOTMA одной из наиболее важных проблем является выбор движения наблюдателя, приводящего к получению качественных оценок [1, 2]. Параметры движения цели оцениваются с помощью линейных регрессионных моделей. Качество получаемых оценок характеризуется их ковариационной матрицей, выражающейся через матрицу плана регрессионной модели. Элементы матрицы плана, в свою очередь, зависят от параметров движения наблюдателя. Задача планирования эксперимента в случае регрессионной модели может быть сформулирована как задача выбора матрицы плана, доставляющей оценки с минимальной дисперсией. Таким образом, задачу планирования эксперимента в рассматриваемом случае можно представить в виде оптимизационной задачи поиска варианта движения наблюдателя, при котором минимизируется ковариационная матрица регрессионных оценок [3].

Будем рассматривать вариант кусочно-линейного движения наблюдателя с постоянной скоростью. При этом траектория наблюдателя может быть представлена в виде последовательности углов его поворота. Подобную оптимизационную задачу имеет смысл решать с помощью генетического алгоритма (ГА) [3], однако в случае большого числа «изломов» (галсов) траектории проявляется одна из частых проблем ГА, состоящая в попадании решения в локальные экстремумы.

Как показал анализ траекторий [3], в оптимальных многогалсовых траекториях часть галсов «сливается» вместе, и фактическое их число оказывается значительно меньше первоначально заданного. Объяснением данному факту является то, что, помимо осуществления маневров, для получения хороших оценок наблюдатель должен также производить измерения из точек, как можно более далеко разнесенных друг от друга. Траектория, удовлетворяющая данному требованию, не может быть сильно изломана. Оптимальные траектории, построенные из большого числа частично «сливающихся» галсов, имеют одно важное свойство – они могут содержать галсы различной длительности, что в ряде случаев приводит к улучшению оценок.

В настоящей работе предлагается модификация ГА, позволяющая находить оптимальные решения в задаче поиска траекторий наблюдателя с различными по длительности галсами.

Планирование эксперимента в задаче BOTMA

В активном эксперименте, каким является задача BOTMA, матрица плана регрессии выбирается так, чтобы повысить качество оценок. Ковариационная матрица управляемася параметрами движения наблюдателя. Элементы ковариационной матрицы тем меньше, чем «больше» информационная матрица. В качестве характеристики матрицы будем рассматривать определитель матрицы.

Обозначим через δ функцию от определителя информационной матрицы. В рассматриваемой постановке задачи в качестве управляющих параметров ограничимся только выбором углов поворота наблюдателя $C_j \in \Omega_C = [\alpha_{\min}, \alpha_{\max}], 0^\circ \leq \alpha_{\min} < \alpha_{\max} \leq 360^\circ$, на заданном числе N_{LEG} галсов. При отсутствии ограничений на угол поворота на галсе естественно выбрать для рассмотрения $\alpha_{\min} = 0^\circ, \alpha_{\max} = 360^\circ$. Формально оптимизационная задача состоит в поиске

$$\arg \max_{(C_1, \dots, C_{N_{LEG}}) \in \Omega_C^{N_{LEG}}} (\delta). \quad (1)$$

Пространство $\Omega_C^{N_{LEG}}$ представляет собой пространство траекторий. Выбор возможных C_j ограничим дискретным множеством, состоящим из точек вида $C_j = j \cdot \Delta C$, где ΔC – шаг дискретизации. Число различных вариантов углов поворота наблюдателя определяется как $N_{ALPH} = \frac{\alpha_{\max} - \alpha_{\min}}{\Delta C}$. Обозначим число галсов N_{LEG} . Тогда общее число траекторий, состоящих из N_{LEG} галсов с N_{ALPH} вариантами возможных углов поворота, равно $N_{ALPH}^{N_{LEG}}$. Каждой траектории сопоставим угол, рассчитываемый по формуле

$$\varphi(C_0, \dots, C_{N_{LEG}-1}) = C_0 + \sum_{i=1}^{N_{LEG}} C_i \cdot N_{ALPH}^{-i} + \frac{\Delta}{2} N_{ALPH}^{-N_{LEG}}.$$

Интерпретируя δ как функцию от φ , изобразим график функции δ (рис. 1).

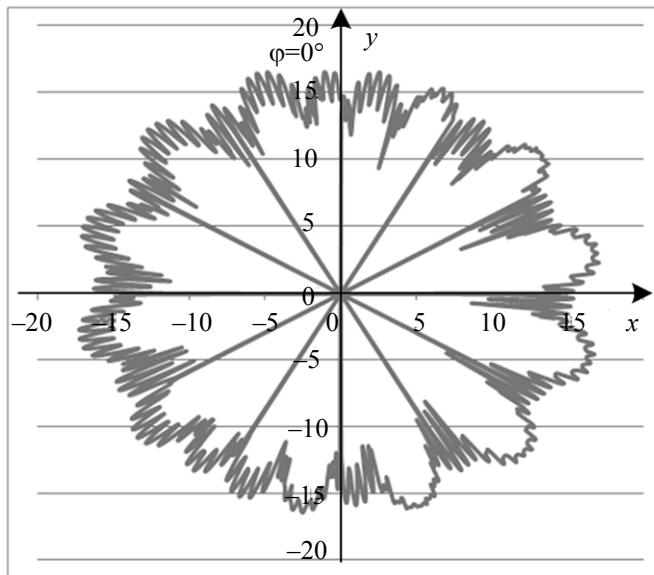


Рис. 1. Способ изображения функции δ при шаге дискретизации 30°

Сформулируем (1) как задачу поиска угла φ , в направлении которого функция δ наиболее сильно удаляется от нуля. Как видно из рис. 1, график функции δ имеет N_{ALPH} корней. Они соответствуют прямолинейным траекториям наблюдателя, на которых определитель информационной матрицы обнуляется. Корни делят весь график функции δ на N_{ALPH} сильно изрезанных «лепестков». Каждый из них отвечает выбору первого угла поворота траектории наблюдателя. Лепестки обладают множеством локальных максимумов, многие из которых имеют близкие значения. Рис. 1 дает представление о том, насколько сложной оказывается задача оптимизации (1).

Генетические алгоритмы и гипотеза строительных блоков

ГА представляет собой оптимизационный алгоритм, имитирующий процесс биологической эволюции [4]. Основной теоремой теории ГА, обосновывающей эффективность алгоритма, является теорема схем (шаблонов) Дж. Холланда, доказанная в 1975 г. [5, 6].

Схемой (шаблоном) H называется подмножество множества генотипов, допустимое в данной популяции, заданное в виде хромосомы с зафиксированными значениями некоторых генов. Число «зафиксированных» генов схемы называется порядком, а расстояние между крайними зафиксированными генами – определяющей длиной. Функцией приспособленности $F(H)$ схемы называется среднее значение функций приспособленности всех ее генотипов [5, 6]. Схемы с функцией приспособленности выше среднего по популяции, малым порядком и малой определяющей длиной принято называть строительными блоками.

Из теоремы схем следует, что шанс увеличить свое представительство в популяции следующего поколения имеется у схем малого порядка, малой определяющей длины и высокой приспособленности – у строительных блоков. В 1989 г. Д. Гольдбергом [6] была высказана гипотеза строительных блоков, состоящая в том, что ГА ведет поиск решения на множестве строительных блоков [7].

С практической точки зрения гипотеза Д. Гольдберга состоит в том, что наибольшая эффективность от применения ГА ожидается в тех задачах, в которых возможно выделение строительных блоков.

В задаче поиска оптимальной траектории (1) хромосомой длины $N_{VAR} = N_{LEG}$ является последовательность углов поворота наблюдателя в моменты смены галса, например, при $N_{VAR} = 6$ хромосома может принять вид $C = (C_0, C_1, C_2, C_3, C_4, C_5) = (270^\circ, 0^\circ, 180^\circ, 0^\circ, 0^\circ, 210^\circ)$. Одна из возможных схем H , которой принадлежит приведенная выше хромосома, будет задаваться выражением $H = (*, 0, *, 0, 0, *)$. Порядок приведенной схемы равен трем, а определяющая длина схемы также равна трем: $5 - 2 = 3$.

В следующем разделе приводится модификация ГА, обеспечивающая решение оптимизационной задачи (1), использующая идею гипотезы строительных блоков.

Генетический алгоритм с эмиссией инtronов

Сформулируем основные отличия предлагаемого метода от классической реализации ГА.

- Интенсивность применения генетических операторов привязывается к мере разнообразия популяции, меняющейся от поколения к поколению (аналогия с методом имитации отжига).
- Вероятность сохранения изменений, возникающих при применении генетических операторов, ставится в зависимость от наблюдаемого изменения функции приспособленности (аналогия с методом случайного восхождения).
- Выделяется ген (инtron), высокое содержание которого характеризует приспособленные хромосомы. На начальном этапе работы алгоритма осуществляется масштабное искусственное добавление данного гена в хромосомы.
- Производится поиск хорошо приспособленных схем (в согласии с гипотезой Д. Гольдберга) путем анализа расположения инtronов в хромосомах – выделения инtronных островов и их дополнений – экзонных островов.
- Вводится ряд новых операторов на экзонных островах, реализующих перебор траекторий-кандидатов.

Для измерения степени разнообразия популяции будем использовать расстояние Хэмминга $d_H(A, B)$, определяемое как число позиций, в которых соответствующие символы двух строк A и B одинаковой длины различны. Обозначим через $d_H(\Omega)$ среднее расстояние Хэмминга по популяции Ω . Используя идею метода имитации отжига [8], интенсивность применения генетических операторов поставим в зависимость от показателя $d_H(\Omega)$.

Предполагая, что выпрямленные траектории будут иметь большую приспособленность, чем сильно изломанные, высажем гипотезу о существовании строительных блоков (или хорошо приспособленных схем), состоящих из последовательностей нулей. При таком рассмотрении нас будет интересовать разделение генов на нулевые (инtronы) и все остальные (экзоны).

Сделаем необходимые замечания относительно терминов «инtron» и «экзон». Термин «инtron» заимствован из биологии, где он обозначает некодирующий участок хромосомы, задающий пространственную структуру хромосомы. Термин «экзон» в биологии обозначает кодирующий участок хромосомы. Термин «инtron» используется также в области, родственной ГА – в генетическом программировании, где он обозначает неэффективный фрагмент кода, изменения в котором не влекут изменения в работе программы [9].

Оператор эмиссии инtronов Ei создадим по аналогии с оператором мутации. Каждый ген C_k хромосомы C может муттировать в инtron (соответствующую новую хромосому обозначим $C^{0,k}$) с вероятностью p_i , зависящей от среднего расстояния Хэмминга по популяции и изменения функции приспособленности. Оператор эмиссии инtronов может быть записан в виде

$$Ei(t+1, C_k \in C) = \begin{cases} 0, p_i(C_k, 0) \\ C_k, q_i(C_k, 0), \end{cases} \quad p_i + q_i = 1,$$

где $q_i = 1 - p_i$ – вероятность того, что данный ген не будет изменен оператором Ei :

$$q_i(C_k, 0) = \left(1 - \frac{d_H(\Omega)}{N_{VAR}}\right) \cdot I_{\{F(C) > F(C^{0,k})\}}.$$

Обозначим через $M_{X,\Omega}$ популяционную матрицу, строками которой будут хромосомы особей, прошедших отбор. Если через N_X обозначить число особей, проходящих отбор, то размерность матрицы $M_{X,\Omega}$ записывается как $N_X \times N_{VAR}$. Далее, матрице $M_{X,\Omega}$ сопоставим матрицу $M_{X,\Omega}^{(i,e)}$, элементами которой являются индикаторы принадлежности к группе инtronов «0» или экзонов «1». Отбор строительных блоков будем производить с помощью матрицы $M_{X,\Omega}^{(i,e)}$. Введем еще одно определение: назовем локусом $loci(j, \Omega)$ позицию гена в хромосоме. Задача отбора блоков состоит в определении локусов, в которых должны быть расположены инtronы. Такие локусы будем называть далее инtronными. Рассмотрим j -й столбец матрицы $M_{X,\Omega}^{(i,e)}$. Он содержит все варианты заполнения локуса $loci(j, \Omega)$ в текущей популяции Ω . Введем в рассмотрение среднее расстояние Хэмминга в локусе j популяции Ω , характеризующее степень определенности относительно того, инtron или экзон должен располагаться в данном локусе:

$$d_H(j, \Omega) = \frac{1}{N_X(N_X - 1)} \sum_{A \neq B \in loci(j, \Omega)} d_H(A, B).$$

Максимальное значение $d_H(j, \Omega)$ равно $N_X/2$ и соответствует ситуации, при которой в данном локусе находится поровну инtronов и экзонов («0» и «1»). Минимальное значение $d_H(j, \Omega)$ равно нулю и отвечает ситуации, в которой в данном локусе находятся либо только инtronы, либо только экзоны. Введем в рассмотрение расстояние $d_{H,t+1}(j, \Omega)$, учитывающее информацию о среднем расстоянии Хэмминга в локусе, содержащуюся в предшествующих поколениях:

$$d_{H,t+1}(j, \Omega) = \frac{t \cdot N_X(N_X - 1) \cdot d_{H,t}(j, \Omega) + \sum_{A \neq B \in loci(j, \Omega)} d_H(A, B)}{(t+1) \cdot N_X(N_X - 1)}.$$

Используем аналогию, подсказанную моделями статистической физики: будем интерпретировать $d_{H,t+1}(j, \Omega)$ как энергию системы частиц, а $d_H(\Omega)$ – как температуру. Обратим внимание на то, что в рассматриваемой системе с ростом времени (числа поколений) температура $d_H(\Omega)$ (при правильно работающем алгоритме) будет снижаться. Наиболее устойчивое состояние системы характеризуется минимумом энергии, что в данном случае соответствует ситуации определенности в распределении инtronов и экзонов в локусах. Дискретизированная энергия системы $E_z = \lfloor d_{H,t+1}(j, \Omega) \rfloor$ (где $\lfloor \rfloor$ – целая часть числа) в каждом локусе может принимать значения из отрезка $[0; N_X/2]$. Обозначим через N_z число частиц, находящихся на энергетическом уровне z , тогда значения N_z/N_X описываются распределением Максвелла-Больцмана:

$$p_{MB}(z) = \frac{N_z}{N_X} = \frac{g_z \exp(-E_z/d_H(\Omega))}{\sum_w g_w \exp(-E_w/d_H(\Omega))}.$$

Будем называть энергетически неустойчивым значение элемента матрицы $M_{X,\Omega}^{(i,e)}(k, j)$, соответствующее локусу $loci(j, \Omega)$ в хромосоме с номером $k \in [1, N_X]$, если данное значение относится к менее многочисленному множеству при разбиении на инtronы-экзоны. Например, если в локусе $loci(j, \Omega)$ большинство составляют инtronы, а $M_{X,\Omega}^{(i,e)}(k, j)$ является экзоном или наоборот, то $M_{X,\Omega}^{(i,e)}(k, j)$ является энергетически неустойчивым значением. Рассматривая последовательно локусы $loci(j, \Omega)$ в хромосомах с номерами $k \in [1, N_X]$, будем изменять энергетически неустойчивые значения $M_{X,\Omega}^{(i,e)}(k, j)$ (находящиеся на энергетическом уровне z) на противоположные с вероятностью $1 - p_{MB}(z) \cdot p_{CF}(k)$, где $p_{CF}(k)$ – дискретное распределение вида

$$p_{CF}(k) = \delta(\varphi_k) \cdot \left(\sum_{u=1}^{N_X} \delta(\varphi_u) \right)^{-1}. \quad (2)$$

Это распределение отдает приоритет в сохранении энергетически неустойчивых состояний более приспособленным хромосомам. Вышеописанная операция, проведенная для всех энергетически неустойчивых элементов матрицы $M_{X,\Omega}^{(i,e)}(k, j)$, преобразует исходную матрицу в некоторую новую матрицу $\tilde{M}_{X,\Omega}^{(i,e)}(k, j)$, с помощью которой далее будет построена маска m_Ω (вектор размерности N_{VAR}) для оценки инtronных локусов:

$$m_\Omega(j) = \begin{cases} 0, & \forall k \in [1; N_X] \tilde{M}_{X,\Omega}^{(i,e)}(k, j) = 0; \\ 1, & \exists k \in [1; N_X]: \tilde{M}_{X,\Omega}^{(i,e)}(k, j) \neq 0. \end{cases}$$

Далее, как только определены инtronные острова, поиск оптимального решения будет происходить путем изменения генов в экзональных локусах. Определим четверку операторов, производящих следующие изменения:

1. оператор очистки – заменяет одиночный экзон на экзональном острове на инtron;
2. оператор сдвига – сдвигает весь экзональный остров на некоторое случайное число позиций вправо или влево;
3. оператор склейки – суммирует управляющие воздействия экзонального острова и помещает суммарное воздействие в один из локусов данного экзонального острова, согласно распределению (2);
4. оператор перемешивания – суммирует управляющие воздействия экзонального острова и распределяет суммарное воздействие по случайному числу локусов данного экзонального острова, согласно распределению (2).

После каждого цикла применения операторов 1–4 запускается оператор уборки мусора, который будет удалять противоположные по направлению управляющие воздействия, например, поворот на 30° и

следующий за ним поворот на минус 30°. В отличие от операторов 1–4, действие оператора уборки мусора не ограничивается границами экзонного острова. Изменения, производимые данными операторами, будем сохранять только в том случае, когда это приводит к увеличению приспособленности особи. Для ускорения работы алгоритма будем применять операторы 1–4 с вероятностью, пропорциональной частоте принятых изменений, произведенных соответствующим оператором.

Укажем изменения, которые необходимо внести в алгоритмы работы операторов рекомбинации и мутации, чтобы их действия были согласованы с алгоритмом отбора строительных блоков. В классическом ГА сохранению блоков большего порядка препятствует оператор мутации. Соответственно, чтобы блоки смогли выжить в ранних поколениях, мутация должна быть низкой в начале и может увеличиваться по мере снижения среднего расстояния Хэмминга по популяции. Действительно, при появлении в ней выживающих из поколения в поколение блоков $d_H(\Omega)$ будет уменьшаться (тем больше, чем выше порядок блоков). В более поздних поколениях (в популяциях с малым $d_H(\Omega)$) увеличение числа мутаций может помочь ускорить отбор более приспособленных блоков, правда, при следующем условии: будут разрешены только мутации, приводящие к увеличению значения функции приспособленности особи. Данное условие аналогично принципу, лежащему в основе метода случайного восхождения [8]. Зададим формулу вероятности мутации особи C (теперь зависящую от времени):

$$p_m(t+1, C) = \left(1 - \frac{d_H(\Omega)}{N_{VAR}}\right) \cdot I_{\{F_{t+1}(C) > F_t(C)\}}.$$

Рассмотрим механизм, который позволяет производить эмиссию инtronов непосредственно в ходе применения оператора рекомбинации. Этот оператор, применяемый совместно с оператором рекомбинации, во избежание путаницы будем называть оператором выпрямления R . Он должен обнулять часть генов в хромосомах. Наиболее приспособленная особь в популяции (после применения генетических операторов) с большей вероятностью будет содержать искомые строительные блоки. По этой причине имеется заинтересованность в распространении ее генотипа. Из всех возможных вариантов оператора выбора пары более всего подходит вариант вожака стада (herd leader breeding [5]), при котором наиболее приспособленная особь образует пары со всеми остальными. Поскольку при скрещивании желательно не разрушать уже полученные блоки, выберем одноточечный вариант скрещивания, при котором родительские хромосомы рвутся в случайной точке и затем комбинируются для образования хромосом потомков.

Суммируя изложенное выше относительно операторов одноточечного скрещивания X и выпрямления R , приведем формальное описание совместного действия этих операторов:

$$(XR) = \begin{cases} X, & \alpha > \kappa_{att} \cdot \frac{d_H(\Omega)}{N_{VAR}}; \\ R, & \alpha \leq \kappa_{att} \cdot \frac{d_H(\Omega)}{N_{VAR}}, \end{cases}$$

где α – равномерно распределенная на $[0, 1]$ величина, а κ_{att} – коэффициент затухания, принимающий неотрицательные значения.

Описанный выше алгоритм будет быстро сходиться к экстремуму, отвечающему найденным блокам с высокой приспособленностью. Для того чтобы увеличить число перебираемых вариантов, реализуем алгоритм как островной [4]. Когда на острове с не самой высокой приспособленностью происходит вырождение, будем удалять популяцию данного острова и переселять наилучшую особь с лучшего острова, а остальные особи будем генерировать случайным образом, оставляя нетронутыми те гены, которые были нулевыми у наилучшей особи. Тем самым будут перенесены на обновленный остров выращенные строительные блоки с лучшего острова, обеспечивая запуск генетического отбора на данном острове с более высокого уровня приспособленности.

Предложенную модификацию ГА будем называть ГА с эмиссией инtronов (IEGA – Intron Emission GA). При этом исходный алгоритм ГА будем обозначать GA.

Результаты моделирования

Приведем результаты работы GA и IEGA. На рис. 2 изображены графики функций приспособленности лучших особей (по поколениям) для GA (без мутации и с мутацией) и IEGA, из которых видно, что модифицированный алгоритм попадает на более высокий уровень приспособленности особей. Результаты работы алгоритмов приведены в таблице. Ниже представлено их сравнение:

- IEGA генерирует в качестве оптимальных траектории со значительно меньшим числом галсов, чем GA (4 против 18–19);
- функция приспособленности особей IEGA выше;
- разброс значений функции приспособленности особей IEGA ниже (среднеквадратическое отклонение (СКО) результатов IEGA в три–пять раз ниже, чем у результатов GA);

- IEGA требует для сходимости в среднем больше поколений, чем GA без мутации. GA с мутацией не сходится за 100 поколений;
- разброс числа различных решений IEGA ниже, чем у GA (16 против 30).

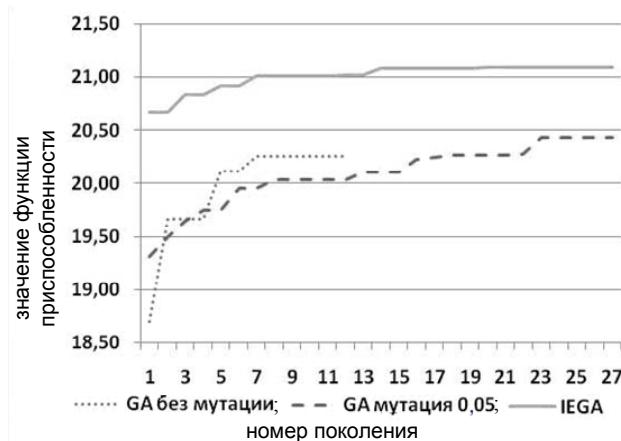


Рис. 2. Графики функций приспособленности лучших особей для GA и IEGA

Алгоритм	Число галсов	δ	СКО(δ)	Среднее число поколений	Общее число решений	Число различных решений
GA (без мутации)	19	19,95	0,26	7	30	30
GA (частота мутации 0,05)	18	20,41	0,15	100		30
IEGA	4	21,06	0,05	20		16

Таблица. Сравнение результатов работы GA и IEGA

Заключение

Предложенный генетический алгоритм с эмиссией инtronов позволяет находить близкие к оптимальным траектории маневрирования наблюдателя, решающего задачу оценивания параметров движущейся цели. Изменения, внесенные в логику работы генетических операторов рекомбинации и мутации, и осуществление отбора на уровне высоко приспособленных схем, увеличили вероятность нахождения глобального оптимума функции приспособленности по сравнению с классическим вариантом генетического алгоритма.

Литература

1. Ince L., Sezen B., Saridogan E., Ince H. An evolutionary computing approach for the target motion analysis (TMA) problem for underwater tracks // Expert Systems with Applications. – 2009. – V. 36. – № 1. – P. 3866–3879.
2. Cadre J.P., Jauffret C. Discrete-Time Observability and Estimability for Bearings-Only Target Motion Analysis // IEEE Transactions on Aerospace and Electronic Systems. – 1997. – V. 33. – № 1. – P. 178–201.
3. Степанов Д.В. Использование генетического алгоритма для нахождения оптимального маневра в задаче N-пеленгов // III Всероссийский конкурс молодых ученых. – Миасс, 2011. – 11 с.
4. Haupt R.L., Haupt S.E. Practical genetic algorithms. – John Wiley & Sons, 2004. – 253 p.
5. Reeves C.R., Rowe J.E. Genetic Algorithms: Principles and Perspectives. – Kluwer Academic Publishers, 2003. – 332 p.
6. Goldberg D.E. Genetic Algorithms in Search, Optimization and Machine Learning. – Addison-Wesley, 1989. – 412 c.
7. Spears W.M. Evolutionary Algorithms: The Role of Mutation and Recombination. – Springer, 2000. – 222 p.
8. Russell S.J., Norvig P. Artificial Intelligence: A Modern Approach. – Prentice Hall, 2003. – 1081 p.
9. Brameier M., Banzhaf W. Linear Genetic Programming. – Springer, 2007. – 315 p.

Степанов Денис Вячеславович – ОАО «Концерн «Аврора», мл. научный сотрудник, denis_v_stepanov@hotmail.com

Шалыто Анатолий Абрамович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, shalyto@mail.ifmo.ru

УДК 004.4'242

ПРИМЕНЕНИЕ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ О ВЫПОЛНИМОСТИ БУЛЕВОЙ ФОРМУЛЫ ДЛЯ ПОСТРОЕНИЯ УПРАВЛЯЮЩИХ КОНЕЧНЫХ АВТОМАТОВ ПО СЦЕНАРИЯМ РАБОТЫ

В.И. Ульянцев, Ф.Н. Царев

Предлагается метод построения управляющих конечных автоматов по сценариям работы. Метод основан на сведении указанной задачи к задаче о выполнимости булевой формулы. Работоспособность метода проверяется на задаче построения автомата управления часами с будильником. На этой задаче построение соответствующего управляющего автомата корректно, а время работы алгоритма составляет меньше секунды на персональном компьютере.

Ключевые слова: конечные автоматы, машинное обучение, задача о выполнимости булевой формулы.

Введение

В последнее время все в более широком кругу задач начинает применяться автоматное программирование, в рамках которого поведение программ описывается с помощью детерминированных конечных автоматов [1]. Для многих задач автоматы удается строить эвристически, однако существуют задачи, для которых такое построение автоматов затруднительно. К задачам этого класса относятся, в частности, задачи об «Умном муравье» [2–4], об управлении моделью беспилотного летательного аппарата [5]. Для построения автоматов в задачах такого типа успешно применяются генетические алгоритмы [6], в том числе на основе обучающих примеров [7]. Недостатком генетических алгоритмов является то, что время их работы весьма велико, и его достаточно трудно оценить аналитически.

Целью настоящей работы является разработка метода построения управляющего конечного автомата, лишенного указанных недостатков.

Постановка задачи

Управляющим конечным автоматом будем называть детерминированный конечный автомат, каждый переход которого помечен событием, последовательностью выходных воздействий и охранным условием, представляющим собой логическую формулу от входных переменных. Автомат получает события от так называемых поставщиков событий (в их роли могут выступать внешняя среда, интерфейс пользователя и т.д.) и генерирует выходные воздействия для объекта управления. При поступлении события автомат выполняет переход в соответствии с охранными условиями и значениями входных переменных. При выполнении перехода генерируются выходные воздействия, которыми он помечен, и автомат переходит в соответствующее состояние. Отметим, что состояния такого автомата не делятся на допускающие и недопускающие.

Формальное определение управляющего автомата дано в [1]. Пример управляющего автомата приведен на рис. 1.

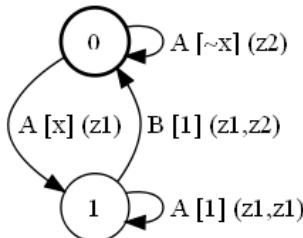


Рис. 1. Пример управляющего автомата

Для данного автомата множество входных событий равно $\{A, B\}$, охранные условия зависят от единственной логической входной переменной x , множество выходных воздействий равно $\{z_1, z_2\}$. Далее состояние автомата с номером 0 будем считать начальным.

В настоящей работе в качестве исходных данных для построения управляющего конечного автомата используется множество сценариев работы. Сценарием работы будем называть последовательность $T_1 \dots T_n$ троек $T_i = \langle e_i, f_i, A_i \rangle$, где e_i – входное событие; f_i – булева формула от входных переменных, задающая охранное условие; A_i – последовательность выходных воздействий. В дальнейшем тройки T_i будем называть элементами сценария.

Будем говорить, что автомат, находясь в состоянии $state$, удовлетворяет элементу сценария T_i , если из $state$ исходит переход, помеченный событием e_i , последовательностью выходных воздействий A_i и охранным условием, тождественно равным f_i как булева формула. Автомат удовлетворяет сценарию работы $T_1 \dots T_n$, если он удовлетворяет каждому элементу данного сценария, находясь при этом в состояниях пути, образованного соответствующими переходами.

В работе [8] разработан метод построения автомата-распознавателя по заданному набору слов. Данный метод основан на сведении поставленной задачи к задаче о выполнимости булевой формулы

(boolean satisfiability problem, SAT) – по набору слов строится логическая формула. Выполняющая подстановка для нее ищется с помощью сторонней программы. Затем на основании полученных значений переменных логической формулы строится искомый автомат-распознаватель. Результаты вычислительных экспериментов показывают более высокую скорость работы этого метода по сравнению с методами, основанными на объединении состояний [9, 10].

В настоящей работе решается задача построения управляющего конечного автомата с заданным числом состояний C по заданному множеству сценариев работы Sc , которым автомат должен удовлетворять. Рассмотрим основные этапы работы предлагаемого алгоритма решения поставленной задачи.

Построение дерева сценариев

Деревом сценариев назовем дерево, каждый переход которого помечен событием, булевой формулой и последовательностью выходных воздействий. Опишем алгоритм построения дерева сценариев по заданному множеству сценариев Sc .

Изначально дерево сценариев состоит из единственной вершины – корня дерева. Затем по очереди добавим в дерево все сценарии работы из Sc . Для каждого из сценариев будем добавлять его элементы в дерево в порядке возрастания их номеров, начиная с первого. При этом будем хранить указатель на текущую вершину дерева v и номер i первого необработанного элемента сценария.

В начале процесса добавления v указывает на корень дерева сценариев, а $i = 1$. На каждом шаге проверяется существование исходящего из вершины v ребра, помеченного событием e_i и логической формулой, совпадающей с f_i как булевой функцией. Если такое ребро не существует, то создается новая вершина дерева u , и в нее направляется ребро, помеченное тройкой $\langle e_i, f_i, A \rangle$. После этого u становится текущей вершиной, а значение i увеличивается на единицу.

Если такое ребро существует, то производится сравнение последовательности A_i и последовательности выходных воздействий A' , которой помечено рассматриваемое ребро. Если $A_i = A'$, то текущей становится вершина, в которую ведет рассматриваемое ребро дерева, а значение i увеличивается на единицу. Если же указанные последовательности не совпадают, то заданное множество сценариев Sc является противоречивым, поэтому работа алгоритма прерывается, и пользователю выводится соответствующее сообщение.

После завершения добавления всех сценариев в дерево производится проверка охранных условий. Для каждой вершины перебираются все пары исходящих из нее ребер. Если существует такая пара ребер, что они помечены одним и тем же событием, а их охранные условия имеют общий выполняющий набор значений входных переменных, то множество сценариев предполагает недетерминированное поведение. Исходя из этого, работа алгоритма прерывается, и пользователю выводится соответствующее сообщение.

На рис. 2 приведен пример дерева сценариев. Сценариям данного дерева удовлетворяет автомат, приведенный на рис. 1.

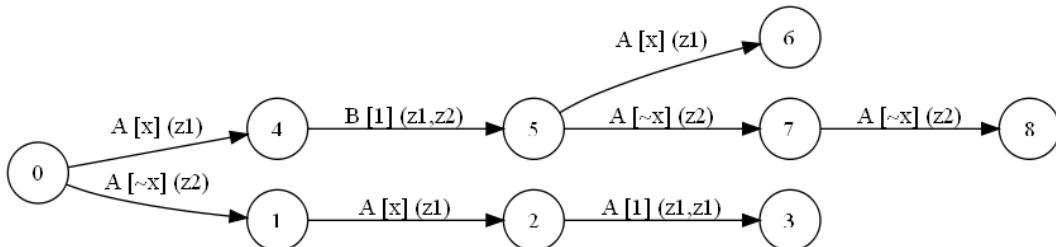


Рис. 2. Пример дерева сценариев

Построение графа совместимости вершин дерева сценариев

Для построения управляющего автомата необходимо «раскрасить» вершины дерева сценариев в заданное число цветов (равное числу состояний, которое задается в качестве одного из параметров алгоритма). При этом вершины одного цвета будут объединены в одно состояние результирующего автомата, а множество исходящих из состояния переходов будет строиться из объединения множеств ребер исходящих из вершин заданного цвета.

Для задания ограничений на раскраску построим так называемый граф совместимости вершин дерева сценариев. Множество вершин этого графа совпадает с множеством вершин дерева сценариев, поэтому в дальнейшем вершины графа и дерева различаться не будут. Ребра графа определяются следующим образом. Вершины графа совместимости u и v соединены ребром (далее такие вершины будем называть несовместимыми), если существует последовательность пар $\langle e_1, values_1 \rangle \dots \langle e_k, values_k \rangle$ событий и наборов значений входных переменных, которая различает соответствующие вершины дерева. Будем говорить, что указанная последовательность различает вершины u и v , если выполняется совокупность следующих условий:

- из вершины u существует путь P_u , ребра которого помечены соответственно событиями $e_1 \dots e_k$ и такими охранными условиями $f_1 \dots f_k$, что наборы значений входных переменных $values_1 \dots values_k$ являются соответственно их выполняющими подстановками;
- аналогичный путь P_v существует из вершины v ;
- для последних ребер путей P_u и P_v верно хотя бы одно из двух условий:
 - пометки этих ребер различаются в части выходных действий;
 - у охранных условий этих ребер есть общий выполняющий набор значений входных переменных, но они не совпадают как булевы функции.

Опишем алгоритм построения графа совместимости. Напомним, что множество вершин этого графа совпадает с множеством вершин дерева сценариев. Основной идеей данного алгоритма является метод динамического программирования [11].

Для каждой вершины дерева сценариев v найдем все несовместимые с ней вершины. Обозначим как $S(v)$ множество вершин, несовместимых с v . Будем вычислять значения функции $S(v)$, начиная с листьев дерева сценариев. Для каждого из листьев u множество $S(u)$ пусто по определению несовместимых вершин.

Покажем, как вычислить значение $S(v)$, если оно уже вычислено для всех значений «детей» вершины v . Переберем все вершины дерева – вершина u входит в множество $S(v)$, если существует пара ребер u (помечено событием e , формулой f_1 и последовательностью действий A_1) и uv (помечено также событием e , формулой f_2 и последовательностью действий A_2) такая, что выполняется одно из трех:

- формулы f_1 и f_2 имеют общий выполняющий набор значений входных переменных, но не совпадают как булевы функции. Тогда $\langle e, values \rangle$, где как $values$ обозначена выполняющая подстановка f_1 , – последовательность, различающая u и v ;
- формулы f_1 и f_2 совпадают как булевы функции, а последовательности A_1 и A_2 не совпадают. Тогда вершины u и v различают такая же последовательность;
- формулы f_1 и f_2 совпадают как булевы функции, и вершина x входит во множество $S(v)$, посчитанное заранее. Тогда существует последовательность $\langle e_1, values_1 \rangle \dots \langle e_k, values_k \rangle$, различающая вершины x и y , а вершины v и u различают последовательность $\langle e, values \rangle \langle e_1, values_1 \rangle \dots \langle e_k, values_k \rangle$.

Время работы этого алгоритма составляет $O(n^2)$ (где за n обозначено число вершин в дереве сценариев), так как каждая пара ребер дерева сценариев в процессе работы алгоритма будет рассмотрена не более одного раза. Такое время работы достижимо, если заранее для каждой пары формул вычислено, равны ли они как булевы функции и имеют ли общий выполняющий набор значений входных переменных. В худшем случае время работы этапа обработки формул составляет $O(2^{2m}n^2)$, где за m обозначено максимальное число входных переменных, использующихся в одном охранном условии. На практике число m не превышает четырех.

Построение булевой КНФ-формулы, задающей требования к раскраске построенного графа

Опишем алгоритм построения конъюнктивной нормальной формы (КНФ) булевой формулы, задающей требования к раскраске построенного графа и выражающей непротиворечивость системы переходов результирующего автомата. Данное построение аналогично построению КНФ-формулы, используемой в работе [8] для построения автомата-распознавателя. Напомним, что в настоящей работе на вход алгоритму подается число C состояний результирующего автомата.

