НАУЧНО-ТЕХНИЧЕСКИЙ
ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО САНКТ-ПЕТЕРБУРГ

ЯНВАРЬ-ФЕВРАЛЬ 2013 ВЫПУСК 1 (83) ISSN 2226-1494

ОБЗОРНАЯ СТАТЬЯ

Адаптивное управление по выходу: проблематика, прикладные задачи и решения — *Бобцов А.А., Никифоров В.О.* — 1

СОВРЕМЕННЫЕ ПРОБЛЕМЫ ТЕОРИИ СИСТЕМ УПРАВЛЕНИЯ

Управление нелинейным параметрически не определенным объектом с входным запаздыванием — *Бобцов А.А., Пыркин А.А.* — 15

Субоптимальное управление нелинейными мультиагентными системами — *Фуртат И.Б.* — 19

Исследование устойчивости динамических систем с использованием свойств потока вектора фазовой скорости через замкнутую выпуклую поверхность — *Фуртат И.Б.* — 23

Гибридный алгоритм идентификации частот мультисинусоидального сигнала — *Бобцов А.А., Ведяков А.А., Колюбин С.А., Пыркин А.А.* — 28

Интегро-дифференцирующие алгоритмы скоростного градиента в задачах кратной синхронизации вибрационных установок — *Фрадков А.Л., Томчина О.П., Галицкая В.А., Горлатов Д.В.* — 30

Робастное управление роботом-манипулятором в дискретном времени — *Гусев С.В.* — 38

Анализ и обработка сигналов электромиограммы — *Зименко К.А., Боргуль А.С., Маргун А.А.* — 41

Внедрение комплексов промышленных манипуляционных роботов в образовательный процесс — *Бобцов А.А., Колюбин С.А., Пыркин А.А.* — 43

Вопросы оптимизации применения беспилотных летательных аппаратов для поиска и слежения объектов — *Абдулов Р.Н., Абдуллаев Н.А., Асадов Х.Г.* — 45

ОПТИЧЕСКИЕ И ОПТИКО-ЭЛЕКТРОННЫЕ СИСТЕМЫ. ОПТИЧЕСКИЕ ТЕХНОЛОГИИ

Гомогенизаторы лазерного излучения, изготовленные методом лазерной гравировки полимерных пленок — *Баля В.К., Денисюк И.Ю., Бараусова Е.В.* — 50

ФОТОНИКА И ОПТОИНФОРМАТИКА

Характеристики метапленок для терагерцового диапазона частот при масштабировании геометрических параметров — *Терехов Ю.Е., Ходзицкий М.К., Белокопытов Г.В.* — 55

Воздействие терагерцового излучения на рост нейритов — *Цуркан М.В., Смолянская О.А., Брянцева Н.Г.* — 60

Фемтосекундное фотонное эхо на неоднородных по размеру квантовых точках оксида цинка при комнатной температуре — *Вашурин Н.С., Попов И.И., Путилин С.Э., Сидорова В.Т., Степанов С.А., Сушенцов Н.И.* — 65

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Погрешности моделирования высоконагруженных систем в GPSS World — *Алиев Т.И.* — 70

Иерархия глобальных структур алгебраической байесовской сети как система графов и гиперграфов — *Фильченков А.А.* — 75

Извлечение и ранжирование ключевых фраз в задаче аннотирования — *Попова С.В., Ходырев И.А.* — 81

Верификация данных в системах отслеживания задач с помощью продукционных правил — *Катериненко Р.С., Бессмертный И.А.* — 86

Агломеративная кластеризация речевых сегментов фонограммы на основе байесовского информационного критерия — *Кудашев О.Ю.* — 90

Программная система и информационный банк метаданных третичных структур белков — *Никитин Т.А., Порозов Ю.Б.* — 94

Анализ частоты повторений rle-блоков в семействах бинарных кодов, наилучших по минимаксному критерию автокорреляционной функции — *Ковылин А.А., Злобин Д.В., Родионов А.Ю.* — 99

Минимальная модификация автоматных программ при изменении сценариев их работы — *Шестаков А.В.* — 104

Фазово-градиентный метод измерения магнитного поля в электромобиле — *Птицына Н.Г., Исмагилов В.С., Копытенко Ю.А., Коробейников А.Г.* — 108

МАТЕРИАЛОВЕДЕНИЕ И НАНОТЕХНОЛОГИИ

Гальваномагнитные свойства тонких пленок висмута, легированного теллуром, полученных методом термического испарения в вакууме — *Комаров В.А., Матвеев Д.Ю., Демидов Е.В., Крушельницкий А.Н.* — 113

Влияние процесса лазерного микроструктурирования на электрофизические параметры системы SiO_2/Si — *Скворцов А.М., Халецкий Р.А., Хуинь Конг Ту* — 119

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

Применение симплекс-метода и метода искусственного базиса при проектировании бортового приборного оборудования — *Дейко М.С., Жаринов И.О.* — 124

Анализ методик проектирования беспроводных сенсорно-актуаторных сетей — *Киров Д.А., Ожиганов А.А.* — 129

Использование методов моделирования эволюции для оптимизации документооборота на предприятии — *Чернышев Ю.О., Басова А.В., Панасенко П.А., Поляков В.И.* — 135

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ

Оценка рисков информационной безопасности в облачных сервисах на основе линейного программирования — *Зикратов И.А., Одегов С.В., Смирных А.В.* — 141

Алгоритм контроля целостности деперсонализированных данных в информационных системах персональных данных — *Гатчин Ю.А., Теплоухова О.А., Куракин А.С.* — 145

Система предотвращения чрезвычайных ситуаций на железной дороге на базе геоинформационных технологий — *Алексеев В.В., Коновалова В.С., Минина А.А.* — 148

Цифровые водяные знаки в графических файлах — *Коробейников А.Г., Кувшинов С.С., Блинов С.Ю., Лейман А.В., Кутузов И.М.* — 152

ЭКОНОМИКА И ФИНАНСЫ. МЕНЕДЖМЕНТ

Методические положения оценки эффективности деятельности телекоммуникационного предприятия — *Цуканова О.А., Лаврентьева Т.С.* . — 158

Инновационная деятельность субъектов банковской системы как форма реализации экономических интересов — *Красноперова Т.Я.* — 161

НОВЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

Архитектура лингвистического тренажера для экспресс освоения навыков общения на иностранном языке — *Киселев В.В., Елисеева О.Е., Ковалёнок Ю.И., Хитрова Ю.Н.* — 166

SUMMARY — 171

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: //ntv.ifmo.ru
E-mail:karmanov@mail.ifmo.ru

ОБЗОРНАЯ СТАТЬЯ

УДК 681.51.015

АДАПТИВНОЕ УПРАВЛЕНИЕ ПО ВЫХОДУ: ПРОБЛЕМАТИКА, ПРИКЛАДНЫЕ ЗАДАЧИ И РЕШЕНИЯ[1]

А.А. Бобцов, В.О. Никифоров

Бобцов Алексей Алексеевич, доктор технических наук, профессор НИУ ИТМО, декан факультета компьютерных технологий и управления. Специалист по системному анализу, управлению и обработке информации в технических системах. Автор более 200 печатных работ, из которых более 80 – в международных журналах. Автор девяти монографий и учебных пособий, обладатель двух патентов. Действительный член Академии навигации и управления движением, член Научного совета РАН по теории управляемых процессов и автоматизации. Член научно-технического совета ассоциации авиаприборостроителей, член двух технических комитетов Международной федерации по автоматическому управлению. Председатель Совета молодых ученых и специалистов при Правительстве Санкт-Петербурга. Член редколлегии журнала «Научно-технический вестник информационных технологий, механики и оптики».

Никифоров Владимир Олегович, доктор технических наук, профессор, проректор НИУ ИТМО, заместитель генерального директора – технический директор ОАО «ЛОМО». Область научных интересов – нелинейные и адаптивные системы автоматического управления, методы компенсации внешних возмущений, адаптивные наблюдатели. Автор более 120 печатных работ, в том числе более 40 – в зарубежных изданиях. Действительный член Академии навигации и управления движением. Автор пяти монографий и шести учебных пособий. Председатель диссертационного совета. Член Санкт-Петербургской региональной группы национального комитета по автоматическому управлению РАН. Главный редактор журнала «Научно-технический вестник информационных технологий, механики и оптики» и российского издания журнала «Control Engineering».

Рассматривается проблема адаптивного управления по выходу параметрически и функционально неопределенными объектами. Наряду с математической постановкой задачи представлены прикладные примеры, иллюстрирующие практические применения обсуждаемой в статье теории. Приведен краткий обзор методов адаптивного управления по выходу как линейными, так и нелинейными системами, представлена обширная библиография, в которой читатель сможет найти детальное описание конкретных алгоритмов и их свойств. Подробно рассмотрен новый подход к задаче управления по выходу – метод последовательного компенсатора.

Ключевые слова: адаптивное управление, робастное управление, управление по выходу, сложные динамические системы, нелинейные системы.

Введение

Задача адаптивного и робастного управления параметрически и функционально неопределенными объектами в условиях возмущений и запаздывания относится к фундаментальным и актуальным проблемам современной теории и практики автоматического управления. Идеализирующие предположения о линейности объекта управления, стационарности его параметров, возможности измерения всех переменных состояния, отсутствии запаздывания и возмущающих воздействий постепенно уходят из рассмотрения в рамках современной теории управления [1–41].

В статье особое внимание уделяется методам адаптивного и робастного управления по выходу, т.е. без измерения производных выходной переменной или всех переменных состояния объекта. Мотивация данных исследований обусловлена тем, что управление по выходу позволяет уменьшить затраты на проектирование, разработку и изготовление датчиков, которые, в свою очередь, могут вносить дополнительные погрешности, связанные с ошибками измерений. Кроме того, для ряда реальных систем не удается обеспечить непосредственное измерение всех переменных состояния объекта управления.

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

К настоящему времени получены решения многих важных задач адаптивного и робастного управления по выходу, часть из которых доведена до инженерных методов синтеза. Однако большинство известных методов адаптивного и робастного управления отличается сложностью процедур синтеза, а также сложностью структуры и высоким динамическим порядком получаемого регулятора. Таким образом, разработка новых методов адаптивного и робастного управления, позволяющих получить более простые и малоразмерные регуляторы, является актуальной задачей современной теории управления.

Целью данной статьи является представление широкому кругу специалистов проблематики адаптивного и робастного управления по выходу параметрически и функционально неопределенными объектами. Будут сформулированы строгие постановки задач управления, а также предложены различные примеры систем управления, в которых актуальна данная проблематика. Также читателю будет представлен краткий обзор методов адаптивного управления по выходу и предложен для более подробного знакомства один из современных подходов – «метод последовательного компенсатора», базирующийся на теореме о пассификации А.Л. Фрадкова [42–45], которая, в свою очередь, расширяет фундаментальный результат теории автоматического управления – лемму Якубовича–Калмана [4, 46].

Задачи адаптивного и робастного управления по выходу линейными объектами

Представим математическую постановку задач адаптивного и робастного управления по выходу линейными объектами в условиях возмущений.

Постановка задачи. Рассматривается линейный стационарный объект управления, представленный математической моделью в форме «вход–выход»

$$\alpha(p)y(t) = b_m\beta(p)[u(t)+\delta(t)], \tag{1}$$

где $p = d/dt$ – оператор дифференцирования; $u(t)$ – управление; $y(t)$ – выходная переменная и $\delta(t)$ – ограниченное возмущающее воздействие; b_m – постоянный коэффициент; $\alpha(p)$ и $\beta(p)$ – нормированные полиномы с неизвестными постоянными коэффициентами a_i и b_j, такие что

$$\alpha(p) = p^n + a_{n-1}p^{n-1} + a_{n-2}p^{n-2} + \ldots + a_1 p + a_0,$$

$$b_m\beta(p) = b_m p^m + b_{m-1}p^{m-1} + \ldots + b_1 p + b_0.$$

Как правило, задача адаптивного управления по выходу решается при следующих допущениях (см. например, [4–6, 9]).

Допущение Д1. Полином $\beta(p)$ гурвицев.

Допущение Д2. Знак коэффициента b_m является известным (без потери общности будем считать, что $b_m > 0$).

Допущение Д3. Прямым измерениям доступна только выходная переменная $y(t)$, но не ее производные.

Допущение Д4. Известны степени полиномов $\alpha(p)$ и $\beta(p)$ – n и m, соответственно.

Допущения (Д1)–(Д4) широко используются в литературе и поэтому получили название *стандартных допущений*. Решение задачи адаптивного управления по выходу при нарушении стандартных допущений (например, при неизвестном знаке коэффициента b_m, неизвестных степенях полиномов $\alpha(p)$ и $\beta(p)$, доступности измерениям зашумленного выходного сигнала и т.п.) является предметом современных исследований.

Наряду с объектом управления рассмотрим эталонный сигнал $y^*(t)$, задающий желаемый характер изменения регулируемой переменной $y(t)$ и удовлетворяющий условию

$$\left|d^i y^*/dt^i\right| \le C_0 < \infty, \tag{2}$$

где $i = \overline{0,\rho}$, $\rho = n - m$ – относительная степень объекта управления (1). Обычно $y^*(t)$ формируется в виде выхода *эталонной модели*

$$y^*(t) = \frac{k}{\alpha^*(p)}g(t), \tag{3}$$

где k – постоянный коэффициент; $\alpha^*(p)$ – гурвицев полином степени ρ; $g(t)$ – ограниченный командный сигнал. Отметим, что использование эталонной модели (3) гарантирует выполнение условий (2).

Теперь сформулируем цели управления.

Цель управления 1 (адаптивное управление). *При условии, что внешнее неизмеряемое возмущение отсутствует* (*т.е. при* $\delta(t) \equiv 0$), *требуется найти такой закон управления в форме обратной связи по выходной переменной* $u = u(y, y^*)$, *чтобы для объекта* (1) *выполнялось целевое условие*

$$\lim_{t\to\infty} \varepsilon(t) = 0,$$

где

$$\varepsilon(t) = y^*(t) - y(t) \tag{4}$$

– ошибка слежения.

Другими словами, при отсутствии внешнего возмущения может быть поставлена и решена задача асимптотического слежения за эталонным сигналом. Наличие неизмеряемого возмущения $\delta(t)$ в общем случае не позволяет решить задачу асимптотического слежения. При этом даже обеспечение устойчивости замкнутой параметрически неопределенной системы становится нетривиальной задачей, решение которой может быть достигнуто с использованием специальных методов *робастного управления* [4, 9, 17].

Цель управления 2 (робастное управление). *При наличии ограниченного возмущения* $\delta(t)$ *требуется найти такой закон управления в форме обратной связи по выходной переменной* $u = u(y, y^*)$, *чтобы для объекта* (1) *выполнялось целевое условие*

$$\left| y^*(t) - y(t) \right| < \Delta \text{ для всех } t \ge t_1, \tag{5}$$

где Δ *– заданная разработчиком точность управления.*

Основные проблемы адаптивного управления по выходной переменной. Даже в условиях отсутствия внешних возмущений задача синтеза адаптивного управления по выходной переменной является нетривиальной, что связано с двумя принципиальными проблемами.

Первая проблема – математическая модель объекта управления должна быть соответствующим образом параметризована, т.е. приведена к специальному виду, позволяющему использовать универсальные алгоритмы адаптации, разработанные для широких классов канонических моделей.

Например (подробнее – см. [4, 9, 23]), с использованием вспомогательных фильтров

$$\dot{\boldsymbol{\upsilon}}_1 = \boldsymbol{\Lambda}\boldsymbol{\upsilon}_1 + \mathbf{e}_{n-1}u, \tag{6}$$

$$\dot{\boldsymbol{\upsilon}}_2 = \boldsymbol{\Lambda}\boldsymbol{\upsilon}_2 + \mathbf{e}_{n-1}y, \tag{7}$$

где $\boldsymbol{\upsilon}_1$ и $\boldsymbol{\upsilon}_2$ – $(n-1)$-мерные векторы состояния фильтров (6), (7); $\boldsymbol{\Lambda}$ – сопровождающая матрица произвольного гурвицева полинома степени $(n-1)$; $\mathbf{e}_i$ – единичный вектор с единицей на i-м месте, модель ошибки слежения (4) (при $\delta(t) \equiv 0$) может быть представлена в виде

$$\varepsilon(t) = \frac{b_m}{\alpha^*(p)}\left(\boldsymbol{\phi}^T(t)\boldsymbol{\theta} + u\right), \tag{8}$$

где $\boldsymbol{\phi}^T = [\boldsymbol{\upsilon}_1^T, \boldsymbol{\upsilon}_2^T, y, g]$ – вектор известных функций (регрессор); $\boldsymbol{\theta}$ – $2n$-мерный вектор неизвестных параметров. Тогда, выбирая управление

$$u = -\boldsymbol{\phi}^T(t)\hat{\boldsymbol{\theta}}(t),$$

где $\hat{\boldsymbol{\theta}}(t)$ – вектор настраиваемых параметров, генерируемый алгоритмом адаптации, для уравнения (8) получаем замкнутую параметризованную модель вида

$$\varepsilon(t) = \frac{b_m}{\alpha^*(p)}\boldsymbol{\phi}^T(t)\tilde{\boldsymbol{\theta}}(t), \tag{9}$$

которая представляет собой линейную регрессионную модель, хорошо изученную в задачах идентификации и адаптации линейных систем (здесь $\tilde{\boldsymbol{\theta}}(t) = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t)$ – вектор параметрических ошибок).

Альтернативный подход предусматривает использование вспомогательных фильтров:

$$\dot{\boldsymbol{\xi}}_n = \mathbf{A}_\Sigma\boldsymbol{\xi}_n + \boldsymbol{\kappa}y,$$

$$\dot{\boldsymbol{\xi}}_n = \mathbf{A}_\Sigma\boldsymbol{\xi}_n + \mathbf{e}_{n-i}y, \qquad 0 \le i \le n-1,$$

$$\dot{\mathbf{v}}_i = \mathbf{A}_\Sigma\mathbf{v}_i + \mathbf{e}_{n-i}u, \qquad 0 \le \mathrm{i} \le m,$$

где $\boldsymbol{\xi}_n$, $\boldsymbol{\xi}_i$ и $\mathbf{v}_i$ – n-мерные векторы состояния фильтров, а вектор коэффициентов обратных связей $\boldsymbol{\kappa}$ выбран так, что матрица $\mathbf{A}_\Sigma = \mathbf{E} - \boldsymbol{\kappa}\mathbf{e}_1^T$ является гурвицевой.

Тогда выходная переменная может быть представлена в виде

$$y = \boldsymbol{\xi}_{n,1} + \boldsymbol{\varphi}^T(t)\boldsymbol{\psi}, \tag{10}$$

где $\boldsymbol{\varphi}(t) = \boldsymbol{\varphi}(\boldsymbol{\xi}_i, \mathrm{v}_i)$ – $(n+m+1)$-мерный вектор известных функций, а $\boldsymbol{\psi}$ – вектор неизвестных параметров (подробнее – см. [4, 9, 17]). Управление в явном виде не представлено в модели (10) и появляется только в выражении для ρ-ой производной выходной переменной. В связи с этим для модели (10) не могут быть использованы универсальные алгоритмы адаптации, что потребовало разработки специальных итеративных процедур синтеза алгоритмов адаптивного управления – так называемого метода *обратного обхода интегратора* [4, 9, 14, 16, 17, 27].

Вторая проблема адаптивного управления по выходу – высокая относительная степень объекта управления, что связано с измерением только выходной переменной, а не всего вектора состояния (или производных выходной переменной).

Для пояснения проблемы высокой относительной степени представим замкнутую параметризованную модель ошибки (9) в пространстве состояний:

$$\dot{\boldsymbol{\eta}} = \boldsymbol{\Gamma}\boldsymbol{\eta} + \mathbf{e}_\rho \boldsymbol{\phi}^T(t)\tilde{\boldsymbol{\theta}}(t), \tag{11}$$

$$\varepsilon = \mathbf{e}_1^T \boldsymbol{\eta},$$

где $\boldsymbol{\eta}$ – ρ-мерный вектор состояния модели (11), $\dfrac{b_m}{\alpha^*(p)} = \mathbf{e}_1^T\left(p\mathbf{I} - \boldsymbol{\Gamma}\right)^{-1}\mathbf{e}_\rho$. Предполагая временно, что вектор состояния $\boldsymbol{\eta}$ доступен прямым измерениям, выберем стандартный алгоритм адаптивной настройки параметров регулятора:

$$\dot{\hat{\boldsymbol{\theta}}} = \gamma\boldsymbol{\phi}\mathbf{e}_\rho^T\mathbf{P}\boldsymbol{\eta}, \tag{12}$$

где γ – произвольный положительный коэффициент (коэффициент адаптации), а симметрическая положительно определенная матрица $\mathbf{P}$ является решением уравнения

$$\boldsymbol{\Gamma}^T\mathbf{P} + \mathbf{P}\boldsymbol{\Gamma} = -\mathbf{Q} \tag{13}$$

с произвольной симметрической положительно определенной матрицей $\mathbf{Q}$. Свойства устойчивости замкнутой системы доказываются с использованием функции Ляпунова вида

$$\mathbf{V}(\boldsymbol{\eta}, \tilde{\boldsymbol{\theta}}) = \frac{1}{2}\boldsymbol{\eta}^T\mathbf{P}\boldsymbol{\eta} + \frac{1}{2\gamma}\tilde{\boldsymbol{\theta}}^T\tilde{\boldsymbol{\theta}},$$

производная которой по времени в силу уравнений замкнутой системы удовлетворяет неравенству

$$\dot{\mathbf{V}}(\boldsymbol{\eta}, \tilde{\boldsymbol{\theta}}) = \frac{1}{2}\boldsymbol{\eta}^T(\boldsymbol{\Gamma}^T\mathbf{P} + \mathbf{P}\boldsymbol{\Gamma})\boldsymbol{\eta} + \tilde{\boldsymbol{\theta}}^T\boldsymbol{\phi}\mathbf{e}_\rho^T\mathbf{P}\boldsymbol{\eta} - \frac{1}{\gamma}\tilde{\boldsymbol{\theta}}^T\dot{\tilde{\boldsymbol{\theta}}} = -\frac{1}{2}\boldsymbol{\eta}^T\mathbf{Q}\boldsymbol{\eta} \le 0.$$

Из последнего выражения следует устойчивость по Ляпунову замкнутой системы и выполнение целевого условия $\boldsymbol{\eta} \to 0$ при $t \to \infty$.

Исследуем теперь условия применимости базового алгоритма адаптации (12) в случае недоступности прямым измерениям вектора состояния параметризованной модели $\boldsymbol{\eta}$. Легко видеть, что данный алгоритм применим, если удастся выбрать матрицу $\mathbf{P}$ таким образом, что

$$\mathbf{e}_\rho^T\mathbf{P} = \mathbf{e}_1^T. \tag{14}$$

В этом случае алгоритм адаптации принимает вид

$$\dot{\hat{\boldsymbol{\theta}}} = \gamma\boldsymbol{\phi}\varepsilon \tag{15}$$

и является физически реализуемым, так как в нем используются только измеряемые сигналы.

Таким образом, условием применимости алгоритма адаптации вида (15) является существование симметрической положительно определенной матрицы $\mathbf{P}$, удовлетворяющей одновременно двум уравнениям (13) и (14). В свою очередь, такая матрица может быть найдена не для всех моделей ошибки (9), а только для моделей со *строго положительно вещественной* (*СПВ*) *передаточной функцией*.

Лемма Якубовича-Калмана [46]. *Симметрическая положительно определенная матрица* $\mathbf{P}$, *являющаяся решением одновременно двух уравнений* (13) *и* (14), *существует только в том случае, если передаточная функция модели ошибки* (9) *является СПВ.*

Рассмотрим произвольную передаточную функцию вида

$$H(s)=\frac{b_m s^m+b_{m-1}s^{m-1}+\cdots+b_0}{s^n+a_{n-1}s^{n-1}+\cdots+a_0},$$

где s – комплексная переменная. Передаточная функция $H(s)$ является СПВ функцией только в том случае, если:

(У1) она не имеет полюсов в области $\mathrm{Re}[s]\geq 0$;

(У2) $\mathrm{Re}[H(j\omega)]>0$ для всех $-\infty<\omega<\infty$;

(У3) $\lim \omega^2\,\mathrm{Re}[H(j\omega)]>0$ при $\omega\to\infty$.

Для пояснения приведенного критерия рассмотрим передаточную функцию апериодического звена первого порядка

$$H_A(s)=\frac{k}{Ts+1}, \tag{16}$$

где $k>0$ – коэффициент усиления, а $T>0$ – постоянная времени. Покажем, что передаточная функция (16) удовлетворяет условиям (У1)–(У2). Действительно, единственный полюс функции (16) $s=-1/T$ лежит вне области $\mathrm{Re}[s]>0$. Как известно, частотный годограф апериодического звена первого порядка полностью лежит в четвертом квадранте комплексной плоскости, что гарантирует выполнение условия (У2). Наконец, рассмотрим частотную передаточную функцию звена

$$H_A(j\omega)=\frac{k}{jT\omega+1}=\frac{k}{T^2\omega^2+1}-j\frac{kT\omega}{T^2\omega^2+1},$$

где ω – угловая частота. Тогда

$$\lim_{\omega\to\infty}\omega^2\left(\frac{k}{T^2\omega^2+1}\right)=\frac{k}{T^2}>0,$$

что означает выполнение условия (У3).

Таким образом, можно сделать вывод, что СПВ функции обладают почти такими же частотными свойствами, что и звено первого порядка. Так, их частотные годографы лежат в правой полуплоскости (а значит, фазовый сдвиг, вносимый динамическим звеном с такой передаточной функцией, не превышает $90°$). Кроме того, скорость убывания вещественной части частотной передаточной функции при $\omega\to\infty$ не быстрее, чем у $1/\omega^2$.

Насколько ограничительным является требование строгой положительной вещественности? Для ответа на этот вопрос учтем, что СПВ передаточная функция не может иметь относительную степень выше единицы (что определяется требованием предельного фазового сдвига в $90°$). Очевидно, что класс динамических систем с единичной относительной степенью является достаточно узким, и в него не попадают математические модели большинства реальных технических устройств. Таким образом, стандартный алгоритм адаптации по выходу (15) может быть использован только в частных случаях, и необходимо получение универсальных алгоритмов адаптации, применимых ко всему классу моделей вида (9). Для объектов с произвольной относительной степенью должны использоваться специальные методы синтеза алгоритмов адаптации или адаптивного управления в целом.

Методы адаптивного и робастного управления линейными стационарными объектами по выходу. На сегодняшний день можно уверенно говорить о том, что задачи адаптивного и робастного управления линейными стационарными объектами в условиях неопределенности параметров математической модели и действии неизвестных возмущений решены. За последние 30 лет опубликовано достаточное число работ, в которых представлены разнообразные методы, позволяющие решать задачу адаптивного управления по выходу. Большинство методов обсуждались в обзоре [5], ряд новых подходов рассмотрен в монографиях [4, 9]. Среди основных методов управления линейными стационарными неопределенными объектами по выходу можно выделить:

– метод расширенной ошибки или схема Монополи [5, 21, 47];
– алгоритм высокого порядка или метод Морза [22];

– использование итеративных процедур синтеза [9, 14, 16, 17, 27, 30];
– метод шунтирования [2];
– робастная модификация алгоритма высокого порядка [4, 9, 26];
– робастная модификация итеративных процедур синтеза [4, 9].

Каждый из указанных методов имеет преимущества и недостатки по сравнению с аналогами. Общее достоинство данных методов заключается в том, что они позволяют синтезировать управление в случае, когда относительная степень исходной модели (1) больше единицы. Однако данные методы имеют два существенных недостатка:
– высокая размерность регулятора;
– сложность практической реализации.

Задачи адаптивного и робастного управления по выходу нелинейными объектами

В данном разделе будет сформулирована математическая постановка задачи адаптивного и робастного управления по выходу нелинейными объектами в условиях возмущений и запаздывания. Будет дан краткий обзор методов адаптивного управления по выходу нелинейными объектами.

Постановка задачи. Рассмотрим нелинейную систему (рис. 1) вида

$$\dot{\mathbf{x}} = A\mathbf{x} + \sum_{i=1}^{r} \mathbf{G}_i \overline{\varphi}_i + \mathbf{B}u + \mathbf{B}w, \tag{17}$$

$$y = \mathbf{H}x, \tag{18}$$

где $\mathbf{x} \in R^n$ – вектор переменных состояния; $\mathbf{A}, \mathbf{B}, \mathbf{G}_i, \mathbf{H}$ – неизвестные матрица и векторы соответственно; $w(t)$ – неизвестное гладкое ограниченное возмущение, т.е. $|w(t)| \le w_0 < \infty$; $\overline{\varphi}_i = \overline{\varphi}_i(y(t-\tau))$ – неизвестная функция, где $\tau \ge 0$ – неизвестное постоянное запаздывание, $y(\vartheta) = \overline{\varphi}(\vartheta)$ для $\forall \vartheta \in [-\tau, 0]$.

Прежде чем формулировать цель управления, рассмотрим ряд допущений (например, [4, 9]).

Допущение Д5. Передаточная функция $W(p) = \mathbf{H}(p\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} = b(p)/a(p)$ – минимально-фазовая, т.е. полином $b(p)$ гурвицев.

Допущение Д6. Относительно неизвестной нелинейности $\overline{\varphi}_i$ будем полагать, что она:
– либо ограничена для любых значений $y(t-\tau)$, т.е.

$$|\overline{\varphi}_i| \le C_0, \tag{19}$$

– либо выполнено секторное ограничение вида

$$|\overline{\varphi}_i| \le C_0 |y(t-\tau)|, \tag{20}$$

– либо относительно нелинейности задана мажоранта вида

$$|\overline{\varphi}_i| \le C_0 \psi(y(t-\tau)), \tag{21}$$

где число $C_0 > 0$ для всех случаев неизвестно, а $\psi(y(t-\tau))$ – известная функция.

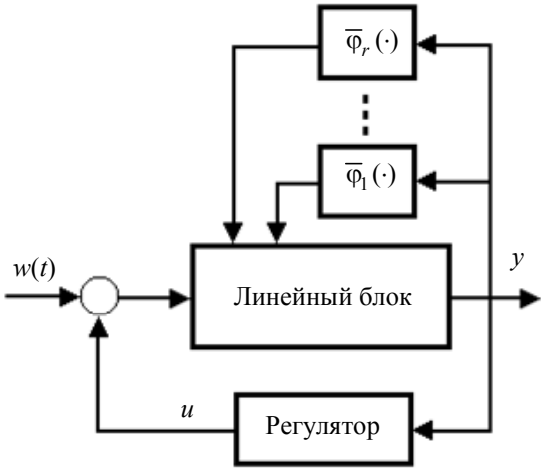


Рис. 1. Структурная схема системы управления (17), (18)

Теперь для представленных допущений сформулируем цели управления.

Цель управления 3 (адаптивное управление). Пусть неизвестное возмущение $w(t) \equiv 0$ и выполнены условия (20), (21). Для системы (17), (18) требуется найти закон управления в форме обратной связи по выходу $u = u(y)$, обеспечивающий выполнение целевого условия

$$\lim_{t\to\infty} |y(t)| = 0 . \qquad (22)$$

Цель управления 4 (робастное управление). Пусть неизвестное возмущение $w(t)$ является гладкой ограниченной функцией и (или) выполнено условие (19). Требуется найти закон управления в форме обратной связи по выходу $u = u(y)$, обеспечивающий выполнение целевого неравенства

$$|y(t)| < \Delta \text{ при } t \ge t_1 , \qquad (23)$$

где число Δ задается разработчиком системы управления.

Краткий обзор методов адаптивного и робастного управления по выходу нелинейными неопределенными объектами в условиях возмущений и запаздывания. Отличительной особенностью задач управления нелинейными системами является невозможность указать универсальные пути их решения. Именно этим вызван неугасающий рост публикаций в области нелинейного управления. Большинство методов обсуждались в обзоре [48] и работах [2, 4, 9, 10, 17]. Особую нишу занимают задачи адаптивного и робастного управления по выходу нелинейными неопределенными объектами. Среди основных методов управления нелинейными объектами можно выделить:

- методы, предусматривающие поиск функции Ляпунова [3, 4, 10, 49, 50];
- геометрические методы, связанные с точной линеаризацией и поиском диффеоморфного преобразования координат [3, 4, 9, 14, 37–40, 48];
- частотные методы, предусматривающие использование критерия Попова, кругового критерия и т.д. [2, 35, 36, 48, 50, 51];
- методы, предусматривающие поиск так называемой управляющей функции Ляпунова [48, 52];
- методы теории пассивных систем [4, 44, 48, 53, 54];
- итеративные процедуры синтеза [3, 16, 28, 48, 55].

Естественно, что перечисленными методами анализа и синтеза не ограничивается современный уровень развития нелинейной теории. Однако указанные методы являются наиболее распространенными при решении многих теоретических и практических задач управления нелинейными объектами.

Также как и при анализе методов адаптивного и робастного управления по выходу линейными стационарными неопределенными объектами, можно говорить о том, что рассмотренные выше методы анализа и синтеза нелинейных систем имеют два существенных недостатка:

- высокая размерность и сложность построения регуляторов;
- сложность практической реализации.

Некоторые прикладные задачи адаптивного и робастного управления

Для широкого круга читателей, не являющихся специалистами в области математической теории управления, авторы предлагают ряд практических примеров адаптивного управления по входу различными классами технических и физических систем.

Адаптивное управление по выходу двигателем постоянного тока. Запишем математическую модель двигателя постоянного тока в виде [19, 20, 56]

$$\frac{d\alpha}{dt} = \omega ,$$

$$J\frac{d\omega}{dt} = -F(t)\omega + K_t i - T(t) ,$$

$$L\frac{di}{dt} = -K_b\omega - R(t)i + u,$$

где α – угол поворота выходного вала (регулируемая переменная); ω – угловая скорость вращения выходного вала; i – ток обмотки возбуждения; u – напряжение якоря (сигнал управления), все или некоторые физические параметры L, $R(t)$, J, $T(t)$, $F(t)$ и K_t изменяются во времени и могут быть неточно заданы. Во многих электроприводах переменные ω и i не измеряются.

Легко видеть, что данная модель двигателя постоянного тока при некоторых допущениях соответствует математическим моделям (1) или (17), (18). Следовательно, задачи управления и методы их решения, представленные выше, могут быть распространены на данный технический объект.

Адаптивное управление углом либрации спутника. В современной отечественной и зарубежной литературе можно найти довольно много работ по исследованию угловых колебаний космических

аппаратов (спутников) и управлению ими [57–59]. В последнее время, помимо традиционных задач управления космическими конструкциями, внимание ученых стали привлекать задачи, связанные с хаотическими колебаниями спутниковых систем. В частности, является актуальной задача управления углом либрации спутника (например, [60, 61]). Математическая модель угла либрации $\phi(t)$ спутника в плоскости орбиты имеет вид

$$C\ddot{\phi} + c\dot{\phi} + 3\omega_c^2\left(B - A\right)\sin\phi\cos\phi + \mu_m i I r^{-3}\left(2\sin\phi\sin\omega_c t + \cos\phi\cos\omega_c t\right) = M_c(t),$$

где c – коэффициент собственного демпфирования спутника; ω_c – значение угловой скорости движения спутника по орбите; A, B – главные моменты инерции спутника ($B > A$); μ_m – магнитная постоянная; I – величина магнитного момента спутника; r, i – радиус и наклонение орбиты; $M_c(t)$ – значение управляющего момента.

Легко показать, что данная математическая модель соответствует, рассмотренной ранее обобщенной математической модели (17), (18).

Адаптивная стабилизация хаоса в цепи Чуа. Одним из классических примеров дифференциальных моделей, описывающих хаотические процессы в электрических и электронных схемах, является система Чуа [60, 62–64], предложенная специалистами по электронным цепям Л. Чуа и Т. Мацумото. В тоже время математическая модель хаотической системы Чуа эквивалентна математической модели (17), (18). Цепь Чуа описывается уравнениями

$$\frac{dU_1}{dt} = \frac{1}{C_1}\left[G\left(U_2 - U_1\right) - \varphi(U_1)\right],$$

$$\frac{dU_2}{dt} = \frac{1}{C_2}\left[G\left(U_1 - U_2\right) + i\right],$$

$$\frac{di}{dt} = \frac{1}{L}\left[-U_2 - R_0 i\right],$$

где $G = \frac{1}{R}$ и $\varphi(U_1) = G_b U_1 + \frac{1}{2}\left(G_a + G_b\right)\left(\left|U_1 + E\right| - \left|U_1 - E\right|\right)$ – характеристика нелинейного диода Чуа.

Управление однозвенным роботом-манипулятором с гибкими связями и незначительным демпфированием. Рассмотрим простой однозвенный робот-манипулятор [65], круговое движение выходного звена которого осуществляется посредством эластичного соединения звена и исполнительного механизма (силового привода). Эластичным соединением между исполнительным механизмом и звеном во многих практических задачах нельзя пренебречь, и опыт показывает, что роботы-манипуляторы, в которых движение передается посредством длинных валов или ременных передач, в которых исполнительный механизм является гармоническим приводом, обладают резонансным поведением в той же области частот, что и частоты, используемые для управления. Эффекты эластичного соединения между исполнительными механизмами и звеньями, на которые обычно ссылаются как на эластичные связи, могут быть смоделированы включением линейной торсионной пружины в каждом соединении между валом исполнительного механизма и концом звена, вокруг которого осуществляется вращение (рис. 2). Используя q_1 и q_2 для обозначения угловых положений вала исполнительного механизма и руки робота-манипулятора соответственно по отношению к фиксированной системе координат, уравнение исполнительного механизма можно записать в следующем виде:

$$I\ddot{q}_1 + MgL\sin q_1 + k(q_1 - q_2) = 0,$$

$$J\ddot{q}_2 - k(q_1 - q_2) = u,$$

где q_1 и q_2 – угловые позиции; u – момент, возникающий на оси исполнительного механизма; k – постоянная упругости торсионной пружины, которая представляет эластичное соединение, а физические параметры I, M, g, L и J являются положительными и постоянными.

Легко видеть, что представленная выше математическая модель робота-манипулятора соответствует обобщенной математической модели нелинейного объекта вида (17), (18).

Адаптивное управление двухэтапным химическим реактором. Еще один не типовой для задач механики и мехатроники объект представлен в данном подразделе. Рассмотрим задачу стабилизации двухэтапного химического реактора, модель которого может быть, например, найдена в работе [66]:

$$\dot{z}_1(t) = -\frac{1}{\varsigma_1}z_1(t) - \upsilon_1 z_1(t) + \frac{1 - R_2}{V_1}z_2(t),$$

$$\dot{z}_2(t) = -\frac{1}{\varsigma_2}z_2(t) - \upsilon_2 z_2(t) + \frac{R_1}{V_2}z_1(t - \tau) + \frac{R_2}{V_2}z_2(t - \tau) + \frac{F_2}{V_2}u,$$

$$y(t) = z_1(t),$$

где $z_1(t)$ и $z_2(t)$ – состав резервуаров; R_1, R_2, ς_i, υ_i и F_2 – некоторые параметры, связанные с химическими процессами в резервуарах и V_i – объем резервуара.

Работая с данной математической моделью, можно обнаружить запаздывание, а после ряда несложных математических преобразований получить систему вида (17), (18).

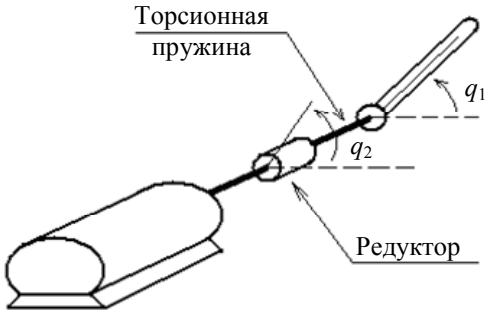


Рис. 2. Однозвенный робот-манипулятор с гибкими связями

Адаптивное управление системой впрыска для инжекторных двигателей внутреннего сгорания. В соответствии с результатами, опубликованными в работах [67–73], усредненная математическая модель системы впрыска имеет следующий вид:

$$\dot{x}_2 = -a_2 x_2 + b_2 u_1,$$

$$y_1 = \frac{1}{x_1 x_3}\left(c_1 x_2 + d_1 u_1\right),$$

$$T_s \dot{\lambda}_s + \lambda_s = \lambda(t - \tau_s),$$

где $x_1 = \omega$ – скорость вращения выходного вала [об/мин]; $x_3 = P_m$ – давление во впускном коллекторе [Па]; $x_2 = \mu_{ff}$ – масса топливной пленки на стенках впускного коллектора; $y_1 = \lambda$ – соотношение воздух/топливо; λ_s – значение соотношения воздух/топливо, выдаваемое датчиком кислорода; $u_1 = \mu_{fi}$ – количество впрыснутого в цилиндр топлива [кг]; τ_s – запаздывание, связанное с расположением датчика кислорода в выпускном коллекторе.

Сформулируем цель управления как решение задачи синтеза регулятора, обеспечивающего выполнение неравенства

$$\left|y_1(t) - y^*(t)\right| \le \varepsilon, \tag{24}$$

где $y^*(t) = 1$, а ε – некоторое, в общем случае малое, число, которое может быть уменьшено за счет выбора закона управления. Физический смысл цели управления (24) заключается в поддержании заданного соотношения воздух/топливо в горючей смеси, впрыскиваемой в цилиндр. Легко видеть, что цель управления вида (24) соответствует представленной ранее задаче управления по выходу линейными стационарными параметрически неопределенными объектами (условие (5)).

Таким образом, все шесть прикладных примеров удовлетворяют обобщенным математическим моделям, рассмотренным в постановочной части.

Метод последовательного компенсатора

Большинство из сформулированных задач (и относящихся к ним приложений) может быть решено с использованием ставших уже классических методов адаптивного управления, рассмотренных подробно в обзорах [3, 5] и монографиях [2, 4, 9]. В данной статье авторы не видят смысла повторять работы [2–5, 9] и предлагают читателю достаточно новый подход к адаптивному управлению по выходу – «метод последовательного компенсатора», базирующийся на теореме о пассификации А.Л. Фрадкова [42–45].

Для простоты рассмотрения ограничимся задачей стабилизации нелинейного объекта управления вида (17), (18) в отсутствии возмущения и запаздывания. Более подробно с данным методом и различными его вариациями можно познакомиться в монографии [61]. Представим нелинейную систему (17), (18) в форме вход–выход

$$y = \frac{b(p)}{a(p)} u + \frac{c(p)}{a(p)} \overline{\varphi}(y), \tag{25}$$

где измеряется выходная переменная $y = y(t)$, но не ее производные; $b(p) = b_m p^m + ... + b_1 p + b_0$, $c(p) = c_r p^r + c_{r-1} p^{r-1} + ... + c_1 p + c_0$ и $a(p) = p^n + a_{n-1} p^{n-1} + ... + a_1 p + a_0$ – полиномы с неизвестными параметрами; $r \le n-1$; передаточная функция $b(p)/a(p)$ имеет относительную степень $\rho = n - m$; полином $b(p)$ гурвицев, коэффициент $b_m > 0$; неизвестная функция $\overline{\varphi}(y)$ удовлетворяет неравенству (20). В качестве цели управления зададимся обеспечением условия (22) (в случае возмущенного движения необходимо было бы выполнить неравенство (23)).

Как и в [45], выберем закон управления следующим образом:

$$u = -\overline{\alpha}(p)(\mu + \kappa)\hat{y}, \tag{26}$$

где число $\mu > \mu_0$ и полином $\overline{\alpha}(\mathrm{p})$ степени $\rho - 1$ выбираются из соображений гурвицевости полинома $a(p) + \mu b(p)\overline{\alpha}(p)$, положительный параметр κ предназначен для компенсации неопределенности $\overline{\varphi}(y)$, а функция $\hat{y}(t)$ является оценкой выхода $y(t)$ и формируется алгоритмом вида

$$\begin{cases} \dot{\xi}_1 = \sigma\xi_2, \\ \dot{\xi}_2 = \sigma\xi_3, \\ ... \\ \dot{\xi}_{\rho-1} = \sigma(-k_1\xi_1 - k_2\xi_2 - ... - k_{\rho-1}\xi_{\rho-1} + k_1 y), \end{cases} \tag{27}$$

$$\hat{y} = \xi_1, \tag{28}$$

где число $\sigma > \mu + \kappa$, а коэффициенты k_i рассчитываются из требований асимптотической устойчивости системы (27) при нулевом входе y.

Очевидно, что закон управления (26)–(28) является практически реализуемым, так как содержит известные или измеряемые сигналы.

Подставляя (26) в уравнение (25), получаем

$$y = \frac{b(p)}{a(p)}[-\overline{\alpha}(p)(\mu+\kappa)\hat{y}] + \frac{c(p)}{a(p)}\overline{\varphi}(y) = \frac{b(p)}{a(p)}[-\overline{\alpha}(p)(\mu+\kappa)y + \overline{\alpha}(p)(\mu+\kappa)\varepsilon] + \frac{c(p)}{a(p)}\overline{\varphi}(y). \tag{29}$$

Проводя несложные преобразования, для (29) имеем

$$a(p)y + \mu\overline{\alpha}(p)b(p)y = b(p)\overline{\alpha}(p)[(\mu+\kappa)\varepsilon - \kappa y] + c(p)\overline{\varphi}(y),$$

принимая обозначения $\gamma(p) = a(p) + \mu\overline{\alpha}(p)b(p)$ и $\overline{\beta}(p) = \overline{\alpha}(p)b(p)$, для системы (29) получаем

$$y = \frac{\overline{\beta}(p)}{\gamma(p)}[-\kappa y + (\mu+\kappa)\varepsilon] + \frac{\mathrm{c}(p)}{\gamma(p)}\overline{\varphi}(y), \tag{30}$$

где невязка (функция отклонений) равна

$$\varepsilon = y - \hat{y}.$$

Теперь представим модель вход–выход (30) в виде модели вход–состояние–выход

$$\dot{\mathbf{x}}' = \overline{\mathbf{A}}\mathbf{x}' + \mathbf{b}(-\kappa y + (\mu+\kappa)\varepsilon) + \mathbf{q}\varphi(y), \tag{31}$$

$$y = \mathbf{c}^T\mathbf{x}', \tag{32}$$

где $\mathbf{x}' \in R^n$ – вектор переменных состояния модели (32); $\overline{\mathbf{A}}$, $\mathbf{b}$, $\mathbf{q}$ и $\mathbf{c}$ – соответствующие матрицы перехода от модели вход–выход к модели вход–состояние–выход, причем можно указать симметрическую положительно определенную матрицу $\overline{\mathbf{P}}$, удовлетворяющую двум следующим матричным уравнениям (подробнее см. [45]):

$$\overline{\mathbf{A}}^T\overline{\mathbf{P}} + \overline{\mathbf{P}}\overline{\mathbf{A}} = -\mathbf{Q}_1, \quad \overline{\mathbf{P}}\mathbf{b} = \mathbf{c},$$

где $\mathbf{Q}_1 = \mathbf{Q}_1^T$ – положительно определенная матрица, причем ее значения зависят от параметра μ и не зависят от параметра κ.

Перепишем (27), (28) в векторно-матричной форме

$$\dot{\xi} = \sigma(\overline{\mathbf{\Gamma}}\xi + \mathbf{d}k_1 y), \quad \hat{y} = \mathbf{h}^T\xi,$$

где

$$\overline{\mathbf{\Gamma}} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -k_1 & -k_2 & -k_3 & \dots & -k_{\rho-1} \end{bmatrix}, \ \mathbf{d} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \text{ и } \mathbf{h} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Введем в рассмотрение вектор отклонений $\overline{\boldsymbol{\eta}} = \mathbf{h}y - \boldsymbol{\xi}$, тогда в силу структуры матрицы $\mathbf{h}$ невязка ε примет вид

$$\varepsilon = y - \hat{y} = \mathbf{h}^T\mathbf{h}y - \mathbf{h}^T\boldsymbol{\xi} = \mathbf{h}^T(\mathbf{h}y - \xi) = \mathbf{h}^T\overline{\boldsymbol{\eta}}.$$

Для производной от $\boldsymbol{\eta}$ получим

$$\dot{\overline{\boldsymbol{\eta}}} = \mathbf{h}\dot{y} - \sigma(\overline{\mathbf{\Gamma}}(\mathbf{h}y - \overline{\boldsymbol{\eta}}) + \mathbf{d}k_1 y) = h\dot{y} + \sigma\overline{\Gamma}\overline{\eta} - \sigma(\mathbf{d}k_1 + \overline{\mathbf{\Gamma}}\mathbf{h})y.$$

Так как $\mathbf{d}k_1 = -\overline{\mathbf{\Gamma}}\mathbf{h}$ (проверяется подстановкой), то

$$\dot{\overline{\boldsymbol{\eta}}} = \mathbf{h}\dot{y} + \sigma\overline{\mathbf{\Gamma}}\overline{\boldsymbol{\eta}}, \tag{33}$$

$$\varepsilon = \mathbf{h}^T\overline{\boldsymbol{\eta}}, \tag{34}$$

где матрица $\overline{\mathbf{\Gamma}}$ в силу расчета коэффициентов k_i модели (27) имеет собственные числа с отрицательной вещественной частью и удовлетворяет уравнению Ляпунова

$$\overline{\mathbf{\Gamma}}^T\mathbf{N} + \mathbf{N}\overline{\mathbf{\Gamma}} = -\mathbf{Q}_2,$$

где $\mathbf{N} = \mathbf{N}^T$ и $\mathbf{Q}_2 = \mathbf{Q}_2^T$ – положительно определенные матрицы.

Условия применимости закона управления (27)–(29) для стабилизации системы (31), (32), (33), (34) приведены в следующей теореме.

Т е о р е м а . *Пусть полином $b(p)$ гурвицев, порядок полинома $c(p)$ $r \le n-1$, число μ выбрано так, что передаточная функция $G(p) = \dfrac{\overline{\beta}(p)}{\gamma(p)}$ СПВ и функция $\varphi(y)$ удовлетворяет следующим допущениям:*

$$\overline{\varphi}(0) = 0,$$

$$-C_0 \le \frac{\overline{\varphi}(y)}{y} \le C_0 \ \textit{для любых} \ y \ne 0,$$

где число $C_0 > 0$. Тогда существует число $\sigma > \mu + \kappa$ такое, что система (31), (32), (33), (34) *экспоненциально устойчива.*

Доказательство теоремы приведено в [45]. Из экспоненциальной устойчивости системы (31), (32) следует выполнение цели управления $\lim\limits_{t\to\infty}|y(t)| = 0$.

Теперь дадим рекомендации по настройке параметров регулятора. Возможным вариантом настройки коэффициентов κ, μ, σ является их увеличение до тех пор, пока не будет выполнено целевое условие следующего вида (например, [61]):

$$|y(t)| < \Delta \ \text{при} \ t \ge t_1,$$

где число Δ задается разработчиком системы управления. Для реализации этой идеи можно воспользоваться алгоритмом настройки вида

$$\tilde{k}(t) = \int_{t_0}^{t} \lambda(\tau)d\tau,$$

где $\tilde{k} = \kappa + \mu$, а функция $\lambda(t)$ рассчитывается следующим образом:

$$\lambda(t) = \begin{cases} \lambda_0 & \text{при } |y(t)| > \Delta, \\ 0 & \text{при } |y(t)| \le \Delta, \end{cases}$$

где число $\lambda_0 > 0$.

Выберем σ следующим образом:

$$\sigma = \sigma_0\tilde{k}^2,$$

где число $\sigma_0 > 0$. Работоспособность данного алгоритма обоснована в монографии [61].

Заключение

Авторы надеются, что им удалось хотя бы частично осветить проблематику адаптивного управления по выходу для различного типа сложных динамических систем. Теоретические методы и алгоритмы управления, представленные в рамках данного обзора, на взгляд авторов, имеют широкие перспективы для их практического использования. Однако современные технические приложения не позволяют ограничиться представленными в статье методами управления. Одной из наиболее серьезных и не имеющих, на взгляд авторов, успешных решений является задача синтеза адаптивного регулятора в условиях, когда сигнал управления содержит запаздывание, т.е. $u = u(t-\tau)$. Для линейных стационарных асимптотически устойчивых параметрически неопределенных объектов эта задача решена, но для неустойчивых и нелинейных систем проблема в целом остается открытой. Другая не менее важная прикладная задача это адаптивное и робастное управление в условиях помех измерений выходной переменной, т.е. рассматривается объект вида (17), (18)

$$\dot{x} = \mathbf{A}x + \sum_{i=1}^{r} \mathbf{G}_i \overline{\varphi}_i + \mathbf{B}u + \mathbf{B}w, \quad y' = y + \chi = \mathbf{H}x + \chi,$$

в котором измеряется не переменная $y = \mathbf{H}x$, а некоторый сигнал $y' = y + \chi$, содержащий помехи измерения χ.

Литература

1. Андерсон Б., Битмид Р., Джонсон К. и др. Устойчивость адаптивных систем. – М.: Мир, 1989. – 263 с.
2. Андриевский Б.Р., Фрадков А.Л. Избранные главы теории автоматического управления с примерами на языке MATLAB. – СПб: Наука, 1999. – 467 с.
3. Дружинина М.В., Никифоров В.О., Фрадков А.Л. Методы адаптивного управления нелинейными объектами по выходу // Автоматика и телемеханика. – 1996. – № 2. – С. 3–33.
4. Мирошник И.В., Никифоров В.О., Фрадков А.Л. Нелинейное и адаптивное управление сложными динамическими системами. – СПб: Наука, 2000. – 549 с.
5. Никифоров В.О., Фрадков А.Л. Схемы адаптивного управления с расширенной ошибкой. Обзор // Автоматика и телемеханика. – 1994. – № 9. – С. 3–22.
6. Никифоров В.О. Адаптивная стабилизация линейного объекта, подверженного внешним детерминированным возмущениям // Известия РАН. Теория и системы управления. – 1997. – № 2. – С. 103–106.
7. Никифоров В.О. Нелинейная система управления с компенсацией внешних детерминированных возмущений // Известия РАН. Теория и системы управления. – 1997. – № 4. – С. 69–73.
8. Никифоров В.О. Робастное управление линейным объектом по выходу // Автоматика и телемеханика. – 1998. – № 9. – С. 87–99.
9. Никифоров В.О. Адаптивное и робастное управление с компенсацией возмущений. – СПб: Наука, 2003. – 282 с.
10. Слита О.В., Никифоров В.О. Ушаков А.В. Управление в условиях неопределенности: неадаптивные и адаптивные алгоритмы. – Saarbrucken: LAP LAMBERT Academic Publisher, 2012. – 292 с.
11. Цыкунов А.М. Робастное управление нестационарными объектами // Автоматика и телемеханика. – 1996. – № 2. – С. 117–125.
12. Egart B. Stability of adaptive controllers. – NY: Springer-Verlag, 1979.
13. Feuer A., Morse A.S. Adaptive control of single-input, single-output linear systems // IEEE Trans. Automat. Contr. – 1978. – V. 23. – № 4. – P. 557–569.
14. Kanellakopoulos I., Kokotovich P.V., Morse A.S. Systematic design of adaptive controllers for feedback linearizable systems // IEEE Trans. Automat. Contr. – 1991. – V. 36. – P. 1241–1253.
15. Kreisselmeier G. Adaptive control of a class of slowly time varying plants // Syst. Control Lett. – 1986. – V. 8. – № 2. – P. 97–103.
16. Krstic M., Kanellakopoulos I., Kokotovich P.V. Adaptive nonlinear control without overparametrization // Syst. Control Lett. – 1992. – V. 19. – P. 177–185.
17. Krstic M., Kanellakopoulos I., Kokotovic P. Nonlinear and adaptive control design. – NY: John Weley and Sons Inc., 1995.
18. Krstic M., Kokotovic P. Adaptive nonlinear output-feedback schemes with Marino-Tomei controller // IEEE Trans. Automat. Contr. – 1996. – V. 41. – № 2. – P. 274–280.
19. Marino R., Tomei P. Robust adaptive regulation of linear time-varying systems // IEEE Trans. Automat. Contr. – 2000. – V. 45. – № 7. – P. 1301–1311.
20. Marino R., Tomei P. Adaptive control of linear time-varying systems // Automatica. – 2003. – V. 39. – № 4. – P. 651–659.
21. Monopoli R.V. Model reference adaptive control with an augmented signal // IEEE Trans. Automat. Contr. 1974. – V. 19. – № 5. – P. 474–484.