В данной формуле будут использоваться следующие логические переменные:

- $x_{v,i}$ (для каждой вершины дерева сценариев v и цвета вершины i – числа от 1 до C) – верно ли, что вершина v имеет цвет i . Напомним, что вершины одного цвета будут объединены в одно состояние результирующего автомата;
- $y_{a,b,e,f}$ (для каждой пары состояний результирующего автомата a и b , каждого события e , каждой формулы f , встречающейся в сценариях) – верно ли, что в результирующем автомате существует переход из состояния a в состояние b , помеченный событием e и формулой f .

КНФ-формула состоит из следующих дизъюнкто:

- $(x_{v,1} \vee \dots \vee x_{v,C})$ (для каждой вершины v дерева сценариев) – накладываем условие существования цвета вершины v ;
- $(\sim x_{v,i} \vee \sim x_{v,j})$ (для каждой вершины v дерева сценариев, цвета i и цвета j , где $i < j$) – накладываем условие, что вершина v не покрашена одновременно в цвета i и j ;
- $(\sim x_{v,i} \vee \sim x_{u,i})$ (для каждой пары несовместимых вершин u и v дерева сценариев и цвета i) – накладываем ограничения на раскраску, задаваемые графиком совместимости;
- $(\sim y_{a,b,e,f} \vee \sim y_{a,d,e,f})$ (для каждой тройки состояний результирующего автомата a , b и d ($b < d$), каждого события e , каждой формулы f) – из состояния a выходит не более одного ребра, помеченного событием e и формулой f ;
- $(y_{a,b,e,f} \vee \sim x_{v,a} \vee \sim x_{u,b})$ (для каждого ребра vu дерева сценариев) – если выполняется:
 - ребро из вершины дерева v в вершину u помечено событием e и формулой f ;

- вершина v покрашена в цвет a ;
 - вершина u покрашена в цвет b ;
- то в результирующем автомате существует переход из состояния a в состояние b , помеченный событием e и формулой f ;
- $(\sim y_{a,b,e,f} \vee \sim x_{v,a} \vee x_{u,b})$ (для каждого ребра vu дерева сценариев) – если выполняется:
 - ребро из вершины дерева v в вершину дерева u помечено событием e и формулой f ;
 - вершина v покрашена в цвет a ;
 - в результирующем автомате существует переход из состояния a в состояние b , помеченный событием e и формулой f ;
- то вершина u покрашена в цвет b .

Запуск сторонней программы, решающей задачу о выполнимости булевой КНФ-формулы

Чтобы найти выполняющий набор для построенной КНФ-формулы, воспользуемся сторонней программой (SAT-solver), решающей задачу SAT. В результате анализа результатов соревнования SAT-Race 2010 [12] была выбрана программа cryptominisat [13]. Эта программа признана победителем в указанном соревновании. Подадим на вход выбранной программе построенную КНФ-формулу, записанную в формате DIAMAX (<http://www.satlib.org/ubcsat/satformat.pdf>).

Если программа не обнаружила выполняющий набор значений переменных, то будем считать, что по данному набору сценариев невозможно построить управляющий автомат с заданным числом состояний. Если же программа обнаружила выполняющий набор, то построим искомый автомат. Для этого на основании полученных значений переменных $x_{v,i}$ определим цвет каждой вершины дерева сценариев. На рис. 3 приведен пример раскраски дерева сценариев, приведенного на рис. 2.

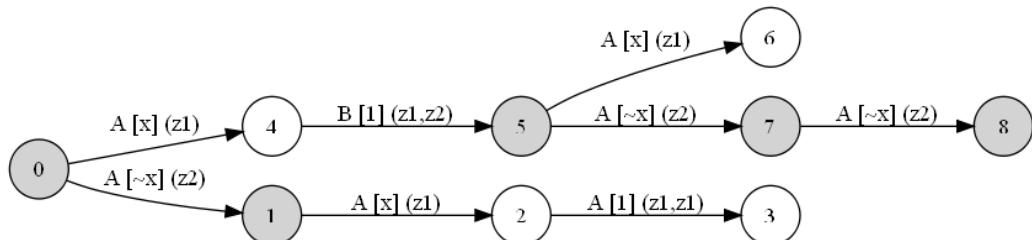


Рис. 3. Пример раскраски дерева сценариев в два цвета

После этого объединим все вершины одного цвета в одно состояние автомата, а начальным состоянием положим состояние, соответствующее цвету корня дерева сценариев. Множество исходящих из состояния переходов построим из объединения множеств ребер, исходящих из вершин заданного цвета.

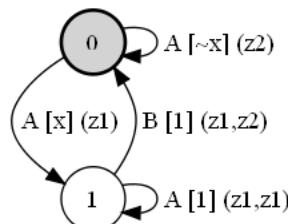


Рис. 4. Автомат, полученный после объединения вершин дерева

Например, после объединения вершин дерева, приведенного на рис. 3, получим автомат, изображенный на рис. 4. Заметим, что данный автомат изоморфен автомата, приведенному на рис. 1. Отметим, что не исключено существование нескольких автоматов, удовлетворяющих сценариям множества Sc .

Экспериментальное исследование

Экспериментальное исследование проводилось на задаче построения автомата управления часами с будильником [1]. Было задано 38 сценариев, аналогичных тестам, приведенным в работе [7]. На основе этих сценариев был построен автомат, изоморфный автомата, построенному вручную в работе [1]. Его построение заняло менее секунды на персональном компьютере с процессором Intel Core 2 Quad Q9400, что позволяет говорить о достаточно высокой производительности разработанного метода.

Заключение

В работе предложен метод построения управляющих конечных автоматов по сценариям работы. Этот метод основан на сведении указанной задачи к задаче о выполнимости булевой формулы. Работоспособность метода проверена на задаче построения автомата управления часами с будильником.

На этой задаче соответствующий управляющий автомат был построен корректно, а время работы алгоритма составляло меньше секунды на персональном компьютере с процессором Intel Core 2 Quad Q9400.

Исследование выполнено в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы».

Литература

1. Поликарпова Н.И., Шалыто А.А. Автоматное программирование. – СПб: Питер, 2011. – 176 с.
2. Angeline P.J., Pollack J. Evolutionary Module Acquisition // Proceedings of the Second Annual Conference on Evolutionary Programming. – 1993. – P. 154–163.
3. Jefferson D., Collins R., Cooper C., Dyer M., Flowers M., Korf R., Taylor C., Wang A. The Genesys System. – 1992 [Электронный ресурс]. – Режим доступа: <http://www.cs.ucla.edu/~dyer/Papers/AlifeTracker/Alife91Jefferson.html>, св. Яз. англ. (дата обращения 25.11.2011).
4. Chambers L. Practical Handbook of Genetic Algorithms. Complex Coding Systems. – Boca Raton: CRC Press, 1999. – V. III – 592 р.
5. Царев Ф.Н., Шалыто А.А. Гибридное управление беспилотными летательными объектами на основе автоматного программирования // 1-я Российская мультиконференция по проблемам управления. Сборник докладов четвертой научной конференции «Управление и информационные технологии». – СПб ГЭТИ «ЛЭТИ», 2006. – С. 138–144.
6. Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы. – М.: Физматлит, 2006. – 368 с.
7. Царев Ф.Н. Метод построения управляющих конечных автоматов на основе тестовых примеров с помощью генетического программирования // Информационно-управляющие системы. – 2010. – № 5. – С. 31–36.
8. Heule M., Verwer S. Exact DFA Identification Using SAT Solvers // Grammatical Inference: Theoretical Results and Applications. 10th International Colloquium. ICGI 2010. – 2010. – P. 66–79.
9. Oncina J., Garcia P. Inferring regular languages in polynomial update time / Pattern Recognition and Image Analysis. Series in Machine Perception and Artificial Intelligence. – Singapore: World Scientific, 1992. – V. 1. – P. 49–61.
10. Oliveira A.L., Marques-Silva J.P. Efficient search techniques for the inference of minimum sized finite state machines // Proceeding of 5th Symposium on String Processing and Information Retrieval. – 1998. – P. 81–89.
11. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. – 2-е изд. – М.: МЦНМО, 2010. – 1296 с.
12. Presentation of SAT-Race results at the SAT'10 conference [Электронный ресурс]. – Режим доступа: <http://baldur.iti.uka.de/sat-race-2010/downloads/SAT-Race-2010-Presentation.pdf>, свободный. Яз. англ. (дата обращения: 25.11.2011).
13. Soos M. CryptoMiniSat [Электронный ресурс]. – Режим доступа: <http://www.msoos.org/cryptominisat2>, св. Яз. англ. (дата обращения: 25.11.2011).

Ульянцев Владимир Игоревич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, ulyantsev@rain.ifmo.ru
Царев Федор Николаевич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, fedor.tsarev@gmail.com

УДК 004.4'242

ПРИМЕНЕНИЕ АДАПТИВНОГО ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ ГЕНЕРАЦИИ КЛЕТОЧНЫХ АВТОМАТОВ

А.В. Тихомиров, А.А. Шалыто

Рассматривается улучшенный алгоритм генерации произвольных клеточных автоматов на основе тестовых наборов при помощи адаптивного генетического алгоритма. Описаны основные отличия и преимущества по сравнению со стандартным генетическим алгоритмом. Произведена апробация на нескольких обучающих примерах.

Ключевые слова: клеточные автоматы, генетические алгоритмы.

Введение

Настоящая работа развивает идеи, описанные в [1]. Для генерации различных клеточных автоматов используется адаптивный генетический алгоритм (ГА), который устраняет многие проблемы стандартного ГА. Алгоритм используется для моделирования многих физических процессов, например: диффузия энергии, разнообразные химические реакции, рост кристаллов и т.д. [2–5].

ГА, описанный в работе [1], имеет недостатки – он не адаптируется под текущее состояние популяции, имеет проблемы со сходимостью и временем работы. Цель настоящей работы – устранить недостатки ГА, разработав адаптивный ГА.

Постановка задачи

Задана квадратная сетка и автомат, управляющий всеми ячейками [6–8]. При этом задаются данные для различных временных шагов, т.е. начальное состояние системы, несколько промежуточных состояний и конечное состояние системы.

Целью работы алгоритма является выращивание автомата, наиболее точно описывающего заданную физическую систему.

Строение хромосомы клеточного автомата

Обычно при использовании генетического подхода используется метод, при котором данные в хромосоме хранятся в виде строки [6]. Однако в рассматриваемой задаче длина хромосомы может варьироваться в зависимости от количества переходов между состояниями, поэтому ген представляет собой не строку, а специализированную структуру.

Количество базовых генов соответствует количеству состояний в клеточном автомате. Таким образом, базовый ген под названием «StateGene» описывает определенное состояние автомата. «StateGene» состоит из генов переходов клеточного автомата, которые называются «TransitionGene». Каждый ген перехода состоит из списка генов условий на переход и генов действий, которые будут осуществляться в случае этого перехода. Длина хромосомы зависит от количества состояний в клеточном автомате, который описывает данная хромосома.

В случае отсутствия генов условий подразумевается, что условие на данный переход всегда является верным.

Начальное поколение состоит из фиксированного числа случайно генерированных автоматов. Все автоматы в поколении имеют одинаковое наперед заданное количество состояний, но количество переходов между состояниями может варьироваться.

Проблема вырождения популяции

Одной из проблем в применении генетического подхода для генерации клеточных автоматов стало вырождение популяции, т.е. такой ситуации, когда выделяется один-единственный генотип, который представляет собой локальный максимум, а затем все элементы популяции проигрывают ему отбор, и вся популяция «забивается» копиями этой особи. Из-за схожего строения хромосом операция скрещивания становится практически бесполезной, и вероятность получить хромосому, которая лучше решает поставленную задачу, сильно падает. В настоящей работе рассматриваются несколько способов преодолеть этот недостаток:

- алгоритм каскадной фитнесс-функции;
- операция инъекции хромосом;
- адаптивный генетический алгоритм.

Все эти модификации позволяют улучшить «сходимость» генетического подхода при генерации клеточных автоматов.

Операция скрещивания

В зависимости от режима работы генератора используется скрещивание двух видов – симметричное и несимметричное. Заметим, что в обоих видах скрещивания производится обмен генами одного типа в соответствующих участках хромосомы.

Особенность симметричного скрещивания заключается в том, что происходит обмен соответствующими участками поддеревьев генов состояний между хромосомами. Особенность несимметричного скрещивания заключается в том, что происходит обмен случайными поддеревьями одного уровня между хромосомами (рис. 1).

Для оптимизации операции скрещивания применяется анализ хромосом на совпадающие гены, т.е. из каждой пары хромосом, участвующих в скрещивании, выделяется список уникальных генов. Именно среди этих генов и производится поиск для обмена. Такая операция позволяет совершать меньшее число скрещиваний и заранее исключать операции, которые приведут к хромосомам, которые уже есть в популяции.

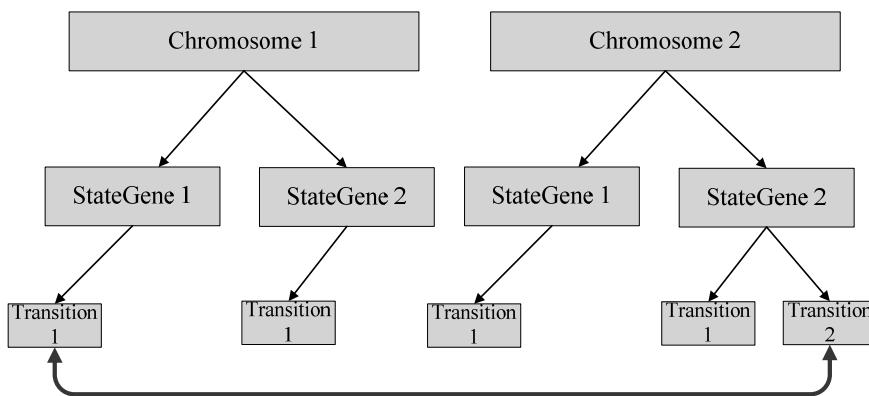


Рис. 1. Пример несимметричного скрещивания хромосом

Операция инъекции хромосом

Для добавления особей, полученных из предыдущих запусков генератора, используется операция инъекции хромосом в новую популяцию. Следует отметить, что в стандартных генетических алгоритмах нет предложенной ниже операции, она была разработана специально для ГА, предложенного в настоящей работе. Цель этой операции – добавление частей хромосом, которые хранятся с предыдущих запусков генетического алгоритма, в новую популяцию, для возможного улучшения особей за счет материала, накопленного за предыдущие запуски.

Ниже предложены основные этапы работы этой операции:

- выбор особи из текущей популяции;
- выбор особи из хранилища результатов предыдущих запусков генератора;
- получение «разности» между хромосомой из хранилища результатов предыдущих запусков генератора и хромосомой из популяции;
- операция скрещивания между разностной хромосомой и хромосомой из популяции.

Фитнесс-функция

Функция приспособленности особи, т.е. клеточного автомата, состоит из нескольких значений. Первое значение характеризует автомат с точки зрения решения поставленной задачи, при моделировании сравниваются эталонные и полученные результаты для каждого шага моделирования. Второе значение характеризует автомат с точки зрения избыточности, в нем хранится «длина автомата» и количество неиспользуемых переходов и условий. Для борьбы с вырождением популяции была применена каскадная фитнесс-функция. Идея заключается в том, что на начальных этапах выращивания клеточных автоматов функция приспособленности учитывает только несколько первых временных срезов входных данных. Когда в популяции есть клеточные автоматы, которые в какой-то степени удовлетворяют текущей фитнесс-функции, то происходит смена на функцию с учетом большего количества слоев.

Ниже рассмотрены «минусы» использования каскадной функции приспособленности:

- дополнительные затраты при смене функции;
- сложность сравнения эффективности особей (особенно это проявляется при сравнении особей из разных «островов»).

Ниже рассмотрены «плюсы» использования каскадной фитнесс-функции:

- прирост производительности (на ранних этапах уменьшается количество этапов проверки клеточного автомата).
- после каждого этапа полученный автомат частично решает поставленную задачу.

Генетический алгоритм

Основной проблемой при совместном использовании генетического подхода и клеточных автоматов являются вырождение и стагнация популяции, поэтому возникла необходимость в разработке специального ГА, который бы учитывал текущее состояние популяции и текущую стадию генерации.

Предложенный ГА состоит из нескольких состояний и, в сущности, представляет собой конечный автомат с правилами переключения между стадиями генератора. Ниже перечислены разработанные состояния генератора (рис. 2):

1. нормальный режим работы генератора клеточных автоматов;
2. форсированный режим работы, который применяется при стагнации процесса генерации;
3. режим доводки решения, который начинает работать при приближении к глобальному минимуму.

Заметим, что из режима доводки решения генератор уже не будет перезапускаться даже в состоянии стагнации.



Рис. 2. Схема генетического алгоритма

Чтобы избежать локальных минимумов функции приспособленности и вырождения популяции, применяются два алгоритма отбора в новое поколение.

1. Происходит отбор первых P особей с лучшей функцией приспособленности. Все они попадают в новое поколение.
2. Если на протяжении достаточно большого числа поколений не происходит улучшения значения функции приспособленности, то при обработке текущего поколения отбрасываются все особи, кроме нескольких, которые имеют наилучшее значение, т.е. наиболее приспособленных. Оставшееся место в новом поколении занимают особи, полученные из особей текущего поколения путем скрещивания и мутаций. Мутация может происходить не в единственном гене хромосомы, как при обычных условиях, а во всех генах хромосомы с некоторой вероятностью.

При отборе первых P особей при одинаковой функции приспособленности выбираются более молодые особи.

Полученные результаты

Для тестового примера была проведена апробация разработанного метода для генерации клеточных автоматов для игр на двумерном поле. Основным отличием таких клеточных автоматов является то, что они не хранят в себе данных и меняют свое состояние только в зависимости от состояний окружающих ячеек.

Далее приведены два примера апробации предложенного адаптивного ГА.

Игра «Жизнь». Игра «Жизнь» является одним из наиболее известных клеточных автоматов для двумерного поля. Она была создана Джоном Хортоном Конвеем в 1970 г. Процесс эволюции поля эмулирует процессы, происходящие при зарождении, развитии и гибели колоний живых организмов. В качестве тестовых примеров на вход генератору подавались наборы полей и эталонные примеры работы (рис. 3).

В силу простого строения автомата и, как следствие, малого числа генов улучшение фитнесс-функции происходит в основном за счет многочисленных мутаций особей. Было установлено, что использование ГА является неэффективным. Применение эволюционной стратегии на всех этапах генерации клеточного автомата позволило улучшить сходимость предлагаемого метода для генерации клеточного автомата «Жизнь». Для сравнения: при использовании эволюционной стратегии генератор перебирает порядка 200 особей, при ГА – более 2000 особей.

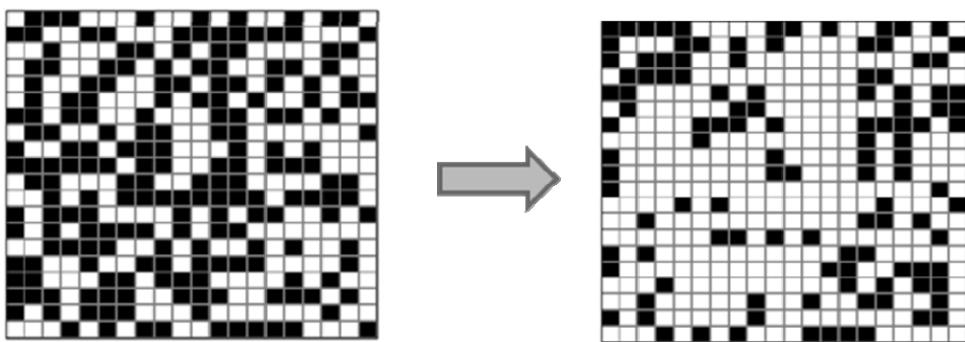


Рис. 3. Пример тестового набора из двух шагов симуляции для игры «Жизнь»

Общий случай игр на двумерном поле. Для тестового примера была поставлена задача получить клеточный автомат, зная только состояния клеток на каждом конкретном шаге расчетов. Для этого брался случайный эталонный автомат, количество состояний которого варьировалось в пределах от пяти до восьми, и на его основе генерировались тестовые наборы. Для ускорения расчетов функции приспособленности все тестовые наборы имели размеры тестового поля 40×40 (рис. 4).

Для генерации автоматов игр на двумерном поле использовались следующие настройки генератора: число островов 1–4; вероятность мутации 0,1; число особей на острове 50; доля элитизма 0,1; период миграции 50.

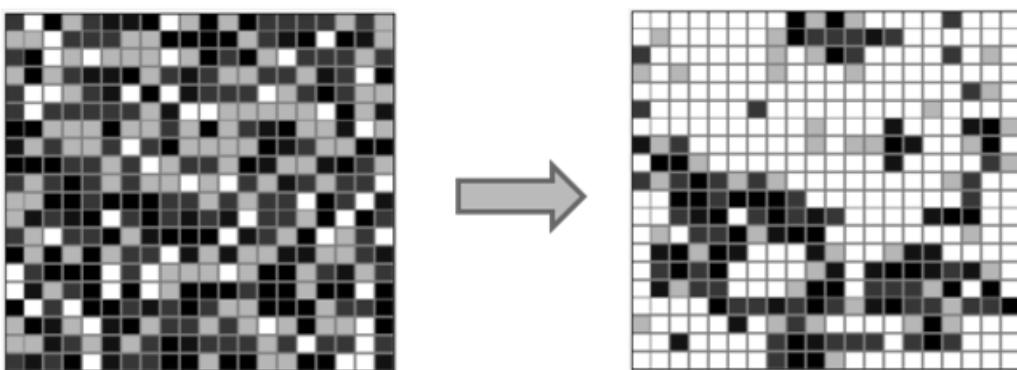


Рис. 4. Пример тестового набора из двух шагов симуляции произвольного клеточного автомата

Быстродействие предложенного способа сильно зависит от начальных настроек, таких как пропорции между генетическими операциями, величина начальной популяции и многих других.

Также было установлено, что на быстродействие описанного в работе ГА влияет наличие точных условий, содержащихся в клеточном автомате. В таблице представлены значения числа вычисления функции приспособленности для двух клеточных автоматов, отличающихся всего одним условием:

- $n(3)=3$ у первого выращиваемого эталона;
- $n(3)<3$ у второго выращиваемого эталона.

	N_{\min}	N_{avg}	N_{\max}
Automaton 1	27106	52315	102411
Automaton 2	23594	30173	44862

Таблица. Сравнение времени работы генератора для генерации эталонов с различием в одном условии (N – число вычислений фитнесс-функции)

Заключение

В работе предложен метод генерации клеточных автоматов произвольного количества состояний, учитывающий текущее состояние популяции. Он позволяет автоматически получать клеточные автоматы, которые довольно точно описывают данные о физической системе на входе. Также этот метод позволяет получать клеточные автоматы, которые затруднительно реализовать эвристическими методами. Алгоритм апробирован на ряде задач, таких как игра «Жизнь» и игры на двумерном поле. Показано, что адаптивный алгоритм умеет приспосабливаться к текущему состоянию генерируемых особей и выбирать оптимальную вариацию генетического алгоритма.

Литература

1. Тихомиров А.В., Шалыто А.А. Применение генетического подхода для генерации клеточных автоматов // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 2. – С. 62–66.
2. Тоффоли Т., Марголос Н. Машины клеточных автоматов. – М.: Мир, 1991. – 280 с.
3. Frish U. Lattice gas hydrodynamics in two and three dimensions // Complex Systems. – 1987. – V. 1. – P. 649–707.
4. Wolfram S. Cellular automaton Fluids // J. Stat. Phys. – 1986. – V. 45. – P. 471–526.
5. Фон Нейман Дж. Теория самовоспроизводящихся автоматов: Пер. с англ. – М.: Мир, 1971. – 384 с.
6. Царев Ф.Н., Шалыто А.А. Применение генетического программирования для генерации автомата в задаче об «кумном муравье» [Электронный ресурс]. – Режим доступа: http://is.ifmo.ru/genalg/_ant_ga.pdf, св. Яз. рус. (дата обращения 25.11.2011).
7. Наумов Л.А. Метод введения обобщенных координат и инструментальное средство для автоматизации проектирования программного обеспечения вычислительных экспериментов с использованием клеточных автоматов. Дис. ... канд. техн. наук: 05.13.12. – СПб, 2007. – 283 с.
8. Скаков П.С. Классификация поведения одномерных клеточных автоматов. Магистерская диссертация [Электронный ресурс]. – Режим доступа: http://is.ifmo.ru/papers/_skakov_master.pdf, св. Яз. рус. (дата обращения 25.11.2011).

Тихомиров Андрей Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, and.tikhomirov@gmail.com

Шалыто Анатолий Абрамович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, shalyto@mail.ifmo.ru

МЕХАНИКА И МЕХАТРОНИКА**6**

УДК 62-294.4

**УСТАНОВКА ДЛЯ ФОРМИРОВАНИЯ АЭРОДИНАМИЧЕСКОГО
ПРОФИЛЯ ГАЗОВЫХ ОПОР**

С.А. Родников

Рассмотрен гироприбор с бесконтактным подвесом чувствительного элемента на базе аэродинамических опор, имеющих специальный аэродинамический профиль, наносимый методом механической обработки. Существующая установка для формирования профиля имеет ряд серьезных недостатков, отрицательно влияющих на профиль опор. Для их устранения предложена новая конструкция установки формирования профиля опор. Проведены исследования, которые позволили ввести дополнительные изменения в конструкцию, выбрать материал притира. Изготовлен комплект опор с аэродинамическим профилем, соответствующим всем геометрическим требованиям для чувствительного элемента гироприбора.

Ключевые слова: аэродинамическая опора, установка для формирования профиля, механическая обработка, абразивный износ.

Введение

Впервые идея использования газа в качестве смазочного вещества была высказана в работе Хирна опубликованной в Париже в 1854 г. Однако планомерные исследования в области газовой смазки начались в СССР и США в 1930-е г.г. в связи с проблемами в станкостроении. Первые лабораторные испытания технического устройства (гироскопа) с газовыми подшипниками были проведены в США в 1932 г., но основы газодинамической теории были сформулированы в начале 1950-х г.г., причем основополагающей можно считать работу советского ученого С.А. Шейнберга. Разработка основ аэродинамической теории подвесов с внешним наддувом была завершена к концу 1950-х г.г., и основной вклад был внесен отечественными учеными профессорами Л.Г. Лойянским и Л.Г. Степанянцем. В теории вибронесущих газовых опор основная заслуга принадлежит зарубежным ученым, таким как Тэйлор, Сэффмен, Ланглуа, Салблю. Также быстрые и практические успехи в применении подшипников были бы невозможны без углубленных теоретических исследований течений вязкого газа в тонком слое, приведенных в трудах В.Н. Константинеску и В.Н. Дроздовича [1, 2].

Среди газовых подшипников сложной микрогеометрии наиболее широкое применение нашли подшипники со спиральными канавками. Именно такие подшипники (опоры) используются «НИИ командных приборов» в комплексах командных приборов для навигации и управления движением ракет.

Аэродинамическая опора (рис. 1) в данных приборах является особо ответственной деталью, к которой предъявляются очень высокие требования по геометрии: некруглость сферической поверхности не больше 0,2 мкм, шероховатость рабочих поверхностей не хуже 0,1 мкм, необходимый для распространения воздушного потока профиль спиральных канавок, а также их глубина. Эти канавки имеют переменную глубину, которая в зоне вершины составляет 0,003 мм, а у основания увеличивается до 0,004 мм. Опоры изготовлены из материала 40ХНЮ-ВИ с покрытием нитридом титана толщиной 15 ± 3 мкм. Особенностью технологии их изготовления является то, что канавки наносятся на опоры с покрытием. При несоответствии полученного профиля требованиям по геометрии покрытие стравливается и наносится заново.

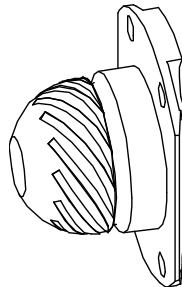


Рис. 1. Аэродинамическая опора

Существующая на предприятии установка для формирования аэродинамического профиля на газовых опорах имеет ряд серьезных недостатков: быстрый износ притира, отсутствие правки кромки притира, большие скорости вращения, делающие процесс формирования профиля сложным в управлении. В связи с некоторыми непродуманными конструкторскими решениями в обеспечении вращения притира возникают погрешности в получаемом профиле, а именно, углубления в крайних

зонах канавок. Все это влияет на качество газовых опор и может привести к снижению точности работы гироприбора в целом.

Ниже представлены результаты проектирования новой установки для формирования профиля, отвечающей современным требованиям и позволяющей исключить недостатки прежней установки.

Преимущества опор с газовой смазкой

Вязкость газа весьма мала по сравнению с жидкостью, соответственно несущая способность газового слоя и сопротивление его вращению шипа во много раз меньше, чем у смазочного слоя несжимаемой жидкости. Эти особенности обуславливают область применения подшипников с газовой смазкой – в основном для опор быстро вращающихся валов, а также направляющих статического типа различных механизмов и приборов (в том числе направляющих тяжелых станков), работающих при малых скоростях скольжения практически без трения [3].

Газовая смазка исключает проблемы, связанные с граничной смазкой и приработкой подшипников: приработка в том смысле, как она производится при жидкостной смазке в подшипниках качения, не имеет места в опорах с газовой смазкой.

Следует подчеркнуть, что в принципе смазкой для газовых опор может служить любой газ или смеси газов. Они мало отличаются по тем свойствам, которые важны для подшипников, но воздушная смазка более практична и экономична. По этой причине, за исключением частных случаев, в качестве смазки используется воздух.

Применение газовой смазки дает следующие преимущества:

- обеспечивается герметичность системы смазки;
- уменьшается риск загрязнения, засорения и выхода из строя смазочной системы;
- исключается необходимость уплотнений валов и различных сложных лабиринтов;
- исключается громоздкое оборудование для хранения, подогрева и охлаждения, нагнетания и откачки жидкостных смазок (регуляторы давления, специальные фильтры и т.п.);
- упрощается и удешевляется конструкция подшипниковых узлов;
- снижаются вибрации и шумы, производимые охлаждающими устройствами.

Следует отметить, что при этом сокращаются габариты и вес всей аппаратуры, упрощается эксплуатация и надолго обеспечивается ее кондиционность [4].

Выбор метода обработки

Для выполнения всех требований к профилю спиральных канавок необходимо применение специальных методов обработки. На предприятии существует возможность выполнения канавок тремя способами:

1. Электрохимическая обработка. При этом возникают трудности с обеспечением разной глубины по длине канавки; существуют проблемы с достижением необходимого профиля канавки;;
2. Нанесение канавок лазером. При экспериментальной отработке этого метода выявлено, что параметры существующего лазера, такие как мощность и диаметр пучка, не позволяют обеспечить необходимую чистоту поверхности дна канавки;
3. Механическая обработка.

Ниже предложена одна из возможных схем получения канавок подшипников с аэродинамическим профилем путем механической обработки.

Прототип установки

Для формирования аэродинамических канавок способом механической обработки на предприятии применяется специальная установка (рис. 2). Установка представляет собой стол, на котором размещен шпиндель 1 с закрепленной на нем опорой и узел притира 7. Вращение притира обеспечивается двигателем с рядом зубчатых колес, передающих вращение и расположенных на узле крепления притира 7. Сам притир 8 подпружинен и прижимается к опоре. Узел притира имеет возможность вертикального перемещения за счет конической передачи 9. Предполагалось предусмотреть возможность регулирования угла расположения опоры 4 относительно основания с целью унификации установки, так как в разных комплексах используются опоры с разным углом наклона профиля канавки,. однако позднее было принято решение ограничиться использованием подставок с необходимыми углами наклона.

На шпинделе 1 располагается каретка 3 для регулировки положения опоры по плоскости. Вращение шпинделя осуществляется двигателем 2. Конструкторское решение размещения двигателя и его реверса является одним из самых больших недостатков данной установки. Реверс осуществляется за счет переключателей 5 и установленного на валу двигателя флагшка 6. При реверсе происходит удар флагшка о переключатель, и возникают ненужные вибрации корпуса. Помимо этого, реверс проходит в течение некоторого времени, создавая простой в работе двигателя, в то время как притир работает. Вследствие этого по краям профиля образуются углубления. Данные отклонения геометрии профиля могут приводить к

образованию воздушных завихрений и нарушению аэродинамических качеств опоры, что, в свою очередь, негативно сказывается на точности работы чувствительного элемента.

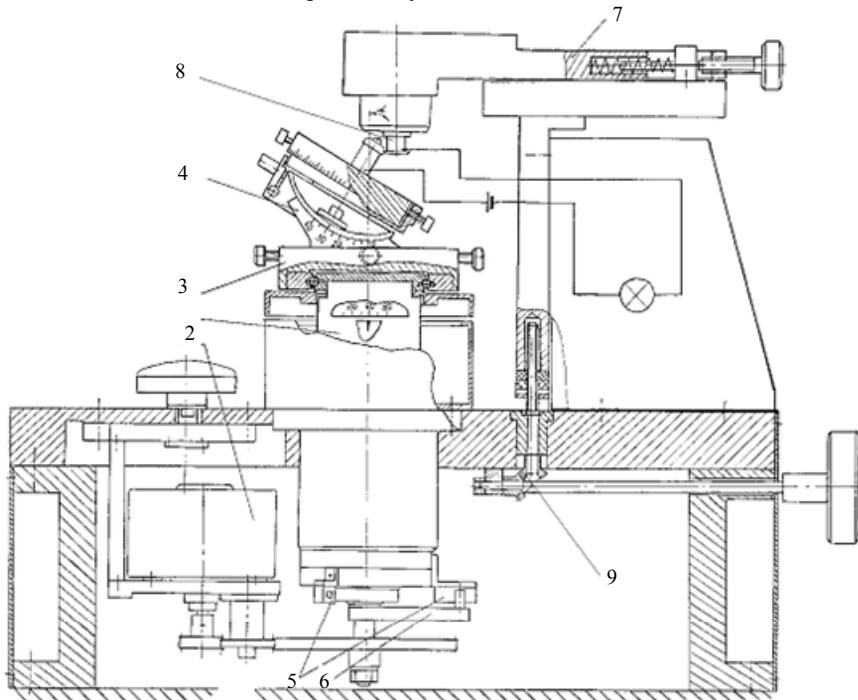


Рис. 2. Установка для формирования профиля: 1 – шпиндель; 2 – двигатель; 3 – каретка; 4 – угловая оправка для опоры; 5 – переключатель; 6 – флагок; 7 – узел притира; 8 – притир; 9 – коническая передача

Усовершенствованная установка для получения аэродинамического профиля

На рис. 3 представлено описание установки, подготовленной для внедрения и проходящей технологическую отработку на предприятии.

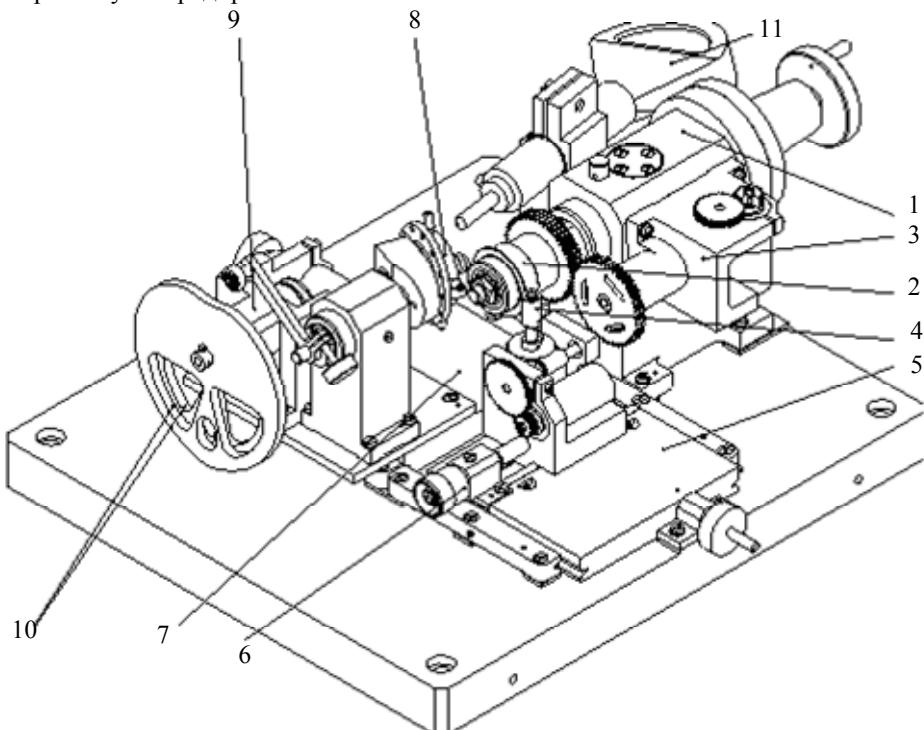


Рис. 3. Модернизированная установка для формирования аэродинамического профиля:
1 – узел крепления притира; 2 – притир; 3 – узел привода притира; 4 – контрпритир; 5 – каретка;
6 – микрометрический винт; 7 – каретка; 8 – узел крепления опоры; 9 – узел возвратно-
вращательного вращения опоры; 10 – кулачки; 11 – микрокатор

Установка предназначена для нанесения на сферическую поверхность опоры спиральных канавок глубиной у основания $0,004 \pm 0,0008$ мм, а в зоне вершины $0,003 \pm 0,0008$ мм.

Узел 1 отвечает за крепление притира 2 и его перемещение вдоль оси на ширину зацепления зубчатых колес. Узел 3 обеспечивает вращение притира с постоянной скоростью 55 об/мин. Сбоку к притиру прижимается контрпритир 4, расположенный на каретке 5. Контрпритир вращается с постоянной скоростью, немного отличной от скорости вращения самого притира. Разности скоростей вращения обеспечивают взаимодействие разных точек притира и контрпритира и, соответственно, более равномерный износ самого притира. Контрпритир для выставки имеет возможность перемещения за счет вращения микрометрического винта 6.

На нижней каретке 7 располагается узел крепления опоры 8, который обеспечивает ее расположение под необходимым углом, под которым будут наноситься канавки, и соосность центра сферы опоры с осью вращения. Также на нижней каретке располагается узел возвратно-вращательного перемещения 9, обеспечивающего вращение опоры на заданный угол. В этом узле предусмотрены два сменных кулачка 10, обеспечивающих нанесение сначала менее глубокой канавки, а затем более глубокой. Скорость вращения кулачков составляет 15 об/мин. Обе каретки с размещенными на них узлами ходят по направляющим и подпружинены для обеспечения контакта притира, с одной стороны, с контрпритиром, а с другой – с опорой. Для обеспечения возможности проведения замеров глубины канавки установлено крепление для микрокатора 11.

Профиль канавки образуется за счет абразивного износа между притиром и опорой, контрпритир с обратной стороны выполняет постоянную правку кромки притира. Для нанесения на опоре канавок по всему диаметру в узле предусмотрен делительный диск с фиксатором. После нанесения профиля каждой канавки опору поворачивают на одно деление диска, пока не будут нанесены все 12 канавок.

В установке можно выделить 6 основных узлов (рис. 4):

1. узел крепления и перемещения притира;
2. узел вращения притира;
3. узел правки кромки притира (контрпритир);
4. узел крепления опоры;
5. узел возвратно-вращательного движения опоры;
6. узел снятия заусенцев и наплы whole с притира.

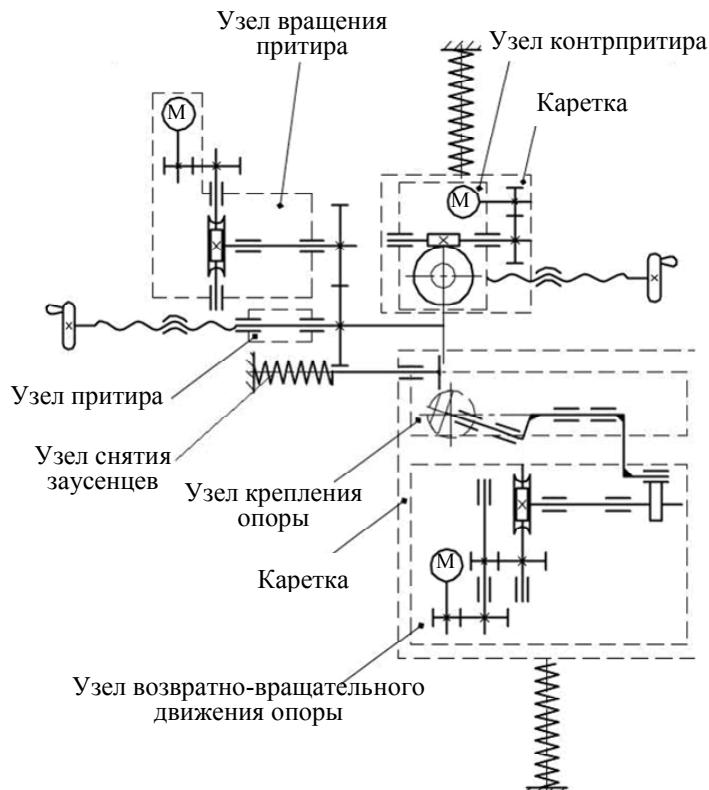


Рис. 4. Кинематическая схема установки

Отличительные особенности модернизированной установки

В предлагаемой конструкции было решено уйти от реверса двигателя. Для решения данной проблемы использована кулачковая передача, которая обеспечивает постоянное вращение двигателя. За счет правильного выбора профиля кулачка задается необходимый угол поворота опоры, причем, так как профиль имеет перепад глубин, необходимо использовать два сменных кулачка. Первым наносится профиль глубиной 3 мкм, затем, после переустановки коромысла, наносится более короткая часть канавки глубиной 4 мкм. Кулачки имеют вырезы с целью облегчения конструкции, а также приближения центра тяжести к оси для более плавного вращения.

Важной отличительной особенностью установки является использование узла контрпритира. Он используется для правки кромки вращающегося притира, что позволяет увеличить срок службы притира и стабильность наносимого профиля.

Было изготовлено и опробовано несколько притиров из разных материалов, таких как сталь 20, латунь ЛС59-1 и чугун СЧ15. В результате проведенных испытаний наилучшее качество канавок получено при формировании их чугунным притиром.

В процессе обработки детали на самом притире по краям могут образовываться наплывы и оставаться заусенцы. Для их устранения введен дополнительный узел, представляющий собой свободно вращающийся на подшипнике круглый закаленный нож с острой кромкой, который придвинут вплотную к торцу притира, сохраняя его свободное вращение, и расположен под притиром.

Общим недостатком существующей и модернизированной установок является сложность позиционирования детали относительно притира. В новой установке это усложняется необходимостью выставки контрпритира, который при неправильном расположении будет задавать другой профиль самого притира и соответственно канавки на опоре. В связи с этим предложено на притире сделать дополнительный буртик, относительно которого и производится выставка.

Заключение

Исходя из возможностей предприятия, наилучшим методом для формирования аэродинамического профиля опор газового подшипника, применяемого в гироприборе с бесконтактным подвесом чувствительного элемента, является метод механической обработки. Основываясь на недостатках в конструкции существующей установки, негативно влияющих на получаемый профиль, предложена кинематическая схема, разработана и изготовлена новая установка для формирования аэродинамического профиля опор. В процессе отработки выявлена необходимость введения дополнительного узла для снятия заусенцев в связи с добавлением в установку контрпритира. Проведено экспериментальное исследование притиров из разных материалов. Выявлено, что лучшие результаты показывает чугунный притир. Изготовлен комплект опор, полностью соответствующий всем геометрическим требованиям, предъявляемым к чувствительным элементам гироприбора.

Новая конструкция установки позволяет устраниТЬ дефекты профиля, получаемые при использовании реверсивного двигателя (углубления в крайних зонах канавок). Таким образом, обеспечивается профиль, максимально приближенный к прямоугольному, заданному конструкторской документацией на опору. Повышено качество профиля, чистота поверхности достигает 0,1 мкм. Уменьшается износ основных узлов, в первую очередь – притира.

Литература

1. Константинеску В.Н. Газовая смазка. – М.: Машиностроение, 1968. – 709 с.
2. Дроздович В.Н. Газодинамические подшипники. – Л.: Машиностроение, Ленингр. отд-ние, 1976. – 208 с.
3. Чернавский С.А. Подшипники скольжения. – М.: Машгиз, 1963. – 244 с.
4. Фролов К.В. Современная трибология. Итоги и перспективы. – М.: ЛКИ, 2008. – 408 с.

Родников Сергей Александрович – НИИ командных приборов, начальник сектора, poguHkoB@mail.ru

УДК 53.082.5+53.086+532.5.011

**МИКРОФЛЮИДНЫЕ ЧИПЫ ДЛЯ ИССЛЕДОВАНИЯ
БИОЛОГИЧЕСКИХ ОБЪЕКТОВ МЕТОДАМИ МИКРОСКОПИИ
ВЫСОКОГО РАЗРЕШЕНИЯ**

И.В. Кухтевич, А.С. Букатин, И.С. Мухин, А.А. Евстратов

Микрофлюидные устройства в сочетании с методами микроскопии высокого разрешения позволяют создавать новые аналитические системы для исследований биологических объектов в естественном состоянии. Сформулированы основные требования, предъявляемые к методам микроскопии при неразрушающем изучении биологических объектов. Рассмотрены и обсуждены варианты фиксации биологических объектов. Приведены примеры экспериментальных образцов микрофлюидных чипов для методов микроскопии высокого разрешения.

Ключевые слова: микрофлюидный чип, биологическая пробы, микро- и наноструктуры, конфокальная микроскопия, микроскопия близкого поля, атомно-силовая микроскопия, литография.

Введение

Концепция создания аналитических систем на микрочиповой платформе (микрофлюидных чипах), предложенная проф. А. Manz в конце 80-х г.г. ХХ в., явилась основой новых высокопроизводительных приборов [1], получивших за рубежом названия «lab-on-a-chip» (лаборатория на чипе) и «micro Total Analysis Systems» (микросистемы «полного» анализа) [2]. Занятость государственных организаций и коммерческих фирм в создании микроаналитических систем и новые возможности современных технологий вызвали прогресс в разработке приборов на микрочиповой платформе, изучении процессов и явлений в микро- и наноразмерном масштабе, создании новых методов синтеза и анализа веществ на микрочипах, исследовании возможностей микрофлюидики [3]. За последние 10 лет по теме микрофлюидика (microfluidic), по данным ISI Web of Science, было издано более 10 тыс. статей, и динамика публикаций непрерывно растет. За рубежом по микроаналитическим системам и смежным направлениям издается значительное число монографий и книг [4]. Современным развитием микрофлюидики является нафлюидика [5], изучающая эффекты в наноразмерных системах, транспорт молекул через наноканалы, взаимодействие с наноструктурами и т.д. Развитие микрофлюидики привело к появлению приборов, в которых осуществлялось воспроизведенное управлениеnano- и пиколитровыми объемами жидкости. В настоящее время микрофлюидные устройства востребованы в научной сфере и в промышленности как новый многообещающий инструмент для развития и создания новых методов и продуктов.

Для получения новых знаний важным является разработка высокочувствительных методов и приборного обеспечения для исследования биологических объектов (клеток, бактерий, вирусов и т.д.) в естественном состоянии. К наиболее эффективным системам при реализации неразрушающих методов анализа следует отнести микрофлюидные чипы (МФЧ), а регистрация изменений, происходящих в объектах под воздействием внешних факторов, может осуществляться разными методами, но достигнуть высокого пространственного разрешения можно, используя методы конфокальной лазерной сканирующей микроскопии (КЛСМ), сканирующей ближнепольной оптической микроскопии (СБОМ), атомно-силовой микроскопии (АСМ). Визуализация в жидкости требует стабильной иммобилизации или фиксации биологических объектов. По этой причине поверхность материала часто подвергают специальной химической обработке, где впоследствии фиксируются биологические объекты. Но это может приводить к разрушению функциональных групп и повреждению структуры изучаемого объекта. Актуальным является разработка альтернативных методов фиксации объекта, в том числе на основе физических методов, не повреждающих биообъект. Микрофлюидные технологии позволяют осуществлять манипуляции с исследуемым объектом, воздействовать на объект, фиксировать его на заданном участке и измерять характеристики объекта, в том числе реологические [6]. Известны системы для исследований биологических объектов (клеток), сочетающие методы микроскопии высокого разрешения (МВР) и микрочиповые технологии [7, 8]. При этом стремятся создать различные гибридные методы и техники, обеспечивающие получение новой информации об объектах.

В настоящей работе приведены некоторые результаты исследований и работ авторов по созданию МФЧ для исследований биологических объектов достаточно больших размеров (1–7 мкм) методами МВР.

Исследование биологических объектов методами МВР

В процессе исследований были сформулированы следующие основные требования, которые необходимо соблюдать при изучении биологических объектов в естественном состоянии методами МВР:

- при наличии зондирующего излучения (светового потока) плотность мощности излучения не должна приводить к негативному воздействию на объект, при этом длительность воздействия излучения должна быть минимизирована;
- если в процессе измерений исследуемый объект подвергается нагреву, то должен быть предусмотрен отвод избыточного тепла от нагреваемого пространства;
- применение красителей и флуорофоров для усиления контраста или дополнительной индикации элементов структуры объекта не должно приводить к изменению функций исследуемого объекта;
- при сканировании объекта с применением методов зондовой микроскопии должно быть обеспечено минимальное физическое воздействие на объект, что может быть получено при использовании специальных зондов и полуконтактных методов исследований, кроме того, необходимо использовать малые амплитуды колебаний зонда;
- для получения достоверной и воспроизводимой информации время сканирования объектов должно быть минимизировано;
- в методах оптической микроскопии для достижения наилучшего разрешения объектив должен быть максимально приближен к изучаемому объекту, что приводит к определенным требованиям к реакционной камере, где размещен исследуемый объект;
- для методов зондовой микроскопии необходимо обеспечить доступ зонда к исследуемому объекту.

При исследовании биологических препаратов процесс подготовки образцов имеет определяющее значение для любого вида микроскопии. Методы и способы фиксации клетки в микрочиповых устройствах подразделяются на иммунологические и неиммунологические (физические). Иммунологические методы, в которых используются высокоспецифичные иммунные реакции между мембранными белками и антителами, реализованы в коммерческих продуктах многих биотехнологических фирм. Недостатками их являются дороговизна реагентов, необходимость специального оборудования и высокие требования к квалификации исследователя. Эти методы непрестанно развиваются, появляются новые методики фиксации объектов на различных подложках. Получила широкое распространение иммобилизация клеток путем включения в различные гели, мембранны, волокна, основанная на химических и физических взаимодействиях. При такой иммобилизации клетки могут сохранять жизнеспособность и в присутствии питательной среды размножаться в приповерхностных слоях гелей. Но такой способ подходит только для методов оптической микроскопии и для КЛСМ.

Приготовление образцов для зондовой микроскопии предполагает иммобилизацию объекта на поверхности подложки, что приводит к дополнительным экспериментальным и методическим трудностям. Например, в случае бактерий вместо такой характеристики, как диаметр бактериальной клетки, появляются две новые – ширина и высота, причем высота меньше ширины из-за «сплющивания» клетки при ее фиксации к поверхности подложки. При получении изображений методом сканирующей зондовой микроскопии общепризнанным методом фиксации является обработка глутаровым альдегидом [9]. Поскольку при обработке глутаровым альдегидом клетка сохраняет воду и не расплывается по подложке, то считается, что ее изображение идентично изображению нативных клеток. При исследовании образцов в жидкости возникает проблема фиксации биологических объектов. В этом случае на поверхность подложки, на которой должны располагаться изучаемые объекты, наносится поли-L-лизин, лак (раствор нитроцеллюлозы и дубутилфталата в этилацетате и бутилацетате) или агароза.

В рамках данного проекта исследован вариант методики модификации стеклянной поверхности микрочипа для селективной фиксации биологического объекта, содержащей стадии активации поверхности, силанизации, обработки глутаровым диальдегидом и иммобилизации чувствительного слоя (белка или других агентов) [10]. Такой метод позволяет реализовать высокоселективную реакцию, например, взаимодействие антиген–антитело, с последующим детектированием образованного комплекса методами МВР.

К физическим методам фиксации биологических объектов в микрочипе следует отнести механические и гидродинамические «ловушки» с микро- и наноразмерными структурами. Суть действия такой «ловушки» заключается в том, что под действием ламинарных потоков изучаемый объект попадает в область, где его движение ограничено или затруднено. Структурами, затрудняющими движения объекта могут являться микро- и наноканалы с размерами меньше размеров объекта. Эти методы можно назвать пассивными, так как при этом воздействие на исследуемый объект минимально. К другим физическим методам следует отнести те, в которых применяются электрические, магнитные, электромагнитные и ультразвуковые поля [11–13]. Эти методы можно определить как активные, так как они в какой-то степени меняют исходные характеристики объекта. Физические методы позволяют создавать быстрые и простые технологии фиксации объектов. Наиболее простыми и эффективными методами для применения в микрочиповых устройствах являются механические и гидродинамические «ловушки» с микро- и нанопористыми мембранами, микро- и наноструктурами в канале или реакторе чипа. В работе уделялось внимание созданию МФЧ с функциональными микро- и наноструктурами для удерживания и фиксации клетки. Другим перспективным методом удержания и фиксации клетки является диэлектрофорез (ДЭ) [14]. Из-за простоты ДЭ широко используется в микрочиповых приборах для разделений бактерий, рако-

вых клеток, стволовых клеток, субпопуляций лейкоцитов. Различные типы клеток могут быть отделены с одной и той же конструкцией электрода при соответствующей частоте для каждого типа клетки. Преимуществом ДЭ является гибкость, полная контролируемость, удобство для автоматизации, а недостатком – нагрев системы и необходимость подбора буферного раствора с требуемой диэлектрической проницаемостью и проводимостью, а это означает, что клетки не могут быть непосредственно отделены в исходном образце.

МФЧ для микроскопии высокого разрешения

Существует множество технологий и методов формирования микро- и наноструктур в различных подложках, но выбор нужной технологии всегда в значительной степени определяется применяемым материалом и характеристиками формируемых структур. Все технологии (например, лазерная микрообработка, кислотное травление, ионно-реактивное травление и т.п.) связаны с активным воздействием на поверхность материала, что приводит к изменению свойств поверхности. Подробно с технологиями и методами формирования микроструктур можно ознакомиться в соответствующих обзорах и книгах [15, 16].

Микроразмерные каналы и реакторы для МФЧ из стекла марки K8 были получены методами фотолитографии и кислотного травления, а сеть наноразмерных каналов между микроканалами – методом ионной литографии [17]. Герметизация микро- и наноканалов осуществлялась разными методами: термического связывания (спекания), с применением полидиметилсилоксановых пленок или с помощью фотоотверждаемых kleев [18].

Были изготовлены и исследованы экспериментальные образцы МФЧ для методов КЛСМ, для СБОМ и АСМ: микрочипы с квадрупольными электродами для дилектрофоретических методов и микрочипы с системой наноканалов для фиксации микрообъектов [19].

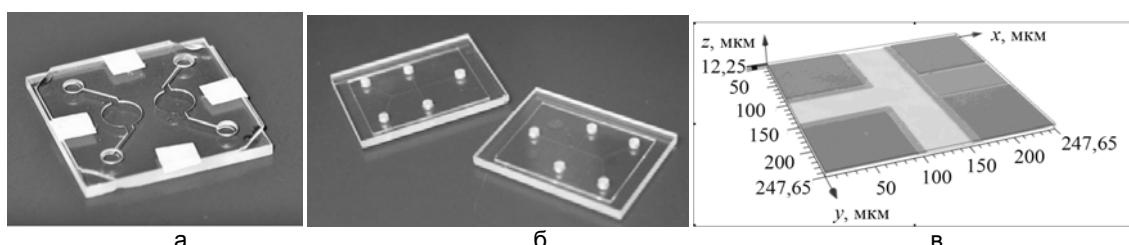


Рис. 1. Микрофлюидные чипы для КЛСМ: (а) – двухканальный микрочип с реакционными камерами (размеры МФЧ 24×24 мм, толщина 3 мм); (б) – чип с топологией «ступенька» (размеры МФЧ 30×40 мм, толщина 3 мм); (в) – область пересечения каналов МФЧ с топологией «ступенька» (изображение получено на КЛСМ Leica TCS SL)

МФЧ (рис. 1, а) предназначен для исследований биологических объектов методами оптической микроскопии и КЛСМ в реакционных камерах диаметром 5 мм и глубиной 20 мкм, имеет два независимых канала и реактора. Герметизация чипа проведена фотоотверждающим kleем Rite Lock. Размеры чипа составляют 24×24 мм, толщина – 3 мм. МФЧ для исследований биологических объектов, изображенный на рис. 1, б, позволяет отделить объекты размером более 5 мкм на ступеньке, сформированной на пересечении каналов (рис. 1, в). Выделенные объекты могут быть перенаправлены в боковые каналы. В микрочипе имеется отдельный транспортный канал, обеспечивающий стабильный поток буферного раствора, канал ввода пробы и реакционный канал (ширина 60 мкм, глубина 15 мкм), где происходит смешивание пробы и буфера. Такая универсальная топология позволяет реализовать широкий круг различных методик при исследовании образцов.

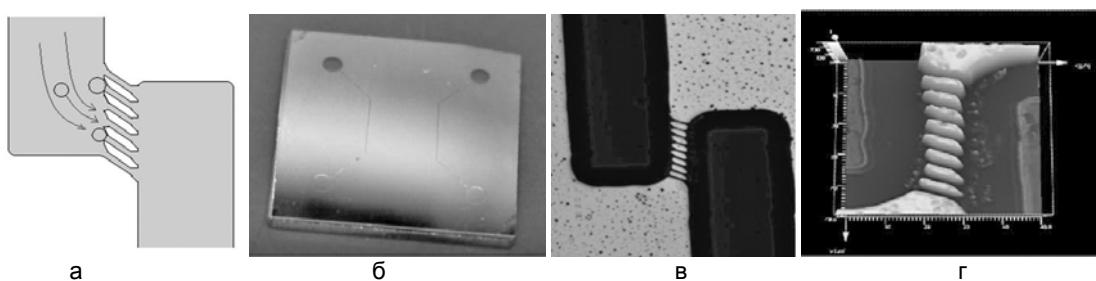


Рис. 2. Принцип действия гидродинамической ловушки (а) и изображения: МФЧ с наноканалами (б); наноканалов, соединяющих микроканалы: (в) – изображение получено на КЛСМ Leica TCS SL, длина волны сканирования 488 нм, (г) – 3D-реконструкция). Размеры МФЧ 24×24 мм, толщина 3 мм

Особенностью МФЧ для СБОМ и АСМ является необходимость открытого доступа зонда к исследуемому объекту. Это подразумевает, что микроканал или реактор, в котором располагается объект в естественной жидкой среде, имеют контакт с воздушной средой. При этом наблюдается испарение жидкости, которое должно быть каким-то способом скомпенсировано. Были выбраны геометрические харак-

теристики канала, обеспечивающие максимальный приток жидкости на порядок больше, чем скорость испарения – 10 нл/с. Другим эффективным методом является применение жидких масел для предотвращения испарения жидкости, но это влияет на процедуру и точность измерений.

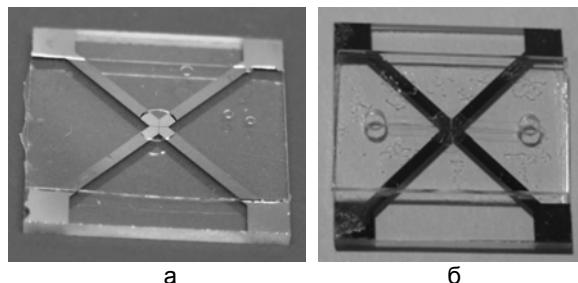


Рис. 3. МФЧ для диэлектрофореза: (а) – с открытой реакционной камерой (для АСМ и СБОМ); (б) – с закрытой реакционной камерой и подводящими каналами (для КЛСМ). Размеры МФЧ 24×24 мм, толщина 1,3 мм

На рис. 2, а, поясняется принцип действия гидродинамической ловушки. МФЧ с системой микроканалов имеет размеры 24×24 мм, толщину около 1,3 мм (рис. 2, б). Между микроканалами (ширины 50 мкм) расположена сеть из 10 наноканалов шириной 260–270 нм (рис. 2, в, г). Две разных конструкции МФЧ для ДЭ приведены на рис. 3. Микрочип (рис. 3, а) с открытой реакционной камерой предназначен для исследований методами АСМ и СБОМ, другой микрочип (рис. 3, б) – с закрытой реакционной камерой и подводящими каналами для буферного раствора и пробы – предназначен для оптической микроскопии и КЛСМ. В последнем случае необходимо использование при измерениях микрообъектива с рабочим расстоянием не менее 0,5 мм.

Все конструкции МФЧ были апробированы на модельных системах – буферных растворах, растворах флуоресцеина и полистирольных латексах размером 3 мкм и 6 мкм. Более подробную информацию о чипе для ДЭ можно найти в [19].

Заключение

Сочетание микрофлюидных технологий и методов микроскопии высокого разрешения позволяет создавать приборы и системы для исследований живых биологических объектов в естественной среде. При этом возникает ряд научных и технических задач, требующих решения. В методах микроскопии высокого разрешения необходимым является:

- фиксация объекта на время измерений;
- создание условий, при которых оказывается минимальное воздействие на изучаемый объект.

Авторами разработаны, изготовлены и апробированы несколько вариантов экспериментальных образцов микрофлюидных чипов для оптической микроскопии, КЛСМ, СБОМ и АСМ, позволяющих осуществлять исследования биологических объектов. Работоспособность микрофлюидных чипов была проверена на тестовых образцах и модельных системах.

Работа проведена при поддержке: АВЦП «Развитие научного потенциала высшей школы (2009–2011)» (РНП 2.1.2/9501); ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (ГК П557, ГК 14.740.11.0451, ГК 14.740.11.1218); Программы У.М.Н.И.К.

Литература

1. Manz A., Gruber N., Widmer H.M. Miniaturized total chemical analysis systems: A novel concept for chemical sensing // Sensors and Actuators B: Chemical. – 1990. – № 1. – P. 244–248.
2. Reyes D., Iossifidis D., Auoux P., Manz A. Micro total analysis systems. 1. Introduction, theory and technology // Analytical Chemistry. – 2002. – № 74. – P. 2623–2636.
3. Encyclopedia of Microfluidics and Nanofluidics / Editor-in-Chief Li Dongqing. – New York: Springer Science & Business Media, 2008. – 2226 p.
4. Microarrays. Preparation, Microfluidics, Detection Methods and Biological Applications / Ed. K. Dill, R.H. Liu, P. Grodzinski. – New York: Springer Science & Business Media, 2009. – 356 p.
5. Nanofluidics. Nanoscience and Nanotechnology / Ed. by J.B. Edel, A.J. deMello. – Cambridge: The Royal Society of Chemistry, Thomas Graham House, 2009. – 198 p.
6. Nagai M., Oishi M., Oshima M., Asai H., Fujita H. Three-dimensional two-component velocity measurement of the flow field induced by the Vorticella picta microorganism using a confocal micro-PIV technique // Biomicrofluidics. – 2009. – V. 3. – P. 014105.
7. Meister A., Gabi M., Behr P., Studer P., Vörös J., Niedermann P., Bitterli J., Polesel-Marais J., Liley M., Heinzelmann H., Zambelli T. FluidFM: combining atomic force microscopy and nanofluidics in a universal liquid delivery system for single cell applications and beyond // Nano Lett. – 2009. – V. 9. – № 6. – P. 2501–2507.

8. Flores S.M., Toca-Herrera J.L. New future of scanning probe microscopy: Combining atomic force microscopy with other surface-sensitive techniques, optical microscopy and fluorescence techniques // Nanoscale. – 2009. – № 1. – P. 40–49.
9. Dufrêne Y.F. Application of atomic force microscopy to microbial surfaces: from reconstituted cell surface layers to living cells // Micron. – 2001. – V. 32. – № 2. – P. 153–165.
10. Евстратов А.А., Есикова Н.А., Рудницкая Г.Е., Антропова Т.В., Анфимова И.Н. Разработка оптического сенсорного элемента для микрофлюидных чипов на основе натриево-бороシリкатного пористого стекла // Научное приборостроение. – 2010. – Т. 20. – № 1. – С. 52–58.
11. Tsutsui H., Ho C.-M. Cell separation by non-inertial force fields in microfluidic systems // Mech. Res. Comm. – 2009. – № 36. – P. 92–103.
12. Евстратов А.А. Физические методы управления движением и разделением микрочастиц в жидкых средах. Ч. 1. Диэлектрофорез, фотофорез, оптофорез, оптический пинцет // Научное приборостроение. – 2005. – Т. 15. – № 1. – С. 8–21.
13. Manneberg O., Vanherberghen B., Önfelt B., Wiklund M. Flow-free transport of cells in microchannels by frequency-modulated ultrasound // Lab Chip. – 2009. – № 9. – P. 833–837.
14. Voldman J. Electrical Forces For Microscale Cell Manipulation // Annu. Rev. Biomed. Eng. – 2006. – № 8. – P. 425–454.
15. Lab-on-a-Chip Technology: Fabrication and Microfluidics / Ed. K.E. Herold and Rasooly A. – Norwich: Caister Academic Press, 2009. – V. 1. – 410 p.
16. Becker H., Gärtner C. Polymer microfabrication technologies for microfluidic systems // Anal. Bioanal. Chem. – 2008. – V. 390. – № 1. – P. 89–111.
17. Евстратов А.А., Мухин И.С., Кухтевич И.В., Букатин А.С. Применение ионной литографии для формирования наноразмерных каналов микрофлюидных чипов в стеклянных подложках // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 4 (68). – С. 59–63.
18. Евстратов А.А., Лукашенко Т.А., Тупик А.Н. Применение фотоотверждаемых оптических клеев для герметизации аналитических микрочипов // Научное приборостроение. – 2010. – Т. 20. – № 1. – С. 29–38.
19. Кухтевич И.В., Букатин А.С., Евстратов А.А., Мухин И.С. Создание аналитической установки для биологических исследований на основе оптического микроскопа Axio Observer D1 и микрочиповых технологий. Ч. 1 // Научное приборостроение. – 2010. – Т. 20. – № 3. – С. 3–8.

Кухтевич Игорь Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, ba@inbox.ru

Букатин Антон Сергеевич

– Санкт-Петербургский академический университет – научно-образовательный центр нанотехнологий РАН, аспирант, antbuk.fiztek@gmail.com

Мухин Иван Сергеевич

– Санкт-Петербургский академический университет – научно-образовательный центр нанотехнологий РАН, аспирант, imukhin@yandex.ru

Евстратов Анатолий Александрович

– Институт аналитического приборостроения РАН, кандидат технических наук, доцент, an_evs@mail.ru

УДК 003.26.09

РАЗРАБОТКА СТЕГАНОАЛГОРИТМА НА БАЗЕ ФОРМАТНЫХ И ПРОСТРАНСТВЕННЫХ ПРИНЦИПОВ СОКРЫТИЯ ДАННЫХ

А.Г. Коробейников, С.С. Кувшинов, С.Ю. Блинов, А.В. Лейман, С.И. Нестеров

Рассмотрена задача разработки стеганоалгоритма на базе форматных и пространственных методов сокрытия данных. Разработаны алгоритмы встраивания и извлечения информации из стегоконтейнера. Предложена общая схема работы системы скрытой передачи.

Ключевые слова: стеганография, стегоконтейнер, дискретное косинусное преобразование, вейвлет-преобразование.

Введение

В настоящее время в связи с бурным развитием информационных технологий возникают задачи как по криптографической защите данных, так и по сокрытию факта передачи информации [1, 2]. Задача сокрытия данных во внешне безобидных контейнерах с целью скрытой передачи встает, например, при необходимости защиты переписки в сети, которая заставляет абонентов сети иногда использовать методы компьютерной стеганографии (КС) для сокрытия факта переписки. Использование цифровой графики в качестве стегоконтейнеров обусловлено следующими причинами:

- высокая степень распространения цифровой графики;
- популярность и простота процессов обмена цифровыми фотографиями и опубликования их в сети Интернет;
- удобный объем контейнера с точки зрения операций работы с файлами (аудиофайлы и видеофайлы, как правило, в среднем имеют больший объем, чем цифровые изображения);
- особенности системы человеческого зрения, не позволяющие визуально определить наличие незначительных изменений контейнера.

В настоящей работе предложены методы для встраивания сообщений в изображения формата JPEG. Рассмотрим основные методы КС, работающие с данным форматом.

Классификация методов стеганографии

Форматные методы. Форматные методы, по мнению специалистов, не относятся к КС в чистом виде, поскольку не связаны с цифровой обработкой сигналов. Они основаны на избыточности форматов компьютерных данных, например, структуре файлов, IP-пакетов. Цифровые изображения также являются сигналами, имеющими, однако, «застывший» характер. С этой точки зрения при работе с алгоритмом, дописывающим в конец файла JPEG-байты файла RAR, нельзя строго говорить о КС изображений, поскольку это не что иное, как форматный метод в КС.

Предлагается следующее определение: форматный метод в КС изображений – это метод, осуществляющий такое преобразование изображения, при котором вносимые изменения не инициируют визуализацию артефактов встраивания данных и в то же время учитываются и используются в процессе декодирования файла изображения программой просмотра изображений в соответствии со спецификацией формата JPEG.

Данное определение не делает классификацию метода зависимой от конкретной программы просмотра при условии соблюдения программами спецификации формата JPEG.

Стеганографический алгоритм, представляемый в работе, использует именно такой форматный подход к сокрытию информации в графическом изображении. Отметим, что сокрытие в формате осуществляется не в потоке собственно сообщения, а в сигнальной и служебной информации, не несущей смысловой нагрузки. Это позволяет говорить о том, что в представляемом комбинированном стегоалгоритме форматная составляющая и исходное сообщение развязаны.

Стеганоалгоритмы пространственной области. Алгоритмы данного типа внедряют информацию в области самого изображения. Их преимуществом является то, что для внедрения нет необходимости выполнять вычислительно громоздкие линейные преобразования изображений. Данные внедряются за счет манипуляций цветовыми составляющими или яркостью.

Стеганоалгоритмы области преобразования. Наиболее популярны в стеганографии следующие преобразования:

1. дискретное косинусное преобразование (ДКП);
2. вейвлет-преобразование (ВП).

ДКП используется в алгоритме сжатия JPEG, что является большим стимулом использования ДКП в стеганографии JPEG. ВП, в свою очередь, – основа сжатия в алгоритме JPEG 2000.

ДКП может применяться как ко всему изображению в целом, так и к отдельным блокам пикселей изображения. Обычно же контейнер разбивается на блоки размером 8×8 пикселей. ДКП применяется к каждому блоку, в результате чего получаются матрицы коэффициентов ДКП, также размером 8×8 [2]. Коэффициенты будем обозначать через $c_b(j, k)$, где b – номер блока, (j, k) – позиция коэффициента внутри блока. Если блок сканируется в зигзагообразном порядке (как это имеет место в JPEG), то коэффициенты будем обозначать через c_{bj} . Коэффициент в левом верхнем углу, $c_b(0,0)$, обычно называется DC-коэффициентом. Он содержит информацию о яркости всего блока. Остальные коэффициенты называются AC-коэффициентами. Иногда выполняется ДКП всего изображения, а не отдельных блоков [3].

Рассмотрим процесс внедрения/извлечения информации в области ДКП на примере алгоритма Koch [4]. В данном алгоритме в блок размером 8×8 осуществляется встраивание 1 бита цифрового водяного знака (ЦВЗ). Известны две реализации алгоритма: псевдослучайно могут выбираться два или три коэффициента ДКП. Предлагается модификация алгоритма с двумя выбираемыми коэффициентами (s_i).

Встраивание информации осуществляется следующим образом: для передачи бита 0 добиваются того, чтобы разность абсолютных значений коэффициентов была бы больше некоторой положительной величины ε , а для передачи бита 1 эта разность делается меньше некоторой отрицательной величины $-\varepsilon$:

$$\begin{cases} |c_b(j_{i,j}, k_{i,1})| - |c_b(j_{i,j}, k_{i,2})| > \varepsilon, & \text{если } s_i = 0, \\ |c_b(j_{i,j}, k_{i,1})| - |c_b(j_{i,j}, k_{i,2})| < -\varepsilon, & \text{если } s_i = 1. \end{cases}$$

Таким образом, исходное изображение искажается за счет внесения изменений в коэффициенты ДКП. Для чтения ЦВЗ в декодере выполняется та же процедура выбора коэффициентов, и решение о переданном бите принимается согласно следующему правилу:

$$\begin{cases} s_i = 0, & \text{если } |c_b(j_{i,j}, k_{i,1})| > |c_b(j_{i,j}, k_{i,2})|, \\ s_i = 1, & \text{если } |c_b(j_{i,j}, k_{i,1})| < |c_b(j_{i,j}, k_{i,2})|. \end{cases}$$

Разработка стеганоалгоритма

Для успешной работы с JPEG-файлом необходимо знать его структуру, а также алгоритм сжатия с потерями [5]. Поведение программ просмотра JPEG напрямую зависит от структуры файла, и внесение в него изменений стеганографическим алгоритмом не должно провоцировать визуализацию факта изменений. Исходя из структуры формата файлов JPEG, можно утверждать, что следующие маркеры определяют сегменты, не участвующие в JPEG-преобразовании и не влияющие на визуализацию изображения, а потому игнорируемые программами просмотра:

- COM;
- APP15;
- DAC;
- DNL;
- SOF2 – SOF10;
- неспецифицированные сегменты.