22. Morse A.S., Isidori A. and T.J. Tarn. High-order parameter tuner for the adaptive control of linear and nonlinear systems // Systems, Models and Feedback: Theory and Applications. – Basel: Birkhauser. – 1992. – P. 339–364.
23. Narendra K.S., Annaswamy A.M. Stable Adaptive Systems. – Englewood Cliffs, NJ: Prentice-Hall, 1989. – 236 p.
24. Narendra K.S., Valavani L.S. Stable adaptive controller design – direct control // IEEE Trans. Automat. Contr. – 1978. – V. 23. – № 4. – P. 570–583.
25. Nikiforov V.O. Adaptive non-linear tracking with complete compensation of unknown disturbances // European Journal of Contr. – 1998. – V. 4. – № 2. – P. 132–139.
26. Nikiforov V.O. Robust high-order tuner of simplified structure // Automatica. –1999. – V. 35. – № 8. – P. 1409–1415.
27. Nikiforov V.O., Voronov K.V. Nonlinear adaptive controller with integral action // IEEE Trans. Automat. Contr. – 2001. – V. 46. – № 12. – P. 2035–2037.
28. Qian C. and Lin W. Output feedback control of a class of nonlinear systems: a nonseparation principle paradigm // IEEE Trans. Automat. Contr. – 2002. – V. 47. – № 10. – P. 1710–1715.
29. Tsakalis K.S., Ioannou P.A. Adaptive control of linear time-varying plants // Automatica. – 1987. – V. 23. – № 4. – P. 459–468.
30. Tsakalis K.S., Ioannou P.A. Adaptive control of linear time-varying plants: A new model reference controller structure // IEEE Trans. Automat. Contr. – 1989. – V. 34. – № 10. – P. 1038–1047.
31. Tsakalis K.S., Ioannou P.A. A new indirect adaptive control scheme for time-varying plants // IEEE Trans. Automat. Contr. – 1990. – V. 35. – № 6. – P. 697–705.
32. Tsakalis K.S., Ioannou P.A. Linear time varying systems: control and adaptation. – Upper Saddle River, NJ: Prentice-Hall, 1993. – 270 p.
33. Zang Z., Bitmead R.R. Transient bounds for adaptive control systems // IEEE Trans. Automat. Contr. – 1994. – V. 39. – № 1. – P. 171–175.
34. Zhang Y., Fidan B., Ioannou P.A. Backstepping control of linear time-varying systems with known and unknown parameters // IEEE Trans. Automat. Contr. – 2003. – V. 48. – № 11. – P. 1908–1925.
35. Arcak M., Kokotovic P. Feasibility conditions for circle criterion design // Systems and Control Letters. – 2001. – V. 42. – № 5. – P. 405–412.
36. Arcak M., Larsen M., Kokotovic P. Circle and Popov criteria as tools for nonlinear feedback design // Automatica. – 2003. – V. 39. – № 4. – P 643–650.
37. Byrnes C.I., Isidori A. Asymptotic stabilization of minimum phase nonlinear systems // IEEE Trans. Automat. Contr. – 1991. – V. 36. – № 10. – P. 1122–1137.
38. Isidori A. A remark on the problem of semiglobal nonlinear output regulation // IEEE Trans. Automat. Contr. – 1997. – V. 42. – № 12. – P. 1734–1738.
39. Isidori A. A tool for semi-global stabilization of uncertain non-minimum-phase nonlinear systems via output feedback // IEEE Trans. Automat. Contr. – 2000. – V. 45. – № 10. – P. 1817–1827.
40. Isidori A., Byrnes C.I. Output regulation of nonlinear systems // IEEE Trans. Automat. Contr. – 1990. – V. 35. – № 2. – P. 131–140.
41. Земляков С.Д., Рутковский В.Ю., Силаев А.В. Настраиваемая функциональная работоспособность адаптивных систем с эталонной моделью // Автоматика и телемеханика. – 1997. – № 6. – С. 125–134.
42. Фрадков А.Л. Синтез адаптивной системы стабилизации линейного динамического объекта // Автоматика и телемеханика. – 1974. – № 12. – С. 96–103.
43. Фрадков А.Л. Квадратичные функции Ляпунова в задаче адаптивной стабилизации линейного динамического объекта // Сиб. мат. журн. – 1976. – № 2. – С. 436–446.
44. Fradkov A.L. Passification of nonsquare linear systems and Yakubovich-Kalman-Popov Lemma // European Journal of Control. – 2003. – № 6. – P. 573–582.
45. Бобцов А.А., Николаев Н.А. Синтез управления нелинейными системами с функциональными и параметрическими неопределенностями на основе теоремы Фрадкова // Автоматика и телемеханика. 2005. – № 1. – С. 118–129.
46. Гелиг А.Х., Леонов Г.А., Якубович В.А. Устойчивость нелинейных систем с неединственным состоянием равновесия. – М.: Наука, 1978. – 400 с.
47. Никифоров В.О. Адаптивное управления без измерения производных выходного сигнала. Часть I. Процедура синтеза // Изв. вузов. Приборостроение. – 1996. – № 8–9. – С. 50–56; Часть II. Результаты моделирования // Изв. вузов. Приборостроение. – 1997. – № 4. – С. 28–33.
48. Kokotovic P., Murat A. Constructive nonlinear control: a historical perspective // Automatica. – 2001. – V. 37. – № 5. – P 637–662.
49. Бобцов А.А. Алгоритмы адаптивного управления нелинейными динамическими объектами с неопределенностями по входу // Известия РАН. Теория и системы управления. – 2003. – № 1. – С. 35–39.
50. Первозванский А.А. Курс теории автоматического управления. – М.: Наука, 1986. – 615 с.

51. Попов В.М. Гиперустойчивость автоматических систем. – М.: Наука, 1970. – 453 с.
52. Bobtsov A.A., Efimov D.V. On robustness property of dynamical systems feedback connection with respect to multiplicative disturbances // IFAC Symposium NOLCOS 2004. – Stuttgart, Germany, 2004. – P. 1457–1462.
53. Полушин И.Г., Фрадков А.Л., Хилл Д.Д. Пассивность и пассификация нелинейных систем. Обзор // Автоматика и телемеханика. – 2000. – № 3. – С. 3–37.
54. Fradkov A., Hill D. Exponential feedback passivity and stabilizability of nonlinear systems // Automatica. – 1998. – V. 34. – № 6. – P. 697–703.
55. Колесников А.А. Синергетическая теория управления. – М.: Энергоатомиздат, 1994. – 344 с.
56. Мирошник И.В., Бобцов А.А. Линейные системы автоматического управления. – СПб: СПб ГИТМО (ТУ), 2000. – 245 с.
57. Белецкий В.В. О либрации спутника // Сборник «Искусственные спутники Земли». – М.: АН СССР, 1959. – № 3. – С. 13–31.
58. Белецкий В.В. Движение искусственного спутника относительно центра масс. – М.: Наука, 1965. – 416 с.
59. Белецкий В.В. Движение спутника относительно центра масс в гравитационном поле. – М.: МГУ, 1975. – 308 с.
60. Андриевский Б.Р., Фрадков А.Л. Управление хаосом: методы и приложения. Часть 2. Приложения // Автоматика и телемеханика. – 2004. – № 4. – С. 3–34.
61. Бобцов А.А. Адаптивное и робастное управление неопределенными системами по выходу. – СПб: Наука, 2011. – 174 с.
62. Амоскин И.В., Арановский С.В., Бобцов А.А., Николаев Н.А. Адаптивная стабилизация хаоса в цепи Чуа // Изв. вузов. Приборостроение. – 2005. – № 12. – С. 8–13.
63. Mascolo S., Grassi G. Controlling chaotic dynamics using backstepping design with application to the Lorenz system and Chua's circuit // International Journal of Bifurcation and Chaos. – 1999. – V. 9. – № 7. – P. 1425–1434.
64. Yang T., Chua L.O. Generalized synchronization of chaos via linear transformations // International Journal of Bifurcation and Chaos. – 1999. – V. 9. – № 1. – P. 215–219.
65. Spong M.W., Vidyasagar M. Robot dynamics and control. – New York: Wiley, 1989. – 336 p.
66. Nguang S.K. Robust stabilization of a class of time-delay nonlinear systems // IEEE Trans. Automat. Contr. 2000. – V. 45. – № 4. – P. 756–762.
67. Герасимов Д.Н., Мигуш С.А., Никифоров В.О. Разработка математической модели инжекторного двигателя // Научно-технический вестник СПбГУ ИТМО. – 2003. – Вып. 10. – С. 10–18.
68. Герасимов Д.Н., Джавахериан Х., Ефимов Д. В., Никифоров В.О. Инжекторный двигатель как объект управления. Часть I: Схема двигателя и синтез математической модели // Известия РАН. Теория и системы управления. – 2010. – № 5. – С. 125–137; Часть II: Задачи автоматического управления двигателем // Известия РАН. Теория и системы управления. – 2010. – № 6. – С. 170–181.
69. Azzoni P., Moro D., Ponti F. Engine and load torque estimation with application to electronic throttle control. – SAE technical paper, 1998. – № 980795.
70. Hendricks E., Chevalier A., Jensen M. Event based engine control: practical problems and solutions. – SAE paper. – 1995. – № 950008.
71. Hendricks E., Jensen M., Chevalier A., Vesterholm T. Conventional event based engine control. – SAE paper. – 1994. – № 940377.
72. Kim Y.-W., Rizzoni G., V. Utkin Automotive engine diagnostics and control via nonlinear estimation // IEEE Control Systems. – 1998. – V. 18. – № 5. – P. 84–99.
73. Stotsky A., Kolmanovsky I. Application of input estimation techniques to charge estimation and control in automotive engines // Control Engineering Practice. – 2002. – № 10. – P. 1371–1383.

Бобцов Алексей Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, декан, bobtsov@mail.ifmo.ru

Никифоров Владимир Олегович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, проректор, nikiforov@mail.ifmo.ru

1 СОВРЕМЕННЫЕ ПРОБЛЕМЫ ТЕОРИИ СИСТЕМ УПРАВЛЕНИЯ

ОТ РЕДАКЦИИ

В марте 2013 года исполняется 50 лет проректору НИУ ИТМО, главному редактору журнала «Научно-технический вестник информационных технологий, механики и оптики», доктору технических наук, профессору Владимиру Олеговичу Никифорову. Коллеги Владимира Олеговича и редакция журнала поздравляют его с юбилеем и желают дальнейших творческих успехов! В данной рубрике журнала собраны научные работы, любезно предоставленные редакции коллегами и учениками Владимира Олеговича.

УДК 681.5.015

УПРАВЛЕНИЕ НЕЛИНЕЙНЫМ ПАРАМЕТРИЧЕСКИ НЕ ОПРЕДЕЛЕННЫМ ОБЪЕКТОМ С ВХОДНЫМ ЗАПАЗДЫВАНИЕМ [1]

А.А. Бобцов, А.А. Пыркин

Рассматривается задача стабилизации неустойчивого нелинейного параметрически не определенного объекта с входным запаздыванием. Предложен подход, позволяющий провести идентификацию параметров объекта и построить алгоритм стабилизации с предиктором переменных состояния.

Ключевые слова: управление в условиях запаздывания, нелинейные системы, идентификация, предиктор.

Введение. Постановка задачи

Работа посвящена анализу и синтезу методов управления нелинейными параметрически не определенными системами, содержащими запаздывание в управляющем сигнале. Данную проблему можно отнести к фундаментальным задачам теории систем, которые до сих пор не нашли универсальных методов решения. Большое количество результатов получено на сегодняшний день для различного типа систем [1–12]. В частности, для линейных устойчивых систем с неизвестными параметрами решена задача слежения за эталонным сигналом [11]. Найдены изящные решения стабилизации неустойчивых линейных объектов [1], а также получено расширение этой задачи на случай неопределенного синусоидального возмущающего воздействия [12]. Однако, насколько известно авторам, разработка методов управления для параметрически не определенных нелинейных систем представляет существенные трудности. В этой работе будем рассматривать нелинейный стационарный объект управления вида

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t-D) + \mathbf{\Psi}(y), \quad y(t) = \mathbf{C}\mathbf{x}(t), \tag{1}$$

где $\mathbf{x} \in \mathbf{R}^n$ – измеряемый вектор переменных состояния; $u(t)$ – скалярная входная переменная; $y(t)$ – скалярная выходная переменная; $D \geq 0$ – известное постоянное запаздывание; $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ – матрицы соответствующих размерностей, содержащие неизвестные параметры; $\mathbf{\Psi}(y) = col\{\psi_1(y(t-\tau_1)), \psi_2(y(t-\tau_2)), ..., \psi_n(y(t-\tau_n))\}$ – известная нелинейная функция; τ_i – положительные константы, причем $\tau_i \geq D$ для всех $i = \overline{1,n}$. Здесь и далее будем полагать, что $u(t-D) = 0$ при $t < D$. Допуская, что нелинейность $\mathbf{\Psi}(y)$ дифференцируема по аргументу $y(\cdot)$ с соответствующим смещением по времени хотя бы $(n-1)$ раз, зададимся вопросом поиска такой функции $u(t)$, чтобы было выполнено условие $\lim\limits_{t\to\infty} y(t) = 0$.

В качестве инструментария для решения указанной задачи будем использовать так называемый предиктор Смита [4] и его расширение на неустойчивые системы, предложенное, в том числе, в [1, 13–15]. В [1] была доказана экспоненциальная устойчивость замкнутой системы с предиктором с помощью аппарата функций Ляпунова, что является крайне полезным при решении поставленной в этой работе задачи.

Предварительные результаты

Из классической теории управления известно, что для системы вида (1) в случае выполнения условий управляемости и при известных параметрах, а также для $\mathbf{\Psi}(y) = 0$ и $D = 0$, можно синтезировать закон управления вида

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

$$u(t) = \mathbf{K}\mathbf{x}(t)\,, \tag{2}$$

где вектор $\mathbf{K}$ таков, что матрица состояния замкнутой системы $\mathbf{A}+\mathbf{B}\mathbf{K}$ является гурвицевой, т.е. все ее собственные числа имеют отрицательную вещественную часть.

Для случая $D>0$ закон управления (2) можно переписать в виде

$$u(t) = \mathbf{K}\mathbf{x}(t+D)\,, \tag{3}$$

где $\mathbf{x}(t+D)$ – значение вектора $\mathbf{x}(t)$ через временной интервал D.

Закон управления вида (3) нереализуем в явном виде, так как вектор $\mathbf{x}(t+D)$ недоступен для прямого измерения. Однако, следуя [1], вектор $\mathbf{x}(t+D)$ можно рассчитать следующим образом:

$$\mathbf{x}(t+D) = e^{\mathbf{A}(t+D)}\mathbf{x}(0) + \int_0^{t+D} \mathrm{e}^{\mathbf{A}(t+D-\tau)}\mathbf{B}u(\tau-D)d\tau =$$

$$= e^{\mathbf{A}D}e^{\mathbf{A}t}\mathbf{x}(0) + e^{\mathbf{A}D}\int_0^{t} e^{\mathbf{A}(t-\tau)}\mathbf{B}u(\tau-D)d\tau + \int_t^{t+D} e^{\mathbf{A}(t+D-\tau)}\mathbf{B}u(\tau-D)d\tau = e^{\mathbf{A}D}\mathbf{x}(t) + \int_{t-D}^{t} e^{\mathbf{A}(t-\tau)}\mathbf{B}u(\tau)d\tau\,.$$

Тогда алгоритм управления, обеспечивающий стабилизацию неустойчивых систем с запаздыванием в канале управления, примет вид

$$u(t) = \mathbf{K}e^{\mathbf{A}D}\mathbf{x}(t) + \mathbf{K}\int_{t-D}^{t} e^{\mathbf{A}(t-\tau)}\mathbf{B}u(\tau)d\tau\,. \tag{4}$$

Теперь, базируясь на результатах данного раздела, рассмотрим решение задачи стабилизации объекта (1).

Основной результат

Будем полагать, что объект управления (1) записан в виде системы дифференциальных уравнений

$$\begin{aligned} &\dot{x}_1(t) = x_2(t) + \psi_1(y(t-\tau_1)) + \theta_1 y(t),\\ &\ldots\\ &\dot{x}_n(t) = u(t-D) + \psi_n(y(t-\tau_n)) + \theta_n y(t),\\ &y(t) = x_1(t), \end{aligned} \tag{5}$$

где θ_i – неизвестные постоянные параметры.

Введем в рассмотрение n линейных фильтров первого порядка для каждой переменной состояния и один фильтр для запаздывающего сигнала управления:

$$\dot{\xi}_i(t) = -\lambda\,\xi_i(t) + \lambda\,x_i(t), \quad i=\overline{1,n}\,, \tag{6}$$

$$\dot{\varphi}_i(y(t-\tau_i)) = -\lambda\,\varphi_i(y(t-\tau_i)) + \lambda\,\psi_i(y(t-\tau_i)), \quad i=\overline{1,n}\,, \tag{7}$$

$$\dot{\xi}_u(t-D) = -\lambda\,\xi_u(t-D) + \lambda\,u(t-D)\,, \tag{8}$$

где $\lambda>0$ – положительный параметр фильтров.

После прямого и обратного преобразования Лапласа в (5) с учетом (6)–(8) получим следующую систему уравнений:

$$\begin{aligned} &\dot{\xi}_1(t) = \xi_2(t) + \varphi_1(y(t-\tau_1)) + \theta_1\xi_1(t) + \varepsilon_1(t),\\ &\ldots\\ &\dot{\xi}_n(t) = \xi_u(t-D) + \varphi_n(y(t-\tau_n)) + \theta_n\xi_1(t) + \varepsilon_n(t), \end{aligned} \tag{9}$$

где $\varepsilon_i(t)$ – экспоненциально затухающие функции времени.

На основе (9) несложно построить алгоритм идентификации неизвестных параметров:

$$\dot{\hat{\theta}}_i(t) = k_i\xi_1(t)\left(\dot{\xi}_i(t) - \xi_{i+1}(t) - \varphi_i(t-\tau_i) - \hat{\theta}_i(t)\,\xi_1(t)\right),\ i=\overline{1,n-1}\,,\ k_i>0\,, \tag{10}$$

$$\dot{\hat{\theta}}_n(t) = k_n\xi_1(t)\left(\dot{\xi}_n(t) - \xi_u(t-D) - \varphi_n(t-\tau_n) - \hat{\theta}_n(t)\,\xi_1(t)\right),\ k_n>0\,. \tag{11}$$

Утверждение. Алгоритм адаптации (10), (11) обеспечивает сходимость оценок параметров $\hat{\theta}_i$ к истинным значениям θ_i.

Доказательство. Рассмотрим ошибки оценивания параметров

$$\tilde{\theta}_i = \theta_i - \hat{\theta}_i\,.$$

Дифференцируя $\tilde{\theta}_i$ и подставляя уравнения (9)–(11), получим

$$\dot{\tilde{\theta}}_i = -k_i\xi_1^{\,2}\tilde{\theta}_i + k_i\xi_1\varepsilon_i\,,\ i=\overline{1,n}\,,\ k_i>0\,. \tag{12}$$

Из (12) нетрудно показать, что все ошибки оценивания $\widehat{\theta}_i$ стремятся к нулю, что гарантирует сходимость оценок $\widehat{\theta}_i$ к истинным значениям параметров объекта управления θ_i, что и требовалось доказать.

Замечание 1. Следует отметить, что за счет увеличения коэффициентов $\lambda > 0$ и $k_i > 0$ можно увеличивать скорость параметрической сходимости.

Теперь продифференцируем переменную $y(t) = x_1(t)$ n раз, последовательно проводя замены переменных:

$$\begin{aligned}
&\dot{y}(t) = \dot{\varsigma}_1(t) = \varsigma_2(t),\\
&\dot{\varsigma}_2(t) = \varsigma_3(t),\\
&\ldots\\
&\dot{\varsigma}_n(t) = u(t-D) + \frac{\partial^{n-1}\psi_1(y(t-\tau_1))}{\partial y(t-\tau_1)^{n-1}} y^{(n-1)}(t-\tau_1) +\\
&\ldots + \psi_n(y(t-\tau_n)) + \theta_1 y^{(n-1)}(t) + \ldots + \theta_n y(t).
\end{aligned} \tag{13}$$

Выберем управление следующим образом:

$$u(t) = u_1(t) - \left(\frac{\partial^{n-1}\psi_1(y(t+D-\tau_1))}{\partial y(t+D-\tau_1)^{n-1}} y^{(n-1)}(t+D-\tau_1) + \ldots + \psi_n(y(t+D-\tau_n)) \right). \tag{14}$$

Подставляя (14) в уравнение (13), получаем

$$\begin{aligned}
&\dot{\varsigma}_1(t) = \varsigma_2(t),\\
&\dot{\varsigma}_2(t) = \varsigma_3(t),\\
&\ldots\\
&\dot{\varsigma}_n(t) = u_1(t-D) + \theta_1 y^{(n-1)}(t) + \ldots + \theta_n y(t).
\end{aligned} \tag{15}$$

Таким образом, получаем линейную стационарную систему. Теперь перепишем (15) в матричном виде:

$$\dot{\boldsymbol{\varsigma}}(t) = \mathbf{G}\boldsymbol{\varsigma}(t) + \mathbf{q}u_1(t-D),$$

$$y(t) = \mathbf{h}^T\boldsymbol{\varsigma}(t),$$

где $\boldsymbol{\varsigma}(t) = \begin{bmatrix} \varsigma_1(t) \\ \varsigma_2(t) \\ \vdots \\ \varsigma_n(t) \end{bmatrix}$, $\mathbf{G} = \begin{bmatrix} 0 & 1 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_n & \theta_{n-1} & \ldots & \theta_1 \end{bmatrix}$, $\mathbf{q} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$, $\mathbf{h}^T = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

Закон управления $u_1(t)$ построим на основе алгоритма (4)

$$u_1(t) = \widehat{\mathbf{K}}(t)e^{\hat{\mathbf{G}}(t)D}\boldsymbol{\varsigma}(t) + \widehat{\mathbf{K}}(t)\int_{t-D}^{t} e^{\hat{\mathbf{G}}(t)(t-\tau)}\mathbf{q}u(\tau)d\tau, \tag{16}$$

где вектор-строка $\widehat{\mathbf{K}}(t)$ определяется из условия гурвицевости матрицы $\mathbf{F} = \hat{\mathbf{G}}(t) + \mathbf{q}\widehat{\mathbf{K}}(t)$ в каждый момент времени.

Поскольку оценки $\widehat{\theta}_i$ сходятся к истинным значениям, то для матрицы $\hat{\mathbf{G}}(t)$ справедливо $\lim_{t\to\infty}(\mathbf{G} - \hat{\mathbf{G}}(t)) = 0$; следовательно, закон управления (16) обеспечивает стабилизацию системы (13) и достижение цели $\lim_{t\to\infty} y(t) = 0$. Таким образом, получен алгоритм стабилизирующего управления (6)–(8), (14), (16) для параметрически не определенного нелинейного объекта управления (1) с постоянным входным запаздыванием.

Замечание 2. Для понимания процедуры настройки вектора $\widehat{\mathbf{K}}$, обеспечивающего гурвицевость матрицы $\mathbf{F} = \hat{\mathbf{G}} + \mathbf{q}\widehat{\mathbf{K}}$, рассмотрим следующий частный случай. Пусть динамический порядок объекта управления равен трем, тогда вектор-строка $\widehat{\mathbf{K}} = \begin{bmatrix} \widehat{K}_1 & \widehat{K}_2 & \widehat{K}_3 \end{bmatrix}$. Рассмотрим характеристический полином $Q(p)$ матрицы $\mathbf{F} = \hat{\mathbf{G}} + \mathbf{q}\widehat{\mathbf{K}}$

$$Q(p) = \det\left[p\mathbf{I} - (\hat{\mathbf{G}} + \mathbf{q}\widehat{\mathbf{K}}) \right]^{-1} = \det\left(\begin{bmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \widehat{\theta}_3 + \widehat{K}_1 & \widehat{\theta}_2 + \widehat{K}_2 & \widehat{\theta}_1 + \widehat{K}_3 \end{bmatrix} \right) =$$

$$= p^3 - (\hat{\theta}_1 + \hat{K}_3)p^2 - (\hat{\theta}_2 + \hat{K}_2)p - (\hat{\theta}_3 + \hat{K}_1),$$

где p – комплексная переменная.

Назначим желаемый характеристический полином вида

$$Q^*(p) = p^3 + 3\omega p^2 + 3\omega^2 p + \omega^3,$$

где $\omega > 0$ – положительный параметр, который определяет быстродействие системы.

Тогда параметры закона управления $\hat{\mathbf{K}} = \begin{bmatrix} \hat{K}_1 & \hat{K}_2 & \hat{K}_3 \end{bmatrix}$ найдем, приравняв коэффициенты полиномов $Q(p)$ и $Q^*(p)$:

$$\hat{K}_3 = -(3\omega + \hat{\theta}_1),\ \hat{K}_2 = -(3\omega^2 + \hat{\theta}_2),\ \hat{K}_1 = -(\omega^3 + \hat{\theta}_3).$$

Очевидно, что подобную процедуру без труда можно проделать для любого динамического порядка объекта управления.

Заключение

В работе рассмотрено решение задачи управления в условиях постоянного запаздывания в управляющем сигнале для класса параметрически не определенных нелинейных систем, описываемых дифференциальными уравнениями вида (1), (5). Получен алгоритм управления вида (6)–(8), (14), (16), позволяющий достигать целевого условия $\lim_{t\to\infty} y(t) = 0$. Развитие данного подхода видится как расширение для случая управления исключительно по выходной переменной $y(t)$, а также для парирования возмущающих воздействий.

Литература

1. Krstic M. Delay compensation for nonlinear, adaptive, and PDE systems. – Birkhauser, 2009. – 466 p.
2. Резван В. Абсолютная устойчивость автоматических систем с запаздыванием. – М.: Наука, 1997. – 216 с.
3. Цыкунов А.М. Адаптивное управление объектами с последействием. – М.: Наука, 1984. – 245 с.
4. Smith O.J.M. A controller to overcome dead time // ISA. – 1959. – V. 6. – P. 28–33.
5. Niculescu S.I., Annaswamy A.M. An adaptive Smith-controller for time-delay systems with relative degree $n \le 2$ // Systems & Control Letters. – 2003. – V. 49. – P. 347–358.
6. Цыпкин Я.З. Устойчивость систем с запаздывающей обратной связью // Автоматика и телемеханика. – 1947. – Т. 7. – № 2, 3. – С. 107–129.
7. Цыкунов А.М. Управление объектами с последействием. – Фрунзе: Илим, 1985. – 108 с.
8. Цыкунов А.М. Адаптивное и робастное управление динамическими объектами по выходу. – М.: ФИЗМАТЛИТ, 2009. – 268 с.
9. Солодовников В.В., Филимонов А.Б. Конструирование регуляторов для объектов с запаздыванием // Техническая кибернетика. – 1979. – № 1. – С. 168–177.
10. Солодовников В.В., Филимонов А.Б. Упреждающее управление линейными стационарными объектами с запаздыванием // Автоматика и телемеханика. – 1982. – № 11. – С. 57–60.
11. Паршева Е.А., Цыкунов А.М. Адаптивное управление объектом с запаздывающим управлением со скалярным входом-выходом // Автоматика и телемеханика. – 2001. – № 1. – С. 142–149.
12. Anton Pyrkin, Andrey Smyshlyaev, Nikolaos Bekiaris-Liberis, Miroslav Krstic, Rejection of Sinusoidal Disturbance of Unknown Frequency for Linear System with Input Delay // American Control Conference. – Baltimore, 2010. – P. 5688–5693.
13. Manitius A.Z., Olbrot A.W. Finite spectrum assignment for systems with delays // IEEE Trans. Autom. Control. – 1979. – V. 24. – P. 541–553.
14. Kwon W.H., Pearson A.E. Feedback stabilization of linear systems with delayed control // IEEE Trans. Autom. Control. – 1980. – V. 25. – P. 266–269.
15. Arstein Z. Linear systems with delayed controls: A reduction // IEEE Trans. Autom. Control. – 1982. – V. 27. – P. 869–879.

Бобцов Алексей Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, bobtsov@mail.ifmo.ru

Пыркин Антон Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, a.pyrkin@gmail.com

УДК 519.7

СУБОПТИМАЛЬНОЕ УПРАВЛЕНИЕ НЕЛИНЕЙНЫМИ МУЛЬТИАГЕНТНЫМИ СИСТЕМАМИ [2]

И.Б. Фуртат

Рассмотрена проблема робастного субоптимального управления мультиагентными системами, модель которых представлена дифференциальными уравнениями с липшицевой нелинейностью. При решении предполагается, что доступны измерению только скалярные выходы подсистем. Получен алгоритм децентрализованного управления, обеспечивающий компенсацию неопределенностей и субминимизацию интегрального критерия качества с заданной точностью.

Ключевые слова: мультиагентная система, робастное управление, децентрализованное управление, оптимальное управление, компенсация возмущений.

Введение

Задача компенсации неконтролируемых возмущений была и остается актуальной проблемой в теории управления. В настоящее время решение этой задачи имеет два основных направления. Первый подход основан на построении инвариантных систем управления, т.е. когда система управления малочувствительна или не реагирует на неконтролируемые возмущения. Такой подход получил широкое применение, например, в H^{∞}-оптимизации [1] или в методе вложения систем [2]. Второй подход основан на динамической компенсации неизвестных воздействий. Суть данного подхода заключается в оценке возмущений и затем выборе структуры и параметров управляющего устройства с целью исключения влияния неопределенностей на объект управления. Так, в [3, 4] внешние возмущения представлены в виде системы дифференциальных уравнений, которые с помощью методов адаптивного и робастного управления компенсируются. В [5], на базе подхода [6], предложена схема робастного субоптимального управления с минимизаций интегрального критерия качества.

В настоящей работе рассматривается обобщение схемы [5] на случай робастного субоптимального децентрализованного управления нелинейными мультиагентными системами по выходу. Для выделения неконтролируемых возмущений и обеспечения субоптимального управления параллельно объекту предлагается ввести вспомогательный контур определенной структуры. Далее возмущения оцениваются и компенсируются с заданной точностью за конечное время. Приводятся результаты моделирования, подтверждающие аналитические выводы и расчеты.