Форматная составляющая разработанного метода предполагает работу как раз с такими сегментами. В качестве подготовительной работы были разработаны и реализованы в виде программ-парсеров алгоритмы разбора (парсинга) файлов JPEG и BMP. Возможности данных модулей позволяют получить структурированное представление блоков, для дальнейшего анализа потенциальных мест для внедрения битов сообщения.

Данная система обеспечивает скрытую передачу сообщения с использованием в качестве контейнера передачи неподвижное изображение формата JPEG. Сообщение встраивается в графический файл-контейнер, далее файл доставляется адресату, адресат извлекает из полученного изображения текст сообщения (рис. 1).

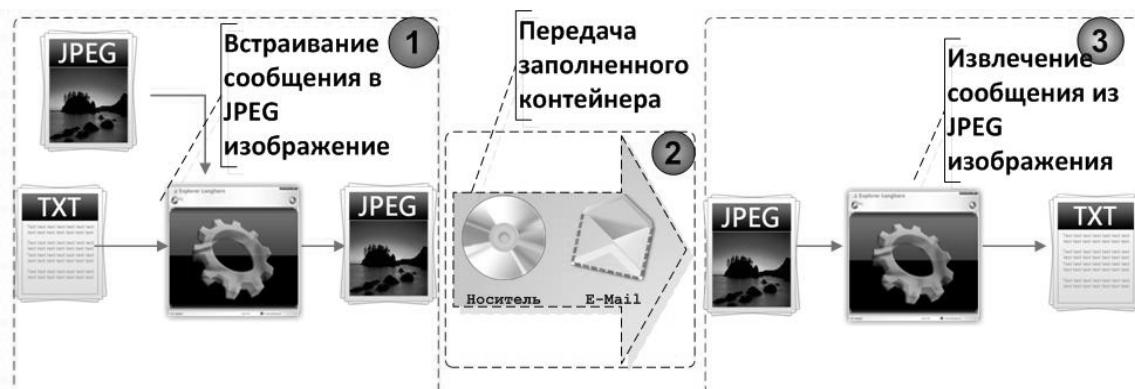


Рис. 1. Общая схема работы системы скрытой передачи

Процессы встраивания и извлечения автоматизированы и выполняются с помощью разработанного приложения. Этап 2 является неконтролируемым, и методы извлечения и встраивания не зависят от канала передачи. В результате получается извлеченное сообщение, идентичное передаваемому, при условии, что передаваемое изображение с внедренным сообщением (стегопосылка) не менялось с момента завершения процесса встраивания данных отправителем до момента начала их извлечения получателем. При несоблюдении указанной целостности программа не гарантирует тот факт, что получатель сможет извлечь полезную для него информацию (сообщение). Схема алгоритма внедрения представлена на рис. 2.

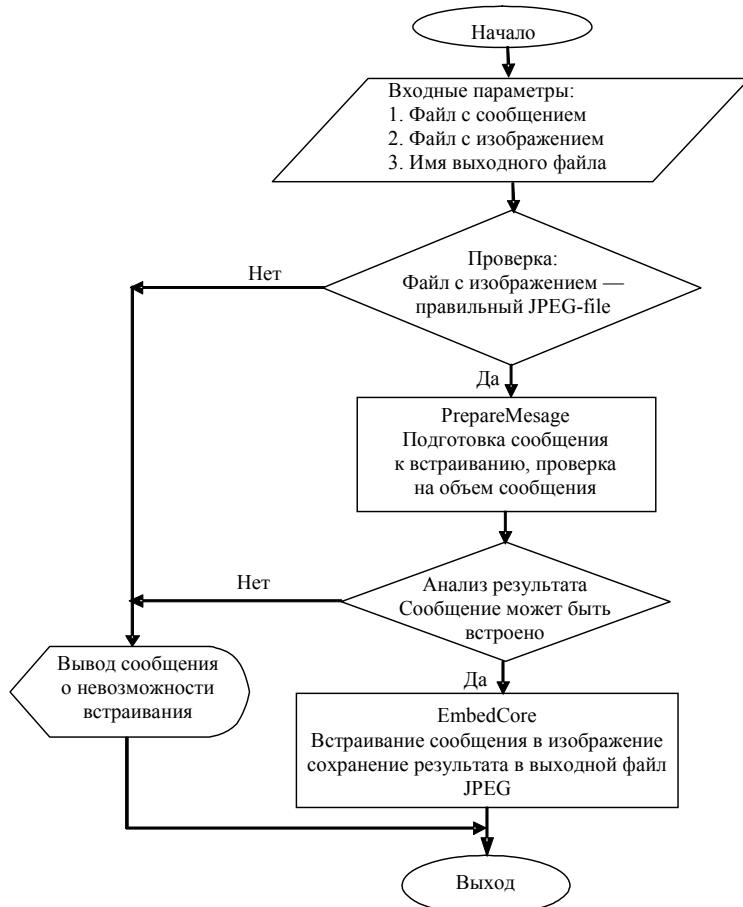


Рис. 2. Общая схема алгоритма внедрения

Система принимает входные данные, проверяет их корректность:

- существование входных файлов на носителе;
- соотношение размеров входных файлов;
- соответствие файла-контейнера формату JPEG.

Алгоритм внедрения включает следующие основные шаги.

На первом этапе производится преобразование потока данных JPEG в поток данных BMP. При этом увеличивается размер потока за счет изменения принципа кодирования информации о цветовых свойствах участков изображения. За счет того, что в формате BMP каждая точка изображения кодируется тремя байтами, отвечающими за вклад основных цветов (R – красного, G – зеленого и B – синего) в целевой цвет точки, изменение размера потока в большую сторону значительно и позволяет встроить необходимый объем информации. Известно, что человеческий глаз наименее чувствителен к изменениям в оттенках синего цвета, поэтому для встраивания используются B-составляющие RGB-структур. Для минимизации объема изменения в алгоритме по умолчанию используется только младший бит такого байта, что до минимума снижает вероятность обнаружения изменения даже на изображениях с большой площадью заливки синего цвета. Далее реализуются режимы, позволяющие использовать 2, 3 и 4 LSB. Простейший способ замены битов – последовательная замена в каждом b-байте.

Разработанный механизм компенсирования потерь при межформатных преобразованиях предназначен для успешного извлечения данных на принимающей стороне. В рамках этого механизма после проведения встраивания выполняется следующая последовательность действий. Измененные b-биты выделяются из потока и копируются во временный буфер – буфер 1. В другой временный буфер (буфер 2) сохраняется информация о том, как встраивались биты сообщения – в каждый байт подряд или через

один, два, т.д. Эта процедура может не выполняться, если порядок встраивания – последовательная замена бита в каждом b -байте. Далее модифицированный поток байтов BMP подвергается JPEG-сжатию с наивысшим коэффициентом качества. Полученный JPEG-поток на этой стадии еще не сохраняется в выходной файл. Для обеспечения возможности последующего извлечения производятся дополнительные действия. Производится попытка извлечения сообщения, которая заключает в себе:

- клонирование потока JPEG во временный поток;
- декодирование временного потока JPEG в поток байтов BMP;
- извлечение модифицированных b -байтов, в которые были встроены биты сообщения с учетом информации из буфера 2.

Далее следует анализ извлеченного сообщения, т.е. сравнение извлеченных b -байтов с b -байтами, сохраненными в буфере 1, и сохранение разностей в новый буфер – буфер 3. Эти разности являются неизбежным результатом потерь при межформатных преобразованиях JPEG – RGB, BMP – JPEG.

При реализации представленного стеганоалгоритма в виде программного обеспечения необходимо учитывать следующие особенности:

- сохранение сообщения в зоне LSB-файла BMP позволяет скрыть большое количество байтов сообщения, однако факт такого сокрытия легко обнаружить;
- сохранение сообщения в игнорируемых сегментах файла JPEG позволяет скрыть большое количество байтов сообщения, однако факт такого сокрытия легко обнаружить;
- количество битов, пригодных для встраивания, зависит от исходного JPEG-файла (характер зависимости определяется JPEG-преобразованием) и линейно зависит от числа LSB, используемых для встраивания.

Заключение

В работе представлен стеганоалгоритм внедрения сообщения с использованием в качестве контейнера передачи неподвижное изображение формата JPEG. Кроме того, разработан механизм компенсирования потерь при межформатных преобразованиях, предназначенный для извлечения данных на принимающей стороне.

Дальнейшее развитие данного направления сопряжено с разработкой систем принятия решений, необходимых для программной реализации стеганографических дополнений к web-браузерам и почтовым клиентам.

Литература

1. Коробейников А.Г., Воробьев А.О., Сидоркина И.Г., Пылин В.В. Анализ криптографической стойкости алгоритмов асимметричного шифрования информации // Изв. вузов. Приборостроение. – 2007. – Т. 50. – № 8. – С. 28–32.
2. Коробейников А.Г., Прохожев Н.Н., Михайличенко О.В., Хоанг З. Выбор коэффициентов матрицы дискретно-косинусного преобразования при построении стеганографических систем // Вестник компьютерных и информационных технологий. – 2008. – № 11. – С. 12–17.
3. Коробейников А.Г., Михайличенко О.В., Прохожев Н.Н. Оценка устойчивости ЦВЗ к внешним воздействиям, внедренных с помощью алгоритмов пространственной области встраивания // Научно-технический вестник СПбГУ ИТМО. – 2008. – Вып. 51. – С. 168–172.
4. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. – Киев: МК-Пресс, 2006. – 288 с.
5. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ-МИФИ, 2003. – 384 с.

**Коробейников Анатолий
Григорьевич**

– Институт земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова РАН, Санкт-Петербургский филиал, доктор технических наук, профессор, зам. директора, Kogobeynikov_A_G@mail.ru

Кувшинов Станислав Сергеевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, ss.kuvshinov@gmail.com

Блинов Станислав Юрьевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, stas-blino@yandex.ru

Лейман Альберт Владимирович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, allxxl@yandex.ru

Нестеров Сергей Игоревич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, nesterov.serge@gmail.com

УДК 654.924

МЕТОДИКА ОЦЕНКИ ВЕРОЯТНОСТИ ОБНАРУЖЕНИЯ НЕСАНКЦИОНИРОВАННОГО ПРОНИКНОВЕНИЯ ОПТИКО-ЭЛЕКТРОННЫМ ИЗВЕЩАТЕЛЕМ

В.В. Волхонский, П.А. Воробьев

Предложена методика оценки вероятности обнаружения оптико-электронным извещателем. Выполнено исследование плотности распределения вероятности расстояния обнаружения несанкционированного проникновения. Получены экспериментальные данные, позволяющие оценить вероятность обнаружения при различных параметрах движения нарушителя.

Ключевые слова: вероятность обнаружения, оптико-электронный извещатель, охраняемый объект.

Введение

Эффективность любой системы безопасности в значительной мере зависит от параметров устройств обнаружения угроз и, в частности, от вероятности обнаружения этими устройствами несанкционированных действий. Одним из наиболее распространенных устройств обнаружения такой угрозы, как несанкционированное проникновение, является оптико-электронный извещатель (следуя терминологии [1]) или пассивный инфракрасный датчик (по принятой в других странах, а также и России, терминологии).

Как известно, вероятность обнаружения несанкционированного проникновения (НП) такими устройствами существенно зависит от параметров движения, в том числе от скорости и направления движения [2]. Наилучшим характеристикам обнаружения соответствует движение поперек диаграммы обнаружения. Такие же условия установлены для проверки подобных устройств в государственном стандарте [1] и в европейском стандарте [3]. Ясно, что в реальных условиях это далеко не всегда соблюдается. В [4] рассматриваются особенности решения задачи синтеза структуры средств обнаружения с точки зрения минимизации влияния на средства обнаружения различных воздействий, применяемых нарушителем, но без учета оценки степени уменьшения вероятности обнаружения НП. Представляется целесообразной разработка методики количественной оценки вероятности обнаружения таких устройств в разных условиях реализации НП. Это позволит корректнее решать задачи анализа и синтеза структуры средств обнаружения рассматриваемых угроз и оценки их эффективности.

Методика оценки вероятности обнаружения

Очевидно, что вероятность обнаружения НП $\hat{P}_{\text{Обн.}}$ будет зависеть от продолжительности нахождения нарушителя в зоне обнаружения, скорости и траектории его движения, условий окружающей среды и других параметров. В соответствии с разделом «Методы испытаний» государственного стандарта [1], «...извещатель должен обнаруживать движение (выдавать извещение о проникновении) стандартной цели (человека), перемещающейся в пределах зоны обнаружения поперечно ее боковой границе в диапазоне скоростей 0,3–3 м/с (0,1–5,0 м/с для извещателей для открытых площадок) на расстояние до 3 м. При этом расстояние между извещателем и целью (человеком) должно оставаться постоянным». Однако стандарт не учитывает, во-первых, требования к другим направлениям движения и, во-вторых, вероятностный характер обнаружения. Таким образом, представляется целесообразным переход к вероятностной методике оценки обнаружения.

Основываясь на вышеупомянутом методе испытаний [1], можно говорить о пройденном нарушителем расстоянии $X_{\text{Обн.}}$ от начала зоны обнаружения X_0 (когда начинается воздействие на контролируемый физический параметр) до точки обнаружения при квазистационарной, т.е. с близкой к постоянной, скорости его движения (v). Оценка $\hat{P}_{\text{Обн.}}$ такой случайной величины, как значение вероятности обнаружения НП $\hat{P}_{\text{Обн.}}$, может быть получена на основе использования закона распределения $F(x_{\text{Обн.}})$ расстояния обнаружения $x_{\text{Обн.}}$ (т.е. пути, пройденного нарушителем по зоне обнаружения). В свою очередь, оценка закона распределения (в частности, плотности $p(x_{\text{Обн.}})$) распределения вероятности $x_{\text{Обн.}}$ и его числовых характеристик может быть получена на основе экспериментальных исследований с последующей аппроксимацией выборочных оценок.

Таким образом, вероятность обнаружения НП $\hat{P}_{\text{Обн.}}$ в некоторой зоне можно характеризовать интегралом от плотности распределения вероятности расстояния обнаружения

$$\hat{P}_{\text{Обн.}} = \int_{X_0}^{X_{\text{Обн.}}^{\max}} p(x_{\text{Обн.}}) dx_{\text{Обн.}}$$

от начала зоны обнаружения X_0 до максимально допустимого расстояния обнаружения $X_{\text{Обн.}}^{\max}$. Это расстояние в соответствии с [1] составляет 3 м.

Заметим, что из выражения для $\hat{P}_{\text{Обн.}}$ следует, что вследствие какого-либо ограничения размеров зоны обнаружения может заметно уменьшиться вероятность обнаружения (увеличиться вероятность пропуска цели), поскольку при этом верхний предел $X_{\text{Обн.}}^{\max}$ интегрирования становится фиксированным $X_{\text{Обн.}}^{\max} = X_{\text{Зон.}}$, определяемым размером $X_{\text{Зон.}}$ зоны обнаружения, ограниченным, к примеру, размерами помещения.

Очевидно, что при этом в любом случае минимальный размер зоны обнаружения не должен быть меньше требуемого, т.е. значение $X_{\text{Зон.}}$ не должно быть меньше, чем $X_{\text{Обн.}}^{\min}$, определяемого техническими параметрами конкретного устройства обнаружения. Таким образом, в идеальном случае $X_{\text{Зон.}} \gg X_{\text{Обн.}}^{\min}$. Например, если пассивный инфракрасный датчик, контролирующий коридор, требует по технической документации для обнаружения пересечения не менее двух пар лучей диаграммы направленности, а маршрут преступника позволяет, двигаясь поперек, пересечь только одну пару лучей, то обнаружение становится практически невозможным.

Также очевидна зависимость закона распределения и (или) его числовых характеристик от окружающих условий [2], характера движения нарушителя и возможных способов его воздействия (как активных, так и пассивных) на систему безопасности, в частности, на средства обнаружения [4]. Например, использование одного из наиболее эффективных способов избежать обнаружения пассивным инфракрасным (ПИК) датчиком – движение с низкой или высокой скоростью (ниже стандартного предела 0,3 м/с или выше 3 м/с). Аналогичная ситуация будет иметь место в случае применения другого пассивного способа «обхода» ПИК датчика – использования плотной одежды, уменьшающей уровень инфракрасного излучения. Применительно к плотности распределения вероятности это может привести к увеличению среднего значения и дисперсии. Следствием будет уменьшение вероятности обнаружения.

Результаты эксперимента

Предложенная методика проверена экспериментально на оптико-электронном извещателе при следующих условиях. В качестве цели использовался человек, одетый в джинсовые штаны и рубашку с длинным рукавом, ростом 180 см и массой тела 75 кг. Температура фона 26°C. Направление движения – тангенциальное по отношению к оси сегментов диаграммы обнаружения и радиальное по отношению к извещателю. Скорость движения цели составляла 0,3; 1,5 и 3 м/с. Среднеквадратическое отклонение реальной скорости движения цели от заданной составила для скоростей движения 0,3; 1,5 и 3 м/с значения 0,00074; 0,037 и 0,019 м/с соответственно.

На рис. 1 приведены гистограммы выборок расстояния обнаружения для тангенциального (рис. 1, а) и радиального (рис. 1, б–г) направлений движения цели. Гистограмма для тангенциального направления движения приведена только для одного значения скорости движения, поскольку результаты для других скоростей были близки. Таким образом, при движении цели в направлении, установленном в стандарте для испытаний, требования [1] выполняются, и выборочная оценка $\hat{P}_{\text{Обн.}}$ вероятности обнаружения равна единице.

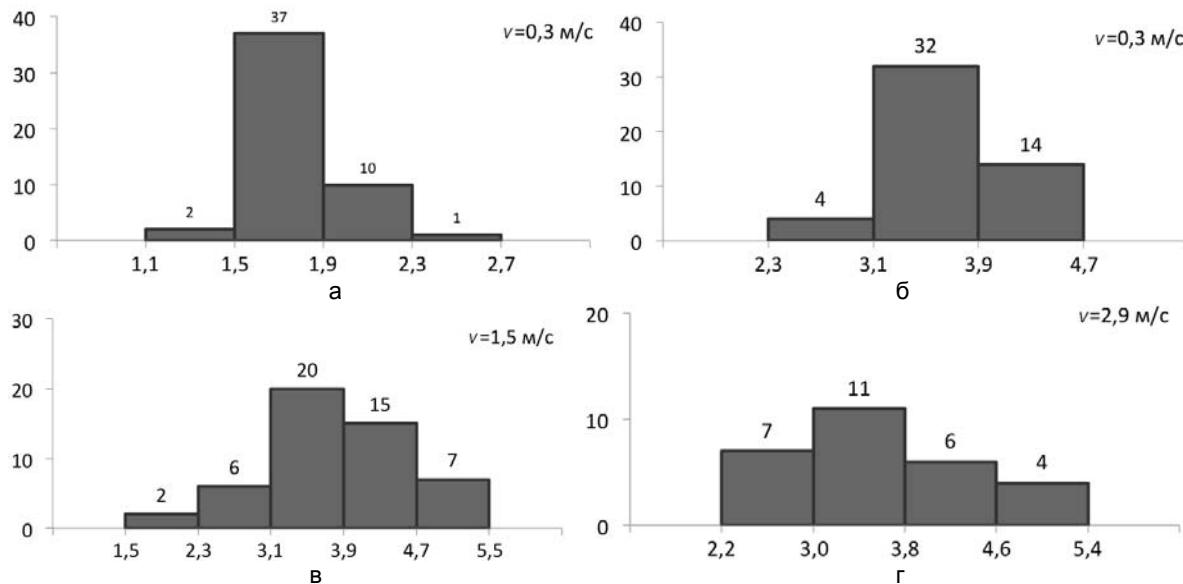


Рис. 1. Гистограммы выборок: тангенциальное направление (а); радиальное направление движения цели (б–г). По горизонтали – расстояние обнаружения, м

При радиальном направлении движения происходит заметное увеличение как среднего значения, так и дисперсии расстояния обнаружения, причем изменение дисперсии более существенно. Рис. 2 иллюстрирует характер зависимостей выборочных оценок $\hat{m} = \varphi(v)$ среднего значения расстояния обнаружения (рис. 2, а) и $\hat{\sigma} = \psi(v)$ – среднеквадратического отклонения при разных скоростях v движения цели (рис. 2, б).

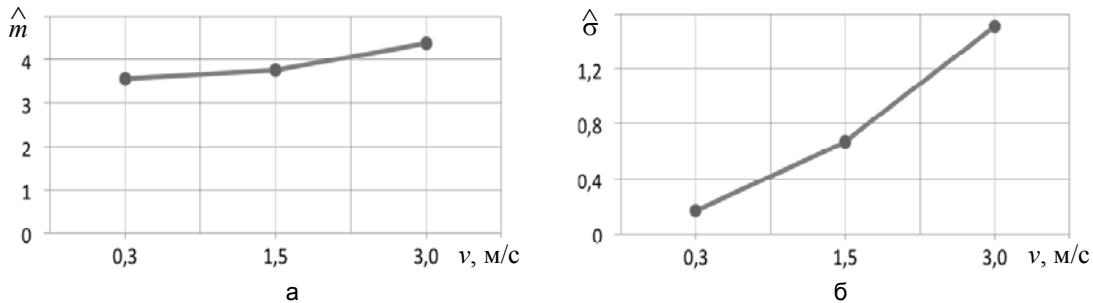


Рис. 2. Характер изменения выборочных оценок: среднего значения (а); среднеквадратического отклонения (б)

Увеличение среднего значения и дисперсии приводит к уменьшению вероятности обнаружения НП. На рис. 3 показаны зависимости оценки $\hat{P}_{\text{Обн.}} = \varphi(x_{\text{Обн.}})$ вероятности обнаружения от расстояния обнаружения для различных скоростей движения цели.

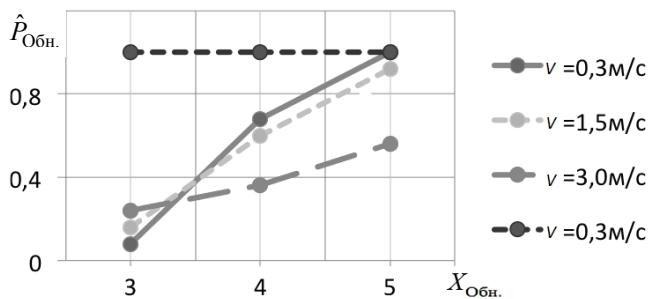


Рис. 3. Влияние расстояния, пройденного целью, на вероятность обнаружения

Таким образом, можно говорить о существенном снижении вероятности обнаружения на заданном стандартом расстоянии в 3 м, пройденном целью. Так, для достижения значения вероятности обнаружения, близкого к единице, цель должна преодолеть практически вдвое большее расстояние по зоне обнаружения, что может не выполняться при ограниченном размере зоны обнаружения. Полученные результаты полностью подтверждают правильность аналитических рассуждений, приведенных выше. Несмотря на то, что экспериментальные результаты соответствуют конкретному образцу извещателя, они свидетельствует о важности выполнения вероятностной оценки для других средств обнаружения на основе предложенной выше методики.

Заключение

Предложена методика оценки эффективности средств обнаружения, позволяющая получить оценки вероятности обнаружения при различных параметрах движения цели. Методика может быть использована для средств обнаружения различного типа.

Получены экспериментальные данные, подтверждающие предложенную методику и позволяющие оценить вероятность обнаружения несанкционированного проникновения оптико-электронным извещателем при различных параметрах и направлениях движения нарушителя.

Показано, что при радиальном движении на извещатель может существенно уменьшиться вероятность обнаружения. Аналогичный эффект имеет место при пространственном ограничении размеров зоны обнаружения.

Показано, что дисперсия вероятности обнаружения заметно возрастает при отклонении направления движения от тангенциального по отношению к оси сегментов диаграммы обнаружения; как следствие, падает вероятность обнаружения. Это свидетельствует о целесообразности уменьшения не только среднего значения, но и дисперсии расстояния обнаружения. Полученные результаты дают возможность

аргументированного выбора расположения на объекте средств обнаружения несанкционированного проникновения и оценки их эффективности при разных параметрах движения нарушителя.

Литература

1. ГОСТ Р 50777–95. Системы тревожной сигнализации. Ч. 2. Требования к системам охранной сигнализации. Раздел 6. Пассивные оптико-электронные инфракрасные извещатели для закрытых помещений и открытых площадок. – Введ. 01.01.1996. – М.: Госстандарт России, 1996. – 25 с.
2. Волхонский В.В. Извещатели охранной сигнализации. Изд. 4-е, доп. и перераб. – СПб: Экополис и культура, 2004. – 272 с.
3. EN 50131-2-2. Alarm systems – Intrusion systems – Part 2-2: Requirements for passive infrared detectors. Approved 2004-05-04. CENELEC. – 38 р.
4. Волхонский В.В., Крупнов А.Г. Особенности разработки структуры средств обнаружения угроз охраняемому объекту // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 4 (74). – С. 131–136.

**Волхонский Владимир
Владимирович**

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, профессор, volkhonski@mail.ru

Воробьев Павел Андреевич

– Государственный научно-исследовательский институт прикладных проблем, инженер-программист, doall88@gmail.com

УДК 694.4: 658.52.011.56

**АВТОМАТИЗИРОВАННАЯ СИСТЕМА ПРОЕКТИРОВАНИЯ
И ИССЛЕДОВАНИЯ ДЕРЕВЯННЫХ КОНСТРУКЦИЙ
НА МЕТАЛЛИЧЕСКИХ ЗУБЧАТЫХ ПЛАСТИНАХ**

М.Н. Павленко, А.Б. Шмидт

Рассматриваются проблемы автоматизированного проектирования деревянных ферм на металлических зубчатых пластинах. Предложен новый подход к созданию систем проектирования ферм и узловых решений, позволяющий провести расчеты наиболее часто используемых конструкций. Внедрены компоненты интерактивной помощи и контроля действий пользователя – бегущая строка сообщений и окно динамических подсказок. Проведено тестирование готового продукта. Система предназначена для проектировщиков конструкций, а также может быть применена в учебном процессе.

Ключевые слова: деревянная ферма, металлические зубчатые пластины, узловые решения.

Введение. Постановка задачи

В мировой практике жилищного домостроения весьма популярен один из современных видов соединений деревянных ферм – металлические зубчатые пластины (МЗП) (рис. 1). Наибольшее применение такие соединения нашли в США, Канаде, Швейцарии, Финляндии. В России повышенный спрос на МЗП связан с задачами жилищного строительства. Выбор в пользу этих соединений обусловлен высокими экономическими и технологическими показателями [1].

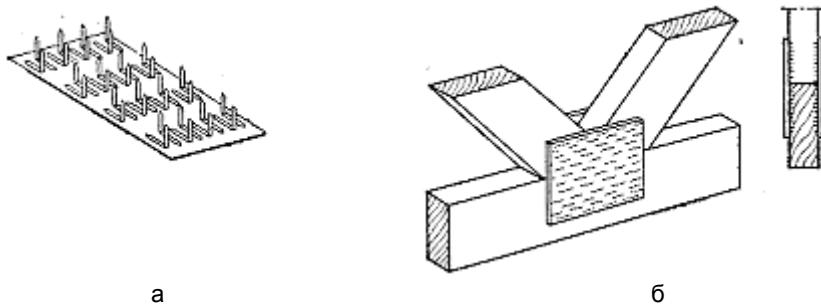


Рис. 1. Металлическая зубчатая пластина: внешний вид МЗП (а); узел (соединение) на МЗП (б)

В зарубежной практике производства и применения МЗП наиболее известна североамериканская компания MiTek [2], которая проектирует, производит и реализует разнообразные строения с использованием МЗП. Компанией разработано собственное программное обеспечение по проектированию таких соединений, ориентированное на использование своих платин МЗП. При этом методика расчета и проектирования соединений на МЗП закрыта для проектировщика и не может использоваться для разработки новых видов соединений на МЗП.

В России исследованиями МЗП в 80-х годах прошлого столетия занимался Центральный научно-исследовательский институт строительных конструкций (ЦНИИСК) им. В.А. Кучеренко. В результате были разработаны пластины МЗП 1.2 и МЗП 2 толщиной 1,2 мм и 2,0 мм и подготовлены «Рекомендации по проектированию конструкций с соединениями на МЗП» [3]. В дальнейшем эти рекомендации использовались в разработке нормативных документов по расчету деревянных конструкций [4, 5]. Одним из последних был составлен свод правил [6].

В основе расчета несущей способности МЗП лежит подбор рабочей площади пластины. Рабочей площадью принято считать площадь пластины на элементе без краевых полос шириной 10 мм (рис. 2) [4]. Однако вручную такой расчет достаточно трудоемок, если учесть большое число узлов на МЗП в решетчатой конструкции. Как правило, это приводит к чрезмерному увеличению запаса прочности.

Среди отечественных разработок по расчету деревянных ферм на МЗП известен продукт АРМ Wood, который, по утверждениям разработчиков [7], соответствует СТО 36554501-002-2006. Но методика расчета МЗП не раскрывается, и проектировщику трудно найти рациональное проектное решение, что снижает положительный эффект применения программы. Кроме того, в упомянутом программном обеспечении рассматриваются узлы только с одной пластиной, что в некоторых типах узла приводит к перерасходу пластин. В примере, приведенном на рис. 3, а, около 40% площади МЗП не задействовано в работе и не учитывается в расчетах. В таком узле целесообразнее устанавливать несколько пластин (рис. 3, б). Расход МЗП на конструкцию в целом снижается до 30%.

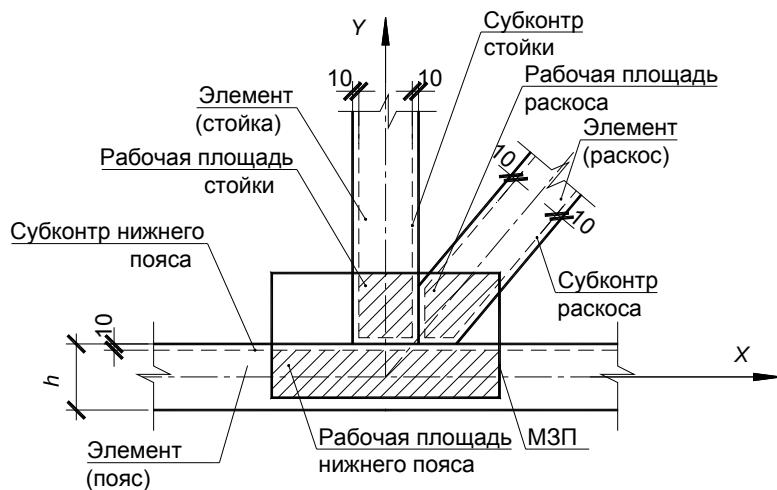


Рис. 2. Рабочая площадь МЗП

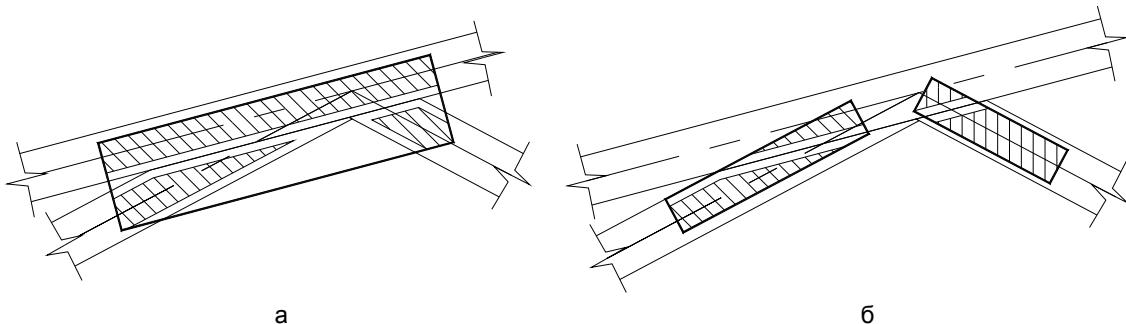


Рис. 3. Варианты размещения пластины в узле: узел с одной пластиной МЗП (а); узел с двумя пластинами МЗП (б)

Вопросы интерактивного определения рабочих площадей пластин МЗП, эффективного их позиционирования, а также снижения трудоемкости работ при проведении вычислений побуждают совершенствовать вычислительные процессы и создавать новые алгоритмы. Целью данной работы является создание автоматизированной системы проектирования деревянных ферм на МЗП с эффективным алгоритмом расчета и конструирования узлов, включая определение рабочих площадей и рациональное размещение МЗП в узле.

Алгоритм определения рабочих площадей пластины

Для автоматизации расчета МЗП, разработанных ЦНИИСК, создана система проектирования деревянных ферм на МЗП «УИИК-Ферма», входящая в состав учебно-информационного исследовательского комплекса (УИИК). В УИИК разработана серия программ по расчету и конструированию деревянных конструкций.

Основными достоинствами «УИИК-Ферма» являются определение рабочих площадей пластины, средства оптимизации расстановки МЗП в узле, мгновенная подсказка в ходе работы с разъяснениями, интуитивно понятный интерфейс, а также новые компоненты помощи и контроля процесса проектирования. Все это позволяет находить рациональное решение по проектированию и исследовать работу узла.

Для решения этих задач разработана библиотека численных методов работы с полигонами, полу-плоскостями и точками, а также создан детерминированный алгоритм определения рабочих площадей.

Ставится задача параметрической оптимизации – определения наиболее подходящих параметров пластины, которые задаются следующим образом:

$$X = \{pl = < d, l, \alpha, dx, dy > \mid 0 \leq d, l \leq 100, 0 \leq \alpha \leq 2\pi, -50 \leq dx, dy \leq 50, A_{pl} \geq 50\},$$

где d – ширина пластины, см; l – длина пластины, см; α – угол поворота пластины, рад; dx – смещение центра пластины по оси OX , см; dy – смещение центра пластины по оси OY , см; $A_{pl} = \sum_e Area(Poly(pl) \cap Poly_{10}(e))$ – рабочая площадь МЗП (рис. 4, а).

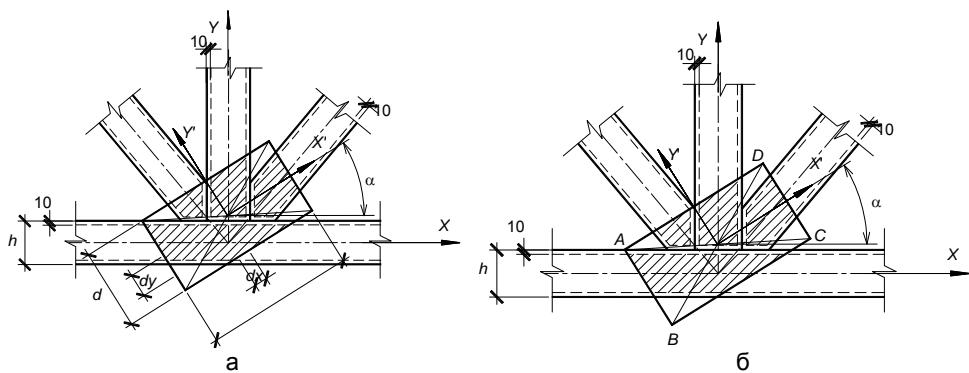


Рис. 4. Схема узла: геометрические параметры МЗП (а); исходный полигон МЗП (б)

Алгоритм определения рабочих площадей пластины представлен ниже.

В качестве исходного полигона используется полигон пластины – $ABCD$
 $X = \{pl = < d, l, \alpha, dx, dy \>\}$ (рис. 4, б).

1. Для нахождения рабочей площади одного элемента в узловом соединении полигон раскоса X_1 уменьшается на полосу шириной 10 мм, параллельную грани стержня:

$$X_{10} = X_1 \cap Poly_{10}(e).$$

2. От полигона X отсекаются полуплоскости, проходящие через границы полигона X_{10} :

$$X_e = X \cap P_{ab} \cap P_{cd} \cap P_{de} \cap P_{ea},$$

где P_{ab} , P_{cd} , P_{de} , P_{ea} – полуплоскости со сторонами ab , cd , de , ea соответственно (рис. 5).

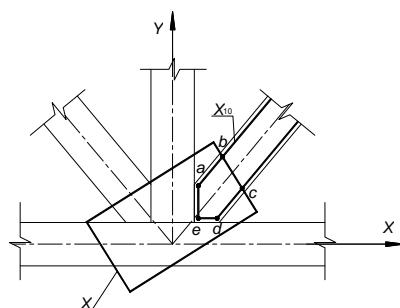


Рис. 5. Полигоны пересечений

3. Разбивая получившийся полигон X_e на треугольники, вычисляется его площадь по формуле Герона, которая и является искомой рабочей площадью МЗП:

$$A_{pl} = \sum_e Area(X'_e).$$

С помощью описанного алгоритма можно моделировать любые узлы на МЗП. Окно компоновки и расчета узла представлено на рис. 6.

Для более рационального расхода пластин разработан функционал размещения двух и более МЗП в узле. Рабочие площади в этом случае вычисляются для всех пластин.

Каждый элемент и МЗП в узле должны быть связаны. С этой целью разработана проверка связности конструкции, реализованная с помощью алгоритма Флойда–Уоршелла на графе. Вершинами графа являются элементы, входящие в узел, а ребрами – пластины. Определяются стержни, для которых выполняется условие минимальных площадей, и соответствующие им вершины в графе соединяются ребром. Если граф связан, условие минимальных площадей и связности узла выполняется.

Для точного и быстрого отображения модели фермы используется интерактивный способ взаимодействия с программой по технологии WYSIWYG (What You See Is What You Get).

Система рассчитана как на опытного проектировщика, так и на слабо подготовленного пользователя. Для этого разработаны и внедрены новые компоненты, направляющие и контролирующие действия проектировщика: окно динамических сообщений, бегущая строка, специализированные поля ввода числовых значений и др.

При нарушении рекомендаций или норм в бегущей строке выдается краткое сообщение с описанием ошибки. Для получения быстрой справки достаточно навести курсор мыши на нужный параметр, и в окне динамических подсказок появится его описание. Для развернутой помощи подготовлено пособие-программа по проблемной части, в которое включены СНиПы, карты районирования, атлас деревянных конструкций и другие справочные материалы.

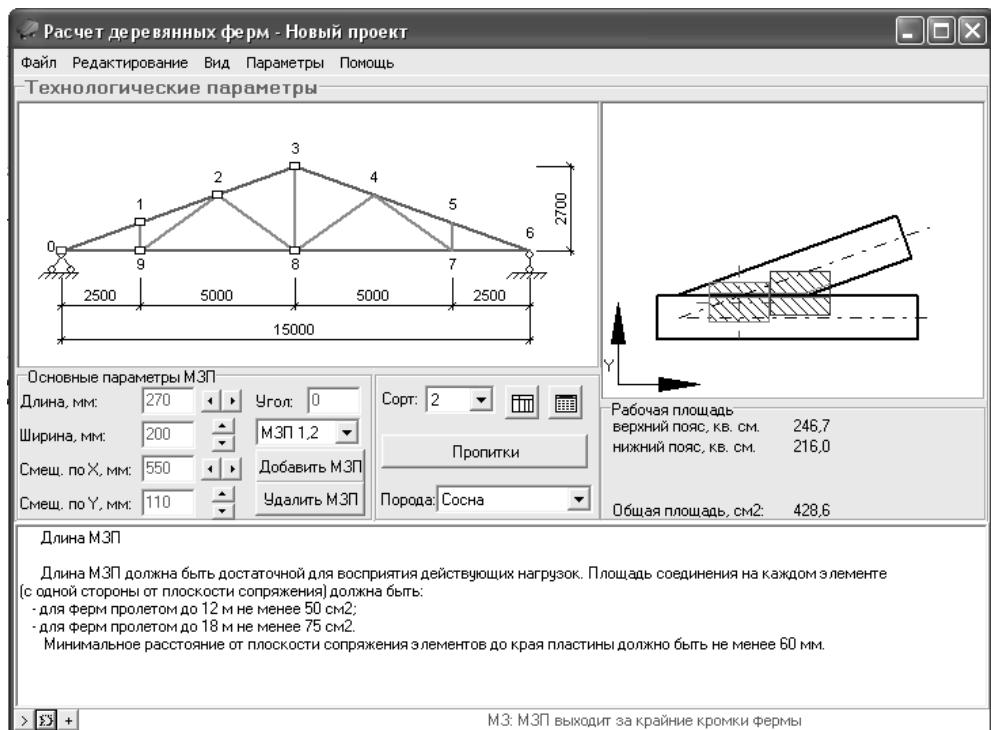


Рис. 6. Окно компоновки и расчета узла

Указанные компоненты, технология WYSIWYG, интерактивный режим взаимодействия с системой являются основополагающими составляющими для всех программ, входящих в УИИК. Графический интерфейс всех систем разработан на языке программирования Object Pascal в среде Delphi. Этот язык позволяет быстро и качественно реализовать интерфейс, не отвлекаясь на детали его реализации и API-функции операционной системы. Создание программного продукта в Delphi выполнено в интегрированной среде разработки IDE (Integrated Development Environment). IDE служит для организации взаимодействия с программистом и включает систему программных средств, содержащих управляющие элементы, с помощью которых можно проектировать интерфейсную часть приложения, писать программный код и связывать его с управляющими элементами.

Данная среда программирования ориентирована на «быструю разработку» программного обеспечения (ПО) – технологию RAD (Rapid Application Development). Это позволяет значительно сократить время разработки диалоговых окон и реализовать поставленную задачу за счет:

- объектно-ориентированного подхода к разработке приложений [8];
- удобного интерфейса работы с базами данных;
- визуальных средств создания диалоговых окон.

Графическая часть реализована в среде AutoCAD компании Autodesk. Для автоматизации вывода чертежей создаются программы на языке Visual Lisp, которые сохраняются в отдельных файлах в виде проектов и загружаются в среде AutoCAD.

Преимуществами предложенного ПО перед продуктами-аналогами являются:

- разделение процесса проектирования на шаги и последовательное их выполнение;
- контроль ошибок ввода информации и быстрая подсказка по их редактированию;
- внедрение новых компонентов интерактивной помощи;
- динамическая визуализация графических элементов;
- развернутый протокол расчета в форме пояснительной записки с необходимыми ссылками и иллюстрациями;
- выполнение графической документации (рабочих чертежей на 60%);
- поддержка нормативно-справочной информации по предмету проектирования;
- составление «Истории расчета» с анализом проведенных сессий расчета;
- наличие экспертной системы.

Заключение

Разработана автоматизированная система проектирования деревянных ферм на металлических зубчатых пластинах. Программа позволяет выполнять прочностной расчет узлов с помощью разработанного алгоритма определения рабочих площадей и рационально размещать металлические зубчатые пластины в узле.

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ ПРОЕКТНЫХ РЕШЕНИЙ ...

В системе «УИИК-Ферма» можно провести расчет нескольких ферм одновременно, сравнить их и выбрать подходящий вариант, отвечающий техническим и экономическим критериям. Система позволяет повысить качество проектной документации и может быть полезна проектировщикам деревянных конструкций на металлических зубчатых пластинах, а также при обучении студентов и в переподготовке специалистов строительных специальностей.

Литература

1. Арленинов Д.К., Буслаев Ю.Н. Конструкции из дерева и пластмасс: Учеб. для техн. вузов. – М.: Изд-во АСВ, 2002. – 280 с.
2. Mitek: [Электронный ресурс]. – Режим доступа: <http://www.mitek.ru/tech-mitek/istoriya-mitek.html/>, св. Яз. рус. (дата обращения 25.11.2011).
3. Рекомендации по проектированию и изготовлению дощатых конструкций с соединениями на металлических зубчатых пластинах / ЦНИИСК им. В.А. Кучеренко. – М.: Стройиздат, 1983. – 40 с.
4. Пособие по проектированию деревянных конструкций (к СНиП II-25-80) / ЦНИИСК им. Кучеренко. – М.: Стройиздат, 1986. – 216 с.
5. СНиП II-25-80. Деревянные конструкции / Госстрой России. – М.: ГУП ЦПП, 2002. – 30 с.
6. СП 64.13330.2011. Деревянные конструкции. – М.: ОАО «ЦПП», 2011. – 87 с.
7. НПЦ АПМ [Электронный ресурс]. – Режим доступа: <http://www.apm.ru/rus/civil/#wood/>, св. Яз. рус. (дата обращения 25.11.2011).
8. Боб Е.Б., Латников А.В. Эволюция методов и технологий программирования // Научно-технический вестник СПбГУ ИТМО. – 2008. – № 48. – С. 191–198.

Павленко Мария Николаевна

– Санкт-Петербургский государственный архитектурно-строительный университет, аспирант, mariya.pavlenko@gmail.com

Шмидт Александр Борисович

– Санкт-Петербургский государственный архитектурно-строительный университет, кандидат технических наук, доцент, зав. кафедрой, ukf@bk.ru

УДК 004.896

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ ПРОЕКТНЫХ РЕШЕНИЙ ДЛЯ ДОКУМЕНТИРОВАНИЯ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

Н.В. Агаркова, Г.О. Артемова, Н.Ф. Гусарова

Работа посвящена оптимизации проектных решений в сфере систем автоматизированного проектирования технологической подготовки производства, в частности в подготовке производственных структур документирования научно-технической информации.

Ключевые слова: САПР, научно-техническая информация, распределение работ, автоматизация, документирование, СППР.

Введение

В настоящее время развитие промышленного производства во всем мире и в России, в частности, проходит в условиях постоянно меняющейся номенклатуры продукции. Для сохранения конкурентоспособности предприятия вынуждены ускорять процесс выпуска продукции, для чего иногда приходится полностью перестраивать технологическую подготовку производства. При этом трудоемкий и ресурсо затратный процесс подготовки документации становится узким местом и требует отдельного внимания в рамках комплексной автоматизации производства.

Кроме технологической документации, в информационном обеспечении приборостроения как наукоемкой отрасли промышленности особое место занимают научно-технические документы – статьи, технические отчеты и описания, буклеты, проспекты, презентации и т.д., которые играют важную роль при взаимодействии с заказчиками и научным сообществом. Процесс создания таких документов – документирование первичного потока научно-технической информации – достаточно редко рассматривается при комплексной автоматизации предприятия, так как ориентирован на малые тиражи (плоть до единичных экземпляров).

В настоящее время на большинстве предприятий приборостроения производство ведется с использованием современных CAD-систем и автоматизированных систем технологической подготовки производства (АСТПП), что обеспечивает достаточный уровень подготовки конструкторской документации в электронном виде. Широкое распространение получили также PDM-системы (ProductDataManagement – управление данными об изделии). Использование в приборостроении PDM-систем, CAD-систем и АСТПП позволяет эффективно решать задачи создания конструкторско-технологической базы данных и эффективного хранения документации.

Однако задача создания и оформления научно-технической документации и ускоренной адаптации производственных мощностей подготовки документации при изменении номенклатуры продукции плохо поддается автоматизации существующими средствами и остается одним из самых трудоемких процессов. Для решения этого вопроса возникает потребность в инструменте, позволяющем поддержать процесс принятия решений при разработке системы подготовки документации.

Идентификация предметной области

Информационное обеспечение играет важную роль на всех этапах жизненного цикла промышленной продукции. В условиях современного производства оно в подавляющем большинстве случаев организовано в виде совокупности документов различного типа. Документ [1] – материальный носитель с зафиксированной на нем в любой форме информацией в виде текста, звукозаписи, изображения и (или) их сочетания, который имеет реквизиты, позволяющие его идентифицировать, и предназначен для передачи во времени и в пространстве в целях общественного использования и хранения.

В информационном обеспечении приборостроения как наукоемкой отрасли промышленности научно-технические документы (статьи, технические отчеты и описания, буклеты, проспекты, презентации и т.д.), которые, согласно принятой классификации [2], относятся к первичным документам¹, но, в отличие от, например, бухгалтерских первичных документов², имеют семантическую, а не фактографическую природу. Следуя определениям [2, 3], технологический процесс создания таких документов можно назвать документированием³ первичного потока научно-технической информации.

Как показывает анализ организации документирования первичного потока информации на предприятиях приборостроения [4], в качестве прототипов здесь используются технологические процессы, продукция которых аналогична по физической природе подготавливаемым документам (например, полиграфия). Для определенности будем рассматривать подготовку печатных документов, которая строится как модификация процесса допечатной подготовки изданий [5, 6]. Структура последнего представлена на рис. 1. Процесс реализуется на компьютере: на экране монитора воспроизводится постепенное заполнение поля будущего структурного элемента издания (например, страницы документа) структурными единицами более низкого уровня (текстом, иллюстрациями и т.д.). На программном уровне в качестве структурных единиц выступают информационные объекты⁴, т.е. программными средствами решается задача сборки.

Однако документирование первичного потока информации имеет существенные отличия от прототипа по характеристикам выпускаемой продукции (документов), представленные в табл. 1 (столбец 1). В этих условиях полномасштабная структура (рис. 1) является избыточной и не позволяет гибко перестраивать систему документирования под нужды конкретного документа.

№	Характеристика документов	Связь с определяющими факторами
1	Малые и сверхмалые тиражи (вплоть до единичных экземпляров)	В документе увеличиваются $\Delta f(z)$ – отклонения от среднестатистического распределения ТПр для документов данного типа;
2	Технологическая разноплановость документов	Увеличиваются $\Delta f(z)$; увеличивается требуемое количество специализированных программно-аппаратных средств
3	Разнообразие программно-аппаратных средств, использованных при подготовке входных ИО	Растет множество входных параметров $Q \rightarrow$ возникают новые ТПр \rightarrow их нужно компенсировать ростом $C_i \rightarrow$ нужно увеличивать c_i
4	Изменение критериев эффективности ДПП со сменой документа	Увеличиваются $\Delta m(z)$ – отклонения от среднестатистического распределения важности ТПр для документов данного типа \rightarrow увеличивается интенсивность труда по их устранению \rightarrow нужно увеличивать c_i
5	Сильные ресурсные ограничения на один документ	Ограничения на $c_i \rightarrow$ нужно уменьшать число ОЗ, приходящихся на один документ

Таблица 1. Характеристики выпускаемых документов и их связь с определяющими факторами

¹ Первичный документ [2] содержит информацию, которая является изложением результатов изучения, исследования, разработок и т.д. и оригинальна по своему характеру, т.е. представляет собой результат непосредственного отражения деятельности автора.

² В бухгалтерии первичный документ составляется в момент совершения хозяйственной операции и является первым свидетельством произошедших фактов [2].

³ Документирование – запись информации на различных носителях по установленным правилам [3]. Документирование – создание документа с использованием различных методов, способов и средств фиксирования информации на материальном носителе [2]

⁴ Информационный объект – представление объекта предметной области в информационной системе, определяющее его структуру, атрибуты, ограничения целостности и, возможно, поведение [7].

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ ПРОЕКТНЫХ РЕШЕНИЙ ...

Таким образом, возникает задача проектирования системы документирования первичного потока информации. В статье рассматривается подход к моделированию процесса с учетом неопределенностей, порождаемых характеристиками документов, и дерево решений для построения системы поддержки принятия проектных решений (СППР).

Концептуализация предметной области

В соответствии с методикой [8] сформулируем исходные данные для синтеза СППР построения системы документирования первичного потока информации. Процесс формируется как последовательность работ, объединенных в группы; номенклатура работ и обобщенная (избыточная) структура их взаимосвязей соответствует рис. 1. Работы выполняются обрабатывающими звеньями, которые характеризуются программно-аппаратным обеспечением, а также знаниями и опытом (компетенцией) сотрудника.

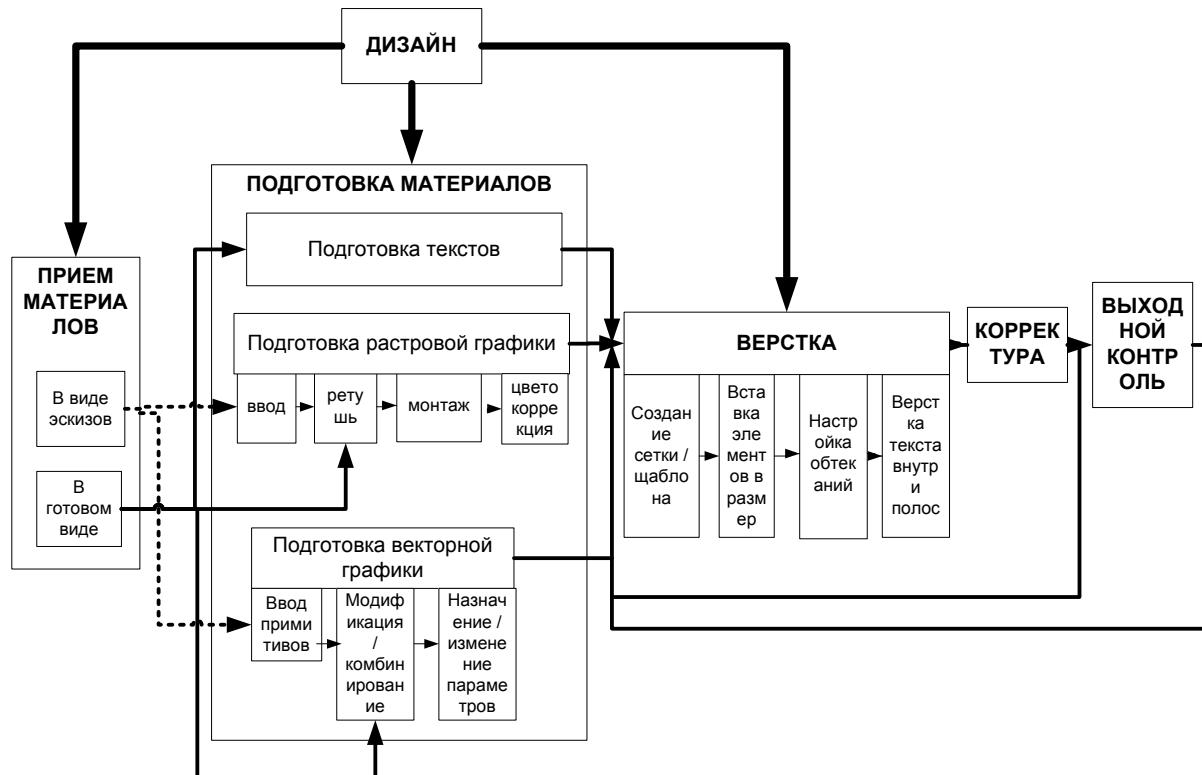


Рис. 1. Обобщенная структура процесса ДПП

Работы состоят из наборов технологических операций, выполняемых над информационными объектами и направленных на решение отдельных технологических задач. Например, для выполнения работы «Верстка текста внутри полосы» может потребоваться решение таких технологических задач, как «редактирование параметров», «модификация», «вставка» и т.п. Специфика программных продуктов такова, что для решения большинства задач могут быть построены разные наборы операций. Их адекватный выбор и реализация являются сложной и слабо формализуемой задачей и в реальной практике рассматриваются как решение технологической проблемы, причем сотрудники с разной компетенцией решают ее по-разному. Некоторые примеры различного уровня сложности представлены в табл. 2.

группа работ	работа	ТЗ	варианты ТО	примеры ТПр
верстка	верстка текста внутри полосы	вставка ИО	CtrlC+CtrlV Paste Import и т.д.	могут изменяться параметры ИО могут изменяться параметры ИО не поддерживается рядом пакетов

Таблица 2. Пример соотношения технологической задачи (ТЗ), технологических операций (ТО) и технологических проблем (ТПр) для отдельной работы

В силу описанных выше отличий процесса документирования первичного потока информации от процесса допечатной подготовки полиграфического издания реализация полного цикла обработки (рис. 1), который поддерживает единую и оптимальную с точки зрения компьютерной обработки структуру формируемого документа [9], для многих документов оказывается нерентабельной. Их приходится формировать из структурных единиц (рисунков, текстов, таблиц и пр.), предоставляемых специалистами

конкретной предметной области в виде информационных объектов со случайной или неадекватной объектной структурой и (или) атрибутами. Их параметры образуют множество внешних параметров Q для процесса. При их дальнейшей обработке возникают дополнительные технологические проблемы [9], решение которых также должно быть включено в состав системы документирования первичного потока. Таким образом, процесс имеет одну точку выхода (готовый документ), но может иметь множество точек входа (источников информации).

В качестве выходных параметров Y , из которых формируется целевая функция, рассматриваются те параметры выпускаемых документов, которые интересуют потребителя. Содержательно их можно обозначить как $Y = \{Y_1, Y_2\}$, где параметр Y_1 характеризует качество документа, а параметр Y_2 – его себестоимость.

Учитывая изложенное, задачу синтеза СППР необходимо рассматривать как задачу структурного синтеза, или задачу принятия решений [8]:

$$ЗПР = < A, K, \text{Мод}, \Pi >, \quad (1)$$

где A – множество альтернатив проектного решения; $K = (K_1, K_2, \dots, K_n)$ – множество критериев (выходных параметров), по которым оценивается соответствие альтернативы поставленным целям; Мод: $A \rightarrow K$ – модель, позволяющая для каждой альтернативы рассчитать вектор критериев; Π – решающее правило для выбора наиболее подходящей альтернативы в многокритериальной ситуации.

Формализация базы знаний СППР

Материал предыдущего раздела показывает, что процесс документирования первичного потока реализуется в условиях существенной неопределенности, связанной, прежде всего, с высокой вариативностью информационных объектов и программных сред их обработки, ответом на которую является использование слабо формализуемых знаний и опыта (компетенции) сотрудников. В результате операционный состав работ и их распределение между обрабатывающими званиями могут динамически изменяться в ходе процесса, что затрудняет формальное задание операндов выражения (1). В этой ситуации вместо единичной работы или операции в качестве элементарной планируемой части процесса будем использовать технологическая проблема. А именно, будем рассматривать процесс как обработку потока технологических проблем, точнее, как последовательность принятия решений по их устранению, а одним из показателей эффективности процесса будем считать количество проблем, устранивших в готовом документе.

Следуя [10, 11], представим множество технологических проблем точками отрезка $[0, Z]$ действительной оси, упорядоченными по возрастанию сложности, причем плотность вероятности $f(z)$ появления проблемы $z \in [0, Z]$ убывает с ростом z . Соответствующую функцию распределения обозначим через $F(z) = \int_0^z f(\zeta) d\zeta$. Тогда параметр Y_1 , характеризующий качество готового документа, можно представить как

$$Y_1 = \int_{Z_0}^z f(z)m(z)dz, \quad (2)$$

где $Z_0 \in [0, Z]$ – подмножество проблем, устранивших в готовом документе, а $m(z)$ – функция, задающая относительную важность проблем различного типа для конкретного документа.

Обозначим подмножество технологических проблем, которые умеет решать i -й сотрудник (его множество компетенций), через $C_i \in [0, Z]$. В общем случае компетенции разных сотрудников могут пересекаться, т.е. $C_i \bigcup C_j \neq 0$, $i \neq j$. Тогда эффективность работы i -го сотрудника с компетенцией C_i может быть представлена через количество ТПр, решенных им за рабочее время t_i^w :

$$V = t_i^w \int_{A_i} f(z) dz. \quad (3)$$

Затраты на содержание i -го обрабатывающего звена ОЗ, куда входят стоимость оборудования и программного обеспечения, а также зарплата сотрудника с компетенцией C_i , равны

$$E_i = c_i \mu(C_i), \quad (4)$$

где $\mu(C_i)$ – длина отрезка C_i [10]. Соответственно, себестоимость готового документа можно представить как

$$Y_2 = \sum_i c_i \mu(C_i). \quad (5)$$

Чтобы учесть характеристики формируемых документов в задаче синтеза (1), на множестве характеристик (табл. 1, столбец 2) был проведен анализ [12], который показал, что определяющими факторами здесь также являются распределения $f(z)$ и $m(z)$. Это содержательно проиллюстрировано в табл. 1 (столбец 3). Таким образом, величины $f(z)$ и $m(z)$ задают множество внешних параметров Q , а характеристики документа (табл. 1) определяют множество ограничений $\{R\}$.

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ ПРОЕКТНЫХ РЕШЕНИЙ ...

Критерий эффективности системы документирования первичного потока формируется на множестве выходных параметров $Y = \{Y_1, Y_2\}$. Тогда в общем виде, без учета размерных коэффициентов, можно представить задачу как

$$\underset{\{R\}}{\text{extr}}(Y_1 - Y_2). \quad (6)$$

Для конструктивного решения задачи (1) необходимо разработать подход к построению альтернатив A . В связи с охарактеризованной выше неопределенностью процесса альтернативы должны задаваться не путем явного перечисления, а в виде кортежа [8]

$$A = < P, \mathcal{E} >, \quad (7)$$

где P – способ синтеза проектных решений, \mathcal{E} – набор элементов, из которых формируются альтернативы. Каждой альтернативе требуется поставить в соответствие вектор управляемых параметров $\{X\}$, причем выбор компонентов вектора $\{X\}$ быть таким, чтобы задача (6) допускала решение в пространстве $\{X\}$, т.е. в постановке

$$\text{extr } F(X), X \in D_x,$$

где $F(X)$ – критериальная функция; $D_x = \{X\}_{\text{доп}}$ – допустимая область решений. В состав $\{X\}$ должны, очевидно, входить параметры автоматизированного рабочего места (АРМ), определенные выше – множество компетенций сотрудников $\{C_i\}$ и множество затрат на содержание АРМ $\{c_i\}$, а в качестве $\{\mathcal{E}\}$ выступают работы, выполняемые конкретным АРМ.

Выражения (2), (5), (6) показывают, что эффективность процесса документирования информации тем выше, чем точнее совпадают распределения технологических проблем $f(z)$ и компетенций сотрудника, работающего на конкретном АРМ. Однако, как показано выше, для конкретного документа $f(z)$ заранее неизвестно, т.е. $\Delta f(z)$ велико, и сотрудник с компетенцией C_i может столкнуться с проблемой, для решения которой требуется более высокая компетенция. В этих условиях целесообразно предусмотреть механизм совместной работы над такими проблемами – например, в режиме консультаций. Поэтому требуется включить в состав $\{X\}$ параметр влияния X_3 , позволяющий компенсировать текущие $\Delta f(z)$ за счет задания структуры взаимодействия между АРМ.

Содержательный анализ процесса позволил выделить типовые варианты упорядочения компетенций по технологическим проблемам и сопоставить каждому из них способ синтеза структуры системы документирования первичного потока информации (табл. 3), т.е. построить множество альтернатив проектного решения $\{A\}$.

В соответствии с общей постановкой задачи синтеза (1) построим для каждой альтернативы правило выбора П и модель Мод. Для определенности при построении модели Мод вместо (2) будем использовать условие $Y_1=Y_2$.

Значение X_3	Характер распределения компетенций по ТПр	Способ синтеза структуры ДПП
$X_3=1$	Подмножество ТПр, решаемых i -м ОЗ, соответствует границам его компетенции A_i	ОЗ полностью решает приходящие к нему ТПр («конвейерная» сборка)
$X_3=2$	В выполняемой i -м ОЗ работе есть ТПр, выходящие за границу его компетенции A_i ; разрешено консультироваться у сотрудника с более высокой компетенцией A_{i+1} .	Допускается частичное перекрытие работ путем уменьшения рабочего времени ($i+1$)-го ОЗ на время консультаций
$X_3=3$	В выполняемой i -м ОЗ работе есть ТПр, выходящие за границу его компетенции A_i ; при этом обработка ТПр может быть разделена на диагностику и решение, которые требуют различных компетенций.	Вводится дополнительный контур обработки ТПр: ОЗ с компетенцией A_{i+1} диагностирует проблему и передает ее сотруднику с компетенцией A_i , а затем, в соответствии с типом проблемы (см. табл. 1), проверяет полученное от A_i решение

Таблица 3. Типовые варианты упорядочения компетенций по ТПр

Вариант $X_3=1$ может быть реализован при условии

$$\langle f(x) \rangle - f(x) \leq \Delta_{\text{доп}}, \quad (8)$$

т.е. если функция $f(z)$ для типового документа несущественно отличается от ее оценки $\langle f(x) \rangle$, получаемой по конкретному документу. В этом случае целесообразно так организовать процесс, чтобы критерий (6) выполнялся для каждого отдельного АРМ: за счет выбора границ интервалов компетенции C_i учесть различную трудоемкость устранения технологической проблемы, т.е. обеспечить $\forall i: t_i^w = \text{const} = t^w$, а за счет выбора c_{i+1}/c_i учесть неравную значимость проблем на смежных интервалах, μ_i и μ_{i+1} . Соответствующая модель Мод₁ имеет вид

$$\frac{\int_{A_i} f(x) dx}{\mu_i} = \frac{c_{i+1}}{c_i} \frac{\int_{A_{i+1}} f(x) dx}{\mu_{i+1}}. \quad (9)$$

Вариант $X_3=2$ целесообразно применять, если не выполняется условие (8), т.е. суммарный интервал компетенций сотрудников имеет зону перекрытия

$$C_i \bigcup C_{i+1} \neq 0. \quad (10)$$

Обозначим эту зону $(C_i+C_{i+1}) \in [0, Z]$. Согласно [10], выполнение работ в зоне (C_i+C_{i+1}) оплачивается i -му сотруднику по ставке c_i , а $(i+1)$ -му сотруднику оплачиваются только консультации. Тогда, вводя обозначения $A_i=AC$, $A_{i+1}=BD$, $A_i \bigcup A_j = BC$, можно рассчитать относительное уменьшение оплаты $(i+1)$ -го сотрудника:

$$\Delta_{i+1} = 1 - \left(\frac{\mu_{BC}\mu_{AB}}{\mu_{BD}^2} \right) \frac{\int_B^D f(x) dx}{\int_A^B f(x) dx}. \quad (11)$$

Задавая допустимое значение Δ_{i+1} , получаем модель Мод₂, которая, с одной стороны, обеспечивает мотивацию консультирующего сотрудника (за счет небольших Δ), а с другой стороны – взаимозаменяемость сотрудников и надежность процесса в целом (за счет увеличения зоны перекрытия компетенций μ_{BC}).

Вариант $X_3=3$ реализуется в случае, когда некоторому типу проблем (обозначим его $BC \in [0, Z]$) придается повышенная важность, т.е. допустимая вероятность пропуска ошибки по этому типу проблем становится меньше, чем по другим типам. Обработка таких проблем требует повышенной интенсивности труда и, соответственно, дополнительной оплаты. Ситуацию можно моделировать заменой функции распределения проблем $f(x)$ на

$$F(x)=f(x) I(x), \forall x: I(x) \geq 1. \quad (12)$$

Если проблемы интервала ВС полностью обрабатываются $(i+1)$ -м АРМ, то вся работа соответствующего сотрудника оплачивается по ставке c_{i+1} ; если же реализуется описанное выше разделение труда, то часть работы, связанная с непосредственным устранением проблемы, выполняется i -м сотрудником упрощенным способом (чаще всего с нарушением объектной структуры результирующего документа [4, 9]) и оплачивается по ставке c_i , т.е. возникает экономия, вычисляемая из (3), (11):

$$\Delta_{i+1} = \int_B^C f(x)(I(x)-1) dx \left(\frac{\mu_{BD}}{\int_B^D f(x) dx} - \frac{\mu_{AB}}{\int_A^B f(x) dx} \right), \quad (13)$$

из которой оплачивается проверка, выполняемая $(i+1)$ -м сотрудником. Задавая допустимое значение Δ_{i+1+1} , из (13) получаем Мод₃. Дальнейшее уточнение модели должно быть связано с оценкой объема контекста, потенциально затрагиваемого переходом к упрощенному решению проблемы.

Таким образом, в результате проведенного анализа выполнена конструктивная постановка задачи принятия решения (1), а именно: определены альтернативы проектного решения в форме (7) (таблица 3); определено множество критериев $K = (K_1, K_2, \dots, K_n)$ (выражения (2), (5), (6)); для каждой альтернативы построены модели Мод (выражения (9), (11), (13)) и решающие правила П (выражения (8), (10), (12)). Это позволяет проводить синтез проектных решений для конкретных условий процесса с учетом парирования неопределенностей, порождаемых характеристиками документов, перераспределением ТПр между сотрудниками и разбиением всего множества ТО на подмножества для соответствующих классов сотрудников.

Реализация базы знаний СППР

Проведенный анализ показывает, что процесс создания системы документирования реализуется в условиях существенной неопределенности, связанной, прежде всего, с влиянием человеческого фактора, что приводит к трудности оценки текущих значений параметров процесса. Возникает потребность в средствах, позволяющих поддержать процесс принятия решений при разработке системы.

Изучение специализированных источников информации показало, что на текущий момент в Российской Федерации не описано и не внедрено ни одного программного продукта для создания системы документирования первичного потока информации. Предлагаемая СППР состоит из трех компонентов – информационного (базы знаний), автоматизационного и выходных данных. Каждый из них имеет свое назначение, состав и методы взаимодействия с другими элементами СППР.

Одним из элементов информационного блока является дерево решений. Дерево содержит сведения, необходимые для проведения и верификации результатов проектирования системы документирования по различным показателям, а также для вынесения экспертных оценок. Кроме того, здесь хранятся значения допустимых отклонений и погрешностей.

Разработанная структурная схема СППР может быть дополнена другими модулями для расширения функциональности системы или для ее перепрофилирования.

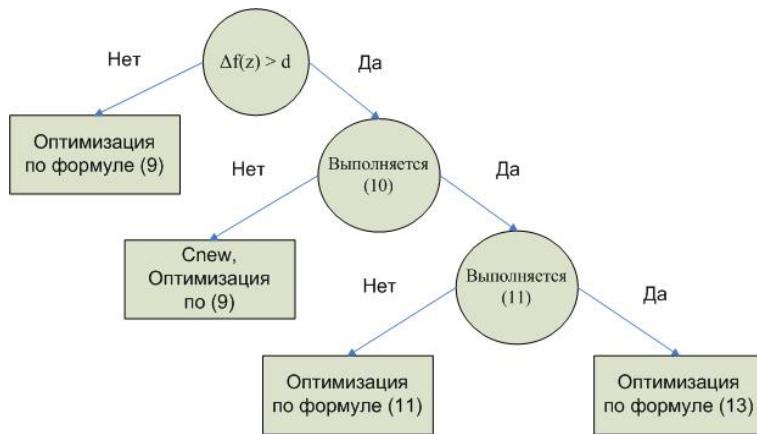


Рис. 2. Дерево решений СППР

Заключение

Выделена система документирования первичного потока научно-технической информации в приборостроении как объект автоматизированного проектирования. Показано, что характерной особенностью этого процесса является текущее перераспределение технологических проблем между автоматизированными рабочими местами. Предложен подход к формализации процесса принятия решений при построении системы документирования, позволяющий проводить синтез проектных решений для конкретных условий.

Литература

1. Федеральный закон № 77-ФЗ «Об обязательном экземпляре документов».
2. Булатова А.Т. Документоведение. Учебное пособие. – М.: Инфра-М, 2005. – 183 с.
3. ГОСТ Р 51141-98. Делопроизводство и архивное дело. Термины и определения.
4. Гусарова Н.Ф., Чернышов А.С., Маятин А.В. Система поддержки принятия решения для предметной области «Полиграфия» на примере настольной издательской системы «InDesign» // Научно-технический вестник СПбГУ ИТМО. – 2005. – Выпуск 20. – С. 248–251.
5. Иванова Т.Б. Компьютерная обработка информации. Допечатная подготовка. Учебное пособие. – СПб: Питер, 2004. – 368 с.
6. Горбачев А.А. Самарин Ю.Н. Методика определения производительности системы допечатной подготовки: Монография. – М.: Август-Принт, 2006. – 112 с.
7. ГОСТ Р ИСО/МЭК 8824-2-2001. Информационная технология. АСН 1. Спецификация информационного объекта.
8. Норенков И.П. Основы автоматизированного проектирования. Учебник для вузов. – М.: МВТУ им. Н.Э. Баумана, 2002. – 336 с.
9. Гусарова Н.Ф., Маятин А.В., Смирнов Ф.А. Обратные задачи в компьютеризированных технологических средах // Научно-технический вестник СПбГУ ИТМО. – 2007. – № 44. – С. 284–294.
10. Garicano L. Hierarchies and the Organization of Knowledge // Journal of Political Economy. – 2000. – V. 108. – № 5. – P. 874–904.
11. Губко М.В. Математические модели оптимизации иерархических структур. – М.: ЛЕНАНД, 2006. – 264 с.
12. Айвазян С.А., Бухштабер В.М., Енуков Е.С. Прикладная статистика. Классификация и снижение размерности: Справочник. Т. 3. – М.: Финансы и статистика, 1989. – 608 с.
13. Мирошник И.В. Теория автоматического управления. Линейные системы. – СПб: Питер, 2005. – 336 с.
14. Справочник по теории автоматического управления / Под ред. Красовского А.А. – М.: Наука. Гл. ред. физ.-мат. лит., 1987. – 712 с.
15. Пупков К.А., Устюжанин А.Д. Идентификация и оценка обученности в динамических человеко-машинных системах // Вестник МГТУ им. Н.Э. Баумана. Приборостроение. – 2003. – № 4. – С. 95–103.