Постановка задачи

Пусть управляемая и наблюдаемая мультиагентная система с перекрестными связями по состоянию задана дифференциальным уравнением

$$\begin{aligned}&\dot{\mathbf{x}}_i(t)=\mathbf{A}_i(t)\mathbf{x}_i(t)+\mathbf{B}_i(t)u_i(t)+\mathbf{\Psi}_i\left(y_i(t),t\right)\mathbf{N}_i(t)+\sum_{j=1,i\neq j}^{k}\mathbf{S}_{ij}(t)\mathbf{x}_j(t)+\mathbf{D}_i(t)f_i(t),\\&y_i(t)=\mathbf{L}_i\mathbf{x}_i(t),\quad \mathbf{x}_i(0)=\mathbf{x}_{0i},\quad i=\overline{1,k},\end{aligned}\tag{1}$$

где $\mathbf{x}_i(t)\in\mathbf{R}^{n_i}$ – вектор состояния i-й подсистемы; $u_i(t)$, $f_i(t)$ и $y_i(t)$ – скалярные вход, внешнее неизвестное ограниченное воздействие и выход соответственно; $\mathbf{A}_i(t)\in\mathbf{R}^{n_i\times n_i}$, $\mathbf{N}_i(t)\in\mathbf{R}^{n_i}$, $\mathbf{S}_{ij}(t)\in\mathbf{R}^{n_i\times n_i}$, $\mathbf{\Psi}_i\left(y_i(t),t\right)\in\mathbf{R}^{n_i\times n_i}$, $\mathbf{B}_i(t)\in\mathbf{R}^{n_i}$, $\mathbf{D}_i(t)\in\mathbf{R}^{n_i}$, $\mathbf{L}_i=[1, 0, \ldots, 0]$; $\boldsymbol{x}_{0i}$ – известные начальные условия; k – количество подсистем. Необходимо синтезировать непрерывный закон управления, обеспечивающий перевод объекта (1) из начального положения $y_i(0)$ в конечное $y_i(t_{fi})$ за заданное время t_{fi}, при этом минимизируя критерий качества

$$J_i=\int_0^{t_{fi}}\left(\tilde{q}_i y_i^2(t)+\tilde{r}_i u_{0i}^2(t)\right)dt,\tag{2}$$

с малой погрешностью [5], $\tilde{q}_i\geq 0$ и $\tilde{r}_i>0$ – весовые коэффициенты, функция $u_{0i}(t)$ определяет оптимальный закон управления [2].

Предположение 1. Элементы матриц $\mathbf{A}_i(t)$, $\mathbf{B}_i(t)$, $\mathbf{N}_i(t)$, $\mathbf{S}_{ij}(t)$ и $\mathbf{D}_i(t)$ – неизвестные ограниченные функции. Известно множество Ξ возможных значений этих функций.

Предположение 2. Выполнены условия: $\mathbf{A}_i(t)=\mathbf{A}_{Ni}+\mathbf{B}_{Ni}\mathbf{c}_i^T(t)$, $\mathbf{B}_i(t)=\mathbf{B}_{Ni}+\mathbf{B}_{Ni}\tau_i(t)$, $\mathbf{D}_i(t)=\mathbf{B}_{Ni}k_i(t)$, $\mathbf{N}_i(t)=\mathbf{B}_{Ni}\omega_i(t)$, $\mathbf{S}_{ij}(t)=\mathbf{B}_{Ni}\boldsymbol{\theta}_{ij}^T(t)$, где $\mathbf{A}_{Ni}\in\mathbf{R}^{n_i\times n_i}$, $\mathbf{B}_{Ni}\in\mathbf{R}^{n_i}$ – известные матрицы, при-

[2] Работа выполнена при поддержке ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы», государственный контракт № 11.519.11.4007.

чем собственные числа $\mathbf{A}_{Ni}$ не лежат в правой полуплоскости комплексной плоскости, $\mathbf{c}_i(t) \in \mathbf{R}^{n_i}$, $\tau_i(t) \in \mathbf{R}$, $k_i(t) \in \mathbf{R}$, $\omega_i(t) \in \mathbf{R}$ и $\theta_{ij}(t) \in \mathbf{R}^{n_i}$ – неизвестные функции.

Предположение 3. Неизвестные элементы матрицы $\mathbf{\Psi}_i(y_i(t), t)$ удовлетворяют глобальному условию Липшица по $y_i(t)$, ограничены по t и являются гладкими функциями.

Предположение 4. Объект (1) – минимально фазовый, последние коэффициенты $\mathbf{B}_i(t)$ и $\mathbf{B}_{Ni}$ – положительные функции и число.

Предположение 5. В системе управления доступны измерению только $y_i(t)$ и $u_i(t)$.

Метод решения

С учетом предположения 2 преобразуем уравнение объекта (1) к виду

$$\dot{\mathbf{x}}_i(t) = \mathbf{A}_{Ni}\mathbf{x}_i(t) + \mathbf{B}_{Ni}u_i(t) + \mathbf{B}_{Ni}\varphi_i(t), \quad y_i(t) = \mathbf{L}_i\mathbf{x}_i(t), \quad \mathbf{x}_i(0) = \mathbf{x}_{0i}, \tag{3}$$

где $\varphi_i(t) = \mathbf{c}_i^T(t)\mathbf{x}_i(t) + \tau_i(t)u_i(t) + \mathbf{\Psi}_i\left(y_i(t), t\right)\omega_i(t) + \sum\limits_{j=1, i\neq j}^{k} \theta_{ij}^T(t)\mathbf{x}_j(t) + k_i(t)f_i(t)$.

Рассмотрим номинальный объект (когда в (3) $\varphi_i(t) \equiv 0$):

$$\dot{\mathbf{x}}_i(t) = \mathbf{A}_{Ni}\mathbf{x}_i(t) + \mathbf{B}_{Ni}u_i(t), \quad y_i(t) = \mathbf{L}_i\mathbf{x}_i(t), \quad \mathbf{x}_i(0) = \mathbf{x}_{0i}. \tag{4}$$

Для номинального объекта (4) критерий качества (2) и оптимальный закон управления $u_{0i}(t)$ определятся в виде [7]

$$J_i = \int\limits_0^{t_{fi}} \left(\mathbf{x}_i^T(t)\tilde{\mathbf{Q}}_i\mathbf{x}_i(t) + \tilde{r}_i u_{0i}^2(t)\right)dt, \quad u_{0i}(t) = \tilde{r}_i^{-1}\mathbf{B}_{Ni}^T\mathbf{z}_{22,i}(t)\mathbf{z}_{12,i}^{-1}(t)\mathbf{x}_i(t) + \bar{u}_{0i}(t), \tag{5}$$

где $\tilde{\mathbf{Q}}_i = \tilde{q}_i \mathrm{diag}\{1, 0, ..., 0\}$, $\bar{u}_{0i}(t) = \tilde{r}_i^{-1}\mathbf{B}_{Ni}^T\left(\mathbf{z}_{21,i}(t) + \mathbf{z}_{22,i}(t)\mathbf{z}_{12,i}^{-1}(t)\mathbf{z}_{11,i}(t)\right)\mathbf{x}_i(t_{fi})$, $\mathbf{z}_i(t) = \left[\mathbf{x}_i^T(t), \mathbf{w}_i^T(t)\right]^T$, $\dot{\mathbf{w}}_i(t) = \tilde{\mathbf{Q}}_i\mathbf{x}_i(t) - \mathbf{A}_{Ni}^T\mathbf{w}_i(t)$, $\dot{\mathbf{z}}_i(t) = \mathbf{D}_i\mathbf{z}_i(t)$, $\mathbf{D}_i = \begin{bmatrix} \mathbf{A}_{Ni} & \tilde{r}_i^{-1}\mathbf{B}_{Ni}\mathbf{B}_{Ni}^T \\ \tilde{\mathbf{Q}}_i & -\mathbf{A}_{Ni}^T \end{bmatrix}$, $\mathbf{Z}_i(\tau_i) = e^{\mathbf{D}_i\tau_i} = \begin{bmatrix} \mathbf{z}_{11,i}(\tau_i) & \mathbf{z}_{12,i}(\tau_i) \\ \mathbf{z}_{21,i}(\tau_i) & \mathbf{z}_{22,i}(\tau_i) \end{bmatrix}$, $\mathbf{z}_i(t) = \mathbf{Z}_i(\tau_i)\mathbf{z}_i(t_{fi})$, $\tau_i = t - t_{fi}$.

Добавим и вычтем в (3) $u_{0i}(t)$ из (5) и преобразуем уравнение (3) к виду

$$\dot{\mathbf{x}}_i(t) = \mathbf{A}_{0i}(t)\mathbf{x}_i(t) + \mathbf{B}_{Ni}\bar{\mathbf{u}}_{0i}(t) + \mathbf{B}_{Ni}u_i(t) + \mathbf{B}_{Ni}\varphi_{1i}(t), \quad y_i(t) = \mathbf{L}_i\mathbf{x}_i(t), \quad \mathbf{x}_i(0) = \mathbf{x}_{0i}, \tag{6}$$

где $\mathbf{A}_{0i}(t) = \mathbf{A}_{Ni} + \tilde{\mathbf{r}}_i^{-1}\mathbf{B}_{Ni}\mathbf{B}_{Ni}^T\mathbf{z}_{22,i}(t)\mathbf{z}_{12,i}^{-1}(t)$, $\mathbf{\Psi}(t) = \varphi(t) - u_{0i}(t)$.

Для выделения этих неопределенностей введем вспомогательный контур [5, 6]

$$\dot{\mathbf{x}}_{ai}(t) = \mathbf{A}_{0i}(t)\mathbf{x}_{ai}(t) + \mathbf{B}_{Ni}u_{0i}(t) + \beta_i\mathbf{B}_{Ni}u_i(t), \quad y_{ai}(t) = \mathbf{L}_i\mathbf{x}_{ai}(t), \quad \mathbf{x}_{ai}(0) = \mathbf{x}_{0i}, \tag{7}$$

где $\beta_i > 0$. Составим функцию $\boldsymbol{\sigma}_i(t) = \boldsymbol{x}_i(t) - \boldsymbol{x}_{ai}(t)$, вычитая из (6) уравнение (7):

$$\dot{\boldsymbol{\sigma}}_i(t) = \mathbf{A}_{0i}(t)\boldsymbol{\sigma}_i(t) + \mathbf{B}_{Ni}\varphi_i(t), \quad \zeta_i(t) = \mathbf{L}_i\boldsymbol{\sigma}_i(t), \quad \boldsymbol{\sigma}_i(0) = 0. \tag{8}$$

Здесь $\boldsymbol{\sigma}_i(t) \in \mathbf{R}^{n_i}$, $\varphi_i(t) = \left(1 - \beta_i\right)u_i(t) + \mathbf{\Psi}_i(t)$. Преобразуем (8) к форме вход–выход:

$$Q_{0i}(p, t)\zeta_i(t) = R_{Ni}(p)\varphi_i(t), \tag{9}$$

где $Q_{0i}(p, t)$, $R_{Ni}(p)$ – линейные дифференциальные операторы, полученные при переходе от (8) к (9) при фиксированном параметре t.

Для компенсации возмущений в (1) функцию $u_i(t)$ зададим в виде

$$u_i(t) = -\beta_i^{-1}R_{Ni}^{-1}(p)Q_{0i}(p, t)\bar{\zeta}_i(t), \tag{10}$$

где $\bar{\zeta}_i(t)$ – оценка $\zeta_i(t)$. Для реализации алгоритма (10) рассмотрим наблюдатель [8]

$$\dot{\boldsymbol{\xi}}_i(t) = \mathbf{G}_{0i}\boldsymbol{\xi}_i(t) + \mathbf{D}_{0i}\left(\bar{\zeta}_i(t) - \zeta_i(t)\right), \quad \bar{\zeta}_i(t) = \mathbf{L}_i\boldsymbol{\xi}_i(t). \tag{11}$$

Здесь $\boldsymbol{\xi}_i(t) \in \mathbf{R}^{n_i}$, $\mathbf{G}_{0i} = \begin{bmatrix} 0 & \mathbf{I}_{n_i-1} \\ 0 & 0 \end{bmatrix}$, $\mathbf{I}_{n_i-1}$ – единичная матрица порядка $n_i - 1$, $\mathbf{D}_{0i} = -\left[d_{1i}\mu^{-1}, d_{2i}\mu^{-2}, ..., d_{n_i i}\mu^{-n_i}\right]^T$, $d_{1i}, d_{2i}, ..., d_{n_i i}$ выбираются из условия гурвицевости матрицы $\mathbf{G}_i = \mathbf{G}_{0i} - \mathbf{D}_i\mathbf{L}_i$, $\mathbf{D}_i = \left[d_{1i}, d_{2i}, ..., d_{n_i i}\right]^T$, $\mu > 0$ – малое число.

Введем вектор отклонений $\bar{\boldsymbol{\eta}}_i(t) = \mathbf{\Gamma}_i^{-1}\left(\boldsymbol{\xi}_i(t) - \boldsymbol{\theta}_i(t)\right)$, где $\mathbf{\Gamma}_i = \mathrm{diag}\left\{\mu^{n_i-1}, \mu^{n_i-2}, ..., \mu, 1\right\}$, $\boldsymbol{\theta}_i(t) = \left[\zeta_i(t), \dot{\zeta}_i(t), ..., \zeta_i^{(n_i)}(t)\right]^T$. Продифференцировав $\bar{\boldsymbol{\eta}}_i(t)$ по времени с учетом уравнения (11), получим: $\dot{\bar{\boldsymbol{\eta}}}_i(t) = \mu^{-1}\mathbf{G}_i\bar{\boldsymbol{\eta}}_i(t) + \bar{\mathbf{b}}_i\zeta_i^{(n_i+1)}(t)$, $\bar{\Delta}_i(t) = \mu^{n_i-1}\mathbf{L}_i\bar{\boldsymbol{\eta}}_i(t)$. Преобразуем предпоследние уравнения к виду

$$\dot{\boldsymbol{\eta}}_i(t) = \mu^{-1}\mathbf{G}_i\boldsymbol{\eta}_i(t) + \mathbf{b}_i\dot{\zeta}_i(t), \quad \bar{\Delta}_i(t) = \mu^{n_i-1}\mathbf{L}_i\boldsymbol{\eta}_i(t). \tag{12}$$

Здесь $\eta_i^l(t)=\overline{\eta}_i^l(t)-\mu^{1+l-n_i}\zeta_i^{(l-1)}(t)$, $l=\overline{2,n_i}$, $\eta_i^k(t)$ и $\overline{\eta}_i^l(t)$ – l-е компоненты векторов $\boldsymbol{\eta}_i(t)$ и $\overline{\boldsymbol{\eta}}_i(t)$, $\eta_i^1(t)=\overline{\eta}_i^1(t)$, $b_i=\left[\mu^{2-n_i},0,\dots,0\right]^T$. Очевидно, что последние два уравнения эквивалентны относительно переменных $\eta_1(t)=\overline{\boldsymbol{\eta}}_i(t)$. Принимая во внимание (11) и (12), преобразуем уравнение (10) к виду

$$\dot{\mathbf{x}}_i(t)=\mathbf{A}_{0i}(t)\mathbf{x}_i(t)+\mathbf{B}_{Ni}\overline{\mathbf{u}}_{0i}(t)+\mu^{n_i-1}\overline{\mathbf{b}}_i\mathbf{g}_i^T\boldsymbol{\Delta}_i(t),\quad y_i(t)=\mathbf{L}_i\mathbf{x}_i(t), \tag{13}$$

где $\boldsymbol{\Delta}(t)=\left[\eta_1^i(t),\dot{\eta}_1^i(t),\dots,\left(\eta_1^i(t)\right)^{(n_i)}\right]^T$, $\mathbf{g}_i$ – вектор, составленный из коэффициентов оператора $Q_{0i}(p,t)$ и записанных в обратном порядке.

Утверждение. Пусть выполнены условия предположений. Тогда существует числа $\beta_i>0$ и $\mu_0>0$ такие, что при $\mu\le\mu_0$ система управления (7), (10), (11) обеспечивает переход объекта (1) из начального положения $y_i(0)$ в конечное $y_i(t_{fi})$ с субминимизацией критерия качества (2).

Пример

Рассмотрим мультиагентную систему, математическая модель которой имеет вид

$$\dot{\mathbf{x}}_i(t)=\begin{bmatrix}\mathbf{O}_{3\times1} & \mathbf{I}_3\\ -a_{0i} & -a_{1i}\quad -a_{2i}\quad -a_{3i}\end{bmatrix}\boldsymbol{x}_i(t)+\begin{bmatrix}\mathbf{O}_{3\times1}\\ r_{0i}\end{bmatrix}u_i(t)+\begin{bmatrix}\mathbf{O}_{3\times1}\\ d_{0i}\end{bmatrix}f_i(t)+$$
$$+\begin{bmatrix}\mathbf{O}_{3\times4}\\ -\varepsilon_{0i}\quad -\varepsilon_{1i}\quad -\varepsilon_{2i}\quad -\varepsilon_{3i}\end{bmatrix}\begin{bmatrix}\mathbf{O}_{3\times1}\\ n_{0i}\end{bmatrix}+\begin{bmatrix}\mathbf{O}_{3\times4}\\ -s_{0i}\quad -s_{1i}\quad -s_{2i}\quad -s_{3i}\end{bmatrix}\mathbf{x}_j(t), \tag{14}$$
$$y_i(t)=\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}\mathbf{x}_i(t),\quad i,j=1,2,\quad i\ne j,$$

где $\mathbf{O}_{i\times j}$ – матрица размерности $i\times j$ с нулевыми элементами. Класс неопределенности Ξ задан неравенствами: $|a_{li}|\le10$, $l=0,1,2,3$, $i=1,2$, $1\le r_{0i}\le4$, $|d_{0i}|\le10$, $|f_i(t)|\le10$, $|s_{li}|\le10$, $j=1,2$. Предполагаются известными начальные условия $\boldsymbol{x}_i(0)=[1,1,1,1]^T$.

Цель управления – квазиминимизация интегрального критерия качества (2) и ограниченность всех сигналов в системе управления. Зададим следующие параметры в (2): $\tilde{q}_i=1$, $\tilde{r}_i=1$, $t_{fi}=10$ с, $y_i(0)=1$ и $y_i(10)=2$.

Сформируем уравнение номинального объекта управления (4) в виде

$$\dot{\mathbf{x}}_i(t)=\begin{bmatrix}\mathbf{O}_{3\times1} & \mathbf{I}_3\\ -1 & -4\quad -6\quad -4\end{bmatrix}\mathbf{x}_i(t)+\begin{bmatrix}\mathbf{O}_{3\times1}\\ 1\end{bmatrix}u_i(t),\ y_{Ni}(t)=\begin{bmatrix}1,0,0,0\end{bmatrix}\boldsymbol{x}_i(t). \tag{15}$$

Для (15) перепишем критерий (2) и сформируем оптимальное управление в виде

$$J_i=\int_0^{10}\left(\mathbf{x}_i^T(t)\mathrm{diag}\left\{1,0,0,0\right\}\mathbf{x}(t)+u_{0i}^2(t)\right)dt,$$
$$u_{0i}(t)=\left[0,0,0,1\right]\left[\mathbf{z}_{22,i}(t)\mathbf{z}_{12,i}^{-1}(t)\mathbf{x}_i(t)+\left(\mathbf{z}_{21,i}(t)+\mathbf{z}_{22,i}(t)\mathbf{z}_{12,i}^{-1}(t)\mathbf{z}_{11,i}(t)\right)\mathbf{x}_i(10)\right].$$

Введем вспомогательный контур (7), где $\mathbf{x}_i(10)=[2,2,2,2]^T$ и $\beta_i=0{,}04$. Тогда

$$\dot{\overline{\mathbf{x}}}_i(t)=\mathbf{A}_{0i}(t)\overline{\mathbf{x}}_i(t)+\begin{bmatrix}1 & 0 & 0 & 0\\ & & \mathbf{O}_{3\times4} & \end{bmatrix}\left(\mathbf{z}_{21,i}(t)+\mathbf{z}_{22,i}(t)\mathbf{z}_{12,i}^{-1}(t)\mathbf{z}_{11,i}(t)\right)\mathbf{x}_i(10)+$$
$$+0{,}04\left[0,0,0,1\right]^T u_i(t),\quad \overline{\mathbf{y}}_i(t)=\left[1,0,0,0\right]\overline{\mathbf{x}}_i(t),\quad \overline{\mathbf{x}}_i(0)=\left[1,1,1,1\right]^T.$$

Пусть в (11) $\boldsymbol{D}_i=[20,150,500,625]^T$, $\mu=0{,}01$. В результате наблюдатель определится в виде

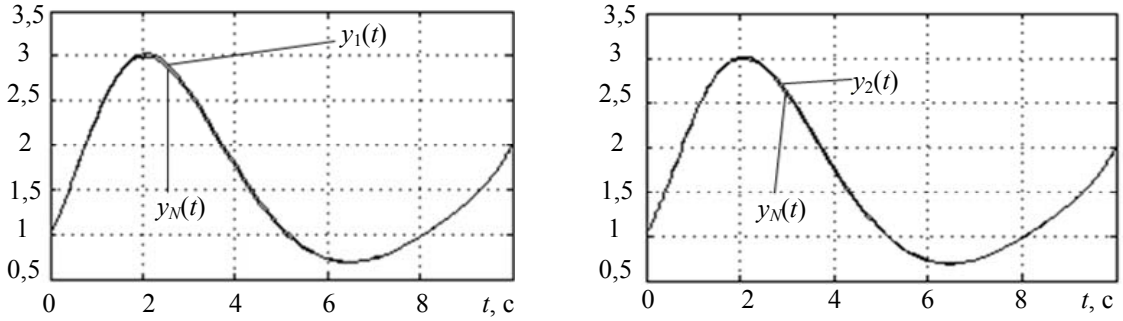
$$\begin{aligned}&\dot{\xi}_{1i}(t)=-\xi_{2i}(t)-2\cdot10^3\left(\xi_{1i}(t)-\zeta_i(t)\right),\quad \dot{\xi}_{2i}(t)=-\xi_{3i}(t)-1{,}5\cdot10^6\left(\xi_{2i}(t)-\zeta_i(t)\right),\\ &\dot{\xi}_{3i}(t)=-\xi_{4i}(t)-5\cdot10^8\left(\xi_{3i}(t)-\zeta_i(t)\right),\quad \dot{\xi}_{4i}(t)=-6{,}25\cdot10^{10}\left(\xi_{4i}(t)-\zeta_i(t)\right),\quad \boldsymbol{\xi}_i(0)=0.\end{aligned} \tag{16}$$

Используя наблюдатель (16), сигнал компенсации (10) можно записать в виде $u_i(t)=-25\mathbf{a}_{04,i}(t)\boldsymbol{\xi}_i(t)$, где $\mathbf{a}_{04,i}(t)=\left(\begin{bmatrix}-1 & -4 & -6 & -4\end{bmatrix}-\mathbf{z}_{4,i}(t)\right)$, $\mathbf{a}_{04,i}(t)$ и $\mathbf{z}_{4,i}(t)$ – четвертые строки матриц $\mathbf{A}_{0i}(t)$ и $\mathbf{z}_{22,i}(t)\mathbf{z}_{12,i}^{-1}(t)$ соответственно.

На рисунке, а, приведены результаты моделирования по выходам первой подсистемы (14) и номинального объекта (15), на рисунке, б, – результаты моделирования по выходам второй подсистемы (14) и номинального объекта (15) при следующих данных:

– в первой подсистеме: $a_{31}=8+\sin t$, $a_{21}=-6+\cos t$, $a_{11}=9+\sin 2t$, $a_{01}=8+2\cos 3t$, $r_{01}=6$, $d_{01}=2+2\sin t$, $\varepsilon_{31}=1+\sin y_1+\sin t$, $\varepsilon_{21}=\cos y_1$, $\varepsilon_{11}=1+\ln\left(1+|y_1|\right)$, $\varepsilon_{01}=\sin 2y_1$, $n_{01}=1+\sin t$, $s_{31}=5+\sin t$, $s_{21}=7$, $s_{11}=3$, $s_{01}=15$, $f_1(t)=2+\sin 0{,}8t$;

– во второй подсистеме: $a_{32} = 2 + \sin t$, $a_{22} = 6 - \cos t$, $a_{12} = -8 + 2\sin t$, $a_{02} = 5 - \cos 3t$, $r_{02} = 5$, $d_{02} = 5 + 2\sin 2t$, $\varepsilon_{32} = 2 + \sin 2y_2$, $\varepsilon_{22} = \cos y_2$, $\varepsilon_{12} = 2 + \ln\left(3 + |y_2| + \sin t\right)$, $\varepsilon_{02} = \cos 2y_2$, $n_{02} = -1 + 3\sin t$, $s_{32} = 15$, $s_{22} = 9 + \sin t$, $s_{12} = 9 - \cos t$, $s_{02} = 10 + \cos 2t$, $f_2(t) = 1 + \cos 1,2t$.



а б

Рисунок. Результаты моделирования: по $y_1(t)$ и $y_N(t)$ (а); по $y_2(t)$ и $y_N(t)$ (б)

Заключение

В работе рассмотрена схема децентрализованного робастного субоптимального управления нелинейными мультиагентными системами. При решении предполагалось, что измерению доступны только скалярные выходы локальных подсистем и запрещен обмен информации между ними. Для выделения неизвестных возмущений и обеспечения субоптимального управления предлагалось ввести вспомогательный контур параллельно объекту. Далее выделенные возмущения оценивались и компенсировались с заданной точностью за конечное время. Результаты аналитических выводов были подтверждены численным моделированием.