Агаркова Наталья Викторовна	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, grazioKISA@yandex.ru
Артемова Галина Олеговна	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, g.o.artemova@gmail.com
Гусарова Наталья Федоровна	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, natfed@list.ru

УДК 004.855.5

МЕТОДИКА ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ ВЫПУСКНИКА

А.С. Пирская

Предлагается новая методика оценивания компетенций выпускника основных образовательных программ в автоматизированной балльно-рейтинговой системе вуза. Разработана модель, устанавливающая связи между компетенциями выпускника и их составляющими компетенциями в дисциплинах. Уровни сформированности компетенций в ней устанавливаются на основе единого тарификатора результатов обучения (компетенций).

Ключевые слова: компетенция выпускника вуза, уровень сформированности компетенции, компетентностная модель выпускника, оценки результатов освоения компетенций, рейтинг студента.

Введение

Федеральный государственный образовательный стандарт высшего профессионального образования (ФГОС ВПО), выполняя функцию социальной нормы качества ВПО, может трактоваться как рамочная модель выпускника вуза данного направления и уровня подготовки. Квалификационные требования ФГОС к уровню профессиональной подготовки выпускника задаются в виде пакета общекультурных (ОК) и профессиональных компетенций (ПК), которыми он должен обладать после завершения одной из основных образовательных программ (ООП) вуза. На их основе выпускающие кафедры вузов с учетом собственных научно-педагогических традиций и потребностей рынка труда самостоятельно разрабатывают требования к выпускникам своих профильных ООП. Такие требования формулируются в виде компетентностной модели выпускника (КМВ) ООП, которая в ходе разработки и реализации ООП систематически модифицируется и уточняется. При разработке КМВ ООП вузы должны устанавливать требования к уровням сформированности компетенций выпускника из соответствующих ФГОС ВПО. В работе использован единый тарификатор результатов обучения (РО) (компетенций) Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (НИУ ИТМО), устанавливающий общие требования к знаниям, умениям и личностным качествам выпускника университета [1]. Предложенная методика оценивания компетенций выпускника ООП ориентирована на использование в балльно-рейтинговой системе (БаРС) вуза. Данная методика может служить основой для модернизации технологии существующей БаРС НИУ ИТМО по обеспечению перехода на новые ООП, реализующие ФГОС ВПО и собственные образовательные стандарты НИУ ИТМО по подготовке магистров.

Методики разработки компетентностной модели выпускника ООП

КМВ ООП представляет собой пакет взаимосвязанных компетенций выпускника ООП и результатов их освоения, ожидаемых при завершении ООП. КМВ ООП описывается на следующих уровнях ее представления:

1. составные компетенции выпускника ООП;
2. составляющие компетенции выпускника ООП;
3. результаты освоения компетенций выпускника ООП.

Методика разработки составных компетенций выпускника ООП. В ряде случаев компетенция из ФГОС ВПО может содержать не одну, а несколько формулировок компетенций, взаимосвязанных совокупным РО. Кроме того, некоторые компетенции из ФГОС ВПО могут дополнительно детализироваться разработчиками КМВ ООП путем уточнения объектов и видов деятельности выпускника [2]. Каждая такая формулировка является составной для компетенции ФГОС ВПО. Методика разработки КМВ ООП включает анализ компетенций ФГОС ВПО, определение составных компетенций выпускника ООП и установление требований к ожидаемому уровню их освоения в соответствии с заявленным профилем подготовки. Разработчики при этом могут устанавливать различные требования к ожидаемому уровню освоения составных компетенций выпускника ООП, реализующей ФГОС ВПО. Для установления требований разработчики ООП могут использовать тарификатор РО (компетенций) [1].

Пример 1. Профессиональная компетенция ФГОС ВПО направления 230400 – «Информационные системы и технологии» (уровень бакалавра) ПК-12 – «Способность разрабатывать средства реализации информационных технологий (методические, информационные, математические, алгоритмические, технические и программные)» – детализируется в КМВ с установленными ожидаемыми уровнями их освоения в ООП с профилем «Автоматизация и управление в образовательных системах» следующим образом:

- способность разрабатывать *методические средства* реализации информационных технологий (32-У2-СЛ1);
- способность разрабатывать *информационные средства* реализации информационных технологий (33-У3-СЛ1);

- способность разрабатывать *математические средства* реализации информационных технологий (32-У2-СЛ1);
- способность разрабатывать *алгоритмические средства* реализации информационных технологий (33-У3-СЛ1);
- способность разрабатывать *технические средства* реализации информационных технологий (32-У2-СЛ1);
- способность разрабатывать *программные средства* реализации информационных технологий (33-У3-СЛ1).

Здесь в скобках указаны заявленные разработчиком данной ООП ожидаемые уровни сформированности компетенций ее выпускника на основе тарификатора РО (компетенций) НИУ ИТМО [1]. Модель, устанавливающая связи между компетенциями ФГОС ВПО и их составными компетенциями выпускника ООП, приведена на рис. 1.

Методика разработки составляющих компетенций выпускника ООП. Составные компетенции выпускника ООП в КМВ (рис. 1) служат основной для разработки компетенций дисциплин (КД), которые по отношению к ним являются составляющими компетенциями выпускника ООП. Среди составных компетенций выпускника ООП в дисциплине выделяются основные, дополнительные и сопутствующие компетенции. Основные компетенции непосредственно связаны с содержанием предметной области изучения дисциплины, а дополнительные компетенции, как правило, являются компетенциями из других предметных областей обучения (из других дисциплин), но частично формируются на содержании данной дисциплины. Разработчики рабочих программ дисциплин могут запланировать формирование сопутствующих компетенций, которые непосредственно не связаны с содержанием данной дисциплины, а являются результатом проявления технологий преподавания и обучения, используемых при реализации программы данной дисциплины.

Методика разработки составляющих компетенций выпускника в дисциплинах ООП схожа с методикой разработки составных компетенций выпускника ООП, она связана с поочередным уточнением объектов и видов деятельности выпускника [2]. При этом для формирования различных составных компетенций выпускника может планироваться общая составляющая их компетенция выпускника в дисциплине. Это обстоятельство связано с тем, что такая составляющая компетенция формируется на общем содержании дисциплины.

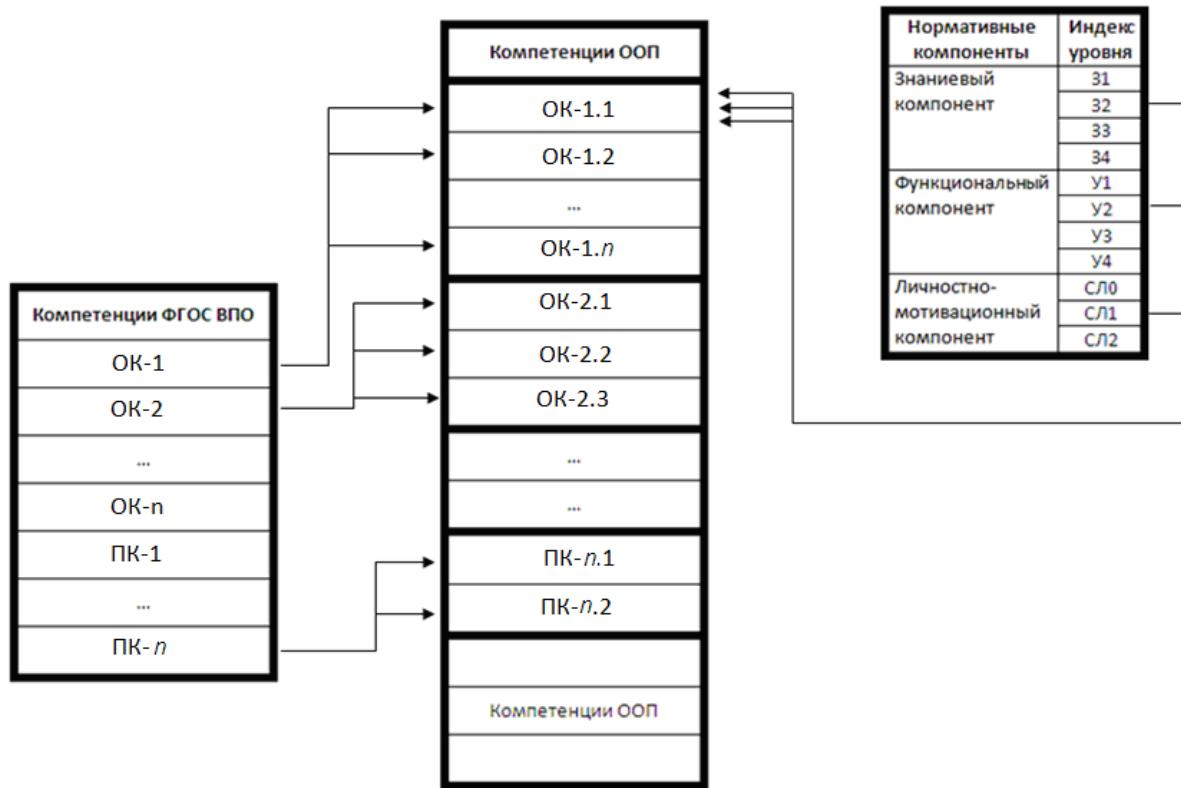


Рис. 1. Компетентностная модель выпускника ООП на основе составных компетенций ФГОС ВПО

Пример 2. При переходе от составных компетенций выпускника ООП (пример 1) к их составляющим в дисциплине «Информатика» разработчик программы этой дисциплины установил бинарное отношение «быть основой» на множествах объектов основной компетенции ОК-6 (практические задачи

информатики) и объектов дополнительной компетенции ПК-12 (средства разработки информационных систем и технологий). Составляющие компетенции выпускника в дисциплине КД представляют собой результат отображения элементов множества практических задач информатики (ОК-6) и множества средств разработки информационных систем и технологий (ПК-12). Установленное таким образом отображение $\varphi: \text{ОК-6} \rightarrow \text{ПК-12}$ (рис. 2) является многозначным, так как каждой практической задаче информатики ОК-6 можно поставить в соответствие различные средства разработки информационных систем и технологий ПК-12.

- способность интерпретировать и визуализировать результаты исследований для разработки методических средств информационных технологий (КД 2);
- способность представлять и кодировать информацию для разработки информационных средств информационных технологий (КД 3);
- способность проектировать цифровые схемы хранения и обработки информации для разработки информационных средств информационных технологий (КД 4);
- способность решать задачи по сортировке и поиску информации для разработки информационных и программных средств информационных технологий (КД 5);
- способность строить вычислительные системы и информационные сети для разработки технических средств информационных технологий (КД 6);
- способность проектировать и создавать мультимедиа приложения для разработки информационных и программных средств информационных технологий (КД 7);
- способность управлять проектами для разработки программных средств информационных технологий (КД 8).

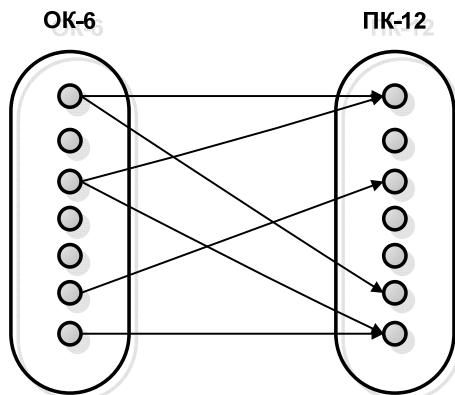


Рис. 2. Модель отношения компетенций выпускника ООП в дисциплине «Информатика»

Причинно-следственные связи, установленные между результатами освоения составляющих компетенций выпускника в дисциплинах ООП, позволяют разработать модель образовательного процесса в виде план-графа [2], определить структуру дисциплины и отобрать для нее содержание образовательного процесса в соответствии с моделью (рис. 2), что позволит в дальнейшем определить еще один важный идентификатор результатов освоения этих компетенций – характеристики учебной нагрузки в дисциплинах.

Методика разработки результатов освоения компетенций выпускника ООП. Результаты освоения компетенций выпускника ООП разрабатываются преподавателями дисциплин ООП и представляют набор знаний, умений и навыков (ЗУН), которые должны быть у студента для формирования соответствующих составляющих и составных компетенций выпускника ООП. ЗУН формулируются в соответствии с ожидаемыми уровнями по тарификатору РО (компетенций), они содержат только те знания, умения и навыки, которые сможет зафиксировать, измерить и оценить преподаватель данной дисциплины. Для этого в рабочей программе дисциплины [1] должны быть соответствующие оценочные средства, обеспечивающие текущий, рубежный и промежуточный контроль всех заявленных результатов освоения компетенций выпускника ООП в дисциплине.

Методика оценивания компетенций выпускника

Методика оценивания компетенций выпускника ООП основана на использовании КМВ ООП, устанавливающей связи компетенций выпускника с их составляющими компетенциями в дисциплинах (рис. 1, 2). При планировании форм и видов контроля результатов освоения компетенций студентом разработчиками дисциплин ООП устанавливаются их максимальные и минимальные оценки [1]. Связь составляющих компетенций выпускника ООП в дисциплине с компетенциями выпускника ООП, реализующей ФГОС, позволяет рассчитать оценки результатов освоения компетенций выпускника для каждого из студентов, а трудоемкость освоения компетенции выпускника ООП – определить нормированные

оценки сформированности компетенций ее выпускника по всем дисциплинам программы для каждого студента.

Предложенная далее методика оценивания компетенций выпускника развивает технологию оценивания РО в БaPC НИУ ИТМО [3, 4] и позволяет получить объективную интегральную оценку освоения компетенций выпускника ООП каждым студентом.

Соответствие заявленных РО и реальных достижений студентов устанавливается с помощью оценочных средств, являющихся неотъемлемой частью рабочей программы дисциплины ООП. На основе сведений о планируемых формах текущего, рубежного и промежуточного контроля результатов освоения дисциплины, а также сведений о запланированной трудоемкости и максимальных оценках разработанных оценочных средств можно определить трудоемкость формирования составляющих компетенций выпускника в дисциплине и максимальные оценки за результаты освоения компетенций дисциплины, что позволяет в итоге определить максимальные оценки компетенций выпускника по всем дисциплинам ООП.

Основой для расчета трудоемкости формирования составляющих компетенций выпускника в дисциплине и максимальной оценки за результаты освоения компетенций дисциплин являются следующие сведения рабочей программы дисциплины:

1. структура содержания дисциплины и его связь с формируемыми компетенциями;
2. виды учебной нагрузки, структура и трудоемкость их содержания;
3. оценочные средства для контроля результатов освоения дисциплины.

Трудоемкость формирования компетенций дисциплин определяется как суммарная трудоемкость учебной нагрузки разделов дисциплины (РД), направленных на формирование соответствующих компетенций выпускника, т.е.

$$T_{\text{КД}} = \sum_{i=1}^n T_{\text{РД}_i},$$

где n – количество разделов дисциплины ООП, направленных на формирование компетенции дисциплины; i – номер текущего раздела дисциплины ООП; $T_{\text{РД}_i}$ – трудоемкость в часах освоения i -го раздела дисциплины ООП.

Трудоемкость освоения раздела дисциплины $T_{\text{РД}_i}$ определяется на основе данных раздела рабочей программы «Виды учебной нагрузки, структура и трудоемкость их содержания» [1].

Максимальные оценки за результаты освоения компетенций дисциплин определяются как сумма оценок по всем видам контроля дисциплины (текущий, рубежный и промежуточный контроль) и оценок личностных качеств студента.

Раздел рабочей программы дисциплины «Формы текущего и рубежного контроля результатов освоения дисциплины» [1] устанавливает соответствие оценочных средств для контроля освоения РД и, следовательно, контроля освоения соответствующей составляющей компетенции выпускника в дисциплине.

Максимальные оценки освоения компетенций дисциплины при прохождении текущего контроля определяются как сумма максимальных оценок, запланированных разработчиками рабочей программы дисциплины за предусмотренные виды работ (тесты, опросы, лабораторные и практические работы и т.д.), т.е.

$$O_{\text{тек.контрольКД}}^{\max} = \sum_{i=1}^m O_i^{\max},$$

где m – количество оценочных средств для контроля результатов освоения компетенции дисциплины; i – номер текущего оценочного средства дисциплины ООП; O_i^{\max} – максимальный балл, запланированный разработчиком рабочей программы за успешное прохождение текущего контроля с использованием i -го оценочного средства.

Максимальные оценки освоения составляющих компетенций выпускника в дисциплине при прохождении рубежного контроля определяются по формуле

$$O_{\text{руб.контрольКД}}^{\max} = \frac{O_{\text{оц.ср-ва}}^{\max}}{K},$$

где K – количество контролируемых компетенций дисциплины (определяется на основе данных раздела рабочей программы «Формы текущего и рубежного контроля результатов освоения дисциплины» [1]);

$O_{\text{оц.ср-ва}}^{\max}$ – максимальный балл, запланированный разработчиком рабочей программы дисциплины за прохождение текущего контроля.

Для расчета максимальных оценок за результаты освоения компетенций дисциплины при прохождении промежуточного контроля используется максимальный балл за результаты освоения компетенций дисциплин при прохождении промежуточного контроля, нормированный по трудоемкости формирования составляющей компетенции выпускника в дисциплине, т.е.

$$O_{\text{пром.контрольКД}}^{\max} = \frac{O_{\text{пром.контроль}}^{\max} \cdot T_{\text{РД}}}{\sum_{i=1}^l T_i},$$

где l – количество РД, направленных на формирование компетенции дисциплины; i – номер текущего раздела дисциплины ООП; $T_{\text{РД}}$ – трудоемкость в часах освоения раздела дисциплины ООП; $O_{\text{пром.контроль}}^{\max}$ – максимальная оценка, которую может получить студент при прохождении промежуточного контроля, запланированная в рабочей программе дисциплины.

Максимальные оценки личностных качеств (ЛК) студента, проявленных при освоении компетенции дисциплины, определяются следующим образом:

$$O_{\text{ЛККД}}^{\max} = \frac{O_{\text{ЛК}}^{\max} \cdot O_{\text{тек.контрольКД}}^{\max}}{\sum_{i=1}^p O_{\text{тек.контрольКД}}^{\max}},$$

где p – количество компетенций дисциплины; i – номер текущей компетенции дисциплины ООП.

Максимальная оценка за результат освоения компетенции дисциплины определяется как сумма максимальных оценок за прохождение текущего контроля, рубежного контроля, промежуточного контроля и личностных качеств студента, т.е.

$$O_{\text{КД}}^{\max} = O_{\text{тек.контрольКД}}^{\max} + O_{\text{руб.контрольКД}}^{\max} + O_{\text{пром.контрольКД}}^{\max} + O_{\text{ЛККД}}^{\max}.$$

В случае если один РД направлен на формирование взаимосвязанных компетенций выпускника, то баллы, которые может набрать студент при освоении этого раздела, должны накапливаться для каждой из взаимосвязанных компетенций. Это связано с тем, что компетенции формируются на едином предметном содержании дисциплин и взаимно дополняют и расширяют себя.

Пример 3. На основе данных рабочей программы дисциплины «Информатика» [1] и предложенных формул были рассчитаны трудоемкость формирования и максимальные оценки за результаты освоения составляющих компетенций выпускника в дисциплине. Установленные связи между компетенциями дисциплин и компетенциями выпускника позволяют определить максимальные оценки за результаты освоения компетенций выпускника ФГОС.

В табл. 1 приведена связь составных компетенций выпускника с формируемыми компетенциями дисциплин ООП, трудоемкость их формирования и планируемые максимальные оценки результата их освоения.

Предложенная методика оценивания компетенций выпускника позволяет рассчитать максимальные оценки за результаты освоения компетенций выпускника и трудоемкость их формирования при изучении дисциплин ООП и составить сводную таблицу (табл. 2) для расчета рейтинга студента, характеризующего сформированность i -ой компетенции выпускника.

Трудоемкость освоения составной компетенции выпускника (T) определяется как сумма трудоемкостей освоения взаимосвязанных составляющих компетенций дисциплин ООП (T_i), т.е.

$$T = \sum_{i=1}^r T_i,$$

где r – количество составных компетенций выпускника; i – номер текущей составной компетенции; T_i – трудоемкость в часах освоения i -ой компетенции дисциплины ООП, запланированная разработчиками дисциплины ООП. Максимальная оценка результата освоения составляющей компетенции выпускника ООП (O^{\max}) определяется по формуле

$$O^{\max} = \frac{\sum_{i=1}^s O_{\text{КД}i}^{\max} \cdot T_i}{T},$$

где i – номер текущей компетенции дисциплины ООП; s – количество составляющих компетенций дисциплины ООП; $O_{\text{КД}i}^{\max}$ – максимальная оценка в баллах результатов освоения i -ой составной компетенции выпускника. Оценка результатов освоения компетенции студентом (O_i^*), при обучении по конкретной ООП рассчитывается следующим образом: в качестве оценок результатов освоения компетенций дисциплин ООП используются сведения накопительной БаRC об оценках, полученных при прохождении студентом форм контроля результатов освоения i -ой компетенции дисциплины, запланированных разработчиком дисциплины.

Рейтинг студента, характеризующий сформированность i -ой компетенции выпускника, определяется по формуле

$$R_i = \frac{O_i^*}{O_i^{\max}} \cdot 100\%,$$

МЕТОДИКА ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ ВЫПУСКНИКА

где O_i^* – оценка результатов освоения i -ой компетенции выпускника студентом; O_i^{\max} – максимальная оценка результатов освоения компетенции выпускника.

Индекс составляющих компетенций выпускника в дисциплине	Трудоемкость формирования КД, час	Максимальные оценки за результаты освоения, балл					Индекс компетенций ФГОС
		Текущий контроль	Рубежный контроль	Промежуточный контроль	Личностные качества	Всего	
КД1	6	2	0	1,18	0,4	3,58	OK-1 OK-6
КД2	21	15	0	4,12	3	22,12	OK-6 ПК-12
КД3	9	5	5	1,76	1	12,76	OK-6 ПК-12
КД4	9	5	5	1,76	1	12,76	OK-6 ПК-12
КД5	29	10	10	5,69	2	27,69	OK-6 ПК-12
КД6	59	31	10	11,57	6,2	58,77	OK-6 ПК-12
КД7	48	22	0	9,41	4,4	35,81	OK-6 ПК-12
КД8	23	10	10	4,51	2	26,51	OK-6 ПК-12
Итого:	204	100	40	40	20	200	

Таблица 1. Составляющие компетенции выпускника в дисциплине «Информатика», трудоемкость и планируемая оценка результата их освоения

Компетенции ФГОС	...		Б2.1.3 – «Информатика»		...	
	Трудоемкость, час	Макс. оценка, балл	Трудоемкость, час	Макс. оценка, балл	Трудоемкость, час	Макс. оценка, балл
OK-1	204	3,58
...
OK-6	6	200
...
ПК-12	188	196,42
...

Таблица 2. Максимальные оценки за результаты освоения и трудоемкость формирования компетенций выпускника в дисциплинах ООП

Предложенная методика оценивания компетенций выпускника в БаРС обучения позволяет проверять уровень освоения студентом компетенции выпускника на протяжении всего периода обучения и разработать модель мониторинга результативности образовательного процесса. На каждом этапе контроля результатов освоения дисциплин ООП следует проверять уровень освоения составной компетенции выпускника и успешное преодоление студентом минимального порогового значения. Если оценка результата освоения одной из составляющих компетенций выпускника ниже порогового значения, то результат освоения компетенции выпускника из пакета компетенций ФГОС на момент итоговой государственной аттестации может оказаться неудовлетворительным. Это ставит задачу разработки модели управления образовательными траекториями выпускника ООП.

Заключение

В работе предложена методика оценивания компетенций выпускника в автоматизированной БаРС вуза. Разработана модель, устанавливающая связи между компетенциями выпускника и их составляющими компетенциями в дисциплинах основных образовательных программ. Предложена методика разработки компетентностной модели выпускника на основе требований ФГОС.

Методика оценивания компетенций выпускника позволяет определить объективные оценки компетенций выпускника ООП к моменту начала итоговой государственной аттестации. Предложенные модели и методики позволяют проводить мониторинг результативности подготовки каждого студента и использовать их для управления его индивидуальной образовательной траекторией, обеспечивающей достижение сформированности каждой компетенции выпускника основных образовательных программ не ниже минимального (порогового) значения.

Литература

1. Лисицына Л.С., Лямин А.В., Шехонин А.А. Разработка рабочих программ дисциплин (модулей) в составе основных образовательных программ, реализующих ФГОС ВПО. Методическое пособие. – СПб: СПбГУ ИТМО, 2011. – 63 с.
2. Лисицына Л.С. Теория и практика компетентностного обучения и аттестаций на основе сетевых информационных систем. – СПб: СПбГУ ИТМО, 2006. – 147 с.
3. Шехонин А.А., Тарлыков В.А. Балльно-рейтинговая система оценивания результатов образования в информационно-образовательной среде вуза // Труды Всероссийской научно-практической конференции с международным участием «Информационные технологии в обеспечении нового качества высшего образования (14–15 апреля 2010 г., Москва, НИТУ «МИСиС»). – М.: Исследовательский центр проблем качества подготовки специалистов НИТУ «МИСиС», 2010. – Кн.1. – С. 168–173.
4. Лямин А.В., Чежин М.С. Обработка и интерпретация результатов обучения при балльно-рейтинговой системе оценивания / Под общ. ред. проф. А.А. Шехонина // Сборник трудов конференции «Оптика и образование – 2010». – СПб: СПбГУ ИТМО, 2010. – С. 23–24.

Пирская Анна Сергеевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, peeerskaya@mail.ru

УДК 67.02, 621.37

ОРГАНИЗАЦИЯ ОБУЧЕНИЯ СТУДЕНТОВ ТЕХНОЛОГИИ ПРОИЗВОДСТВА ЭЛЕКТРОННОГО УСТРОЙСТВА В ЕДИНОЙ ИНФОРМАЦИОННОЙ СРЕДЕ

А.С. Воробьев, Ю.В. Федосов

Предложен учебный пример технологического процесса в единой информационной среде на примере создания печатной платы в лаборатории поверхностного монтажа. Рассмотрены теоретические основы и практическая реализация цикла подготовки производства и изготовления изделия. Особое вниманиеделено работе с автоматизированным оборудованием.

Ключевые слова: технология поверхностного монтажа, обучение, подготовка производства.

Введение

В последние годы на фоне развития промышленного производства современной техники особое место занимает бурно развивающийся сектор электронного оборудования. Растет степень интеграции микросхем и микросборок, повышается количество компонентов в изделиях и плотность их упаковки. Вместе с повышением компактности изделий ужесточаются и требования, предъявляемые к ним. В связи с этим ручной монтаж повсеместно заменяется на автоматизированный. В этом процессе важную роль приобретает подготовка компетентных специалистов, способных решать нетривиальные технологические задачи. Преподавание технологии электромонтажа имеет свои особенности. В последнее время технологическое оборудование постоянно усложняется. Вследствие этого учащимся становится все труднее представить себе принципы его работы, особенно, если они раньше не имели опыта общения с таким оборудованием. Из-за увеличения объема производимых оборудованием операций увеличивается количество необходимых приемов и нюансов практической работы с ним, поэтому студентам приходится изучать все больше и больше его особенностей, которые невозможно описать на лекциях: нужно непосредственно взаимодействовать с оборудованием, чтобы понять, как оно работает.

Постановка задачи

Развитие современного образования в России является непрерывным процессом и требует от учащихся и обучающих новых инновационных подходов к обучению с целью сформировать новые компетенции, отвечающие последним веяниям науки и техники. Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики успешно осуществляет ряд стратегических разработок по государственным программам и проектно-конструкторским работам. На кафедре технологии приборостроения (ТПС) введена в эксплуатацию лаборатория поверхностного монтажа изделий электронной техники. На кафедре реализуется модель современного производства, когда разработка и изготовление изделий происходит в единой информационной среде. С целью оптимизации

процесса обучения необходимо сформулировать для учащихся конкретное представление о производственно-технологическом процессе электромонтажа.

Удобно рассмотреть в качестве учебного примера создание печатной платы (ПП), спроектированной студентом в рамках учебного процесса по смежной дисциплине, например, конструированию радиоэлектронной аппаратуры.

Рассмотрим ход работы в единой информационной среде, состоящей из программного комплекса Altium Designer (trial-версии) и управляющей программы установки Mechatronika M60, на примере проектирования и изготовления платы для блока подсчета объектов, проходящих через считающее устройство (рис. 1). Устройство имеет возможность запрограммировать отображение их количества (например, подсчет объектов осуществляется поштучно по 80 штук). Изготовление платы включает в себя заказ и приобретение самой платы, а также компонентов, автоматизированный монтаж компонентов на плату, операции сушки, очистки и проверки изделия. Цель – создание сборки платы.

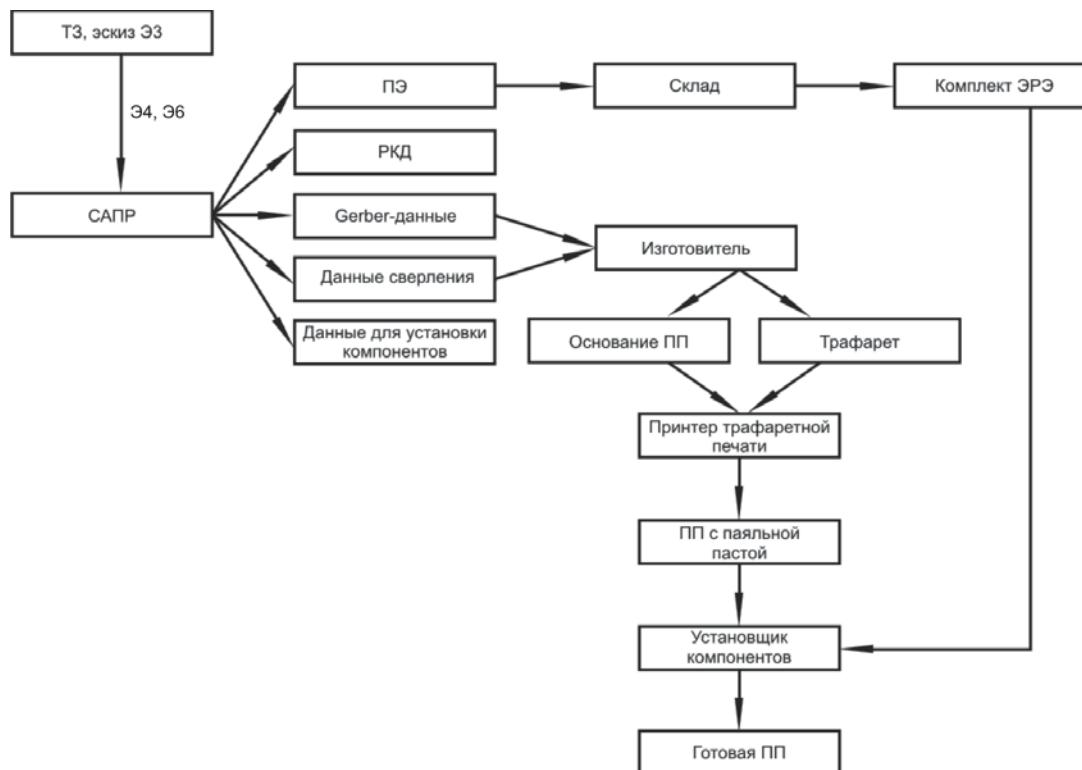


Рис. 1. Алгоритм работы в единой информационной среде: ПП – печатная плата, ПЭ – перечень элементов; РКД – рабочая конструкторская документация; Э3 – схема электрическая принципиальная; Э4 – схема электрическая соединений; Э6 – схема электрическая структурная; ЭРЭ – электрорадиоэлементы

Ход работ

Как правило, схема работы состоит в следующем. По заранее написанному техническому заданию осуществляется создание схемы Э6 (рис. 1). Затем определяются отдельные элементы схемы, их функции и назначение. Далее создается эскиз схемы Э4 и подбираются элементы по каталогу. Если необходимые элементы не удается подобрать, то производится редактирование параметров и функций элементов схемы, после чего цикл повторяется. Затем на основе имеющейся схемы соединений создается схема Э3 и заносится в память персонального компьютера (ПК) средствами системы автоматизированного проектирования (САПР). Далее, на основе имеющейся схемы в САПР Э3, конструктор в автоматизированном режиме создает ПП и размещает на ней компоненты схемы. Если какие-либо компоненты отсутствуют в базе данных ПК, их необходимо создать. Трассировка ПП производится либо автоматически, либо в полуавтоматическом режиме. При этом положение компонентов на плате, возможно, корректируется (рис. 2). Далее схема проходит симуляцию с целью проверки ее физических параметров и совместности компонентов, и, если ее результаты неудовлетворительны, то осуществляется повторный цикл разработки с момента размещения компонентов. Если данные симуляции все равно остаются неудовлетворительными даже после нескольких циклов размещения, то осуществляется повторный подбор компонентов.

Когда данные симуляции удовлетворяют заданным требованиям, происходит подготовка данных для изготовления ПП. В производство передаются данные сверления ПП, данные топологии ПП, данные контактных площадок для нанесения пасты на ПП и данные о положении компонентов ПП для автомата-установщика.

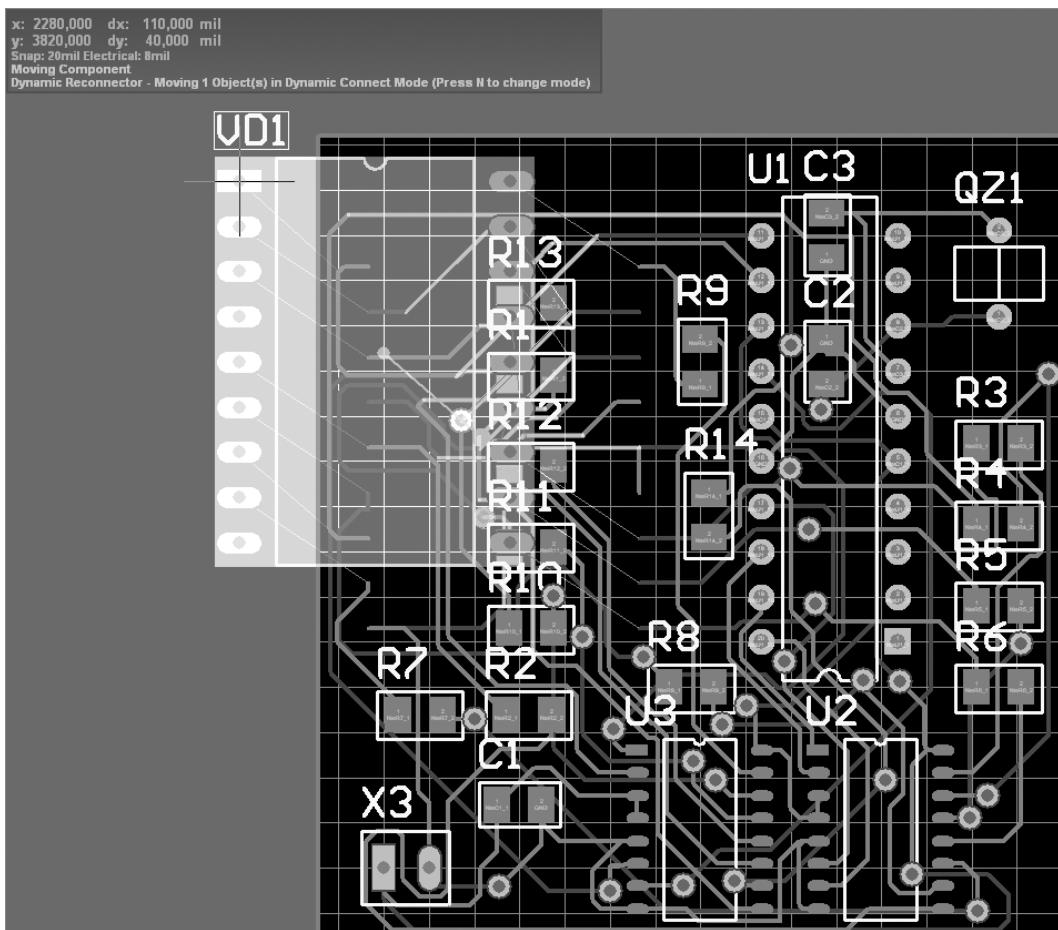


Рис. 2. Пример работы с ПП

Рассмотрим подготовку данных для производства ПП на примере работы лаборатории поверхностного монтажа и компании «Резонит».