Приложение

Доказательство утверждения. Рассмотрим новую переменную $\boldsymbol{\varepsilon}_i(t) \in \mathbf{R}^{n_i}$, равную разности фазовых переменных (13) и (4), причем на вход номинального объекта (4) подано оптимальное управление (5). В результате получим

$$\dot{\boldsymbol{\varepsilon}}_i(t) = \mathbf{A}_{Ni}\boldsymbol{\varepsilon}_i(t) + \beta_i^{-1}\mu^{n_i-1}\bar{\mathbf{b}}_i\mathbf{g}_i\boldsymbol{\Delta}_i(t), \quad \boldsymbol{\varepsilon}_i(0) = 0 .$$

Преобразуем последнее уравнение и уравнение (12) к виду

$$\dot{\boldsymbol{\varepsilon}}_i(t) = \boldsymbol{A}_{Ni}\boldsymbol{\varepsilon}_i(t) + \mu_2^{n_i-1}\bar{\mathbf{b}}_i\mathbf{g}_i\boldsymbol{\Delta}_i(t), \quad \mu_1\dot{\boldsymbol{\eta}}_i(t) = \mathbf{G}_i\boldsymbol{\eta}_i(t) + \mu_2\mathbf{b}_i\dot{\zeta}_i(t), \tag{П.1}$$

где $\mu_1 = \mu_2 = \mu$. Воспользуемся первой леммой [9], согласно которой рассмотрим систему (П.1) при $\mu_2 = 0$. Так как собственные числа матрицы $\mathbf{A}_{Ni}$ не лежат в правой полуплоскости комплексной плоскости (предположение 2) и $\mathbf{G}_i$ гурвицева, то решение (П.1) при $\mu_2 = 0$ глобально устойчиво. Значит, согласно [9], $\boldsymbol{\eta}_i(t)$, $\boldsymbol{\varepsilon}_i(t)$, $\boldsymbol{\Delta}_i(t)$ и $\dot{\zeta}_i(t)$ ограничены. Доказательство ограниченности остальных функций аналогично [5]. В соответствии с леммой [9] система (П.1) диссипативна. Значит, $\sup\limits_t\left|\dot{\zeta}_i(t)\right| < k_1$, $\sup\limits_t\left|\boldsymbol{\Delta}_i(t)\right| < k_2$, $k_1 < \infty$, $k_2 < \infty$. Однако диссипативность (П.1) не гарантирует ее асимптотическую устойчивость как сингулярно-возмущенной системы. Покажем, что при $\mu_2 > 0$ обеспечивается малость величины $|\boldsymbol{\varepsilon}_i(t)|$. Выберем функцию Ляпунова в виде

$$V = \boldsymbol{\varepsilon}_i^T(t)\mathbf{P}_i\boldsymbol{\varepsilon}_i(t) + \boldsymbol{\eta}_i^T(t)\mathbf{H}_i\eta_i(t), \tag{П.2}$$

где $\mathbf{P}_i = \mathbf{P}_i^T > 0$, $\mathbf{H}_i = \mathbf{H}_i^T > 0$, и вычислим полную производную от нее вдоль траекторий (П.1) при $\mu_1 = \mu_2 = \mu_0$:

$$\dot{V} = -\boldsymbol{\varepsilon}_i^T(t)\mathbf{Q}_{1i}\boldsymbol{\varepsilon}_i(t) + 2\mu_0^{n_i-1}\boldsymbol{\varepsilon}_i^T(t)\mathbf{P}_i\bar{\mathbf{b}}_i\mathbf{g}_i\boldsymbol{\Delta}_i(t) - \mu_0^{-1}\boldsymbol{\eta}_i^T(t)\mathbf{Q}_{2i}\boldsymbol{\eta}_i(t) + 2\mu_0\boldsymbol{\eta}_i^T(t)\mathbf{H}_i\mathbf{b}_i\dot{\zeta}_i(t) . \tag{П.3}$$

Здесь $\mathbf{A}_{Ni}^T\mathbf{P}_i + \mathbf{P}_i\mathbf{A}_{Ni} = -\mathbf{Q}_{1i}$, $\mathbf{G}_i^T\mathbf{H}_i + \mathbf{H}_i\mathbf{G}_i = -\mathbf{Q}_{2i}$, $\mathbf{Q}_{1i} = \mathbf{Q}_{1i}^T > 0$, $\mathbf{Q}_{2i} = \mathbf{Q}_{2i}^T > 0$. Оценим в (П.3) второе и четвертое слагаемые:

$$2\mu_0^{n_i-1}\boldsymbol{\varepsilon}_i^T(t)\mathbf{P}_i\bar{\mathbf{b}}_i\mathbf{g}_i\boldsymbol{\Delta}_i(t) \le 2\mu_0^{n_i-1}\boldsymbol{\varepsilon}_i^T(t)\mathbf{P}_i\bar{\mathbf{b}}_i\mathbf{g}_i\left(\mathbf{P}_i\bar{\mathbf{b}}_i\mathbf{g}_i\right)^T\boldsymbol{\varepsilon}_i(t) + 2\mu_0^{n_i-1}k_2^2;$$

$$2\mu_0\boldsymbol{\eta}_i^T(t)\mathbf{H}_i\mathbf{b}_i\dot{\zeta}_i(t) \le 2\mu_0\boldsymbol{\eta}_i^T(t)\mathbf{H}_i\mathbf{b}_i\mathbf{b}_i^T\mathbf{H}_i\boldsymbol{\eta}_i(t) + 2\mu_0k_1^2.$$

Воспользовавшись оценками, перепишем (П.3): $\dot{V} \le -\boldsymbol{\varepsilon}_i^T(t)\mathbf{Q}_{3i}\boldsymbol{\varepsilon}_i(t) - \boldsymbol{\eta}_i^T(t)\mathbf{Q}_{4i}\boldsymbol{\eta}_i(t) + \pi_i$, где $\mathbf{Q}_{3i} = \mathbf{Q}_{1i} - 2\mu_0^{n_i-1}\mathbf{P}_i\bar{\mathbf{b}}_i\mathbf{g}_i\left(\mathbf{P}_i\bar{\mathbf{b}}_i\mathbf{g}_{ii}\right)^T$, $\mathbf{Q}_{4i} = \mathbf{Q}_{2i} - \mu_0^2\mathbf{H}_i\mathbf{b}_i\mathbf{b}_i^T\mathbf{H}_i$, $\pi_i = 2\mu_0\left(k_1^2 + \mu_0^{n_i-2}k_2^2\right)$. Очевидно, что всегда

существует число $\mu_0 > 0$, обеспечивающее $\mathbf{Q}_{3i} > 0$ и $\mathbf{Q}_{4i} > 0$. Оценим производную функции Ляпунова в виде $\dot{V} \le -\chi_i V + \pi_i$, где $\varpi_i = \min\left\{\frac{\lambda_{\min}(\mathbf{Q}_{3i})}{\lambda_{\max}(\mathbf{P}_i)}, \frac{\lambda_{\min}(\mathbf{Q}_{4i})}{\lambda_{\max}(\mathbf{H}_i)}\right\}$. Решив последнее неравенство, получим $V \le e^{-\varpi_i t}V(0) + \varpi_i^{-1}\left(1 - e^{-\varpi_i t}\right)\pi_i$. Тогда в силу структуры (П.2) $\left|\boldsymbol{\varepsilon}_i(t)\right|^2 \le \lambda_{\min}(\mathbf{P}_i)\left(e^{-\varpi_i t}V(0) + \left(1 - e^{-\varpi_i t}\right)\varpi_i^{-1}\pi_i\right)$ для $t \in [0, t_{fi}]$, причем значение правой части последнего неравенства зависит от величины μ_0. Очевидно, что уменьшением числа μ_0 можно уменьшить значение $|\boldsymbol{\varepsilon}_i(t)|$, а значит, уменьшить погрешность, возникающую при субминимизации критерия (2), что подтверждено результатами моделирования.

Литература

1. Методы классической и современной теории автоматического управления. Теория оптимизации автоматического управления / Под ред. К.А. Пупкова и Н.Д. Егупова. – М.: Издательство МГТУ им. Н.Э. Баумана. – 2004. – Т. 4. – 744 с.
2. Буков В.Н. Вложение систем. Аналитический подход к анализу и синтезу матричных систем. – Калуга: Издательство научной литературы Н.Ф. Бочкаревой, 2006. – 720 с.
3. Никифоров В.О. Адаптивное и робастное управление с компенсацией возмущений. – СПб: Наука, 2003. – 282 с.
4. Бобцов А.А. Алгоритм робастного управления неопределенным объектом без измерения производных регулируемой переменной // Автоматика и телемеханика. – 2003. – № 8. – С. 82–96.
5. Фуртат И.Б. Робастное субоптимальное управление линейными нестационарными объектами по выходу // Мехатроника, автоматизация, управление. – 2009. – № 7. – С. 7–12.
6. Цыкунов А.М. Алгоритмы робастного управления с компенсацией ограниченных возмущений // Автоматика и телемеханика. – 2007. – № 7. – С. 103–115.
7. Теория автоматического управления. Ч. 2. Теория нелинейных и специальных систем автоматического управления / Под ред. А.А. Воронова. – М.: Высшая школа, 1986. – 504 с.
8. Atassi A.N., Khalil H.K. A separation principle for the stabilization of class of nonlinear systems // IEEE Trans. Automat. Control. – 1999. – V. 44. – № 9. – P. 1672–1687.
9. Брусин В.А. Об одном классе сингулярно возмущенных адаптивных систем. 1 // Автоматика и телемеханика. – 1995. – № 4. – С. 119–127.

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

УДК 519.7

ИССЛЕДОВАНИЕ УСТОЙЧИВОСТИ ДИНАМИЧЕСКИХ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ СВОЙСТВ ПОТОКА ВЕКТОРА ФАЗОВОЙ СКОРОСТИ ЧЕРЕЗ ЗАМКНУТУЮ ВЫПУКЛУЮ ПОВЕРХНОСТЬ[3]

И.Б. Фуртат

Проведено исследование устойчивости автономных динамических систем с использованием потока вектора фазовой скорости через замкнутую выпуклую поверхность. Получены условия на знак потока вектора фазовой скорости, обеспечивающие устойчивость динамических систем.

Ключевые слова: поток векторного поля, устойчивость, вторая теорема Ляпунова.

Введение

Одним из важных этапов при исследовании дифференциальных уравнений или системы дифференциальных уравнений является выяснение характера поведения траекторий их решений. При специальных структурах дифференциальных уравнений можно найти их точное решение [1, 2]. Однако чаще встречаются уравнения, точное решение которых получить достаточно сложно [3, 4]. Тогда для выяснения характера поведения можно использовать следующие методы: метод функций Ляпунова [1–4], метод функций Четаева [1], устойчивость по Лагранжу [1], метод абсолютной устойчивости [5], дивергентные условия устойчивости [6–8] и т.д. Наиболее распространенными из используемых методов являются методы функций Ляпунова и абсолютной устойчивости. Для линейных систем использование данных методов довольно хорошо изучено в [3–5], но для нелинейных систем это зачастую трудная задача.

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

Так, при использовании метода функций Ляпунова, если не удается подобрать квадратичную функцию, то приходится линеаризовать систему в окрестности исследуемой точки и оценить влияние нелинейных членов на устойчивость по первому приближению [5]. Применение метода абсолютной устойчивости требует, чтобы нелинейная система была представлена в форме Лурье [5].

Для данной работы особый интерес представляет дивергентный подход, предложенный в [6–8]. При выяснении характера поведения решения дифференциального уравнения в окрестности точки равновесия используется знак дивергенции вектора фазовой скорости. Чтобы решение системы в окрестности точки равновесия было неустойчивым, необходима и достаточна неотрицательность дивергенции вектора фазовой скорости [6, 7], для асимптотической устойчивости необходимым условием является ее отрицательность [6, 8]. В [8] рассматриваются условия асимптотической устойчивости нелинейных динамических систем второго порядка. Но для выяснения асимптотической устойчивости автор снова прибегает к использованию аппарата квадратичных функций, соответствующих исследуемой области.

В работе рассмотрено доказательство второй теоремы Ляпунова и получены условия асимптотической устойчивости нелинейных и, как частный случай, линейных стационарных дифференциальных уравнений, представленных в векторной нормальной форме, с помощью нахождения знака потока вектора фазовой скорости через замкнутую поверхность.

Асимптотическая устойчивость нелинейных систем

Пусть динамический объект описывается векторной нелинейной системой дифференциальных уравнений в виде

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

где $\mathbf{x}(t) \subset \mathbf{X} \in \mathbf{R}^n$ – вектор состояния, $t \in T \subset R_+$; $\mathbf{X}$, T – открытые множества; $\boldsymbol{f} : \mathbf{D}_x \to \mathbf{R}^n$ – n-мерная гладкая (класса $C^n, n = \overline{1,\infty}$) вектор-функция, заданная в односвязной области $\mathbf{D}_x = \mathbf{X} \subset \mathbf{R}^n$. Также область $\mathbf{D}_x$ является окрестностью некоторой точки равновесия $\boldsymbol{x} = \boldsymbol{x}_0 \in \mathbf{X}$ системы (1). Векторному полю $\dot{\boldsymbol{x}}$ поставим в соответствие скалярное поле потока вектора $\dot{\boldsymbol{x}}$ через выпуклую ($n-1$)-мерную поверхность $S = \left\{\boldsymbol{x} \in \mathbf{R}^n : \mathbf{x}^T \mathbf{P} \mathbf{x} = 1\right\}$, где $\mathbf{P} \in \mathbf{R}^{n \times n}$ – положительно определенная матрица, обозначаемое как $\oint\limits_S \dot{\mathbf{x}}^T(t)\mathbf{N}dS$, $\mathbf{N}$ – внешний вектор нормали к поверхности S. В дальнейшем в качестве поверхности S ради простоты будем рассматривать ($n-1$)-мерную сферу радиуса $R_0 > 0$. Ниже доказана вторая теорема Ляпунова и получены условия, накладываемые на поток вектора фазовой скорости системы (1) через поверхность S.

Теорема 1. Равновесное состояние $\boldsymbol{x} = \boldsymbol{x}_0 = 0$ системы (1) асимптотически устойчиво, если в окрестности точки $\boldsymbol{x}_0 = 0$ существует функция Ляпунова $V(\boldsymbol{x})$ такая, что

$$w_1(\boldsymbol{x}) \le V(\boldsymbol{x}) \le w_2(\boldsymbol{x}), \tag{2}$$

а ее полная производная по времени является отрицательно определенной функцией

$$\dot{V}(\boldsymbol{x}) \le -w_3(\boldsymbol{x}),$$

где $w_1(\boldsymbol{x})$, $w_2(\boldsymbol{x})$ и $w_3(\boldsymbol{x})$ – положительно определенные функции.

Доказательство. Рассмотрим полную производную по времени от функции Ляпунова $V(\boldsymbol{x})$ вдоль траекторий системы (1)

$$\dot{V}(\boldsymbol{x}) = \frac{\partial V}{\partial \boldsymbol{x}} \dot{\boldsymbol{x}} = \frac{\partial V}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}).$$

Если в качестве функции Ляпунова выбрать некоторую выпуклую ($n-1$)-мерную поверхность S, то $\dfrac{\partial V}{\partial \boldsymbol{x}} = \mathrm{grad} V(\boldsymbol{x})$ есть вектор нормали $\boldsymbol{N}$ к S в точке $\boldsymbol{x}$. Тогда по условию теоремы 1

$$\left[\frac{\partial V}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x})\right]^T = \dot{\boldsymbol{x}}^T(t)\boldsymbol{N} \le -w_3(\boldsymbol{x}).$$

Рассмотрим окрестность $g^t \delta_x$ точки равновесия $\boldsymbol{x} = \boldsymbol{x}_0 = 0$ системы (1) с радиусом $R_\delta(t)$, где g^t – оператор сдвига вдоль траекторий (1). Обозначим через $V_\delta(t)$ евклидов объем области $g^t \delta_x$, в которую переводит область δ_x фазовый поток векторного поля $\boldsymbol{f}(\boldsymbol{x})$ в момент времени t. Воспользуемся следствием формулы Лиувилля–Остоградского [1, 2], учитывая гладкость правой части (1):

$$dV_\delta(t)/dt = \int\limits_{g^t \delta_x} \mathrm{div} \boldsymbol{f} dV,$$

где div$\boldsymbol{f}$ – скалярное поле дивергенции векторного поля $\boldsymbol{f}(\boldsymbol{x})$, dV – евклидов объем элемента. Система (1) будет устойчива, если преобразование потока за положительное время уменьшает объем по каждой компоненте вектора $\boldsymbol{x}$, т.е.

$$dV_{\delta i}(t)/dt = \int\limits_{g^t\delta_{x_i}} \mathrm{div} f_i \, dV_i < 0, \tag{3}$$

где $\dot{x}_i = f_i(\boldsymbol{x})$, $i = \overline{1, n}$. Используя обратный переход по формуле Гаусса–Остоградского [9] для (3) и с учетом условий теоремы 1, имеем

$$\int\limits_{g^t\delta_{x_i}} \mathrm{div} f_i \, dV_i = \oint\limits_{g^t\delta_{x_i}} \dot{x}_i \mathbf{N}_{\delta i} dS_{\delta i} \le -\oint\limits_{g^t\delta_x} w_3(\boldsymbol{x}) dS_{\delta i} < 0, \; i = \overline{1, n}, \tag{4}$$

где $\mathbf{N}_{\delta i}$ – вектор внешней нормали к поверхности δ_x; S_{xi} – площадь поверхности δ_x. Если δ_x, в которой выполнено (4), совпадает с областью определения (1) D_x, то система (1) устойчива во всей своей области определения D_x.

Для исследования характера устойчивости системы (1) без привлечения аппарата функций Ляпунова введем регулярное отображение $\varphi : X \to \Xi$ [10], которое преобразует систему (1) к диагональному виду

$$\dot{\xi}_i(t) = h_i(\xi_i(t)), \; i = \overline{1, n},$$

или, в векторной форме,

$$\dot{\boldsymbol{\xi}}(t) = \mathbf{h}(\boldsymbol{\xi}(t)). \tag{5}$$

Здесь $\boldsymbol{\xi}(t) = \left[\xi_1(t), \xi_2(t), ..., \xi_n(t)\right]^T \in \Xi \subset \mathbf{R}^n$, Ξ – открытое множество; $\mathbf{h} = \left[h_1, h_2, ..., h_n\right]^T$ – гладкая вектор-функция (соответствующая классу гладкости (1)) в односвязной области $D_\xi = \Xi \subset R^n$, где $\varphi : D_x \to D_\xi$; D_ξ – окрестность точки равновесия $\boldsymbol{\xi}_0 = \varphi^{-1}\left(\boldsymbol{x}_0\right)$. Без потери общности предположим, что $\boldsymbol{\xi}_0 = 0$. Такое преобразование координат введено для выявления зависимости одних переменных в системе (1) от других. На самом деле вводить его не обязательно, однако при вычислении потока все равно придется учесть взаимосвязь переменных в системе (1), что эквивалентно введению данного регулярного преобразования координат. Теорема 1 доказана.

Теорема 2. Пусть для системы (1) существует диффеоморфизм $\varphi : X \to \Xi$, преобразующий (1) в (5). Тогда система (1) устойчива в ($n-1$)-мерном шаре S_x, если для системы (2) выполнено

$$\oint\limits_{S_\xi} \dot{\xi}_i(t) N_i dS < 0, \; i = \overline{1, n}, \tag{6}$$

где $S_\xi = \varphi^{-1}\left(S_x\right)$, $S_\xi = \left\{\boldsymbol{\xi} \in R^n : \boldsymbol{\xi}^T \boldsymbol{\xi} = R_0^2\right\}$.

Доказательство. Рассмотрим окрестность $g^t\delta_\xi$ точки равновесия $\boldsymbol{\xi} = \boldsymbol{\xi}_0 = 0$ системы (5) с радиусом $R_\delta(t)$, где g^t – оператор сдвига вдоль траекторий (5). Пусть $V_\delta(t)$ – евклидов объем области $g^t\delta_\xi$, в которую переводит область δ_ξ фазовый поток векторного поля $\boldsymbol{h}(\boldsymbol{\xi})$ в момент времени *t*. Из следствия формулы Лиувилля–Остоградского [1, 2] и с учетом гладкости правой части (5) $dV_\delta(t)/dt = \int\limits_{g^t\delta_\xi} \mathrm{div}\mathbf{h} \, dV$, где div$\mathbf{h}$ – скалярное поле дивергенции векторного поля $\mathbf{h}$; dV – евклидов объем элемента. Система (5) будет устойчива, если преобразование потока с течением времени уменьшает объем по каждой компоненте вектора $\boldsymbol{\xi}$, т.е.

$$dV_{\delta i}(t)/dt = \int\limits_{g^t\delta_{\xi_i}} \mathrm{div} h_i \, dV_i < 0, \; i = \overline{1, n}. \tag{7}$$

Используя обратный переход по формуле Гаусса–Остоградского [9] для (3), перепишем (7) в виде

$$\int\limits_{g^t\delta_{\xi_i}} \mathrm{div} h_i \, dV_i = \oint\limits_{g^t\delta_{\xi_i}} h_i \mathbf{N}_{\delta i} dS_{\delta i} < 0, \; i = \overline{1, n}, \tag{8}$$

где $\mathbf{N}_{\delta i}$ – вектор внешней нормали к поверхности δ_ξ; $S_{\delta i}$ – площадь поверхности δ_ξ. Если δ_ξ, в которой выполнено (8), совпадает с областью определения (5) D_ξ, то в силу диффеоморфности преобразования φ система (1) устойчива во всей своей области определения D_x. Если наступает некоторый момент времени $t = t_1$, в котором скорость изменения хотя бы одного фазового объема $dV_{\delta i}(t)/dt\big|_{t=t_1}$, $i = \overline{1, n}$, меняет знак с минуса на плюс и далее с течением времени сохраняет положительное значение, то происходит расширение части фазового объема, и система (1) становится неустойчивой. Область $\delta_\xi = S_\xi$ с радиусом

$R_\delta(t)\big|_{t=t_1} = R_0$, для которой выполнено (6), и является областью устойчивости системы (2). Для получения области устойчивости системы (1) необходимо взять прямое преобразование $S_x = \varphi\left(S_\xi\right)$. Теорема 2 доказана.

Как видно из доказательства, если воспользоваться формулой Гаусса–Остоградского, то вместо формулы (6) можно использовать формулу следующего вида:

$$\oint_{S_\xi} \dot{\xi}_i(t) N_i dS = \int_V \mathrm{div} h_i \, dV_i < 0, \; i = \overline{1, n},$$

причем знак второго интеграла зависит только от знака $\mathrm{div} h_i$ в соответствующей области V.

Результат, использующий знак дивергенции, был получен в [6–8], однако система вида (1) не преобразовывалась к диагональному виду. Вследствие этого автор смог получить только необходимые условия асимптотической устойчивости. В качестве необходимого и достаточного условия в [8] для нелинейных систем второго порядка автор использовал знак дивергенции и квадратичные функции, определенные в исследуемой области. В настоящей работе получены условия устойчивости, основанные только на вычислении потока вектора фазовой скорости через выпуклую поверхность. Для этого, как отмечалось ранее, система (1) преобразовывалась к диагональному виду (5), чтобы при вычислении потока были учтены связи переменных в системе, иначе пришлось бы это учесть путем выражения одних переменных через другие. Следовательно, при нахождении дивергенции, согласно [6–8], мы получим не истинную дивергенцию вектора фазовой скорости. Для того чтобы использовать знак дивергенции при исследовании устойчивости систем без привлечения аппарата квадратичных функций, необходимо при нахождении дивергенции также учесть взаимосвязь переменных системы.

Асимптотическая устойчивость линейных систем

Рассмотрим доказательство второй теоремы Ляпунова и условия асимптотической устойчивости для линейных объектов. Пусть задана система линейных дифференциальных уравнений, записанная в форме

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t), \tag{9}$$

где $\mathbf{x}(t) \in \mathbf{R}^n$ – вектор состояния уравнения (9); $\mathbf{A} \in \mathbf{R}^{n\times n}$ – матрица с постоянными элементами. Хорошо известно, что равновесным состоянием системы (9) является точка $\boldsymbol{x} = 0$.

Теорема 3. Равновесное состояние $\boldsymbol{x} = \boldsymbol{x}_0 = 0$ системы (9) асимптотически устойчиво, если в окрестности точки $\boldsymbol{x}_0 = 0$ существует функция Ляпунова $V(\mathbf{x}) = \mathbf{x}^T\mathbf{H}\mathbf{x}$, $\mathbf{H} = \mathbf{H}^T > 0$ такая, что ее полная производная по времени есть отрицательно определенная функция $\dot{V}(x) = -\mathbf{x}^T\mathbf{Q}\mathbf{x}$, $\mathbf{A}^T\mathbf{H} + \mathbf{H}\mathbf{A} = -\mathbf{Q}$, $\mathbf{Q} = \mathbf{Q}^T > 0$.

Доказательство. Определим вектор нормали $\mathbf{N}$ к поверхности $S = V(\mathbf{x}) = \mathbf{x}^T\mathbf{H}\mathbf{x}$ как $\mathbf{N} = \mathrm{grad}\left(V\right)/\left|\mathrm{grad}\left(V\right)\right| = \mathbf{H}\mathbf{x}/\left|\mathbf{H}\mathbf{x}\right|$ и рассмотрим поток вектора фазовой скорости (9) через поверхность V: $\oint_S \dot{\boldsymbol{x}}^T \boldsymbol{N} dS = \oint_S \frac{1}{\left|\mathbf{P}\mathbf{x}\right|}\mathbf{x}^T\mathbf{A}^T\mathbf{H}\mathbf{x} dS$.

Преобразовав $\mathbf{A}^T\mathbf{H}^T$ к диагональной матрице согласно [5], получим $\mathbf{x}^T\mathbf{A}^T\mathbf{H}\mathbf{x} = 0,5\boldsymbol{x}^T\left(\mathbf{A}^T\mathbf{H}^T + \mathbf{H}\mathbf{A}\right)\mathbf{x} = -0,5\boldsymbol{x}^T\mathbf{Q}\mathbf{x}$, где $\mathbf{Q} = \mathrm{diag}\left\{q_i\right\} > 0$, $i = \overline{1, n}$. В результате этого получено уравнение Ляпунова $\mathbf{A}^T\mathbf{H}^T + \mathbf{H}\mathbf{A} = -0,5\mathbf{Q}$, которое будет выполнено, только если матрица $\mathbf{A}$ – гурвицева [3–5]. Перепишем последние интегралы поэлементно

$$\oint_S \frac{1}{\left|\mathbf{x}\right|}\mathbf{x}^T\mathbf{A}^T\mathbf{H}^T\mathbf{x} dS = -\oint_S \frac{0,5}{\left|\mathbf{x}\right|}\mathbf{x}^T\mathbf{Q}\mathbf{x} dS = -\sum_{i=1}^{n}\oint_S \frac{q_i}{\left|\mathbf{x}\right|}x_i^2 dS.$$

Из свойств матрицы $\mathbf{Q}$, положительной определенности функций $\left|\mathbf{x}\right|^{-1}x_i^2 > 0$ и доказательства теоремы 1 имеем, что $-\oint_S \frac{q_i}{\left|\boldsymbol{x}\right|}x_i^2 dS < 0$, $i = \overline{1, n}$, а значит, система (9) будет устойчива, так как преобразование потока за положительное время уменьшает объем по каждой компоненте вектора $\boldsymbol{x}$. Теорема 3 доказана.

Аналог теоремы 2 для линейных систем можно предложить, если рассмотреть регулярное преобразование координат $\Gamma : \boldsymbol{x}(t) \to \boldsymbol{\xi}(t)$, $\Gamma \in R^{n\times n}$, приводящее систему (9) к виду $\dot{\boldsymbol{\xi}}(t) = \mathbf{D}\boldsymbol{\xi}(t)$, $\mathbf{D} = \mathrm{diag}\left\{d_i\right\} \in \mathbf{R}^{n\times n}$, $i = \overline{1, n}$. Однако это слишком узкий класс объектов. Для того чтобы рассмотреть теорему 2 для всего класса линейных объектов, ниже предложена теорема 4.

Теорема 4. Если для системы (9) можно указать матрицу $\mathbf{H} > 0$, $\mathbf{H} \in \mathbf{R}^{n \times n}$, такую, что для новой системы

$$\dot{\mathbf{x}}(t) = \mathbf{HAx}(t) \tag{10}$$

выполнено

$$\oint_S \dot{x}_i(t) N_i dS < 0, \; i = \overline{1, n}, \tag{11}$$

где $S = \left\{ \boldsymbol{x} \in R^n : \boldsymbol{x}^T \boldsymbol{x} = R_0^2 \right\}$, то система (9) имеет асимптотически устойчивую точку равновесия $\boldsymbol{x} = 0$.

Доказательство. Найдем поток векторного поля системы (10) через поверхность S с вектором нормали $\mathbf{N} = \mathrm{grad}(S) / \left| \mathrm{grad}(S) \right| = \mathbf{x} / |\mathbf{x}|$:

$$\oint_S \dot{\mathbf{x}}^T \mathbf{N} dS = \oint_S (\mathbf{Ax})^T \mathbf{x} / \mathbf{x} \, dS = \oint_S \frac{1}{\mathbf{x}} \mathbf{x}^T \mathbf{A}^T \mathbf{H}^T \mathbf{x} dS. \tag{12}$$

Запишем квадратичную форму $\mathbf{x}^T \mathbf{A}^T \mathbf{H}^T \mathbf{x}$ в виде $\mathbf{x}^T \mathbf{A}^T \mathbf{Hx} = 0,5 \mathbf{x}^T \left(\mathbf{A}^T \mathbf{H}^T + \mathbf{HA} \right) \mathbf{x} = -0,5 \mathbf{x}^T \mathbf{Qx}$, где $\mathbf{Q} = \mathrm{diag}\{q_i\} > 0$, $i = \overline{1, n}$. Уравнение Ляпунова $\mathbf{A}^T \mathbf{H}^T + \mathbf{HA} = -0,5\mathbf{Q}$ будет выполнено в случае гурвицевости матрицы ***A*** [3–5]. Запишем теперь (12) поэлементно:

$$\oint_S \frac{1}{|\mathbf{x}|} \mathbf{x}^T \mathbf{A}^T \mathbf{H}^T \mathbf{x} dS = -\oint_S \frac{0,5}{|\mathbf{x}|} \mathbf{x}^T \mathbf{Qx} dS = -\sum_{i=1}^{n} \oint_S \frac{q_i}{|\mathbf{x}|} x_i^2 dS.$$

Из условий выбора матрицы **Q**, положительной определенности функций $|\boldsymbol{x}|^{-1} x_i^2 > 0$ и доказательства теоремы 2 получим, что $-\oint_S \frac{q_i}{|\mathbf{x}|} x_i^2 dS < 0$, $i = \overline{1, n}$. Значит, (11) выполнено. Теорема 4 доказана.

Заключение

В работе получены условия устойчивости нелинейных и линейных динамических стационарных систем с использованием потока вектора фазовой скорости. Достоинством предложенного подхода является то, что для исследования характера поведения решений системы дифференциальных уравнений достаточно знать знак потока вектора фазовой скорости в заданной области через замкнутую выпуклую поверхность. Существенным недостатком является наличие регулярного преобразования координат, которое зачастую получить достаточно сложно. Если же не применять регулярное преобразование координат, то при вычислении потока по определению возникают сложные интегралы, связанные с зависимостью одних переменных в системе от других. При использовании формулы Гаусса–Остоградского возникает проблема вычисления дивергенции фазовой скорости из-за зависимости переменных в системе друг от друга.