Передача данных в производство

В лаборатории была произведена подготовка технологических данных ПП в САПР Altium Designer согласно технологической инструкции, полученной от предприятия-производителя.

Для подготовки данных топологии и сверления ПП в программе необходимо задать конфигурацию Gerber-файлов [1] (выбрать «File → Fabrication Outputs → Gerber Files»). В закладке «General» необходимо задать единицы измерения и формат вывода Gerber-файлов:

- единицы измерения – миллиметры;
- формат вывода – количество цифр в координатах площадок, линий и т.п. до и после десятичной запятой.

Далее нужно сконфигурировать набор слоев, необходимых для изготовления печатной платы. В закладке «Layers» необходимо выбрать следующие слои:

- TopSolder – верхний слой, формирующий вскрытия в паяльной маске под последующее финишное покрытие контактных площадок и других открытых областей;
- TopLayer – слой топологии на верхней стороне печатной платы;
- BottomLayer – слой топологии на нижней стороне печатной платы;
- BottomSolder – нижний слой, формирующий вскрытия в паяльной маске под последующее финишное покрытие контактных площадок и других открытых областей.

Поле «Include unconnected mid-layer pads» (включая неподключенные площадки на внутренних слоях) для производства ПП следует оставить заполненным. Закладка «Drill Drawing» пропускается, так как на производстве сверление выполняется на станках с ЧПУ, и в картах нет необходимости.

В закладке «Apertures» необходимо поставить галочку в поле «Embedded apertures (RS274X)».

В результате список используемых апертур (набор примитивов) для формирования рисунка печатной платы будет располагаться в начале каждого Gerber-файла.

Далее появляется поле с готовой трассировкой топологии ПП в выбранных слоях. Для сохранения результатов необходимо выбрать «File → Export → Gerber». Должен быть отмечен пункт «RS-274X» [2, 3]. Для формирования файла с программой сверления выбрать «File → Fabrication Outputs → NC Drill

Files». Здесь необходимо задать единицы измерения (миллиметры) и формат вывода (количество цифр в координатах отверстий до и после десятичной запятой). Если в качестве единиц измерения выбраны миллиметры, то для сетки 1 мкм применим формат вывода 4:3; для сетки 0,1 мкм – 4:4; в остальных случаях достаточно выбрать формат 4:2. Остальные параметры можно оставить без изменений.

Для сохранения результатов необходимо выбрать «File → Export → Save Drill → Select layer».

В результате проделанной работы в папке проекта сформируются все необходимые файлы для производства печатной платы и файлы отчетов (Name.EXTPR – отчет о наборе Gerber-файлов, Name.DRR – отчет о программе сверления).

После подтверждения производителем полноты и достоверности данных проекта был заполнен бланк заказ-наряда, в котором указаны необходимые файлы и параметры ПП. В заказ-наряде был обговорен ряд пунктов технологического процесса (золочение, металлизация отверстий, маска, обработка контура ПП), а также характеристики слоев, включенных в проект. Далее был запрошен векторный рисунок верхнего слоя ПП для проверки масштаба фотошаблона со стороны заказчика. После сверки масштаба был подтвержден запуск ПП в производство.

После изготовления основания ПП на нее устанавливаются компоненты схемы. Для установки компонентов на ПП необходимо передать информацию по компонентам на станок. Рассмотрим ход трансляции данных на примере станка Mechatronica M60. Чтобы передать в его память координаты компонентов, нужно:

- в САПР Altium Designer выбрать пункт File, затем «Assembly outputs → Generates pick and place files». Отметить пункт «CSV», затем нажать OK;
- в основном окне управляющей программы станка Mechatronica M60 создать новый файл и определить координаты платы («Pcb Position → Set → Teach In → значок перекрестия») [4]. Далее измеряется высота платы относительно вакуумного захвата («Surface → Measure → значок перекрестия → OK»). Затем задаются реперные точки «First Fiducal → Train → значок перекрестия → OK. Second Fiducal → Train → значок перекрестия → OK». После этого тестируется качество определения реперных точек «First Fiducal → Test → OK. Second Fiducal → Test → OK».

Далее определяются положения компонентов в питателях («Set pickup position → OK → Teach in → OK») и измеряются высоты компонентов относительно вакуумного захвата («Set pickup height → OK → Measure → OK») (таблица).

O15	O14	O13	O12
Res 1,5 K	0,68 μF	15 pF	

Таблица. Отображение типов используемых компонентов в окне программы установщика

Задав типоразмер захватываемого компонента на строчке питателя (пункт «Load component»), нужно выбрать компонент из библиотеки. Если данный компонент в библиотеке отсутствует, то следует выбрать похожий на него и изменить его параметры по щелчку правой кнопки мыши. Для задания координат расстановки компонентов на вкладке «Program» в «Parts placement» нужно выбрать пункт «Import Protel CSV». Компоненты расставляются в автоматическом режиме («Package assignment» → «Assign package → выбрать типоразмер → «Get PCB offset»). На любом этапе работы можно изменить компоненты (Feeders → «Change components») или расположение компонентов в питателях («Edit»).



Рис. 3. Работа на автоматизированном оборудовании

Нанесение пасты на ПП производится посредством специального трафаретного принтера. Для работы с заданной ПП принтеру требуется трафарет, созданный для этой ПП. В случае с рассматриваемым примером трафарет был создан по субподряду. Для получения данных о контактных площадках по пункту меню «Fabrication outputs → Gerber files» на вкладке «Layers» был выбран слой разработанной ПП «Top Solder». После этого осуществляется сборка ПП на автоматизированном оборудовании (рис. 3).

Для завершения сборки необходимо оплавление паяльной пасты в конвекционной печи, промывка в ультразвуковой ванне и визуальный контроль качества пайки на стереоувеличителе. Эти операции являются типовыми и не привязаны к конкретному оборудованию.

Заключение

В результате проделанной работы был сформулирован подробный алгоритм формирования сборки печатной платы, предназначенный для обучения навыкам печатного монтажа и позволяющий полностью осознать данный технологический процесс. Подобная схема построения технологического процесса в единой информационной среде на примере создания печатной платы в лаборатории поверхностного монтажа является наглядным примером работы в единой информационной среде. Тщательный подбор примеров позволяет ускорить и упростить ввод исходных данных, что очень важно для процесса обучения. Технология микропроцессорных изделий очень быстро меняется, каждый месяц появляется новое оборудование, новые материалы и технологии. Все сведения о технологии, рассказанные студентам, могут измениться, когда студенты получат дипломы и приступят к работе. В связи с этим для решения технологических задач очень важно умение работать с автоматизированным оборудованием и использовать информационную среду.

Литература

1. Указания по подготовке данных из САПР в Gerber-формате: официальный сайт компании «Резонит» [Электронный ресурс]. – Режим доступа: <http://rezonit.ru/urgent/howto/>, свободный. Яз. рус. (дата обращения – 30.11.2011).
2. Сабуний А.Е. Altium Designer. Новые решения в проектировании электронных устройств. – М.: СОЛОН-ПРЕСС, 2009. – 432 с.
3. Суходольский В.Ю. Сквозное проектирование функциональных узлов РЭС на печатных платах в САПР Altium Designer 6. Ч. 1, 2: Учебное пособие. – СПб: СПбГЭТУ «ЛЭТИ», 2008. – 148 с.
4. Инструкция к станку. Установщик компонентов MECHATRONIKA. – М50/60. – СПб: ЗАО НПФ «Диполь», 2006.

Воробьев Анатолий Сергеевич	– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, anatoliy.vorobyev@gmail.com
Федосов Юрий Валерьевич	– ЗАО «МНС», аспирант, yf01@yandex.ru

УДК 621.38, 372.862, 377.5

КОМПЛЕКТЫ ВИЗУАЛЬНЫХ И РАСЧЕТНЫХ ЛАБОРАТОРНЫХ МОДЕЛЕЙ ПО ОСНОВАМ ЦИФРОВОЙ ЭЛЕКТРОННОЙ ОБРАБОТКИ ИНФОРМАЦИИ

В. В. Новиков, В. В. Генович, Д. М. Гриншпун

Рассмотрены вопросы разработки комплексов компьютерных моделей электронных устройств, входящих в состав виртуального лабораторного комплекса по основам электронной обработки информации. Вводятся понятия визуальных и расчетных моделей в рамках комплекса, комплексов визуальных и расчетных моделей, рассматриваются этапы их разработки и вопросы составления математических описаний моделей – определения применяемых способов аппроксимации и методов составления математических описаний. Анализируются требования к визуальным и расчетным моделям с технической и педагогической сторон, приводится результат разработки комплексов.

Ключевые слова: виртуальный лабораторный комплекс, основы электронной обработки информации, визуальные модели, расчетные модели.

Введение

Представленный в настоящей работе виртуальный лабораторный комплекс предназначен для математического моделирования и визуального представления физических процессов в электронных узлах формирования цифровых сигналов и их логических преобразований.

Разработка комплекса (далее – Комплекса) ведется с 2008 г. Первые результаты, продемонстрированные на «V Всероссийской межвузовской конференции молодых ученых» в 2008 г. и на конференции «Научный потенциал-XXI век» в 2009 г. [1, 2], содержали разработки по тематическим разделам «Формирование цифровых сигналов»; «Принципиальная реализация логических элементов»; «Синтез комбинационных схем» и состояли из нескольких разработанных лабораторных работ, включавших лабора-

торные установки (компьютерные модели имитируемых устройств) и методические указания по их выполнению. При этом компьютерные модели имитируемых устройств делились на два типа: имитационные и функциональные.

Педагогический эксперимент привел к выводу о целесообразности тематического расширения и разработки единой концепции виртуальных лабораторных комплексов по основам цифровой электронной обработки информации (далее – Концепция). Она была представлена на конференциях «Телематика'2010» и «Современные достижения в науке и образовании» в 2010 г. [3, 4]. В соответствии с Концепцией Комплекс дополнен еще двумя типами моделей: расчетными и визуальными. Под расчетной понимается модель исследуемого объекта, предназначенная для контроля знаний учащихся и проведения лабораторных работ; под визуальной – модель исследуемого объекта, наглядно демонстрирующая принцип его функционирования на уровне физических процессов; под комплектом расчетных моделей (КРМ) и комплектом визуальных моделей (КВМ) – набор расчетных и визуальных моделей по тематическим разделам:

- формирование цифровых сигналов (ФЦС);
- логические элементы (на биполярных транзисторах) (ЛЭ 1);
- логические элементы (на полевых структурах) (ЛЭ 2);
- трехстабильные логические элементы (ТСЭ);
- комбинационная/некомбинационная логика (КЛ/НКЛ);
- магистральные структуры (МС);
- программируемые структуры (ПС).

Результаты первых этапов разработки представлены на конференции «Телематика'2010» [5].

Обоснование разработки

Разработка осуществляется на основе выработанных экспертным методом требований, подробно изложенных в Концепции [3]. Так, визуальные модели (ВМ) должны:

- отображать схемы изучаемых электронных узлов;
- функционировать на основе их математического описания;
- обеспечивать интерактивный режим использования – реакцию на задаваемые пользователем посредством имитаторов приборных органов управления воздействий (установку параметров электронных узлов и входных сигналов) и динамическое отображение результатов моделирующего расчета при помощи элементов индикации, имитирующих физические измерительные приборы;
- наглядно отображать физические процессы, протекающие в электронных устройствах.

Расчетные модели (РМ) также должны отображать схемы изучаемых электронных узлов, функционировать на основе их математического описания и обеспечивать интерактивный режим использования, но в отличие от ВМ, интерактивность достигается вводом и выводом непосредственно числовых значений вместо имитации приборного управления и индикации. Кроме того, должен производиться вывод дополнительных параметров, избыточных с точки зрения получения конечных значений, но необходимых для обоснования математических описаний и путей решения – напряжения и силы токов в каждом узле и цепи устройства.

Кроме того, должна быть учтена специфика целевой аудитории – студенты и учащиеся образовательных уровней среднего профессионального образования (ОУ СПО), начального профессионального образования (НПО), школьники старших классов информационно-технологического профиля. Это определяет предпочтительно линейный (кусочно-линейный) метод аппроксимации нелинейных характеристик полупроводниковых приборов.

Сопоставительный анализ аналогичных решений комплексов лабораторных и демонстрационных работ [6–14] показал, что все представленные программные продукты обладают возможностью интерактивности и индикации измерительных показаний. Однако визуализацию на уровне физических процессов реализуют только отдельные модели пакета «Открытая физика», в которых отсутствуют модели полупроводниковых элементов, и ни один аналог не поддерживает линейную аппроксимацию характеристик полупроводниковых элементов, основываясь на математическом описании системой дифференциальных уравнений, что исключает эффективное применение целевой аудиторией указанного образовательного уровня.

Все вышеизложенное привело к двухэтапной разработке оригинального комплекса:

- математические описания работы электронных устройств;
- создание расчетных и визуальных моделей устройств на основе их математических описаний.

Разработка математического описания

Согласно Концепции, математическое описание всех типов моделей в рамках одного тематического раздела и одного образовательного уровня должны быть одинаковы, что позволяет разрабатывать единые математические модели (ММ). Образовательный уровень целевой аудитории определил следующий способ аппроксимации характеристик полупроводниковых приборов:

1. Вольтамперная характеристика p-n-перехода аппроксимирована ступенчатой функцией: падение напряжения при прямом включении составляет 0,6 В; обратный ток равен нулю.
2. В линейном режиме работы биполярного транзистора коэффициент передачи по току $\beta = \frac{i_k}{i_b}$ (i_k, i_b – ток коллектора и базы) принимается равным константе.
3. В режиме насыщения биполярного транзистора минимальная разность потенциалов между коллектором и эмиттером U_{ce} равна 0,4 В.

Приведенные численные значения соответствуют характеристикам маломощных высокочастотных кремниевых полупроводниковых диодов и транзисторов.

Составление ММ состоит из следующих этапов:

- определение областей режимов работы имитируемого электронного узла (сочетаний состояний входящих в него полупроводниковых приборов);
- исключение из полученного множества режимов невозможных при задаваемых значениях параметров цепи;
- определение условий работы каждого рабочего режима;
- разработка ММ для рабочих режимов.

Разработка ВМ

Приняты следующие способы визуализации: движение потока «положительно заряженных частиц», направление движения которых совпадает с направлением тока, или потока электронов, направление движения которых обратно направлению тока.

Для разработки графического интерфейса применен программный пакет Adobe Flash, для моделирования и программной визуализации – встроенный язык программирования ActionScript 3.0.

Схема функционирования визуальных моделей включает в себя три этапа.

1. Ввод данных с помощью приборных органов управления: параметров имитируемого электронного узла и входных сигналов.
2. Расчет, состоящий из трех шагов:
 - расчет всех рабочих режимов модели;
 - проверка на непротиворечивость условиям работы режимов;
 - вывод результатов расчета непротиворечивого режима.
3. Визуализация – определение путей движения заряженных частиц в электронной схеме и плотности их потока; вывод на монитор.

Разработка РМ

В качестве программной среды используется продукт, не требующий специальных знаний и навыков программирования, но обеспечивающий реализацию математического описания модели на уровне расчета – MS Excel. Схема функционирования РМ, как и ВМ, состоит из тех же этапов, за исключением режима ввода и вывода данных, которых на порядки больше, и поэтому используются упрощенные поля форм.

Каждая модель представляет собой стандартную книгу MS Excel, состоящую из титульного листа и листов ММ режимов. Ячейки листов разделяются на два типа: ячейки с данными и параметрами и ячейки с формулами. Задачи титульного листа:

- обеспечение ввода параметров и входных сигналов модели;
- передача параметров и входных сигналов на листы ММ;
- вывод результатов проверок каждого рабочего режима.

Задачей листов ММ является расчет промежуточных и выходных значений.

В соответствии с этим титульный лист включает в себя следующие таблицы:

- параметров схемы;
- входных сигналов;
- вектора проверок режимов.

Листы ММ моделей режимов включают в себя таблицы:

- параметров схемы;
- входных сигналов;
- промежуточных значений;
- входных и выходных значений;
- результатов проверок режимов.

Вектор проверок режимов формируется на основе результатов проверок режима по каждому листу ММ.

Пример разработки

В качестве примеров представлены интерфейсные окна визуальной (рис. 1) и расчетной (рис. 2) моделей двухкаскадного узла формирования двухтактных сигналов в схемотехнике транзисторно-транзисторной логики (ТТЛ). Назначение этих моделей – демонстрация принципа работы.

Входные данные: сопротивления R_1, R_2, R_3, R_4 и напряжение U_{bx} , напряжение электропитания E .
Выходные данные: напряжения U_{y1} и U_{y2} .

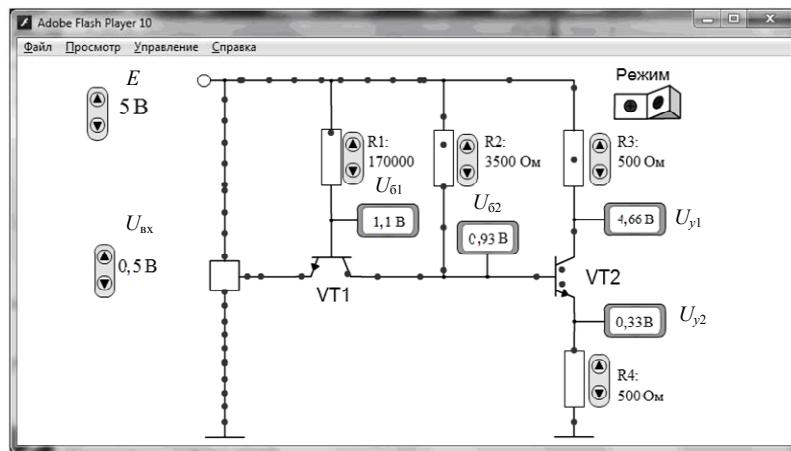


Рис. 1. Иллюстрация окна ВМ «Двухкаскадный узел формирования двухтактных сигналов в схемотехнике ТТЛ»

В соответствии с тематическим разделом (формирование цифровых сигналов) наибольший интерес представляют переключательные свойства моделируемого электронного устройства, непосредственно зависящие от переходных режимов полупроводниковых элементов. В связи с этим важно оценить адекватность функционирования разработанных моделей в переходных режимах.

В качестве метода оценки было применено сопоставление представленной выше модели двухкаскадного узла формирования двухтактных сигналов в схемотехнике ТТЛ (модель А) с аналогичной, построенной в среде наиболее близкого аналога Multisim 11 (моделью Б). В первом случае использовалась модель транзистора KT315, во втором случае – зарубежный аналог 2N2712.

На рис. 3 представлена временная диаграмма входного сигнала, на рис. 4 – укрупненные временные диаграммы выходных сигналов.

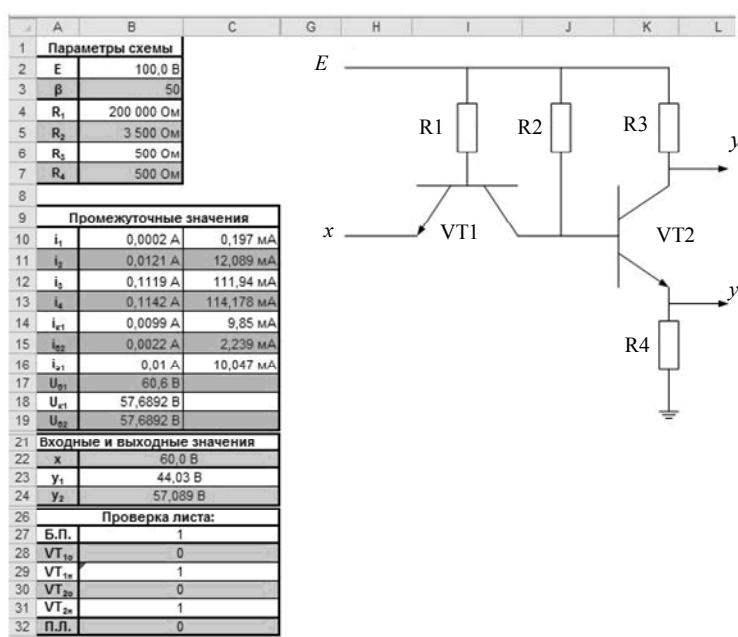


Рис. 2. Фрагмент окна РМ «Двухкаскадный узел формирования двухтактных сигналов в схемотехнике ТТЛ»

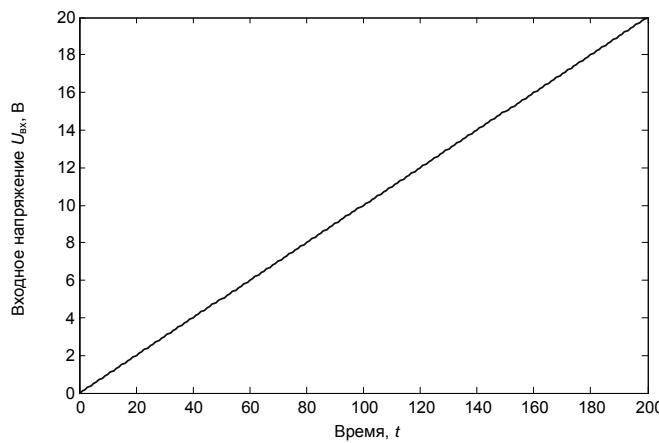


Рис. 3. Временная диаграмма входного сигнала

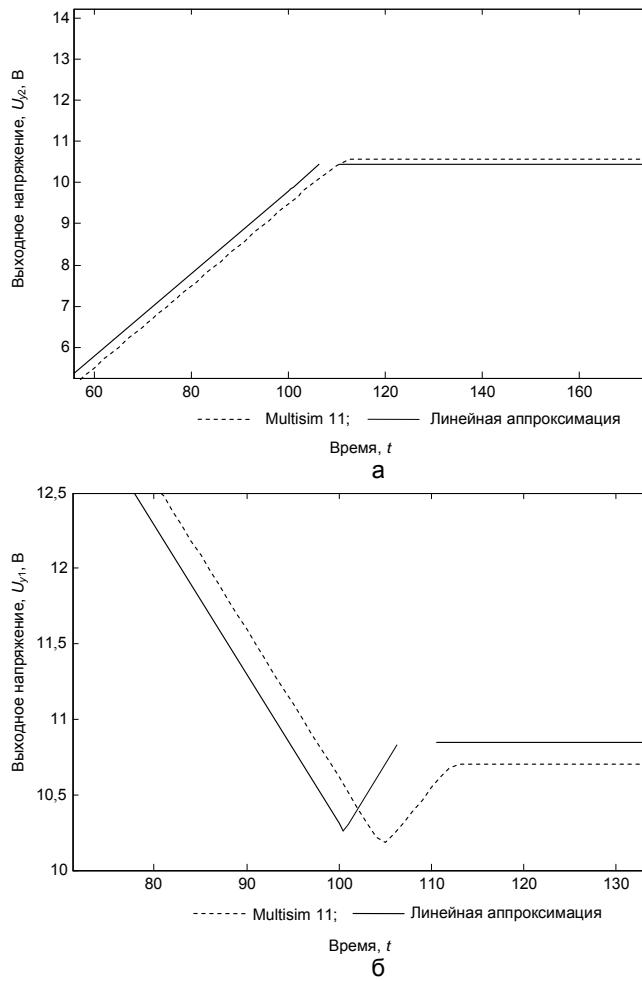


Рис. 4. Временная диаграмма переходных режимов: сигнал U_{y2} (а); сигнал U_{y1} (б)

Сопоставление временных диаграмм показывает практически полное соответствие результатов модели А и модели Б на линейных участках. В то же время на одном из переходных участков виден разрыв, вызванный неопределенностью режима модели А, что говорит о том, что вектор режима в зоне разрыва не определен, т.е. ни одно из условий существующих рабочих режимов этой модели не выполняется. Вызван этот эффект применением упрощенного способа аппроксимации нелинейной характеристики транзисторов, о чем говорилось выше. В то же время переход на более совершенные способы (описание с помощью системы дифференциальных уравнений) не будет соответствовать образовательному уровню ОУ СПО ввиду сложности такой ММ. В связи с этим принято следующее решение: методика выполнения соответствующей лабораторной работы основывается на представленной модели, а полученный эффект представляется студентам в демонстрационном режиме.

Заключение

В настоящее время разработаны следующие визуальные и расчетные модели:

1. Транзисторный каскад с общим эмиттером, соответствующий разделу ФЦС;
2. Двухкаскадный узел формирования двухтактных сигналов в схемотехнике ТТЛ, соответствующий разделу ФЦС;
3. Транзисторный каскад с общей базой, соответствующий разделу ФЦС;
4. Логический элемент «НЕ» (ТТЛ), соответствующий разделу ЛЭ 1;
5. Трехстабильный элемент «НЕ» (ТТЛ), соответствующий разделу ТСЭ.

В ближайшее время планируется проведение полноценного опробования визуальных и расчетных моделей в учебном процессе факультета среднего профессионального образования Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики в рамках дисциплины «Компьютерное моделирование аппаратной обработки данных» с введением в соответствующую рабочую учебную программу. Предполагается дальнейшее расширение комплексов моделей.

Литература

1. Новиков В.В. Виртуальный лабораторный комплекс по основам полупроводниковой цифровой электроники // Научно-технический вестник СПбГУ ИТМО. – 2008. – Вып. 52. – С. 129–133.
2. Новиков В.В. Виртуальный лабораторный комплекс по основам полупроводниковой цифровой электроники // Сборник тезисов «Научный потенциал-XXI». – М., 2009. – С. 117.
3. Гриншпун Д.М. Виртуальные лабораторные комплексы по основам цифровой электроники // Труды XVII Всероссийской научно-методической конференции «Телематика'2010». – СПб, 2010. – Т. 2. – С. 340–342.
4. Гриншпун Д.М. Виртуальные лабораторные комплексы по основам цифровой электронной обработки информации как средство повышения качества ИТ-образования // Сборник трудов IV Международной научной конференции «Современные достижения в науке и образовании». – Будва, 2010. – С. 12.
5. Новиков В.В., Гриншпун Д.М. Интерактивный визуально-демонстрационный виртуальный лабораторный комплекс по основам цифровой электроники // Труды XVII Всероссийской научно-методической конференции «Телематика'2010». – СПб, 2010. – Т. 2. – С. 342.
6. Гамбург К.С. Виртуальные стеновые лабораторные работы как инновационная форма контекстного обучения. Дис. ... канд. пед. наук. – М., 2006. – 186 с.
7. Лапшина И.В. Виртуальная информационно-образовательная лаборатория в профессиональной подготовке студентов. Дис. ... канд. пед. наук. – Ставрополь, 2002. – 188 с.
8. Медведева О.А. Развитие познавательной деятельности старшеклассников посредством виртуальной информационно-образовательной лаборатории. Дис. ... канд. пед. наук. – Карабаевск, 2006. – 186 с.
9. Ревинская О.Г. Методика проектирования и проведения компьютерных лабораторных работ для изучения теоретических моделей явлений и процессов в курсе физики технического вуза. Дис. ... канд. пед. наук. – Томск, 2006. – 229 с.
10. Саватеев Д.А. Компьютерное моделирование в изучении физических основ электромагнитных явлений в курсах общей физики и специальных дисциплин технического вуза. Дис. ... канд. пед. наук. – СПб, 2007. – 158 с.
11. Официальный сайт программной среды виртуального моделирования Edison 5 [Электронный ресурс]. – Режим доступа: <http://www.edisonlab.com>, свободный, Яз. англ. (дата обращения 08.11.2011).
12. Официальный сайт программной среды виртуального моделирования Yenka Technology [Электронный ресурс]. – Режим доступа: http://www.yenka.com/en/Yenka_Technology/, свободный, Яз. англ. (дата обращения 08.11.2011).
13. Официальный сайт программной среды виртуального моделирования Electronic Workbench Multisim [Электронный ресурс]. – Режим доступа: <http://www.ni.com/multisim/>, свободный, Яз. англ. (дата обращения 08.11.2011).
14. Официальный сайт разработчиков пакета «Открытая физика» [Электронный ресурс]. – Режим доступа: <http://www.physicon.ru>, свободный. Яз. рус. (дата обращения 08.11.2011).

Новиков Василий Викторович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, novikov.vz@gmail.com

Генович Владимир Витальевич

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, genovich.v.v@gmail.com

Гриншпун Дмитрий Михайлович

– Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, декан, dm.grinshpun@gmail.com

УДК 681.7.063

МЕТОД ПОНИЖЕНИЯ КОЭФФИЦИЕНТА ОТРАЖЕНИЯ ВОЛОКОННЫХ БРЭГГОВСКИХ РЕШЕТОК С ПОМОЩЬЮ ЭФФЕКТА ФОТОХРОМИЗМА

С.В. Варжель, А.В. Куликов, В.С. Брунов, В.А. Асеев

Представлены результаты апробирования метода понижения коэффициента отражения волоконных брэгговских решеток II типа с помощью эффекта фотохромизма, возникающего в результате их облучения сфокусированным излучением KrF эксимерного лазера.

Ключевые слова: волоконная брэгговская решетка, фотохромизм, эксимерный лазер.

Один из методов получения высокоэффективных отражательных брэгговских решеток в волноводах из германо-силикатного стекла – это облучение сквозь фазовую маску оптического волокна (OB) одиночным импульсом эксимерного лазера с плотностью энергии, близкой к порогу разрушения материала ($\sim 1 \text{ Дж}/\text{см}^2$). При такой плотности энергии в OB возникает решетка показателя преломления с глубиной модуляции, достигающей 6×10^{-3} , такая волноводная структура называется волоконной брэгговской решеткой (BBR) II типа и характеризуется коэффициентом отражения, близким к 100% [1]. Часто для практического применения требуются решетки также и с меньшей эффективностью. Для этого был предложен и апробирован метод понижения коэффициента отражения BBR II типа, в основе которого лежит эффект фотохромизма, возникающий в результате облучения записанных решеток сфокусированным излучением KrF эксимерного лазера.

Установка для облучения BBR II типа сфокусированным лазерным излучением отличается от схемы записи брэгговских решеток, представленной авторами в работе [2], тем, что из нее убрана фазовая маска. В результате облучения участка OB с BBR II типа сфокусированным излучением эксимерного лазера идет постепенное уменьшение ее коэффициента отражения до необходимого значения. Изменение частоты следования лазерных импульсов и плотности энергии излучения, сфокусированного на поверхность OB, позволяет регулировать скорость уменьшения эффективности решетки (чем меньше эта скорость, тем больше точность достижения требуемого коэффициента отражения). На рисунке продемонстрирован процесс уменьшения коэффициента отражения BBR II типа при ее облучении сфокусированным пучком эксимерного лазера.

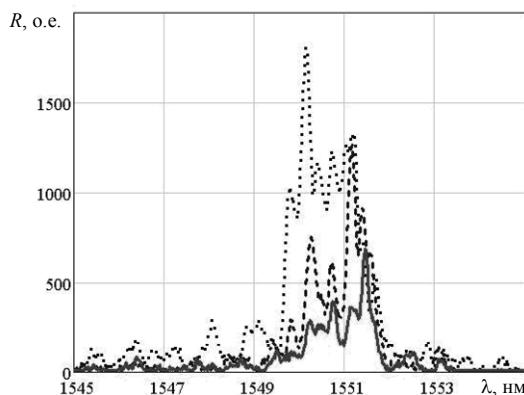


Рисунок. Спектры отражения решетки до (пунктирная линия) и после (сплошная линия) облучения сфокусированным излучением эксимерного лазера. Штриховой линией обозначен промежуточный спектр

В ходе эксперимента коэффициент отражения решетки был уменьшен от исходных 100% до 35–40%. Наличие двух пиков отражения на рисунке связано с тем, что в эксперименте используется двухлучепреломляющее OB [2].

В настоящей работе авторами предложен метод понижения коэффициента отражения BBR II типа, который позволяет достигать требуемого коэффициента отражения с точностью, зависящей от условий облучения (частота следования лазерных импульсов, плотность энергии излучения). Этот способ может быть положен в основу технологии создания чувствительных элементов фазовых интерферометрических датчиков различных физических величин на основе BBR.

1. Reekie L., Archambault J.-L., Russell P.St.J. 100% reflectivity fibre gratings produced by a single excimer laser pulse // OSA/OFC. – 1993. – Paper PD14. – P. 327–330.
2. Варжель С.В., Куликов А.В., Асеев В.А., Брунов В.С., Калько В.Г., Артеев В.А. Запись узкополосных волоконных брэгговских отражателей одиночным импульсом эксимерного лазера методом фазовой маски // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 5. – С. 27–30.

Варжель Сергей Владимирович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, vsv187@gmail.com

Куликов Андрей Владимирович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, a.kulikov86@gmail.com

Брунов Вячеслав Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, brunov@oi.ifmo.ru

Асеев Владимир Анатольевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, научный сотрудник, aseev@oi.ifmo.ru

УДК 681.5 + 531

ПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ ТЕНЗОРОВ ИНЕРЦИИ ТЕЛ НА СФЕРИЧЕСКИХ ДВИЖЕНИЯХ С МЕДЛЕННЫМ СОБСТВЕННЫМ ВРАЩЕНИЕМ

С.Н. Шаховал, Г.И. Мельников

Предлагается метод идентификации тензора инерции тела, использующий полупрограммное движение тестируемого тела с быстрым неравномерным реверсивно-симметричным вращением вокруг вертикальной оси прецессии и сравнительно медленным согласованным вращением вокруг наклонной подвижной оси собственного вращения.

Ключевые слова: тензор инерции, собственное вращение, реверсивно-симметричное движение, параметрическая идентификация.

Элементы тензора инерции определяются либо на вращениях вокруг шести неподвижных осей, либо на прецессии, дополненной одним осевым вращением. При этом применяется реверсивно-симметричный характер движения [1, 2]. В одном из вариантов шесть осевых вращений связаны переходными процессами угловой выставки тела относительно неподвижной оси, что характеризуется снижением точности и дополнительными затратами времени. В связи с этим в данной работе предлагается вариант, в котором осуществляется двухосная полупрограммная реверсивно-симметричная прецессия тестируемого тела, при условии, что угловая скорость собственного вращения тела на порядок меньше угловой скорости прецессионного движения вокруг вертикальной оси (рисунок). При таком условии весьма медленное собственное вращение обеспечивает на протяжении полного собственного оборота совпадение с вертикальной осью прецессии всех линий вокруг наклонной оси собственного вращения кругового конуса, связанного условно с телом (подвижного аксиода), что заменяет пять переходных процессов смены осей. Собственная угловая скорость тела вносит лишь малые изменения в величину и направление (угол γ на рисунке) вектора угловой скорости сферического движения тела. Вращения вокруг двух осей предполагаются синхронными реверсивно-симметричными, содержащими симметричное движение в обратном направлении.