Литература

1. Арнольд В.И., Ильяшенко Ю.С. Обыкновенные дифференциальные уравнения. – М.: Наука, 1985. – Т. 1. – 149 с.
2. Эрроусмит Д., Плейс К. Обыкновенные дифференциальные уравнения. Качественная теория с приложениями. – М.: Мир, 1986. – 243 с.
3. Ла-Салль Ж., Лефшец С. Исследование устойчивости прямым методом Ляпунова. – М.: Мир, 1964. – 168 с.
4. Ляпунов А.М. Общая задача устойчивости движения. – М.: Физматгиз, 1959. – 472 с.
5. Методы классической и современной теории автоматического управления. Т. 1: Математические модели, динамические характеристики и анализ систем автоматического управления / Под ред. К.А. Пупкова, Н.Д. Егупова. – М.: Издательство МГТУ им. Н.Э. Баумана, 2004. – 656 с.
6. Жуков В.П. Об одном методе качественного исследования устойчивости нелинейных систем // Автоматика и телемеханика. – 1978. – № 6. – С. 11–15.
7. Жуков В.П. Необходимые и достаточные условия неустойчивости нелинейных автономных динамических систем // Автоматика и телемеханика. – 1990. – № 12. – С. 59–65.
8. Жуков В.П. Дивергентные условия асимптотической устойчивости нелинейных динамических систем второго порядка // Автоматика и телемеханика. – 1999. – № 7. – С. 34–43.
9. Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. – М.: ФИЗМАТЛИТ, 2001. – Т. 3. – 662 с.
10. Мищенко А.С., Фоменко А.Т. Курс дифференциальной геометрии и топологии. – М.: Факториал Пресс, 2000. – 448 с.

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

УДК 681.51.015

ГИБРИДНЫЙ АЛГОРИТМ ИДЕНТИФИКАЦИИ ЧАСТОТ МУЛЬТИСИНУСОИДАЛЬНОГО СИГНАЛА[4]

А.А. Бобцов, А.А. Ведяков, С.А. Колюбин, А.А. Пыркин

Рассматривается задача идентификации неизвестных частот мультисинусоидального сигнала на примере двух гармоник. На базе метода каскадной редукции предложен алгоритм идентификации, позволяющий ускорять процесс параметрической сходимости оценок неизвестных частот мультисинусоидального сигнала.

Ключевые слова: синусоидальный сигнал, редукция, идентификация.

Рассмотрим мультисинусоидальный сигнал, представленный в виде суммы двух гармоник

$$y(t) = A_1 \sin(\omega_1 t + \varphi_1) + A_2 \sin(\omega_2 t + \varphi_2), \tag{1}$$

где $A_1; \omega_1; \varphi_1; A_2; \omega_2; \varphi_2$ – неизвестные постоянные параметры. Ставится задача синтеза алгоритма идентификации неизвестных параметров ω_1, ω_2 – частот мультисинусоидального сигнала $y(t)$. Очевидно, что такая задача решается не впервые (например, [1–3]). Однако проблема быстродействия сходимости настраиваемых параметров к истинным значениям для мультисинусоидального сигнала, насколько известно авторам, ранее не обсуждалась. В основном предлагается решение данной проблемы и анализируется устойчивость полученных алгоритмов. В этой работе предлагается новый алгоритм идентификации частот сигнала вида (1) и аналогично [4] даются конкретные рекомендации по ускорению процессов параметрической сходимости.

Осуществим параметризацию модели (1) следующим образом:

$$p^4 y(t) = \theta_1 p^2 y(t) + \theta_2 y(t), \tag{2}$$

где $p = d/dt$, $\theta_1 = -\omega_1^2 - \omega_2^2$, $\theta_2 = -\omega_1^2\omega_2^2$ – неизвестные параметры, подлежащие идентификации.

Введем новые обозначения:

$$z(t) = \frac{p^4}{(p+\lambda)^4} y(t), \ \varsigma_1(t) = \frac{p^2}{(p+\lambda)^4} y(t), \ \varsigma_2(t) = \frac{1}{(p+\lambda)^4} y(t), \tag{3}$$

где $\lambda > 0$ – некоторый выбираемый при синтезе коэффициент. Тогда, используя преобразования (2)–(3), для модели (1) имеем

$$z(t) = \theta_1 \varsigma_1(t) + \theta_2 \varsigma_2(t) + \varepsilon(t) = \boldsymbol{\varsigma}^T \boldsymbol{\theta}, \tag{4}$$

где $\boldsymbol{\theta} = col\{\theta_1, \theta_2\}$, $\boldsymbol{\varsigma} = col\{\varsigma_1, \varsigma_2\}$ и $\varepsilon(t)$ – экспоненциально затухающее слагаемое, обусловленное ненулевыми начальными условиями.

Аналогично [4], пренебрегая $\varepsilon(t)$, можно воспользоваться алгоритмом идентификации вида

$$\dot{\hat{\boldsymbol{\theta}}} = -k\boldsymbol{\varsigma}\boldsymbol{\varsigma}^T \hat{\boldsymbol{\theta}} + k\boldsymbol{\varsigma} z, \tag{5}$$

где $\hat{\boldsymbol{\theta}}$ – оценка вектора $\boldsymbol{\theta}$, а $k > 0$ – некоторый коэффициент, либо задаваемый при синтезе, либо настраиваемый в процессе работы.

В скалярном случае за счет увеличения коэффициента $k > 0$ можно достигать увеличения быстродействия параметрической сходимости [4]. Однако в более общем случае, как, например в (5), увеличение коэффициента $k > 0$ не позволит гарантировать увеличения быстродействия параметрической сходимости. Решение данной проблемы может быть найдено с использованием гибридной схемы настройки параметров, базирующейся на методе каскадной редукции [5]. Преобразуем (4), следуя данному методу. Для этого последовательно умножим (4) на ς_1 и проинтегрируем полученное уравнение, т.е.

$$z\varsigma_1 = \theta_1 \varsigma_1^2 + \theta_2 \varsigma_2 \varsigma_1, \ \int_0^t z\varsigma_1 d\tau = \theta_1 \int_0^t \varsigma_1^2 d\tau + \theta_2 \int_0^t \varsigma_2 \varsigma_1 d\tau.$$

Введем обозначения $\xi_1 = \int_0^t z\varsigma_1 d\tau$, $\xi_2 = \int_0^t \varsigma_1^2 d\tau$ и $\xi_3 = \int_0^t \varsigma_2 \varsigma_1 d\tau$ и последовательно сначала разделим на ξ_2, а затем продифференцируем последнее соотношение. Тогда получаем

$$\dot{\xi}_1 \xi_2^{-1} - \xi_1 \dot{\xi}_2 \xi_2^{-2} = \theta_2 (\dot{\xi}_3 \xi_2^{-1} - \xi_3 \dot{\xi}_2 \xi_2^{-2})$$

или

$$\dot{\xi}_1 \xi_2 - \xi_1 \dot{\xi}_2 = \theta_2 (\dot{\xi}_3 \xi_2 - \xi_3 \dot{\xi}_2). \tag{6}$$

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

Введем следующие обозначения: $\bar{z} = \dot{\xi}_1\xi_2 - \xi_1\dot{\xi}_2$ и $\bar{\varsigma}_2 = \dot{\xi}_3\xi_2 - \xi_3\dot{\xi}_2$. Тогда уравнение (6) примет вид

$$\bar{z}(t) = \theta_2\bar{\varsigma}_2 . \tag{7}$$

Из (7) легко получить алгоритм идентификации параметра θ_2

$$\dot{\hat{\theta}}_2 = -\gamma_2\bar{\varsigma}_2^2\hat{\theta}_2 + \gamma_2\bar{\varsigma}_2\bar{z} , \tag{8}$$

где $\gamma_2 > 0$ – некоторый параметр, увеличение которого позволяет повысить скорость сходимости $\hat{\theta}_2$ к θ_2. Пренебрегая в уравнении (4) экспоненциально затухающим слагаемым $\varepsilon(t)$, для идентификации параметра θ_1 будем использовать алгоритм вида

$$\dot{\hat{\theta}}_1 = -\gamma_1\varsigma_1^2\hat{\theta}_1 + \gamma_1\bar{\varsigma}_1 z' , \tag{9}$$

где $z' = z - \hat{\theta}_2\varsigma_2$ и $\gamma_1 > 0$ – некоторый параметр, увеличение которого, как и в предыдущем случае, позволяет повысить скорость сходимости $\hat{\theta}_1$ к θ_1.

Таким образом, получаем гибридную схему идентификации $\theta_1 = -\omega_1^2 - \omega_2^2$ и $\theta_2 = -\omega_1^2\omega_2^2$, подразумевающую использование редуцированной модели (7) и параметрическую настройку с использованием алгоритмов (8)–(9). На основе оценок $\hat{\theta}_1$ и $\hat{\theta}_2$ нетрудно получить оценки частот двух гармоник исходного сигнала $\hat{\omega}_1$ и $\hat{\omega}_2$, решив квадратное уравнение. Для ускорения процесса параметрической сходимости оценок неизвестных частот мультисинусоидального сигнала (1) необходимо увеличивать коэффициенты λ, γ_1 и γ_2.

Для иллюстрации работоспособности предлагаемой схемы идентификации вида (8)–(9) приведем результаты компьютерного моделирования алгоритмов оценивания частот сигнала $y(t) = \sin t + 6\sin(2t - 2)$ при $\lambda = 1$ и $\gamma_1 = \gamma_2 = 10$ (рисунок).

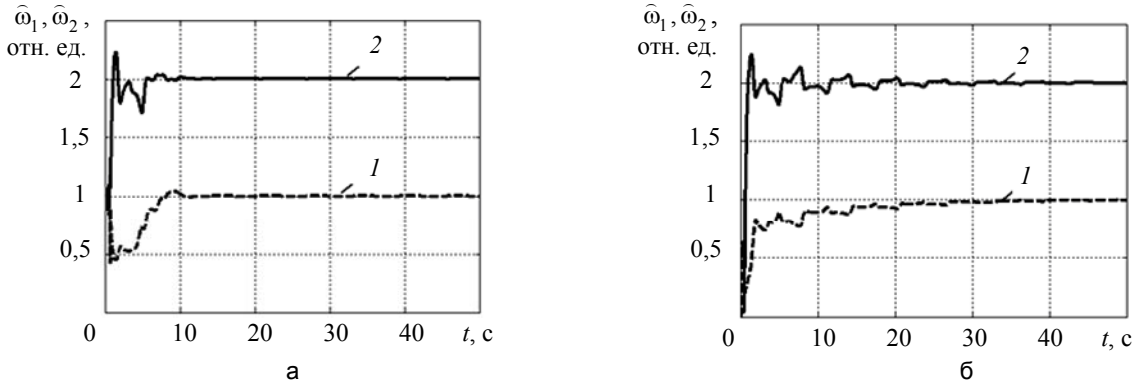


Рисунок. Результаты моделирования алгоритмов оценивания частот сигнала $y(t)$ ($\hat{\omega}_1$ – кривая 1; $\hat{\omega}_2$ – кривая 2): оценка частот алгоритмом (8)–(9) (а); оценка частот алгоритмом [4] (б)

Из рисунка видно, что новый алгоритм оценивания частот (8)–(9) на базе метода каскадной редукции работает быстрее и точнее при одинаковых настроечных коэффициентах λ и γ, чем метод, опубликованный в [4]. В заключение следует отметить, что предлагаемый подход к идентификации частот сигнала вида (1) предусматривает дальнейшее изучение, поскольку при синтезе алгоритма идентификации не были учтены помехи в измерении, а также было опущено экспоненциально затухающее слагаемое $\varepsilon(t)$, влияющее на качество процессов.

Литература

1. Obregón-Pulido G., Castillo-Toledo B. and A.A. Loukianov. Globally Convergent Estimator for n–Frequencies // IEEE Transactions on Automatic Control. – 2002. – V. 47. – P. 857–863.
2. Xia X. Global Frequency Estimation Using Adaptive Identifiers // IEEE Transactions on Automatic Control. – 2002. – V. 47. – P. 1188–1193.
3. Marino R., Tomei P. Global Estimation of n Unknown Frequencies // IEEE Transactions on Automatic Control. – 2002. – V. 47. – P. 1324–1328.
4. Aranovskiy S., Bobtsov A., Kremlev A., Nikolaev N., Slita O. Identification of frequency of biased harmonic signal // European Journal of Control. – 2010. – № 2. – P. 129–139.
5. Арановский С.В., Бобцов А.А., Пыркин А.А. Каскадная редукция в задачах идентификации // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 3 (79). – С. 149–150.

Бобцов Алексей Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, bobtsov@mail.ifmo.ru

Ведяков Алексей Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, инженер, vedyakov@gmail.com

Колюбин Сергей Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, s.kolyubin@gmail.com

Пыркин Антон Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, a.pyrkin@gmail.com

УДК 681.5

ИНТЕГРО-ДИФФЕРЕНЦИРУЮЩИЕ АЛГОРИТМЫ СКОРОСТНОГО ГРАДИЕНТА В ЗАДАЧАХ КРАТНОЙ СИНХРОНИЗАЦИИ ВИБРАЦИОННЫХ УСТАНОВОК[5]

А.Л. Фрадков, О.П. Томчина, В.А. Галицкая, Д.В. Горлатов

Описываются интегро-дифференцирующие алгоритмы скоростного градиента, частными случаями которых являются пропорциональные, интегральные и пропорционально-интегральные алгоритмы скоростного градиента. Вводится определение кратной приближенной частотно-координатной синхронизации, синтезируются интегро-дифференцирующие алгоритмы управления кратной синхронизацией вибрационных установок. Приводятся результаты сравнительного компьютерного исследования пропорционального и пропорционально-интегрального алгоритмов, показывающие, что предложенные алгоритмы обладают рядом преимуществ перед уже известными.

Ключевые слова: интегро-дифференцирующие алгоритмы, скоростной градиент, кратная синхронизация, вибрационная установка, вибротранспортирование.

Введение

При проектировании вибрационных установок, осуществляющих грохочение, дробление, вибротранспортирование сыпучих материалов и другие технологические операции, повышение производительности в значительной мере определяется стабильностью синхронного режима вращения вибровозбудителей. Традиционный подход базируется на явлении самосинхронизации, открытом и изученном И.И. Блехманом [1]. Дополнительные технологические возможности использования виброустановок предоставляет режим кратной синхронизации, при котором вибровозбудители вращаются со средними угловыми скоростями, кратными друг другу. Кратный синхронный режим, вносящий асимметрию в систему, способствует возникновению и усилению эффекта вибрационного перемещения, особенно для таких трудноосуществимых технологических процессов, как транспортирование пылевидных, влажных и липких грузов. При этом синхронность обеспечивает максимальную скорость вибротранспортирования. Кроме того, наличие двух различных частот вращения роторов позволяет транспортно-технологическим машинам осуществлять одновременно вибротранспортирование, возбуждаемое низкой частотой, а также просеивание и сепарацию сыпучих материалов, осуществляемые за счет большей кратной частоты. Но при рассмотрении задач кратной синхронизации стабильная самосинхронизация может и не наблюдаться. Условия на соотношение масс и на взаимное расположение роторов и несущей платформы, обеспечивающее кратную самосинхронизацию, были получены в работах [2–4]. Однако в ряде практически важных случаев эти условия не выполняются, и режим кратной самосинхронизации оказывается неустойчивым.

В связи с этим разработка новых подходов к решению задачи обеспечения стабильного кратного синхронного вращения вибровозбудителей является актуальной технической задачей. Одним из перспективных подходов является подход, основанный на управляемой синхронизации. Управлению синхронизацией колебательных механических систем посвящен целый ряд работ [5–7]. В настоящей работе рассматривается подход, основанный на применении интегро-дифференцирующих алгоритмов скоростного градиента [8, 9], синтезируемых из условия обеспечения стабильного кратного синхронного режима. Предложены и исследованы новые алгоритмы управления кратной синхронизацией, обеспечивающие улучшенные характеристики процессов в системе.

Интегро-дифференцирующие алгоритмы скоростного градиента

В данном разделе, следуя [8], описывается общий подход к синтезу нелинейных систем управления – так называемый метод скоростного градиента, предложенный в конце 70-х годов прошлого века первона-

[5] Работа выполнена при поддержке ФЦП «Кадры» (соглашения 8846, 8855),Программы 10 ОЭММПУ РАН, и РФФИ (грант 11-08-01218).

чально для задач адаптивного управления [10]. Метод основан на использовании функций Ляпунова и требует задания цели управления как уменьшения значений некоторой скалярной целевой функции (функционала) до заданной величины.

Описано несколько типов алгоритмов скоростного градиента: алгоритмы в дифференциальной и конечной формах, построенные по локальному или интегральному целевому функционалу. Завершается раздел описанием интегро-дифференцирующих алгоритмов, при синтезе которых используются как дифференциальные, так и интегральные операторы. Эти алгоритмы будут использоваться в последующих разделах для управления синхронизацией вибрационных установок.

Исходными данными для синтеза алгоритма управления являются уравнение объекта управления (далее называемого для краткости объектом) и цель управления. Будем считать, что объект задан уравнением

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \mathbf{u}, t)\,, \tag{1}$$

где $\mathbf{x} \in \mathbf{R}^n$ – вектор состояния объекта; $\mathbf{u} \in \mathbf{R}^m$ – вектор управления (настраиваемых параметров); вектор-функция $\mathbf{F}(\cdot)$ определена при всех $\mathbf{x} \in \mathbf{R}^n$; $\mathbf{u} \in \mathbf{R}^m$, $t \geq 0$, кусочно-непрерывна по t и непрерывно дифференцируема по $\mathbf{x}$, $\mathbf{u}$.

Цель управления зададим в виде

$$Q_t \leq \Delta\,, \quad t > t_*, \tag{2}$$

где Q_t – целевой функционал (постоянные Δ, t_* могут быть заданы или нет в зависимости от конкретной задачи). Будем различать два типа целевых функционалов: локальный и интегральный. Локальным называется функционал вида

$$Q_t = Q(\mathbf{x}(t), t),$$

где $Q(\mathbf{x}, t)$ – скалярная функция n+1 переменных. Функционал вида

$$Q_t = \int_0^t R(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau) d\tau,$$

где $R(\mathbf{x}, \mathbf{u}, t)$ – скалярная функция n+m+1 переменных, будем называть интегральным.

Ставится задача: определить алгоритм управления

$$\mathbf{u}(t) = U\left[\mathbf{x}(\tau), \mathbf{u}(\tau)\right], \quad 0 \leq \tau \leq t, \tag{3}$$

так, чтобы в системе (1), (3) достигалась цель управления (2) и все ее траектории оставались ограниченными при $t \geq 0$.

Отметим прежде всего, что сформулированная задача весьма общая и включает как частные случаи различные типы задач управления и адаптации. Например, в качестве цели управления (2) может фигурировать как исходная цель управления, так и некоторая вспомогательная цель, выражающая требования к синтезируемой системе. Уравнение (1) также может относиться к различным системам. Более того, в одной и той же системе управления модель (1) может использоваться для описания различных частей системы так, что переменные x, u могут иметь различный смысл. Введение такого унифицированного описания позволяет единообразно ставить и решать задачи синтеза как основного контура, так и контура адаптации в адаптивных системах.

Важную роль в задачах синтеза нелинейных и адаптивных систем играет класс объектов, линейных по входам:

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t)\mathbf{u}\,. \tag{4}$$

Уравнение (4) можно переписать также в виде

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{x}, t) + \sum_{i=1}^{m} \mathbf{B}_i(\mathbf{x}, t)\mathbf{u}_i\,, \tag{5}$$

где $\mathbf{u}_i$ – компоненты вектора $\mathbf{u} \in \mathbf{R}^m$; $\mathbf{B}_i(\mathbf{x}, t) \in \mathbf{R}^n$ – столбцы матрицы $\mathbf{B}(\mathbf{x}, t)$.

Пусть цель управления (2) задана при помощи локального целевого функционала $Q_t = Q(\mathbf{x}(t), t)$. Для построения алгоритма управления (3) вычислим скалярную функцию $\dot{Q}(t) = \omega(\mathbf{x}, \mathbf{u}, t)$ – скорость изменения Q_t в силу уравнения объекта (1):

$$\omega(\mathbf{x}, \mathbf{u}, t) = \frac{\partial Q(\mathbf{x}, t)}{\partial t} + [\nabla_{\mathbf{x}} Q(\mathbf{x}, t)]^T F(\mathbf{x}, \mathbf{u}, t).$$

Затем найдем градиент $\omega(\mathbf{x}, \mathbf{u}, t)$ по входным переменным

$$\nabla_{\mathbf{u}} \omega(\mathbf{x}, \mathbf{u}, t) = \left[\frac{\partial \omega}{\partial \mathbf{u}}\right]^T = \left[\frac{\partial \mathbf{F}}{\partial \mathbf{u}}\right]^T \nabla_{\mathbf{x}} Q(\mathbf{x}, t)]^T\,.$$

Наконец, зададим алгоритм изменения $\mathbf{u}(t)$ дифференциальным уравнением

$$\frac{d\mathbf{u}}{dt} = -\mathbf{\Gamma} \nabla_{\mathbf{u}} \omega(\mathbf{x}, \mathbf{u}, t), \tag{6}$$

где $\mathbf{\Gamma} = \mathbf{\Gamma}^T > 0$ – симметрическая положительно определенная матрица, например $\mathbf{\Gamma} = \mathrm{diag}\{\gamma_1,\ldots,\gamma_m\}$, $\gamma_i > 0$. Алгоритм (6) естественно назвать алгоритмом скоростного градиента (АСГ), поскольку в нем изменение $\mathbf{u}(t)$ происходит пропорционально градиенту скорости изменения Q_t. Интегрируя (6) на промежутке [0, t], можно записать АСГ в форме (3) с интегральным оператором $U(\cdot)$.

Происхождение алгоритма (6) можно объяснить следующим образом. Для достижения цели управления (2) желательно изменять $\mathbf{u}(t)$ в направлении уменьшения Q_t. Однако Q_t не зависит от $\mathbf{u}(t)$, и найти такое направление затруднительно (в частности, это связано с нахождением функций чувствительности). Вместо этого можно пытаться уменьшить Q_t, стремясь к выполнению неравенства $\dot{Q}(t) < 0$, означающего, в свою очередь, уменьшение Q_t. Функция $\dot{Q}(t) = \omega(\mathbf{x},\mathbf{u},t)$ уже явно зависит от $\mathbf{u}$, что и позволяет написать алгоритм (6). Можно также рассматривать АСГ как непрерывный аналог или «идеализированный» вариант дискретного градиентного алгоритма, поскольку при малом шаге дискретизации градиент целевой функции, совпадающий с градиентом ее приращения, приближается по направлению к градиенту скорости изменения целевой функции в силу объекта [8].

В качестве примера выпишем АСГ для задачи регулирования линейного по входам объекта управления (4) при целевой функции

$$Q(\mathbf{x},t) = \frac{1}{2}[\mathbf{y} - \mathbf{y}_*(t)]^T \mathbf{H}[\mathbf{y} - \mathbf{y}_*(t)], \tag{7}$$

где $\mathbf{y} = \mathbf{G}(\mathbf{x},t) \in \mathbf{R}^1$; $\mathbf{y}_*(t) \in \mathbf{R}^1$ – задающее воздействие (желаемая траектория выхода); $\mathbf{G}(\mathbf{x},\ t)$ – гладкая вектор-функция; $\mathbf{H}$ – $l\times l$-матрица. Скорость изменения $Q(\mathbf{x},t)$ будет равна

$$\omega(\mathbf{x},\mathbf{u},t) = [\mathbf{y} - \mathbf{y}_*(t)]^T \mathbf{H}[\mathbf{C}\mathbf{A}(\mathbf{x},t) + \mathbf{C}\mathbf{B}(\mathbf{x},t)\mathbf{u} - \dot{\mathbf{y}}_*(t)], \tag{8}$$

где $\mathbf{C} = \mathbf{C}(\mathbf{x},t) = \partial\mathbf{G}(\mathbf{x},t)/\partial x$, а скоростной градиент и АСГ примут вид соответственно

$$\nabla_u \omega(\mathbf{x},\mathbf{u},t) = \mathbf{B}(\mathbf{x},t)^T \mathbf{C}^T \mathbf{H}[\mathbf{y} - \mathbf{y}_*(t)], \tag{9}$$

$$\frac{d\mathbf{u}}{dt} = -\mathbf{\Gamma}\mathbf{B}(\mathbf{x},t)^T \mathbf{C}^T \mathbf{H}[\mathbf{y} - \mathbf{y}_*(t)]. \tag{10}$$

В качестве матрицы усиления $\mathbf{\Gamma}$ часто берется диагональная ($\mathbf{\Gamma} = \mathrm{diag}\{\gamma_i\}$) или скалярная ($\mathbf{\Gamma} = \gamma\mathbf{I}$) матрица ($\gamma_i$, γ – положительные числа). Алгоритм (10) при $\mathbf{B}(\mathbf{x},t)$ = const и $\mathbf{C}(\mathbf{x},t)$ = const представляет собой хорошо известный интегральный закон регулирования.

Алгоритм управления (3) может конкретизироваться не только в дифференциальной, но и в конечной форме. АСГ в конечной форме имеет вид

$$\mathbf{u}(t) = \mathbf{u_0} - \mathbf{\Gamma}\nabla_{\mathbf{u}}\omega(\mathbf{x}(t),\mathbf{u}(t),t), \tag{11}$$

где $\mathbf{u_0}$ – некоторое начальное (опорное) значение управления. Введем более общую структуру

$$\mathbf{u}(t) = \mathbf{u_0} - \gamma\psi(\mathbf{x}(t),\mathbf{u}(t),t), \tag{12}$$

где $\gamma > 0$ – скалярный множитель шага (коэффициент усиления), а вектор-функция $\psi(\mathbf{x},\mathbf{u},t)$ удовлетворяет условию псевдоградиентности

$$\psi(\mathbf{x},\mathbf{u},t)^T \nabla_{\mathbf{u}}\omega(\mathbf{x},\mathbf{u},t) \ge 0, \tag{13}$$

и назовем (12) алгоритмом скоростного псевдоградиента.

Запись (12) является, вообще говоря, уравнением относительно $\mathbf{u}(t)$. Всюду в дальнейшем будем предполагать, что это уравнение однозначно разрешимо при данном $\mathbf{u}_0$ и любых $\mathbf{x}$, t. Для этого достаточно, например, чтобы функция $\psi(\mathbf{x},\mathbf{u},t)$ не зависела явно от управления или удовлетворяла условию Липшица

$$\left|\psi(\mathbf{x},\mathbf{u},t) - \psi(\mathbf{x},\bar{\mathbf{u}},t)\right| \le L\left|\mathbf{u} - \bar{\mathbf{u}}\right|, \quad L < 1.$$

Очевидно, частным случаем (12) является алгоритм (11), поскольку $\psi(\cdot) = \mathbf{\Gamma}\nabla_{\mathbf{u}}\omega(\cdot)$ удовлетворяет (13). Однако (12) включает и другие структуры, например, знаковый алгоритм

$$\mathbf{u}(t) = \mathbf{u}_0 - \gamma\,\mathrm{sign}\nabla_{\mathbf{u}}\omega(\mathbf{x}(t),\mathbf{u}(t),t), \tag{14}$$

где знак вектора понимается покомпонентно: для $\mathbf{x} = \mathrm{col}(x_1,\ \ldots\ ,\ x_m)$ имеем

$\mathrm{sign}\,\mathbf{x} = \mathrm{col}\,(\mathrm{sign}\ x_1,\ \ldots\ ,\ \mathrm{sign}\,x_m)$.

Выпишем алгоритм скоростного градиента в конечной форме для линейного объекта (4) и квадратичной целевой функции $Q(x) = 0,5(\mathbf{y} - \mathbf{y}_*)^2$ при $\mathbf{u} \in \mathbf{R}^1$, $\mathbf{y} \in \mathbf{R}^1$. Считая $\mathbf{u_0}$= 0, получим алгоритм (11) в виде

$$\mathbf{u}(t) = -\gamma(\mathbf{y}(t) - \mathbf{y}_*),$$

представляющем собой классический пропорциональный закон регулирования. Алгоритм (14) принимает вид

$$\mathbf{u}(t) = -\gamma\,\mathrm{sign}(\mathbf{y}(t) - \mathbf{y}_*),$$

т. е. задает релейный закон регулирования. Выпишем алгоритм (14) для задачи (7), (13) с квадратичной целевой функцией и аффинным по входу объектом управления:

$$\mathbf{u}(t) = \mathbf{u}_0 - \gamma\,\mathrm{sign}[\mathbf{B}(\mathbf{x},t)^T \mathbf{C}\mathbf{H}(\mathbf{y} - \mathbf{y}_*(t))]. \tag{15}$$

Выписанные выше алгоритмы можно использовать во всевозможных сочетаниях. Наиболее употребительными из них являются алгоритм с локально-интегральным функционалом

$$\frac{d\mathbf{u}}{dt} = -\mathbf{\Gamma}\nabla_{\mathbf{u}}[\omega(\mathbf{x},\mathbf{u},t) + \delta R(\mathbf{x},\mathbf{u},t)] \tag{16}$$

(очевидно, (16) является АСГ по отношению к функционалу

$$Q(\mathbf{x},\mathbf{u},t) + \delta\int_0^t R(\mathbf{x},\mathbf{u},\tau)d\tau\,)$$

и АСГ в конечно-дифференциальной форме

$$\frac{d[\mathbf{u} + \gamma\psi(\mathbf{x},\mathbf{u},t)]}{dt} = -\mathbf{\Gamma}\nabla_{\mathbf{u}}\omega(\mathbf{x},\mathbf{u},t)\ .$$

Наконец, наиболее общая форма АСГ имеет вид

$$\frac{d[\mathbf{u} + \gamma\psi(\mathbf{x},\mathbf{u},t)]}{dt} = -\Gamma\nabla_{\mathbf{u}}[\omega(\mathbf{x},\mathbf{u},t) + \delta R(\mathbf{x},\mathbf{u},t)]\ . \tag{17}$$

Алгоритм (17) становится более наглядным в частном случае, если взять

$$\psi(\mathbf{x},\mathbf{u},t) = \nabla_{\mathbf{u}}\omega(\mathbf{x},\mathbf{u},t), \quad R(\mathbf{x},\mathbf{u},t) = \frac{1}{2}\left|\mathbf{u} - \bar{\mathbf{u}}\right|^2, \quad \Gamma = \beta\mathbf{I},$$

где $\bar{u}$ – некоторый постоянный вектор. Тогда $\nabla_{\mathbf{u}}R(\mathbf{x},\mathbf{u},t) = \mathbf{u} - \bar{\mathbf{u}}$, и, меняя местами вторые слагаемые в (17), алгоритм можно представить в виде

$$\frac{d\mathbf{u}}{dt} + \delta(\mathbf{u} - \bar{\mathbf{u}}) = -\beta\nabla_{\mathbf{u}}\omega - \frac{\gamma d(\nabla_{\mathbf{u}}\omega)}{dt}\ . \tag{18}$$

Соотношение (18) задает преобразование вектора $\nabla_{\mathbf{u}}\omega(\mathbf{x},\mathbf{u},t)$ в вектор поправки к управлению $\mathbf{u} - \bar{\mathbf{u}}$. Очевидно, этот закон является линейным и может быть описан передаточной матрицей

$$\mathbf{W}(p) = -\frac{\beta + \gamma p}{p + \delta}\mathbf{I}\ , \tag{19}$$

т.е. алгоритму (18) соответствует матричное интегро-дифференцирующее звено (18) [8, 9].

Применение интегро-дифференцирующего алгоритма для управления синхронизацией двухроторной виброустановки

Расчетная схема для синтеза уравнений динамики двухроторного вибрационного стенда [7] представлена на рис. 1.

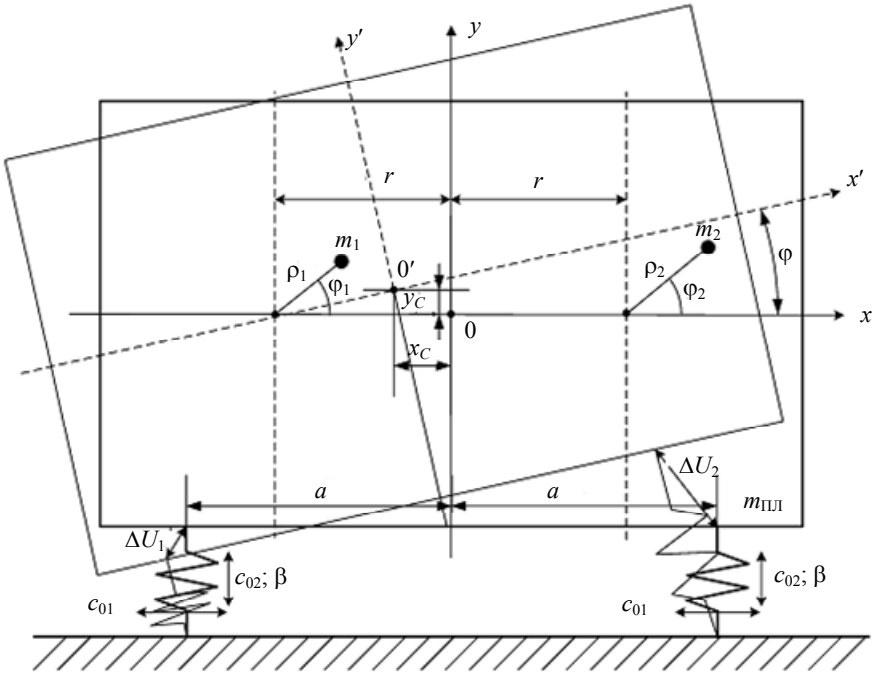


Рис. 1. Расчетная схема двухроторного вибрационного стенда

На рисунке используются следующие обозначения: m_i – масса i-го неуравновешенного ротора; ρ_i – эксцентриситет i-го ротора; $i = 1, 2$; $m_{ПЛ}$ – масса платформы; x_C, y_C – текущие координаты центра масс

платформы – точки 0′; φ – угол поворота платформы; φ_1, φ_2 – углы отклонения неуравновешенных роторов от горизонтального положения; r – расстояние между центром платформы и точками крепления неуравновешенных роторов; β – коэффициент демпфирования пружин. Четыре упругих опоры, на которых закреплена платформа вибрационного стенда, рассматриваются в динамической модели вертикального движения как две пружины с соответствующей эквивалентной жесткостью c_{0i}. В модели не учитывается скручивание пружин. Система координат, связанная с платформой, обозначена $0'x'y'$. Абсциссы крепления пружин $x_{пi} = \pm a$.

Уравнения динамики двухроторной виброустановки, синтезируемые в виде уравнений Лагранжа второго рода, имеют следующий вид [11]:

$$
\begin{gathered}
m_0\ddot{x}_c - \ddot{\varphi}m\rho(\sin(\varphi+\varphi_1)+\sin(\varphi+\varphi_2)) - \ddot{\varphi}_1 m\rho\sin(\varphi+\varphi_1) - \ddot{\varphi}_2 m\rho\sin(\varphi+\varphi_2) - \\
-\dot{\varphi}^2 m\rho(\cos(\varphi+\varphi_1)+\cos(\varphi+\varphi_2)) - \dot{\varphi}_1^2 m\rho\cos(\varphi+\varphi_1) - \dot{\varphi}_2^2 m\rho\cos(\varphi+\varphi_2) - \\
-2\dot{\varphi}\dot{\varphi}_1 m\rho\cos(\varphi+\varphi_1) - 2\dot{\varphi}\dot{\varphi}_2 m\rho\cos(\varphi+\varphi_2) + 2c_{01}x_c + \beta\dot{x}_c = 0; \\
m_0\ddot{y}_c + \ddot{\varphi}m\rho(\cos(\varphi+\varphi_1)+\cos(\varphi+\varphi_2)) - \ddot{\varphi}_1 m\rho\cos(\varphi+\varphi_1) - \ddot{\varphi}_2 m\rho\cos(\varphi+\varphi_2) - \\
-\dot{\varphi}^2 m\rho(\sin(\varphi+\varphi_1)+\sin(\varphi+\varphi_2)) - \dot{\varphi}_1^2 m\rho\sin(\varphi+\varphi_1) - \dot{\varphi}_2^2 m\rho c\sin(\varphi+\varphi_2) - \\
-2\dot{\varphi}\dot{\varphi}_1 m\rho\sin(\varphi+\varphi_1) - 2\dot{\varphi}\dot{\varphi}_2 m\rho\sin(\varphi+\varphi_2) + m_0 g + 2c_{02}y_c + \beta\dot{y}_c = 0; \\
-\ddot{x}_c m\rho(\sin(\varphi+\varphi_1)+\sin(\varphi+\varphi_2)) + \ddot{y}_c m\rho(\cos(\varphi+\varphi_1)+\cos(\varphi+\varphi_2)) + \\
+\ddot{\varphi}\left(J + J_1 + J_2 - 2rm\rho(\cos\varphi_1 - \cos\varphi_2)\right) + \\
+\ddot{\varphi}_1(J_1 - rm\rho\cos\varphi_1) + \ddot{\varphi}_2(J_2 + rm\rho\cos\varphi_2) + \\
\dot{\varphi}_1^2 rm\rho\sin\varphi_1 - \dot{\varphi}_2^2 rm\rho\sin\varphi_2 + 2rm\rho\dot{\varphi}\dot{\varphi}_1\sin\varphi_1 - 2rm\rho\dot{\varphi}\dot{\varphi}_2\sin\varphi_2 + \\
+m\rho g(\cos(\varphi+\varphi_1)+\cos(\varphi+\varphi_2)) + c_{03}\varphi + \beta\dot{\varphi} = 0; \\
-\ddot{x}_c m\rho\sin(\varphi+\varphi_1) + \ddot{y}_c m\rho\cos(\varphi+\varphi_1) + \ddot{\varphi}(J_1 - rm\rho\cos\varphi_1) + \\
+\ddot{\varphi}_1 J_1 - \dot{\varphi}^2 rm\rho\sin\varphi_1 + m\rho g\cos(\varphi+\varphi_1) + \mathrm{k}_c\dot{\varphi}_1 = M_1; \\
-\ddot{x}_c m\rho\sin(\varphi+\varphi_2) + \ddot{y}_c m\rho\cos(\varphi+\varphi_2) + \ddot{\varphi}(J_2 + rm\rho\cos\varphi_1) + \\
\ddot{\varphi}_2 J_2 + \dot{\varphi}^2 rm\rho\sin\varphi_2 + m\rho g\cos(\varphi+\varphi_2) + k_c\dot{\varphi}_2 = M_2;
\end{gathered}
\tag{20}
$$

где M_1, M_2 – управляющие моменты.

Частотная синхронизация понимается как точное совпадение угловых скоростей несбалансированных роторов: $\omega_s = \omega_r$; $s, r = 1, \ldots, k$. Более общий случай – это кратная частотная синхронизация, когда скорости вибровозбудителей пропорциональны [12]:

$\omega_i = n_i \cdot \omega^*$ $(i = 1, \ldots, k)$

для некоторых целых n_i, где $\omega^* > 0$ – синхронная частота.

На практике для случая кратной синхронизации скоростей (средних скоростей) вибровозбудителей имеет смысл рассматривать приближенную синхронизацию:

$$\left|\omega_s - \frac{n_s}{n_r}\cdot\omega_r\right| \le \varepsilon, \tag{21}$$

где $\varepsilon > 0$, численно может быть выбрано $\varepsilon = 0{,}05\,\omega^*$, по аналогии с заданной точностью при традиционном определении времени переходного процесса. Однако соотношение (21) само по себе недостаточно для синхронизации, поскольку его выполнение не запрещает накопления приведенной ошибки синхронизации по фазам (приведенного сдвига фаз). В связи с этим возникает необходимость наложить дополнительные требования на обобщенные координаты (фазы) системы.

Координатная синхронизация, хорошо изученная в физике и механике [1, 7, 12], возникает, когда выходы или некоторые фазовые координаты одной из подсистем вибровозбудителей совпадают с соответствующими координатами других подсистем при всех $t \ge t_0$.

Понятие *кратной координатной синхронизации*, сформулированное в [12, 13], предполагает, что фазы вибровозбудителей φ_i, $i = 1, \ldots, k$ удовлетворяют тождествам

$$\frac{\varphi_s}{n_s} - \frac{\varphi_r}{n_r} = L_{sr}; \quad s, r = 1, \ldots, k. \tag{22}$$

В практических задачах соотношение (22) может выполняться лишь приближенно, т.е. должно быть заменено на

$$\left|\frac{\varphi_s}{n_s} - \frac{\varphi_r}{n_r} - L_{sr}\right| < \varepsilon; \quad s, r = 1, \ldots, k. \tag{23}$$

Будем говорить, что в системе имеет место *приближенная частотно-координатная синхронизация*, если одновременно выполняются соотношения (21) и (23) для некоторых $\varepsilon > 0$, L_{sr}.

Для обеспечения кратного синхронного режима вращения неуравновешенных роторов для системы (20) предлагается выбрать целевой функционал в виде

$$Q(z) = 0,5\left\{(1-\alpha)\left(H - H^*\right)^2 + \alpha\left(\dot{\varphi}_1 / n_1 \pm \dot{\varphi}_2 / n_2\right)^2\right\}, \tag{24}$$

где $0 < \alpha < 1$ – весовой коэффициент, H – полная механическая энергия системы (20).

В соответствии со схемой скоростного градиента, положив $\beta = 0$, $\delta = 0$ в (19), получим из (24) пропорциональный алгоритм (П-алгоритм) управления кратной синхронизацией [14]:

$$\begin{cases} M_1 = -\gamma_1\left\{(1-\alpha)\left(H - H^*\right)\dot{\varphi}_1 + \dfrac{\alpha}{J_1 n_1}\left(\dfrac{\dot{\varphi}_1}{n_1} \pm \dfrac{\dot{\varphi}_2}{n_2}\right)\right\}, \\ M_2 = -\gamma_2\left\{(1-\alpha)\left(H - H^*\right)\dot{\varphi}_2 \pm \dfrac{\alpha}{J_2 n_2}\left(\dfrac{\dot{\varphi}_1}{n_1} \pm \dfrac{\dot{\varphi}_2}{n_2}\right)\right\}, \end{cases} \tag{25}$$

а положив $\delta = 0$, получим пропорционально-интегральный алгоритм (ПИ-алгоритм) управления:

$$\begin{cases} M_1 = -\gamma_1\left\{(1-\alpha)\left[\left(H - H^*\right)\dot{\varphi}_1 + \int\left(H - H^*\right)\dot{\varphi}_1 dt\right] + \dfrac{\alpha}{J_1 n_1}\left(\dfrac{\dot{\varphi}_1}{n_1} \pm \dfrac{\dot{\varphi}_2}{n_2}\right) + \dfrac{\alpha}{J_1 n_1}\left(\dfrac{\varphi_1}{n_1} \pm \dfrac{\varphi_2}{n_2} + C_1\right)\right\}, \\ M_2 = -\gamma_2\left\{(1-\alpha)\left[\left(H - H^*\right)\dot{\varphi}_2 + \int\left(H - H^*\right)\dot{\varphi}_2 dt\right] \pm \dfrac{\alpha}{J_2 n_2}\left(\dfrac{\dot{\varphi}_1}{n_1} \pm \dfrac{\dot{\varphi}_2}{n_2}\right) \pm \dfrac{\alpha}{J_2 n_2}\left(\dfrac{\varphi_1}{n_1} \pm \dfrac{\varphi_2}{n_2} + C_2\right)\right\}. \end{cases}$$

Для упрощения реализации ПИ-алгоритма при синтезе отбросим интегралы от первых членов в правых частей (25) и запишем упрощенный ПИ-алгоритм в виде

$$\begin{cases} M_1 = -\gamma_1\left\{(1-\alpha)\left(H - H^*\right)\dot{\varphi}_1 + \dfrac{\alpha}{J_1 n_1}\left(\dfrac{\dot{\varphi}_1}{n_1} \pm \dfrac{\dot{\varphi}_2}{n_2}\right) + \dfrac{\alpha}{J_1 n_1}\left(\dfrac{\varphi_1}{n_1} \pm \dfrac{\varphi_2}{n_2} + \Delta\varphi_{1\text{зад}}\right)\right\}, \\ M_2 = -\gamma_2\left\{(1-\alpha)\left(H - H^*\right)\dot{\varphi}_2 \pm \dfrac{\alpha}{J_2 n_2}\left(\dfrac{\dot{\varphi}_1}{n_1} \pm \dfrac{\dot{\varphi}_2}{n_2}\right) \pm \dfrac{\alpha}{J_2 n_2}\left(\dfrac{\varphi_1}{n_1} \pm \dfrac{\varphi_2}{n_2} + \Delta\varphi_{2\text{зад}}\right)\right\}. \end{cases} \tag{26}$$

Обоснованием введенного упрощения является малость отброшенных интегралов после переходного процесса по скоростям, который заканчивается достаточно быстро. Поскольку интеграл от функции $\left(\dot{\varphi}_1/n_1 \pm \dot{\varphi}_2/n_2\right)$ равен, вообще говоря, $\left(\varphi_1/n_1 \pm \varphi_2/n_2 + C\right)$, в (26) выбрано $C = \Delta\varphi_{i\text{зад}}$. Задавая $\Delta\varphi_{i\text{зад}}$, можно влиять на установившееся значение приведенного сдвига фаз $\Delta\varphi(\infty)$ и, в конечном итоге, на параметры траектории платформы.

Результаты сравнительного компьютерного исследования П-алгоритма управления синхронизацией (25) и упрощенного ПИ-алгоритма (26) представлены на рис. 2 и рис. 3 соответственно. На графиках используется обозначение для скоростей роторов $\dot{\varphi}_i = \omega_i$. Как видно из рисунков, графики скоростей роторов $\dot{\varphi}_1, \dot{\varphi}_2$ и кратной разности скоростей $\Delta\dot{\varphi} = \dot{\varphi}_1/2 - \dot{\varphi}_2$ в обоих случаях идентичны. Аналогичен и характер изменения кратного сдвига фаз $\Delta\varphi = \varphi_1/2 - \varphi_2$, однако время координатной синхронизации, определяемое по графику сдвига фаз, для ПИ-алгоритма (26) почти в четыре раза меньше, чем для П-алгоритма (25).

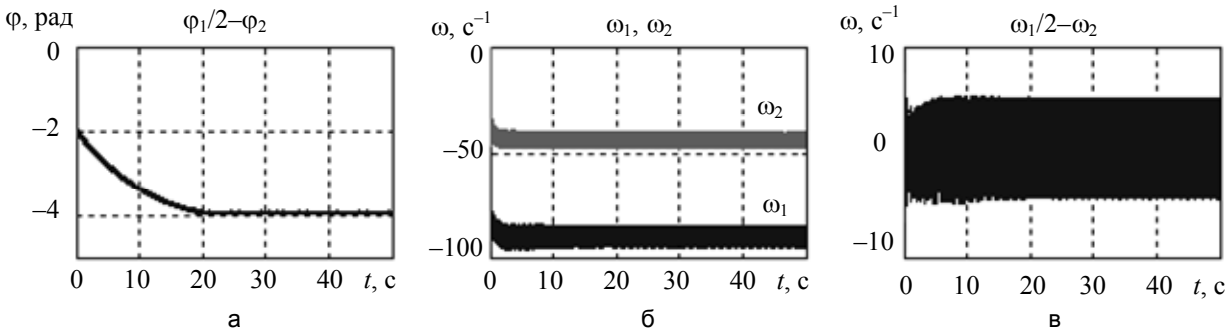


Рис. 2. Динамика двухроторной модели с П-алгоритмом управления при $n_1 = 2$, $n_2 = 1$: график кратной разности фаз роторов $\Delta\varphi = \varphi_1/2 - \varphi_2$ (а); графики изменения скоростей роторов $\dot{\varphi}_1, \dot{\varphi}_2$ (б) и их кратной разности $\Delta\dot{\varphi} = \dot{\varphi}_1/2 - \dot{\varphi}_2$ (в)

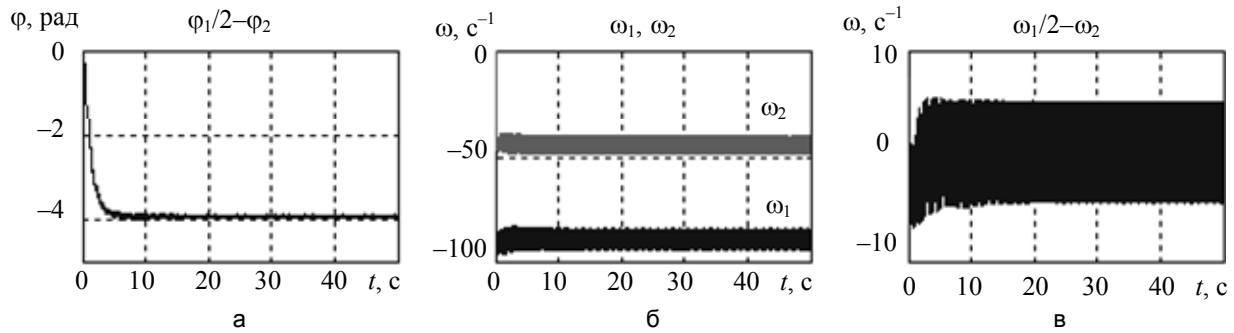


Рис. 3. Динамика двухроторной модели с ПИ-алгоритмом управления при $n_1 = 2$, $n_2 = 1$: график кратной разности фаз роторов $\Delta\varphi = \varphi_1 / 2 - \varphi_2$ (а); графики изменения скоростей роторов $\dot{\varphi}_1$, $\dot{\varphi}_2$ (б) и их кратной разности $\Delta\dot{\varphi} = \dot{\varphi}_1 / 2 - \dot{\varphi}_2$ (в)

Вторым существенным преимуществом ПИ-алгоритма кратной синхронизации (26) по сравнению с П-алгоритмом (25) является возможность управления траекторией платформы в вертикальной плоскости. Известно, что скорость вибротранспортирования определяется установившимся сдвигом фаз роторов, и, например, при $n_1 = 1$, $n_2 = 1$ скорость максимальна в случае нулевого установившегося сдвига фаз $\Delta\varphi = 0$. Но во избежание заторов на разгрузочном конце платформы часто приходится уменьшать скорость вибротранспортирования. Это можно осуществить с помощью изменения установившегося сдвига фаз $\Delta\varphi$. При этом изменяется и угол наклона траектории платформы $\varphi(t)$. Однако, применяя алгоритм (25), можно варьировать лишь отношение скоростей роторов (n_1/n_2). Например, для двукратной синхронизации возможны два варианта: $n_1/n_2 = 1/2$ или $n_1/n_2 = 2/1$. Вид траекторий платформы для указанных случаев заданного соотношения скоростей представлен на рис. 4. В случае же алгоритма (26) задача управления наклоном траектории может быть решена путем задания соответствующих сдвигов фаз $\Delta\varphi_{i\text{зад}}$ в выражениях для управляющих моментов [11]. На рис. 5 представлены график изменения приведенного сдвига фаз $\Delta\varphi(\infty) = \varphi_1 - \varphi_2/2$ (рис. 5, б) и траектория центра масс платформы y_C (x_C) при заданных $\Delta\varphi_{1\text{зад}} = 0$; $\Delta\varphi_{2\text{зад}} = 4$ рад (рис. 5, а).

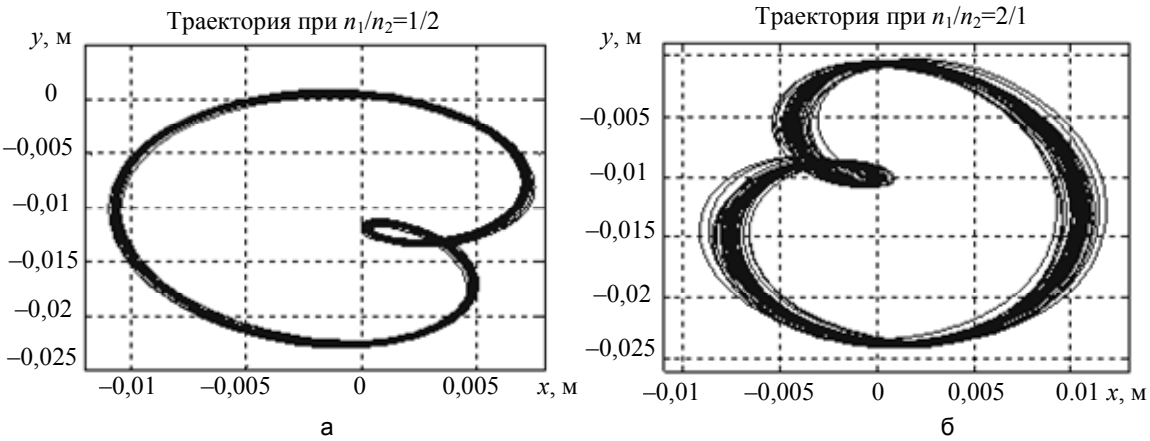


Рис. 4. Траектории центра масс платформы при управляющем алгоритме (25): $n_1/n_2 = 1/2$ (а); $n_1/n_2 = 2/1$ (б)

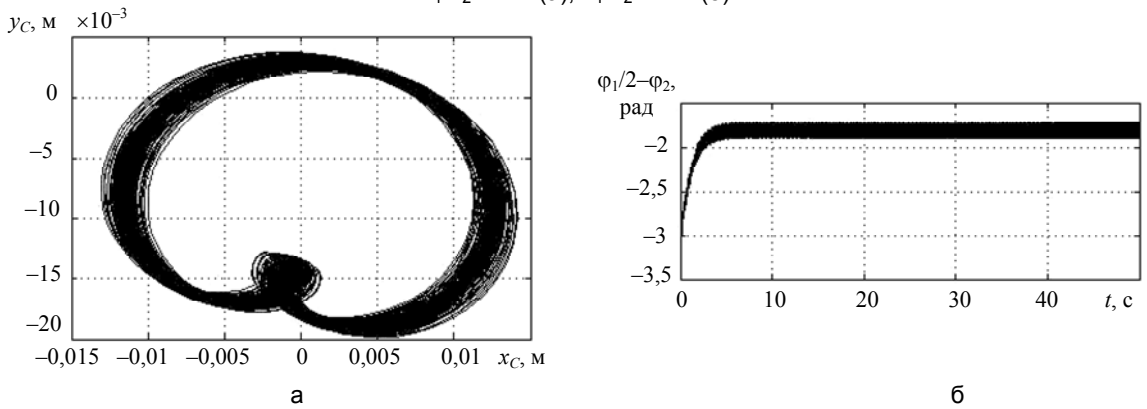


Рис. 5. Траектории центра масс платформы (а) и приведенного сдвига фаз (б) в системе с алгоритмом управления (26): $\Delta\varphi_{1\text{зад}} = 0$; $\Delta\varphi_{2\text{зад}} = 4$ рад

Наконец, в ходе моделирования установлено, что алгоритм (26) существенно менее чувствителен к выбору параметров алгоритма α, γ.

Заключение

В работе синтезированы интегро-дифференцирующие алгоритмы управления процессом кратной синхронизации двухроторных вибрационных установок. Представлено определение приближенной кратной частотно-координатной синхронизации. Сравнительное компьютерное исследование пропорционального и пропорционально-интегрального алгоритмов управления кратной синхронизацией показывает, что предложенные алгоритмы обладают рядом преимуществ перед уже известными. Именно, существенно уменьшается время синхронизации, появляется возможность управления траекторией платформы в вертикальной плоскости и снижается чувствительность алгоритма к выбору параметров.

Литература

1. Блехман И.И. Вибрационная механика. – М.: Наука, 1994. – 440 с.
2. Барзуков О.П. Кратная синхронизация в системе слабосвязанных объектов с одной степенью свободы // Прикладная математика и механика. – 1972. – Т. 36. – С. 225–231.
3. Ярошевич Н.П. К теории синхронизации механических вибровозбудителей, связанных с линейной колебательной системой // Проблемы машиностроения и надежности машин. – 2003. – № 4. – С. 3–10.
4. Блехман И.И., Ярошевич Н.П. Кратные режимы вибрационного поддержания вращения неуровновешенных роторов // Известия АН СССР. Машиноведение. – 1986. – № 6. – С. 62–67.
5. Blekhman I.I., Fradkov A.L., Nijmeijer H. and Pogromsky A.Yu. On self-synchronization and controlled synchronization // Systems and Control Letters. – 1997. – V. 31. – P. 299–305.
6. Мирошник И.В., Никифоров В.О. Методы координации в задачах планирования и управления пространственным движением манипуляционных роботов // Анализ и управление нелинейными колебательными системами / Под ред. Г.А. Леонова и А.Л. Фрадкова. – СПб: Наука, 1998. – С. 215–236.
7. Андриевский Б.Р., Блехман И.И., Борцов Ю.А., Гаврилов С.В., Коноплев В.А., Лавров Б.П., Поляхов Н.Д., Томчина О.П., Фрадков А.Л., Шестаков В.М. Управление мехатронными вибрационными установками. – СПб: Наука, 2001. – 278 с.
8. Мирошник И.В., Никифоров В.О., Фрадков А.Л. Нелинейное и адаптивное управление в сложных динамических системах. – СПб: Наука, 2000. – 549 с.
9. Фрадков А.Л. Интегро-дифференцирующие алгоритмы скоростного градиента // ДАН СССР. – Т. 288. – № 4. – С. 832–835.
10. Фрадков А.Л. Схема скоростного градиента и ее применение в задачах адаптивного управления // Автоматика и телемеханика. – 1979. – № 9. – С. 90–101.
11. Галицкая В.А., Томчина О.П. Пропорционально-интегральный энергоскоростной алгоритм управления кратной синхронизацией роторов вибрационной установки // Информатика и системы управления. – 2012. – № 3. – С. 158–168.
12. Blekhman I.I., Fradkov A.L. On general definitions of synchronization // Selected topics in vibrational mechanics. – Singapore: World Scientific. – 2004. – P. 179–188.
13. Томчина О.П., Томчин Д.А., Горлатов Д.В. Алгоритм управления кратной синхронизацией для виброустановки с роторами, имеющими разные массоинерционные характеристики // Современное машиностроение. Наука и образование: материалы 2-й Международной научно-практической конференции. – СПб: Изд-во Политехн. ун-та, 2012. – С. 752–762.
14. Кудрявцева И.М., Томчина О.П. Алгоритм кратной синхронизации для двухроторного вибрационного стенда с нестационарной нагрузкой // Информатика и системы управления. – 2009. – № 3 (21). – С. 34–44.