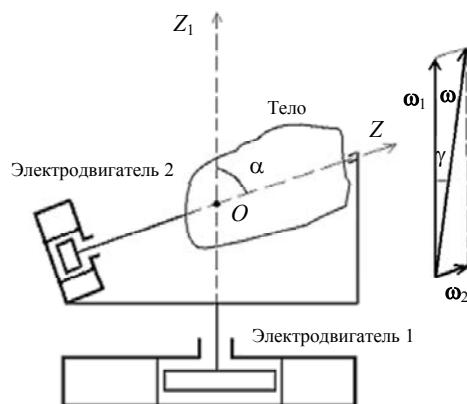


Рисунок. Устройство для определения тензора инерции тела

Пусть $\omega_1 = \dot{\psi} k_1$ – переменная угловая скорость прецессии тела вокруг вертикальной оси Oz_1 , меняющая направление при обратном движении; ϵ_1 – угловое ускорение прецессии; φ – угол поворота тела вокруг наклонной оси собственного вращения тела; $\omega_2 = \dot{\varphi} k$ – угловая скорость вращения вокруг наклонной оси. Предполагаем $\omega_1 >> \omega_2$. Тогда в первом приближении угловую скорость ω_2 вращения вокруг наклонной оси можно не учитывать. Нас интересует только изменение с течением времени углового положения тела относительно оси Oz_1 . Пусть в процессе испытания тело совершило два оборота замедленного вращения вокруг оси Oz в положительном направлении и два оборота симметричного разгона в обратном направлении. Считаем, что ввиду симметричности пары движений работы сил трения в обоих направлениях равны. Работы силы тяжести неуравновешенного тела на полном обороте по углу φ равны нулю ввиду вертикальности оси прецессии Oz_1 . Программные движения [3] осуществляются двумя электродвигателями (рисунок). Назначим на угловом интервале $0 \leq \varphi \leq 4\pi$ пять полных оборотов,

отсчитываемых от следующих пяти угловых значений: $\varphi_0 = 0$, $\varphi_1 = h$, ..., $\varphi_4 = 4h$; $h = \frac{2\pi}{5}$. На этих оборотах имеем пять реверсивно-симметричных вращений (в одну и обратную стороны). Эти движения удовлетворяют парам уравнений энергии (в которых пренебрегаем угловой скоростью собственного вращения):

$$(J + I_k)(\omega_{1,k+1}^2 - \omega_{1,k}^2) = 2A_k + 2V_k, \quad (J + I_k)(\omega_{1,k}^2 - \omega_{1,k+1}^2) = 2A'_k + 2V'_k. \quad (1)$$

В уравнениях (1) $\omega_{1,k}$ ($k = 0, \dots, 4$) – узловые значения угловой скорости прецессии; J – приведенный момент инерции устройства; I_k – моменты инерции тела относительно оси Oz_1 в угловой позиции k (относительно мгновенной оси, проведенной в теле при узловом значении $\varphi = \varphi_k$); A_k , A'_k – полезные работы электродвигателей на движении в одном и обратном направлениях; V_k и $V'_k = V_k$ – отрицательные работы диссипативных сил трения и аэродинамического сопротивления. Из (1) получаем расчетную формулу для пяти осевых моментов инерции тела:

$$I_k = \frac{A'_k - A_k}{\omega_{1,k}^2 - \omega_{1,k+1}^2} - J, \quad k = 0, \dots, 4. \quad (2)$$

Эти оси расположены в теле на круговом конусе, описанном вокруг собственной оси Oz с углом $\alpha \approx 63,4^\circ$. Формулы (2) являются приближенными, так как в них не учитывается угловая скорость собственного вращения ω_2 . Точная формула, очевидно, имеет следующий вид:

$$I_k = \frac{A'_k - A_k}{\omega_k^2 - \omega_{k+1}^2} - J,$$

где $\omega_k = \omega_{1,k} + \omega_{2,k}$; $\omega_k^2 = \omega_{1,k}^2 + \omega_{2,k}^2 + 2\omega_{1,k}\omega_{2,k} \cos \alpha_k$; угол α_k – угол между $\omega_{1,k}$ и $\omega_{2,k}$. При этом осевые моменты инерции тела I_k определяются относительно пяти мгновенных осей, направленных по векторам ω_k , на подвижном аксониде. Дополнительно определяют в эксперименте момент инерции относительно собственной оси вращения Oz . Затем по шести моментам инерции находят матрицу тензора инерции в заданной точке O тела. Малые погрешности формулы (2) оцениваются с использованием способа [4].

Работа выполнена при финансовой поддержке гранта РФФИ проект № 10-08-01046.

1. Мельников В.Г. Энергетический метод параметрической идентификации тензоров инерции тел // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 1 (65). – С. 59–63.
2. Мельников В.Г., Едачев А.С., Мельников Г.И., Шаховал С.Н. Метод определения тензора инерции на программных движениях // Изв. Самарского научного центра РАН. – 2010. – Т. 12 (33). – № 1 (2). – С. 445–448.
3. Мельников В.Г. Использование программных движений для идентификации тензора инерции и центра масс твердого тела // Изв. вузов. Приборостроение. – 2007. – Т. 50. – № 8. – С. 33–36.
4. Шаховал С.Н. Исследование матричных алгебраических уравнений, определяющих тензор инерции через осевые моменты инерции // Научно-технический вестник СПбГУ ИТМО. – 2008. – Вып. 47. – С. 196–201.

Шаховал Сергей Николаевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, shakhovalsergey@gmail.com

Мельников Геннадий Иванович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, melnikov@mail.ifmo.ru

УДК 531

**ИДЕНТИФИКАЦИЯ ТЕНЗОРА ИНЕРЦИИ ТЕЛА НА РЕВЕРСИВНО-СИММЕТРИЧНЫХ
ПРЕЦЕССИЯХ В ОГРАНИЧЕННОМ УГЛОВОМ ИНТЕРВАЛЕ**
В.Г. Мельников, Р.Ю. Кравчук, Г.И. Мельников, С.Н. Шаховал

Решается задача определения тензора инерции на основании прецессионного двухосного реверсивно-симметричного сферического движения тела при ограничениях на два угловых интервала. Получены расчетные формулы для инерционных параметров.

Ключевые слова: матрица тензора инерции, угол прецессии, угол собственного вращения, реверсивно-симметричная прецессия.

Известен способ параметрической идентификации пяти осевых моментов инерции тела на реверсивно-симметричном прецессионном двухосном движении тела вокруг неподвижной точки. Обобщим этот способ на случай отсутствия полных оборотов по углу собственного вращения. Пусть твердое тело вместе со сцепленной с ним системой $Oxyz$, собственной осью Oz с известным центром масс C , собственным углом поворота φ совершает сферическое движение вокруг неподвижной точки O . При этом за рассматриваемый конечный интервал времени мгновенная ось вращения тела описывает вокруг собственной оси Oz тела

часть кругового конуса в некотором секторе $[\varphi_0 = 0, \varphi_5]$ при $\varphi_6 = 5\delta \leq \pi$. Равномерной сетке узловых значений сопоставим шесть положений в теле мгновенной оси вращения с ортами, заданными строчными матрицами

$$\mathbf{e}_i = [e_{i1}, e_{i2}, e_{i3}] = [\cos \varphi_i, \sin \varphi_i, h]r, \quad r = \sin \beta, \quad h = ctg \beta, \quad i = 0, 1, \dots, 5.$$

Пусть на реверсивно-симметричной прецессии [1–4] определены пять моментов инерции тела I_1, \dots, I_5 . Имеем выражения осевых моментов инерции через компоненты тензора инерции:

$$I_i = J_x e_{i1}^2 + J_y e_{i2}^2 + J_z e_{i3}^2 + 2J_{xy} e_{i1} e_{i2} + 2J_{yz} e_{i2} e_{i3} + J_{xz} e_{i1} e_{i3}, \quad \text{или}$$

$$I_i = r^2 \left(\frac{1}{2} J_x (1 + \cos 2\varphi_i) + \frac{1}{2} J_y (1 - \cos 2\varphi_i) + J_z h^2 + 2J_{xy} \sin 2\varphi_i + 2h (J_{yz} \sin \varphi_i + J_{xz} \cos \varphi_i) \right).$$

Приводя подобные члены, получаем окончательные выражения для пяти осевых моментов,

$$I_i = r^2 (X^2 + Y \cos 2\varphi_i + Z \sin 2\varphi_i + F \cos \varphi_i + H \sin \varphi_i), \quad i = \overline{1, 5}, \quad (1)$$

с пятью инерционными коэффициентами, составленными из шести компонент тензора инерции:

$$X = (J_x + J_y) / 2 + J_z h^2, \quad Y = (J_x - J_y) / 2, \quad Z = J_{xy}, \quad F = 2J_{xy} h, \quad H = 2J_{yz} h.$$

Выражения (1) объединяются в матричное строчное выражение

$$\mathbf{VA} = \mathbf{I} \text{ при } \mathbf{I} = [I_1, \dots, I_5] r^2, \quad \mathbf{V} = [X \ Y \ Z \ F \ H], \quad (2)$$

$$\text{где } A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ \cos(2\varphi_1) & \cos(2\varphi_2) & \cos(2\varphi_3) & \cos(2\varphi_4) & \cos(2\varphi_5) \\ \sin(2\varphi_1) & \sin(2\varphi_2) & \sin(2\varphi_3) & \sin(2\varphi_4) & \sin(2\varphi_5) \\ \cos(\varphi_1) & \cos(\varphi_2) & \cos(\varphi_3) & \cos(\varphi_4) & \cos(\varphi_5) \\ \sin(\varphi_1) & \sin(\varphi_2) & \sin(\varphi_3) & \sin(\varphi_4) & \sin(\varphi_5) \end{pmatrix}.$$

Из уравнения (2) находим вектор-строку пяти неизвестных $\mathbf{V} = [X \ Y \ Z \ F \ H]$, а также центробежные моменты инерции тела и момент инерции относительно оси, соответствующий углу $\varphi_0 = 0$:

$$\mathbf{V} = \mathbf{IA}^{-1}, \quad J_{xy} = Z, \quad J_{zx} = \frac{1}{2h} F, \quad J_{yz} = \frac{1}{2h} H, \quad I_0 = (X + Y + Z)r^2.$$

В случае $\beta = 63,4^\circ$ на круговом конусе равномерно распределены пять осей виртуального икосаэдра, сцепленного с телом, и на реверсивно-симметричной прецессии экспериментально находятся пять моментов инерции, распределенных в секторе конуса [1–4]. Тогда моменты инерции относительно пяти осей икосаэдра вычисляются по формуле, аналогичной (1), в которой следует заменить углы $\varphi_i = i\delta$ на $\varphi'_i = i2\pi/5$. При этом вектор-строка \mathbf{V} вычисляется по прежней формуле $\mathbf{V} = \mathbf{IA}^{-1}$, но с новыми значениями строчной матрицы \mathbf{I} . Момент инерции тела относительно шестой оси икосаэдра определяется отдельно на осевом реверсивно-симметричном вращении энергетическим методом, как указано в цитированных работах. Расчетные формулы компонент тензора инерции через найденные шесть моментов инерции приведены в [1–2].

Работа выполнена при финансовой поддержке РФФИ (грант № 10-08-01046).

- Мельников В.Г., Едачев А.С., Мельников Г.И., Шаховал С.Н. Метод определения тензора инерции на программных движениях // Изв. Самарского научного центра РАН. – 2010. – Т. 12 (33). – № 1 (2). – С. 445–448.
- Мельников В.Г. Энергетический метод параметрической идентификации тензоров инерции тел // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 1 (65). – С. 59–63.
- Мельников В.Г. Идентификация компонент тензора инерции и координат центра масс тела на реверсивно-симметричных прецессиях // Вестник СПбГУ. Сер.1. Математика, механика и астрономия. – 2010. – Вып. 3. – С. 97–104.
- Патент 2436055. Способ определения тензора инерции тела и устройство для его осуществления / опубл. 10.12.2011. Бюл. № 34. – 17 с.

Мельников Валентин Геннадьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, зав. кафедрой, melnikov@mail.ifmo.ru

Кравчук Раиса Юрьевна – ОАО «Ростелеком», инженер, ggarotta@gmail.com

Мельников Геннадий Иванович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, melnikov@ifmo.ru

Шаховал Сергей Николаевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, shakhovalsergey@gmail.com

1. OPTICAL AND OPTICAL ELECTRONIC SYSTEMS. OPTICAL TECHNOLOGIES STRUCTURE FORMATION PRINCIPLES OF INTERFERENCE BEAM SPLITTERS

L. Gubanova, E. Putilin

The methodology of interference beam splitters construction, formed by symmetric cells of dielectric layers is considered. The methodology of short-wave and long-wave interference beam splitters formation is given. The impact analysis of symmetric cells number and their structure on output parameters is considered.

Keywords: interference coatings, beam splitters, symmetric cells, dielectric layers, dichroic elements.

Lyudmila Gubanova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Associate professor, La7777@mail.ru

Edvard Putilin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, eputilin @ yandex.ru

CHOICE OF MULTIWAVE AEROSOL LIDAR PARAMETERS FOR REMOTE ATMOSPHERE SOUNDING N. Volkov

The paper deals with description of lidar configuration for simultaneous measurements of aerosol extinction and backscattering coefficients at several wavelengths in 355 and 1064 nm spectral range. Multiwave lidar parameters for aerosol monitoring at heights from 1 up to 10 km are estimated.

Keywords: aerosol, lidar, Raman scattering.

Nikolay Volkov – Moscow State University of Geodesy and Cartography (MIIGAiK), postgraduate, nn-volkov@yandex.ru

INVARIANT MODEL FOR ESTIMATION OF THE ATMOSPHERE TRANSMITTING EFFICIENCY AT OBJECTS MONITORING IN THE OPTICAL SPECTRAL RANGE

A. Demin, M. Moiseeva

Estimation problem of the atmosphere influence on radiation in the middle and distant infra-red range for the increasing of results reliability at aircraft monitoring is considered. The analytical *a posteriori* model of atmosphere transmittance is proposed. The general structure of such model is shown. Identification results of the atmosphere transmitting efficiency value on the basis of experimental data are shown.

Keywords: analytical model, identification, the atmosphere transmitting efficiency, objects monitoring, experimental data.

Anatoly Demin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, dav_60@mail.ru

Maria Moiseeva – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, mim14@mail.ru

2. PHOTONICS AND OPTICAL INFORMATICS

SPECTRUM RESEARCH OF DNA MOLECULE IN THE TERAHERTZ FREQUENCY DOMAIN

M. Tsurkan, E. Sobakinskaya, O. Smolyanskaya, V. Bespalov, V. Vaks, N. Balbekin

Spectrum research of degraded DNA of herring in the range 0.1-1 THz was done. DNA native absorption lines were obtained. Lines, responsible for interaction with water and conformational changes, were identified.

Keywords: terahertz radiation, spectroscopy, DNA spectrum, absorption lines.

Maria Tsurkan – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, junior researcher, tsurkan.maria@yandex.ru

Ekaterina Sobakinskaya – Institute for Physics of Microstructures RAS (IPM RAS), engineer, katja@ipm.sci-nnov.ru

Olga Smolyanskaya – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, o_smolyanskaya@mail.ru

Victor Bespalov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Deputy dean for science, Deputy department head, Professor, D.Sc., Professor, victorbespaloff@gmail.com

Vladimir Vaks – Institute for Physics of Microstructures RAS (IPM RAS), Head of department, PhD, senior researcher, vax@ipm.sci-nnov.ru

Nikolay Balbekin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, Godnikolay@yandex.ru

COMBINED METHOD OF RED BLOOD CELLS COUNTING ON BLOOD SMEAR IMAGES**A. Dyrnaev, A. Potapov**

The problem of red blood cells counting on images obtained by microscope often appears within computer-based analysis of blood products. Segmentation method of those images is presented in the paper, and it is based on Canny edge detector with following circle identifying with Hough transform. Segmentation results refinement with binary mask of objects and background is used in the article. The proposed method is resistant to structure, texture and blood cells color changing, and also it allows detecting cells that overlap each other even on noisy images.

Keywords: blood cells counting, digital microscopy, image segmentation.

Alexey Dyrnaev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, dirnalex@yandex.ru

Alexey Potapov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., pas.aicv@gmail.com

QUANTUM CRYPTOGRAPHY SYSTEM WITH A SINGLE PHOTON SOURCE BASED ON THE SPONTANEOUS PARAMETRIC SCATTERING EFFECT**V. Egorov, I. Latypov, A. Rupasov, A. Gleim, S. Chivilikhin**

A scheme of a single photon source for quantum informatics applications based on the spontaneous parametric scattering effect is proposed and a quantum cryptography setup using it is presented. The system is compared to the alternative ones that operate with attenuated classic light.

Keywords: quantum cryptography, spontaneous parametric scattering, biphotons, sidebands.

Vladimir Egorov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, egorovvl@gmail.com

Ilnur Latypov – Zavoisky Physical-Technical Institute (ZPhTI) of the Kazan Scientific Center of the Russian Academy of Sciences, junior scientific researcher, ilnur.latypov@gmail.com

Andrey Rupasov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, sadbender@yandex.ru

Artur Gleim – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, aglejm@yandex.ru

Sergey Chivilikhin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, senior scientific researcher, sergey.chivilikhin@gmail.com

3. ANALYSIS AND SYNTHESIS OF COMPLEX SYSTEMS**IDENTIFICATION OF THE LINEARLY VARYING FREQUENCY FOR THE SINUSOIDAL SIGNAL****S. Aranovskiy, A. Bobtsov, A. Pyrkin**

Identification problem of the linearly varying frequency for the sinusoidal signal is considered. Practical importance of this problem is shown. Dynamic model of the signal is proposed and identification algorithm is designed. Illustrative numerical example is also presented.

Keywords: identification, linearly varying frequency, sinusoidal signal.

Stanislav Aranovskiy – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, assistant, senior scientific researcher, PhD, s.aranovskiy@gmail.com

Alexey Bobtsov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Dean, Department head, D.Sc., Professor, bobtsov@mail.ifmo.ru

Anton Pyrkin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, assistant, junior scientific researcher, PhD, a.pyrkin@gmail.com

IMPROVED TIME-FREQUENCY ANALYSIS ALGORITHM OF BARELY VISIBLE RADAR STATIONS SIGNALS, BASED ON WINDOWED FOURIER TRANSFORM**A. Korotkov, A. Mukhin**

The problem of time-frequency analysis for barely visible radar stations signals, based on windowed Fourier transform was considered. Improved algorithm, based on integration of Fourier transform data was suggested. Algorithm application gives us the possibility to take into account characteristics of such signals, especially with low signal/noise ratio.

Keywords: time-frequency analysis, integration, barely visible radar stations.

Andrey Korotkov – Research Institute “Vector”, JSC, engineer, Andrey.k0206516@mail.ru

Anatoliy Mukhin – Mozhayskiy Military Space Academy, postgraduate, tolik516@yandex.ru

APPLICATION OF QUALITATIVE EXPONENTIAL INSTABILITY CONDITIONS FOR DYNAMIC PROCESSES ESTIMATION**E. Rabysh, V. Grigoriev, S. Bystrov, A. Sporygin**

Dynamic quality estimations of transient processes were found on the basis of Lyapunov direct method and conditions for qualitative exponential instability. As a result, efficient procedures for analytical analysis of unstable continuous and discrete dynamical systems can be created.

Keywords: qualitative exponential instability, quality estimations, continuous and discrete systems, behavior analysis of unstable systems, loss of control.

Eugene Rabysh – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, Rabysh@yandex.ru

Valeriy Grigoriev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, grigvv@yandex.ru

Sergey Bystrov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, sbystrov@mail.ru

Anatoly Sporyagin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Postgraduate, Avsporyagin@yandex.ru

SERVO ACTUATOR CONTROL SYSTEM WITH NON-RIGID EXECUTIVE AXLE

A. Abdullin, V. Tolmachev

The article presents mathematical simulation results of electric drive tracking processes installed on azimuth axle of large telescope TI-3,12 in Altai Optical-laser Center. We propose two structural solutions of control system. Efficiency of second mass additional feedback usage is shown. Feedback usage gave the possibility to extend smooth velocity tracking range and reduce tracking root-mean-square errors.

Keywords: servo actuator, executive axle, two-mass mechanism, position control, speed subsystem.

Artur Abdullin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, artur.abdullin@gmail.com

Valery Tolmachev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, tolmachev@ets.ifmo.ru

ANALYSIS OF ENERGY PERFORMANCE OF MULTILEVEL SEMICONDUCTOR CONVERTERS OF ELECTRIC DRIVE SYSTEMS

K. Mikheev, V. Tomasov

The energy performance of power subsystems of AC drives, built on the base of different topologies of multilevel voltage-fed converters was analyzed. A special calculation subsystem “*Measurements*” was developed in the *Matlab* package using *Simulink* and *SimPowerSystems* libraries, which allows analyzing and comparing the energy performance of multilevel converters.

Keywords: power subsystem, energy performance, multilevel converters, electrical drive, sinusoidal pulse-width modulation, PWM, simulation.

Kirill Mikheev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, mikhirill@yandex.ru

Valentin Tomasov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, PhD, Associate professor, tomasov@ets.ifmo.ru

4. COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES SIMULATION ENVIRONMENT OF REAL-TIME AGENT-BASED SYSTEMS

M. Panteleev, N. Kokhtenko, S. Lebedev

The problem of program platform developing for agent system simulation is considered. The platform is intended for experimental research and testing of intelligent agent's behavior models under real-time constraints and command opposition conditions. The requirements to such simulation platform are defined. The architecture of the simulation platform and its main functions are proposed.

Keywords: intelligent agents, multiagent systems, simulation engine, real time.

Michail Panteleev – Saint Petersburg State Electrotechnical University (LETI), Associate professor, PhD, Associate professor, mpanteleyev@gmail.com

Nickolay Kokhtenko – Saint Petersburg Electrotechnical University (LETI), postgraduate, nvkokhtenko@gmail.com

Sergey Lebedev – Saint Petersburg Electrotechnical University (LETI), postgraduate, lebedev.sv.etu@gmail.com

EFFECTIVE INFERENCE IN PRODUCTION RULES MODEL WITH COLUMN-ORIENTED DATABASES

R. Katerinenko

The idea of column-oriented databases for inference in production rules model of knowledge is presented. We will show benefits of column-oriented databases application in comparison with traditional row-oriented databases for logic inference purposes in production rules. The results of comparative experiment on KDB+ and Oracle databases are also given.

Keywords: production rules model, inference, column-oriented database, vector programming languages.

Roman Katerinenko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, innuWorks@gmail.com

ONTOLOGY EXTRACTION FROM WIKI-SYSTEMS

V. Shestakov

The paper is devoted to the approach of ontology extraction from Wiki-systems, and its application both in development and support of ontology content-based Wiki-systems and other possible uses. An approach implementation as a client application is described in detail.

Keywords: Wiki, ontologies, data extraction, Pywikipediabot, Semantic MediaWiki.

Vladimir Shestakov – Novosibirsk State University, postgraduate, zfc@ngs.ru

KEY PHRASES EXTRACTION

S. Popova, I. Khodyrev

An approach to topic detection and annotation with keywords and key phrases for narrow-domain short text collections is studied. Algorithms which allow extracting keywords and key phrases based on the results of clustering are described. A formula for words interdependence estimation is suggested, and key phrases extraction algorithm on its base is shown. Experimental results are given.

Keywords: keywords, key phrases, annotation, clustering, narrow-domain short text analysis, information retrieval.

Svetlana Popova – Saint Petersburg State University, postgraduate, spbu@bk.ru

Ivan Khodyrev – OLIMP, programmer, kivan.mih@gmail.com

IMPLEMENTATION ASPECTS OF SPEECH RECOGNITION APPLICATIONS IN CUSTOMER-SERVICE CALL CENTERS

P. Balakshin, G. Petrov

The article provides a summary of current issues in call centers of large companies. Advantages and disadvantages of speech recognition use in call center work are presented. FCR (First Call Resolution) metric and its importance are defined. The features of FCR calculation in the applications with automated speech recognition are discussed.

Keywords: speech recognition, phone customer-service, call center, FCR.

Pavel Balakshin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, assistant, pbalakshin@gmail.com

Grigory Petrov – Enkata Technologies, Inc., Director of implementation department, PhD, gpetrov@enkata.com

5. PROGRAMMING TECHNIQUES

FITNESS FUNCTION CHOOSING FOR GENETIC ALGORITHMS WITH REINFORCEMENT LEARNING

A. Afanasyeva, M. Buzdalov

This paper describes an optimization problem with one target function to be optimized and several supporting functions that can be used to speed up the optimization process. A method based on reinforcement learning is proposed for choosing a good supporting function during optimization using genetic algorithm. Application results of this method to a model problem solution are shown.

Keywords: genetic algorithms, machine learning, reinforcement learning.

Arina Afanasyeva – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, afanasyevrina@gmail.com

Maxim Buzdalov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, mbuzdalov@gmail.com

AUTOMATIC SOLUTION MODIFICATIONS FOR TESTING SOLUTION CHECKERS

A. Akhi, A. Stankevich, A. Shalyto

The article deals with the method for testing solution checkers used in programming contests by making automatic modifications to solution source code. Successful application of the method to regional ACM ICPC NEERC 2011 and Russian Code Cup 2011 contests is described. Real bugs in solution checkers which were to be used for testing participants' solutions but were found by the method and fixed before actual events are shown.

Keywords: programming contests, contest programming problems, solution checkers, testing.

Anton Akhi – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, anton.akhi@gmail.com

Andrey Stankevich – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, stankev@rain.ifmo.ru

Anatoly Shalyto – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department Head, D.Sc., Professor, shalyto@mail.ifmo.ru

TREES APPLICATION FOR MULTI-DIMENSIONAL ARRAY MASS UPDATES

A. Bannykh

The method of mass operations defining for a subclass of multi-dimensional data structures for a narrow problem class is suggested. The method is applicable in case if multi-dimensional structure elements belong to an abelian group and allows effectively computing the sum of all values in a region and effectively adding a value to any region in multi-dimensional structure.

Keywords: data structures, mass operations, multi-dimensional arrays, segment tree, Fenwick tree.

Anton Bannykh – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, anton.bannykh@gmail.com

GENETIC ALGORITHM APPROACH TO OBSERVER'S OPTIMAL TRAJECTORY DESIGN

D. Stepanov, A. Shalyto

Modified genetic algorithm was offered as a tool for observer's optimal trajectory generation in BOTMA (Bearing-Only Target Motion Analysis). Stochastic ascent and annealing simulation concepts were used for the purpose of arising genetic algorithms convergence and local extremum problems solution. For speeding-up the algorithm, we perform selection, combining schemata with higher fitness.

Keywords: genetic algorithms, stochastic ascent method, annealing simulation method, Holland's Schema Theorem, Goldberg building blocks hypothesis, Hamming distance, BOTMA.

Denis Stepanov – JSC «Concern «Aurora», junior research fellow, denis_v_stepanov@hotmail.com

Anatoly Shalyto – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department Head, D.Sc., Professor, shalyto@mail.ifmo.ru

SCENARIO-BASED FINITE STATE MACHINE INDUCTION USING SAT-SOLVERS

V. Ulyantsev, F. Tsarev

The article deals with the method of finite state machine induction by test scenarios using SAT-solvers. Successful application of the method to the induction of alarm clock finite state machine is shown. Induction of appropriate finite state machine is consistent and algorithm operation time is less than one second of CPU time on average personal computer.

Keywords: finite state machines, machine learning, SAT.

Vladimir Ulyantsev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, ulyantsev@rain.ifmo.ru

Fedor Tsarev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, fedor.tsarev@gmail.com

GENETIC PROGRAMMING APPLICATION FOR CELL AUTOMATA GENERATION

A. Tikhomirov, A. Shalyto

Improved algorithm for arbitrary cellular automata generation based on test samples using genetic algorithms is discussed. Issues arising from applying a classic genetic algorithm to the problem are described. Algorithm is tested on a few training examples.

Keywords: cell automata, genetic programming.

Andrey Tikhomirov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, and.tikhomirov@gmail.com

Anatoly Shalyto – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department Head, D.Sc., Professor, shalyto@mail.ifmo.ru

6. MECHANICS AND MECHATRONICS

AIRFOIL FORMATION SETTING FOR GAS BEARINGS

S. Rodinkov

We consider gyroscopic devices with contactless suspension of sensitive element based on aerodynamic bearings with a special aerodynamic profile patterned by machine processing. The current setting for the profile formation has some serious flaws that undermine the bearings of the profile. A new design instrument for the profile formation intended to their elimination was made. Research was conducted to introduce additional changes in the design, select lap material. A set of bearings with an aerodynamic profile was produced meeting all the requirements for the geometric sensor of gyroscopic devices.

Keywords: aerodynamic bearing, airfoil formation setting, machine processing, abrasive wear.

Sergei Rodinkov – FSUE “Scientific Research Institute of Command Devices”, Head of sector, poguHkoB@mail.ru

7. MATERIALS TECHNOLOGY AND NANOTECHNOLOGIES MICROFLUID CHIPS FOR BIOLOGICAL OBJECTS RESEARCH BY METHODS OF HIGH-RESOLUTION MICROSCOPY

I. Kukhtevich, A. Bukatin, I. Mukhin, A. Evstrapov

Microfluid devices coupled with methods of high-resolution microscopy have allowed creating new analytical systems for studies of biological objects in natural state. The main requirements for microscopy methods for the

nondestructive study of biological objects were formulated. Alternatives for fixation of biological objects were reviewed and discussed. Experimental models of microfluid chips for high-resolution microscopy techniques were shown.

Keywords: microfluid chip, biological sample, micro- and nanostructures, confocal laser scanning microscopy, near-field microscopy, atomic-force microscopy, lithography.

Igor Kukhtevich – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, ba@inbox.ru

Anton Bukatin – Saint Petersburg Academic University – Nanotechnology Research and Education Center RAS, postgraduate, antbuk.fiztek@gmail.com

Ivan Mukhin – Saint Petersburg Academic University – Nanotechnology Research and Education Center RAS, postgraduate, imukhin@yandex.ru

Anatoly Evstrapov – Establishment of the Russian Academy of Sciences Institute for Analytical Instrumentation RAS, chief of laboratory, PhD, Associate professor, an_evs@mail.ru

8. INFORMATION SECURITY METHODS AND SYSTEMS

STEGANOALGORITHM CREATION BASED ON FORMAT AND SPATIAL PRINCIPLES OF THE DATA HIDING

A. Korobeynikov, S. Kuvshinov, S. Blinov, A. Leyman, S. Nesterov

Steganoalgorithm creation problem on the base of format and spatial methods of the data hiding is considered. Algorithms of information embedding and extraction from stegocontainer are developed. The general scheme of system operation of the latent transfer is offered.

Keywords: steganography, stegocontainer, discrete cosine transformation, wavelet transformation.

Anatoly Korobeynikov – Saint Petersburg Branch Foundation of the Russian Academy of Sciences “Institute for Earth Magnetism, Ionosphere and Radiowaves n.a. N.V.Puchkov RAS”, Deputy director for science, D.Sc., Professor, Korobeynikov_A_G@mail.ru

Stanislav Kuvshinov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, ss.kuvshinov@gmail.com

Stanislav Blinov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, stasblino@yandex.ru

Albert Leyman – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, allxxl@yandex.ru

Sergey Nesterov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, nesterov.serge@gmail.com

ESTIMATION METHOD OF UNAUTHORIZED INTRUSION PROBABILITY DETECTION BY PASSIVE INFRARED DETECTOR

V. Volkhonskiy, P. Vorobyov

Estimation method of probability detection of intrusion detectors was proposed. Density of probability distribution of detection distance was analyzed for passive infrared detector. Experimental data were obtained for probability detection estimation for various parameters of target movement.

Keywords: probability of detection, passive infrared detector, protected object.

Vladimir Volkhonskiy – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, PhD, Associate professor, volkhonski@mail.ru

Pavel Vorobyov – State Scientific Research Institute of Applied Problems, software engineer, doall88@gmail.com

9. COMPUTER-AIDED DESIGN SYSTEMS

AUTOMATED DESIGN AND RESEARCH SYSTEM OF WOODEN STRUCTURES WITH METAL GEAR PLATES

M. Pavlenko, A. Shmidt

The problems of computer-aided design of wooden trusses with metal gear plates are considered. A new approach to design systems building of trusses and nodal solutions, allowing calculations to the most commonly used designs, is suggested. Components of the online help and user activity monitoring are implemented: scrolling text messages and a window of dynamic hints. Ready product has been tested. The system is intended for designers of structures, and can also be used in the educational process.

Keywords: wooden truss, metal gear plates, nodal solutions.

Mariya Pavlenko – Saint Petersburg State Architectural Building University, postgraduate, mariya.pavlenko@gmail.com

Alexander Shmidt – Saint Petersburg State Architectural Building University, Head of Department, PhD, Associate professor, ukf@bk.ru

DESIGN DECISION SUPPORT SYSTEM FOR SCIENTIFIC AND TECHNICAL INFORMATION DOCUMENTATION

N. Agarkova, G. Artemova, N. Gusarova

The article deals with design decisions optimization for workflow computer-aided design systems, in particular, workflow automation of scientific and technical information documentation.

Keywords: CAD, scientific and technical information, works distribution, automation, documentation, DSS.

Natalya Agarkova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, assistant, graziokisa@gmail.com

Galina Artemova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, assistant, glaya@inbox.ru

Natalya Gusarova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, natfed@list.ru

10. NEW TECHNOLOGIES IN EDUCATION METHODOLOGY OF GRADUATE COMPETENCES EVALUATION

A. Pierskaya

The article is devoted to the methodology of graduate competences evaluation at the university automated point-ranking system. The developed model establishes connections between graduate competences and their component competences at disciplines. Levels of formed competences are established on the base of common rater of education results (competences).

Keywords: graduate competence, level of formed competence, graduate competence model, result evaluation of competence mastering, student's rate.

Anna Pierskaya – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, peerskaya@mail.ru

ELECTRONIC DEVICE DESIGN AND MANUFACTURING IN A UNIFIED INFORMATION ENVIRONMENT FOR THE STUDENTS' EDUCATIONAL PROCESS

A. Vorobyev, Y. Fedosov

An educational example of technological workflow building in unified information environment of printed circuit board creation in a surface mounting laboratory is proposed in the article. Theoretical basement and implementation of production and preproduction flow are considered. Special attention is given to operation with automated equipment.

Keywords: surface mounting technology, education, preproduction.

Anatoly Vorobyev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, anatoliy.vorobyev@gmail.com

Yury Fedosov – JSC “MNS”, postgraduate, yf01@yandex.ru

SETS OF THE VISUAL AND COMPUTATIONAL LABORATORY MODELS ON THE BASIS OF DIGITAL ELECTRONIC INFORMATION PROCESSING

V. Novikov, V. Genovich, D. Grinshpun

The article analyzed the issues of sets development for computer models of electronic devices that are part of a virtual laboratory complex on the basis of electronic data processing. The concepts of visual and computational models are introduced, sets of visual and computational models, the stages of their development and questions of models mathematical descriptions - the definition of used approximation methods and methods for mathematical descriptions. Requirements for visual and computational models from the technical and pedagogical sides are analyzed; the result of the sets development is given.

Keywords: virtual laboratory complex, the basics of electronic data processing, visual models, computational models.

Vasily Novikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, novikov.vz@gmail.com

Vladimir Genovich – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, genovich.v.v@gmail.com

Dmitry Grinshpun – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Dean, dgr1@yandex.ru

11. BRIEF REPORTS

REFLECTION COEFFICIENT DECREASING METHOD OF FIBER BRAGG GRATINGS BY THE EFFECT OF PHOTOCHROMISM

S. Varzhel, A. Kulikov, V. Brunov, V. Aseev

The article presents testing results of reflection coefficient decreasing method of II type fiber Bragg gratings by the effect of photochromism arising while irradiation of the fiber with a focused radiation of KrF excimer laser.

Keywords: fiber Bragg grating, photochromism, excimer laser.

Sergey Varzhel – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, vsv187@gmail.com

Andrey Kulikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, a.kulikov86@gmail.com

Vyacheslav Brunov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, brunov@oi.ifmo.ru

Vladimir Aseev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, scientific researcher, aseev@oi.ifmo.ru

PARAMETRIC IDENTIFICATION OF BODIES' INERTIA TENSORS ON SPHERICAL MOTIONS WITH SLOW SPIN

S. Shakhoval, G. Melnikov

Identification method of body's inertia tensor based on semiprogram movement of the tested body with fast reverse-symmetric rotation around vertical axis of precession and relatively slow coordinated rotation around inclined moving spin axis is offered.

Keywords: inertia tensor, spin, reverse-symmetric motion, parametric identification.

Sergey Shakhoval – ZAO «Ericsson Corporation AO», data base administrator, shakhovalsergey@gmail.com

Gennady Melnikov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, melnikov@ifmo.ru

INERTIA TENSOR IDENTIFICATION ON REVERSE-SYMMETRIC MOTIONS IN A FINITE ANGULAR RANGE

V. Melnikov, R. Kravchuk, G. Melnikov, S. Shakhoval

The problem of inertia tensor determination on a precessional twin-shaft reverse-symmetric motion of objects with angular constrains is solved. Some design formulas for inertia parameters are obtained.

Keywords: inertia tensor matrix, precession angle, eigenrotation angle, reverse-symmetric precession.

Vitaly Melnikov – Saint Petersburg National Research University Of Information Technologies, Mechanics And Optics, Department Head, PhD, Associate Professor, Melnikov@Mail.Ifmo.Ru

Raisa Kravchuk – “Rostelecom”, Engineer, Ggarotta@Gmail.Com

Gennady Melnikov – Saint Petersburg National Research University Of Information Technologies, Mechanics And Optics, Professor, D.Sc., Professor, Melnikov@Ifmo.Ru

Sergey Shakhoval – ZAO «Ericsson Corporation AO», data base administrator, shakhovalsergey@gmail.com