Фрадков Александр Львович – ФГБУН Институт проблем машиноведения Российской академии наук (ИПМаш РАН), доктор технических наук, профессор, зав. лабораторией, fradkov@mail.ru

Томчина Ольга Петровна – Санкт-Петербургский государственный политехнический университет, кандидат технических наук, доцент, otomchina@mail.ru

Галицкая Виктория Андреевна – Санкт-Петербургский государственный политехнический университет, аспирант, Kvika86@mail.ru

Горлатов Дмитрий Владимирович – Санкт-Петербургский государственный политехнический университет, аспирант, dprog@mail.ru

УДК 62-50

РОБАСТНОЕ УПРАВЛЕНИЕ РОБОТОМ-МАНИПУЛЯТОРОМ В ДИСКРЕТНОМ ВРЕМЕНИ[6]

С.В. Гусев

Рассматривается проблема управления движением робота-манипулятора в условиях неопределенности относительно его динамических характеристик и при наличии недоступных измерению возмущений. Предложен линейный регулятор для стабилизации программных движений робота-манипулятора, действующий в дискретном времени. Показано, что при уменьшении интервала дискретизации регулятор может обеспечить стабилизацию программного движения с любой наперед заданной точностью.

Ключевые слова: робот-манипулятор, робастное управление, линейный регулятор, дискретное время.

Введение

Существует несколько подходов к решению проблемы управления движением робота-манипулятора в условиях неопределенности: адаптивное управление [1–6], робастное управление [7–9], управление с помощью скользящих режимов [10, 11], управление с помощью нейронных сетей [12, 13]. При этом большинство публикаций посвящено управлению в непрерывном времени. В то же время как измерения, так и управление в современных роботах осуществляются с помощью цифровых устройств, действующих в дискретном времени. По этой причине построение регуляторов в дискретном времени представляет очевидный практический интерес.

Существенно меньше публикаций касается управления роботом-манипулятором в дискретном времени [1, 14–17]. В этих работах строятся нелинейные регуляторы. Цель данной работы – предложить линейный робастный регулятор, действующий в дискретном времени и использующий минимум информации о динамике робота.

Постановка задачи

Динамика манипулятора описывается уравнениями Лагранжа

$$\mathbf{A}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{b}(\dot{\mathbf{q}},\mathbf{q}) = \mathbf{u} + \mathbf{w}(t), \tag{1}$$

где $\mathbf{q} \in \mathbf{R}^n$ – вектор обобщенных координат; $\dot{\mathbf{q}}$ и $\ddot{\mathbf{q}}$ – его первая и вторая производные; $\mathbf{u} \in \mathbf{R}^n$ – вектор обобщенных моментов, развиваемых соответствующими приводами; $\mathbf{w} \in \mathbf{R}^n$ – возмущение; симметричная матрица $\mathbf{A}(\mathbf{q}) \in \mathbf{R}^{n\times n}$ – матрица кинетической энергии; $\mathbf{b}(\dot{\mathbf{q}},\mathbf{q}) \in \mathbf{R}^n$. Относительно возмущения $\mathbf{w}(t)$ предполагается только, что оно ограничено:

$$|\mathbf{w}(t)| \le C.$$

В частности, $\mathbf{w}(t)$ может включать разрывные силы трения. Предположим, что матрица кинетической энергии $\mathbf{A}(\mathbf{q})$ при всех $\mathbf{q}$ удовлетворяет неравенствам

$$\alpha\mathbf{I} \le \mathbf{A}(\mathbf{q}) \le a\mathbf{I}, \tag{2}$$

где $\alpha, a > 0$; $\mathbf{I}$ – единичная матрица размера $n\times n$. Неравенства (2) понимаются как неравенства для квадратичных форм. При построении регулятора будет использована только константа α, знание величин a и C не требуется.

Предполагается, что задано программное движение $\mathbf{p}(t) \in \mathbf{R}^n$, которое имеет непрерывные и ограниченные первую и вторую производные. Целью управления является отслеживание программного движения с заданной точностью.

Особенностью данной работы является то, что управление и измерения происходят в дискретном времени. Пусть задано некоторое $\delta > 0$, определяющее интервал дискретизации времени. Предполагается, что в дискретные моменты $t_k = k\delta$, $k = 0,1,2,\ldots$, измеряются обобщенные координаты $\mathbf{q}_k = \mathbf{q}(t_k)$ и скорости $\dot{\mathbf{q}}_k = \dot{\mathbf{q}}(t_k)$. Вычисление управления $\mathbf{u}(t)$, используемого на интервале $[t_k, t_{k+1}) = [k\delta, (k+1)\delta)$, также осуществляется в моменты t_k. Управление имеет следующий вид:

$$\mathbf{u}(t) = \begin{cases} \mathbf{u}_k^1, \text{ где } t \in [k\delta, k\delta + \delta/2), \\ \mathbf{u}_k^2, \text{ где } t \in [k\delta + \delta/2, (k+1)\delta). \end{cases} \tag{3}$$

Здесь $\mathbf{u}_k^1$ – управление на первой половине, а $\mathbf{u}_k^2$ – управление на второй половине интервала $[t_k, t_{k+1})$.

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

Рассмотрим векторы $\mathbf{X}_k = \begin{pmatrix} \dot{\mathbf{q}}(t_k) \\ \mathbf{q}(t_k) \end{pmatrix}$, $\mathbf{Z}_k = \begin{pmatrix} \dot{\mathbf{p}}(t_k) \\ \mathbf{p}(t_k) \end{pmatrix}$, $k = 0,1,2,\ldots$, определяющие значения реального и программного фазового вектора системы в момент t_k. Роль нового управления в системе, функционирующей в дискретном времени, будет играть вектор $\mathbf{U}_k = \begin{pmatrix} \mathbf{u}_k^1 \\ \mathbf{u}_k^2 \end{pmatrix}$. При построении стабилизирующего управления рассмотрим два случая: управление при отсутствии и при наличии начального рассогласования между программным и реальным движением.

Управление при отсутствии начального рассогласования

Закон управления определим соотношением

$$\mathbf{U}_k = \gamma \mathbf{G}(\delta)(\mathbf{X}_k - \mathbf{Z}_k), \tag{4}$$

где $\gamma > 0$, $\mathbf{G}(\delta) = \begin{pmatrix} -3\delta^{-1}\mathbf{I} & -4\delta^{-2}\mathbf{I} \\ \delta^{-1}\mathbf{I} & 4\delta^{-2}\mathbf{I} \end{pmatrix}$ – матрица размера $2n \times 2n$.

Теорема 1. *Пусть*

$$\gamma < 2\alpha. \tag{5}$$

Найдутся такие константы $\Delta > 0$ *и* $K > 0$, *что для любого начального состояния* $\mathbf{X}_0$ *и для всех* $\delta \le \Delta$ *неравенства*

$$|\mathbf{q}(t) - \mathbf{p}(t)| \le K\delta^2, \quad |\dot{\mathbf{q}}(t) - \dot{\mathbf{p}}(t)| \le K\delta, \quad |\mathbf{u}(t)| \le K,$$

выполнены на решениях замкнутой системы управления (1), (3), (4) *при всех* $t \ge 0$, *при условии*, *что реальное состояние совпадает с программным в начальный момент*, *т.е.*

$$\mathbf{X}_0 = \mathbf{Z}_0. \tag{6}$$

В принципе условие (6) не является ограничительным, так как всегда можно построить удовлетворяющее ему программное движение. Однако обычно программное движение задано заранее, и средства для его изменения не предусмотрены в системе управления. Исходя из этого, представляет интерес построить управление, обеспечивающее стабилизацию программного движения без предположения (6).

Управление при наличии начального рассогласования

Пусть вектор $\mathbf{r}(t) \in \mathbf{R}^n$ есть решение устойчивого линейного дифференциального уравнения

$$\ddot{\mathbf{r}}(t) + d_1\dot{\mathbf{r}}(t) + d_0\mathbf{r}(t) = 0 \tag{7}$$

с начальными данными

$$\mathbf{r}(0) = \mathbf{q}(0) - \mathbf{p}(0), \dot{\mathbf{r}}(0) = \dot{\mathbf{q}}(0) - \dot{\mathbf{p}}(0). \tag{8}$$

Вектор $\mathbf{r}(t)$ представляет собой желаемую невязку между реальным и программным движением в переходном процессе. В силу устойчивости уравнения (7) $\mathbf{r}(t) \to 0$ при $t \to \infty$.

Пусть $\mathbf{Y}_k = \begin{pmatrix} \dot{\mathbf{r}}(t_k) \\ \mathbf{r}(t_k) \end{pmatrix}$. Из (7) следует, что

$$\mathbf{Y}_k = \mathbf{T}\mathbf{Y}_{k-1}, k = 1,2,\ldots, \tag{9}$$

где $\mathbf{T} = \exp\left(\delta \begin{pmatrix} -d_1\mathbf{I} & -d_0\mathbf{I} \\ \mathbf{I} & 0 \end{pmatrix}\right)$. Начальное условие (8) принимает вид

$$\mathbf{Y}_0 = \mathbf{X}_0 - \mathbf{Z}_0. \tag{10}$$

Определим закон управления соотношением

$$\mathbf{U}_k = \gamma \mathbf{G}(\delta)(\mathbf{X}_k - \mathbf{Z}_k - \mathbf{Y}_k), k = 0,1,2,\ldots, \tag{11}$$

где векторы $\mathbf{Y}_k$ вычисляются рекуррентно в силу (9), (10).

Теорема 2. *Пусть выполнено неравенство* (5). *Найдутся такие константы* $\Delta > 0$ *и* $K > 0$, *что для любого начального состояния* $\mathbf{X}_0$ *и для всех* $\delta \le \Delta$ *неравенства*

$$|\mathbf{q}(t) - \mathbf{p}(t) - \mathbf{r}(t)| \le K\delta^2, |\dot{\mathbf{q}}(t) - \dot{\mathbf{p}}(t) - \dot{\mathbf{r}}(t)| \le K\delta, |\mathbf{u}(t)| \le K,$$

выполнены на решениях замкнутой системы управления (1), (3), (9)–(11) *при всех* $t \ge 0$.

Меняя коэффициенты d_0, d_1 уравнения (7), можно задавать желаемый (например, неколебательный) характер переходного процесса в замкнутой системе. В частности, за счет роста перерегулирования

можно сделать время переходного процесса сколь угодно малым, и, наоборот, за счет увеличения времени переходного процесса можно уменьшить перерегулирование.

Заключение

Предложен регулятор для стабилизации программных движений робота-манипулятора, учитывающий специфику современных цифровых систем управления, – регулятор, действующий в дискретном времени. Регулятор является линейным, он использует дискретизованные по времени значения обобщенных координат и скоростей робота и строит кусочно-постоянное управление. Для расчета регулятора необходимо знать единственный параметр робота-манипулятора – нижнюю оценку собственных чисел матрицы кинетической энергии. Это обеспечивает робастность регулятора к изменениям параметров робота, что важно во многих практических задачах. В дальнейшем представляет интерес рассмотреть возможность построения аналогичного регулятора, использующего только измерения обобщенных координат робота.

Литература

1. Гусев С.В., Якубович В.А. Алгоритм адаптивного управления роботом-манипулятором // Автоматика и телемеханика. – 1980. – № 9. – С. 101–111.
2. Craig J., Hsu P., Sastry S. Adaptive control of mechanical manipulators // Proceedings of the IEEE International Conference on Robotics and Automation. – 1986. – V. 3. – P. 190–195.
3. Slotine J.-J.E., Li W. On the Adaptive Control of Robot Manipulators // The International Journal of Robotics Research. – 1987. – V. 6. – P. 49–59.
4. Midleton P., Goodwin G.C. Adaptive computed torque control for rigid link manipulations // System and Control Letters. – 1988. – V. 10. – P. 9–16.
5. Cheah C.C., Liu C., Slotine J.-J.E. Adaptive Tracking Control for Robots with Unknown Kinematic and Dynamic Properties // The International Journal of Robotics Research. – 2006 – V. 25. – P. 283–296.
6. Huang A.C., Chien M.C. Adaptive Control of Robot Manipulators: A Unified Regressor-Free Approach. – World Scientific. – 2010. – 262 p.
7. Gusev S.V. Linear stabilization of nonlinear system program motion // Systems and Control Letters. – 1988. – V. 11. – P. 409–412.
8. Qu Z., Dorsey J. Robust PID control for robots // International Journal of Robotics and Automation. – 1991. – V. 6. – P. 228–235.
9. Herman P., Franelak D. Robust tracking controller with constraints using generalized velocity components for manipulators // Transactions of the Institute of Measurement and Control. – 2008. – V. 30. – P. 101–113.
10. Пятницкий Е.С., Дунская Н.В. Стабилизация управляемых механических и электромеханических систем // Автоматика и телемеханика. – 1988. – № 12. – С. 40–51.
11. Garcна-Rodrнguez R., Parra-Vega V. Cartesian sliding PID control schemes for tracking robots with uncertain Jacobian // Transactions of the Institute of Measurement and Control. – 2012. – V. 34. – P. 448–462.
12. Barambones O., Etxebarria V. Robust neural control for robotic manipulators // Automatica. – 2002. – V. 38. – P. 235–242.
13. Jiang Z.-H., Ishita T. A Neural Network Controller for Trajectory Control of Industrial Robot Manipulators // Journal of Computers. – 2008. – V. 3. – № 8. – P. 1–8.
14. Middleton R.H. Adaptive control for robot manipulators using discrete time identification // IEEE Transactions on Automatic Control. – 1990. – V. 35. – P. 633–637.
15. Yang S.-P., Woo P.-Y., Wang R. Discrete-Time Model Reference Adaptive Controller Designs for Robotic Manipulators // Proceedings of the American Control Conference. – 1993. – P. 1145–1149.
16. Sun F.C., Sun Z.Q., Zhang R.J., Chen Y.B. Discrete-time tracking control of robotic manipulators based on dynamic inversion using dynamic neural networks // Proceedings of the IEEE International Symposium on Intelligent Control. – 2000. – P. 333–338.
17. Corradinia M.L., Fossib V., Giantomassib A., Ippolitib G., Longhib S., Orlandob G. Discrete time sliding mode control of robotic manipulators: Development and experimental validation // Control Engineering Practice. – 2012. – V. 20. – P. 816–822.

Гусев Сергей Владимирович – Санкт-Петербургский государственный университет, кандидат физ.-мат. наук, ст. научный сотрудник, доцент, gusev@ieee.org

УДК 681.51

АНАЛИЗ И ОБРАБОТКА СИГНАЛОВ ЭЛЕКТРОМИОГРАММЫ[7]

К.А. Зименко, А.С. Боргуль, А.А. Маргун

Представлен метод обработки и идентификации сигналов электромиограммы для последующего использования в управлении устройствами реабилитации. Метод основан на фильтрации высокочастотной составляющей сигнала, улучшающей соотношение сигнал/шум, использовании вейвлет-преобразования для предварительной обработки сигнала и классификации типа движения с использованием обучаемой искусственной нейронной сети. Достигнутая точность определения типа движения составила 94%.

Ключевые слова: электромиограмма, нейронная сеть, обработка сигналов, вейвлет-преобразование.

На сегодняшний день актуальна задача социальной адаптации инвалидов и лиц, страдающих нарушениями опорно-двигательного аппарата. Электромиография – метод исследования биоэлектрических потенциалов, возникающих в скелетных мышцах животных и человека при возбуждении мышечных волокон. Электромиография используется для оценки функционального состояния двигательного аппарата при восстановлении нарушенной двигательной функции, в ортопедии и протезировании. Информация, содержащаяся в сигнале электромиограммы (ЭМГ), может использоваться при создании и использовании активных протезов, экзоскелетов и других устройств, управление которыми происходит посредством изменения биопотенциалов мышц. Анализ ЭМГ-сигналов позволяет определить планируемое движение человека для приведения в действие исполнительного механизма (например, электродвигателя протеза или экзоскелета).

Успешность реализации предложенных идей во многом определяется степенью достоверности расшифровки биопотенциалов мышцы при планируемом движении. Уровень сигнала в диапазоне частот до 500 Гц лежит в пределах 0–10 мВ [1]. Сегодня точной классификации типа движения препятствует низкое отношение сигнал/шум в измерительной системе. Большинство шумов, присутствующих в сигнале, высокочастотны: шумы электронных устройств, окружающее электромагнитное излучение, паразитные взаимонаводки соединительных проводов электродов. Искажения сигнала могут возникать из-за побочного влияния сигналов электрической активности сердца, сдвигов электродов относительно назначенного положения, изменения биопотенциалов на референте при использовании референтной схемы измерения. Наименее искаженный и достоверный сигнал ЭМГ находится в диапазоне частот 15–150 Гц. Для получения такого сигнала авторами применен фильтр низких частот, выполненный на аппаратном уровне.

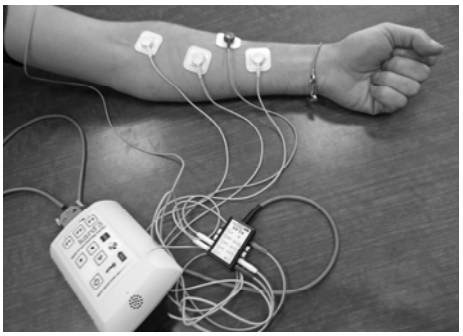
Для успешной классификации типа движения рассмотрена возможность применения нечеткой логики, метода слепого разделения источников, генетических алгоритмов и др.

Из существующего многообразия способов классификации был выбран метод искусственных нейронных сетей [2] благодаря малому времени обучения. Наилучший результат показал однослойный перцептрон. Использование многослойных сетей затруднено из-за влияния на работу сети высокого уровня шумов сигнала. Для выбранной нейронной сети применена регуляризируемая функция издержек, при нахождении минимума которой использовался метод сопряженных градиентов с алгоритмом Полака–Рибери для нахождения направления поиска [3]. Суть алгоритма Полака–Рибери состоит в следующем: генерируется последовательность так называемых рабочих векторов **g** и направлений поиска **h**, таких, что **h** взаимно сопряжены. Причем доказано, что если функция издержек, которую необходимо минимизировать, может быть выражена квадратичной формой, то минимизация вдоль первых **h** направлений приведет к точному минимуму. Совместно с методом поиска начального размера шага использовался критерий остановки Вульфа–Пауэлла.

Для подтверждения работоспособности предложенного алгоритма проведена серия экспериментов. На мышцы предплечья операторов устанавливались четыре электрода, пятый электрод крепился на плечо и выступал в качестве референта (рис. 1, а). Данные с электродов снимались с помощью девятиканального усилителя KARDi3/9, предназначенного для длительной регистрации электрофизиологических сигналов высокого разрешения в полосе частот 0–350 Гц, и обрабатывались на персональном компьютере. Операторами выступали пять здоровых мужчин в возрасте от 19 до 24 лет. Операторы выполняли четыре типа движения: сгибание кисти вперед–назад и вращение кистью по и против часовой стрелки. Для обучения нейронной сети каждый из операторов выполнял по пять раз каждое из определяемых движений. После обучения классификатор должен был распознавать аналогичные движения оператора с точностью не менее 90%.

Достигнутая на первом этапе точность определения типа движения составила 87%, что недостаточно для управления устройствами реабилитации. Повышению точности определения типа движения способствует предварительная обработка сигнала с использованием Фурье- или вейвлет-преобразования.

[7] Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (государственный контракт № 14.740.11.1080).

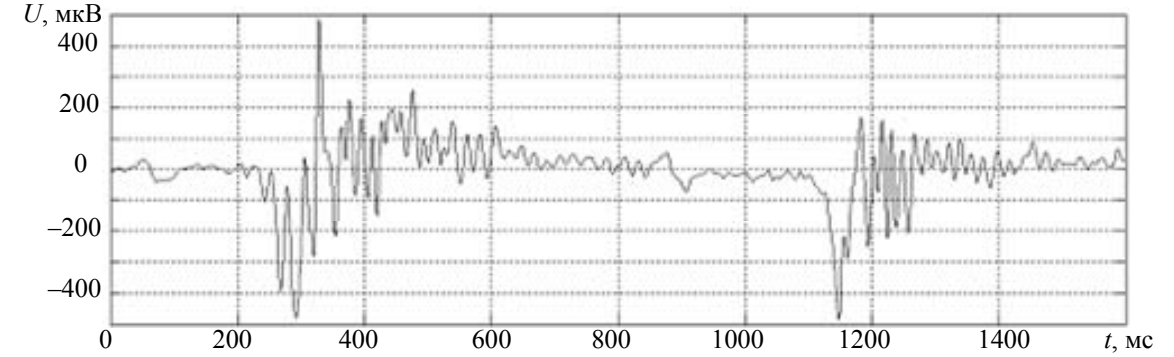


а

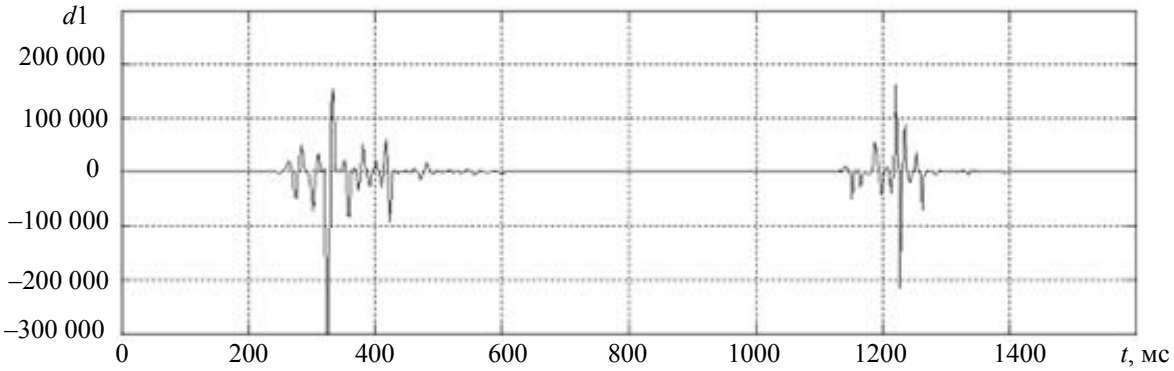
б

Рис. 1. Схема измерения биопотенциалов (а) и внешний вид регистратора биопотенциалов Kardi3/9 (б)

В связи с тем, что время на обучение нейронной сети и, соответственно, время определения типа движения должно быть минимальным, авторы остановились на вейвлет-анализе с использованием вейвлета Хаара. В отличие от преобразования Фурье, анализирующая функция которого покрывает всю временную ось, двухпараметрическая анализирующая функция одномерного вейвлет-преобразования хорошо локализована и во времени, и по частоте. Помимо этого, при декомпозиции сигнала использование детализирующей функции позволяет уменьшить размерность сигнала вдвое, а также избавиться от низкочастотных шумов, вызванных изменением потенциала на референте и побочным влиянием сигналов электрической активности сердца. Уменьшение размерности сигнала позволяет минимизировать размеры нейронной сети и время ее обучения, что является важным фактором при управлении устройствами реабилитации. Тип планируемого движения определялся по данным, полученным за последние 200 мс, что при частоте дискретизации сигнала 1000 Гц с учетом уменьшения размерности после вейвлет-преобразования составляет 100 отсчетов для каждого из четырех электродов. На рис. 2 представлен сигнал ЭМГ до и после обработки с использованием вейвлет-преобразования.



а



б

Рис. 2. Сигнал ЭМГ по одному из каналов до (а) и после обработки (б);
*d*1 – коэффициент детализации (отн. ед.)

Комбинация фильтрации сигнала по частоте, вейвлет-преобразования исходного сигнала и его классификации на основе искусственной нейронной сети позволила повысить точность распознавания движений до 94%. Предложенный метод обработки сигнала электромиограммы не вносит значительных временных задержек, что актуально для систем, используемых в реабилитации.

Алгоритм обладает высокой скоростью обучения, сравнительно прост в применении и позволяет достичь требуемой точности классификации движений оператора.

Литература

1. Rohtash Dhiman et al. Detecting the useful electromyogram signals–extracting, conditioning and classification // IJCSE. – Aug.–Sep. 2011. – V. 2. – № 4. – P 634–637.
2. Borgul A.S., Margun A.A., Zimenko K.A., Kremlev A.S., Krasnov A.Y. Intuitive Control for Robotic Rehabilitation Devices by Human-Machine Interface with EMG and EEG Signals // 17th international conference on Methods and Models in Automation and Robotics (MMAR 2012). Proceedings. Międzyzdroje: IEEE Xplore digital library, 2012. – P. 308–311.
3. Лемешко Б.Ю. Методы оптимизации. – Новосибирск: Изд-во НГТУ, 2009. – 126 с.

Зименко Константин Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, kostyazimenko@gmail.com

Боргуль Александр Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, borgulalexandr@gmail.com

Маргун Алексей Анатольевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, alexeimargun@gmail.com

УДК 378.162, 004.896, 681.51

ВНЕДРЕНИЕ КОМПЛЕКСОВ ПРОМЫШЛЕННЫХ МАНИПУЛЯЦИОННЫХ РОБОТОВ В ОБРАЗОВАТЕЛЬНЫЙ ПРОЦЕСС[8]

А.А. Бобцов, С.А. Колюбин, А.А. Пыркин

Рассматривается использование программно-аппаратных комплексов на базе манипуляционных роботов в учебном процессе при проведении практических работ по современным технологиям робототехники и теории автоматического управления.

Ключевые слова: манипуляционный робот, система управления, образовательные приложения.

В настоящее время манипуляционные роботы широко используются в различных отраслях промышленности, включая производство автомобилей, теплового оборудования, микроэлектронной техники и даже лесозаготовку. Благодаря интеллектуализации систем управления промышленными манипуляторами достигается полная автоматизация различных операций, начиная с перемещения грузов, упаковки и сортировки и заканчивая сваркой, покраской, обработкой поверхностей и сборкой сложных технологических узлов. На фоне снижения потребности в персонале, занятом физическим трудом, спрос на специалистов, способных настраивать и обслуживать сложные робототехнические линии, со стороны компаний неуклонно растет. Очевидно, что одним из важнейших условий успешного внедрения передовых технологий является своевременная подготовка кадров соответствующей квалификации. Задача ведущих университетов здесь состоит в разработке и адаптации новых образовательных программ, и залогом достижения этой цели является наличие современного оборудования.

Благодаря программе развития НИУ ИТМО кафедра систем управления и информатики (СУиИ) была укомплектована двумя комплексами промышленных манипуляционных роботов. Один из них включает шестиосевой манипулятор Kawasaki FS06N с контроллером D77, переносным пультом управления с сенсорным дисплеем и дополнительно установленным электромеханическим параллельным двухпальчиковым захватом Schunk EGN-100. Второй комплекс представляет собой робототехническую ячейку с интегрированной системой технического зрения немецкой компании FESTO на базе компактного манипулятора Mitsubishi RV-3SDB с двухпальчиковым пневматическим схватом. Системы имеют богатый интерфейс, включая возможность подключения Ethernet, что облегчает сопряжение с имеющимся компьютерным оборудованием и внедрение в образовательный процесс. Кроме того, за счет интеграции с дополнительными внешними сенсорами, в том числе многокомпонентными датчиками сил и моментов и системами технического зрения на базе нескольких камер, функционал комплексов может быть существенно расширен.

Эти комплексы могут использоваться как в образовательном процессе при выполнении студентами лабораторных работ, например, в рамках магистерских курсов «Динамика робототехнических систем»,

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

«Управление неполноприводными роботами» и «Современная теория систем управления», разрабатываемых кафедрой СУиИ, так и в полноценных научных исследованиях. Промышленные приложения, которые можно имитировать с помощью данных комплексов, включают сварку сложных профилей, сортировку деталей, пакетирование и сборочные операции с одновременным автоматическим визуальным контролем.

Внедрение подобного оборудования в учебный процесс способствует формированию у студентов таких ключевых профессиональных компетенций, как, например, способность применять современные теоретические и экспериментальные методы разработки математических моделей исследуемых объектов и процессов, относящихся к профессиональной деятельности по направлению подготовки, и способность самостоятельно приобретать и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности.

Важной составляющей использования подобных лабораторных комплексов в учебных целях является специализированное программное обеспечение.

Системы управления промышленных роботов, как правило, поддерживают модульное программирование. В этом случае от студента требуется разбить задачу на элементарные составляющие, записать последовательность соответствующих инструкций, включая циклы и ветвления, на поддерживаемом высокоуровневом языке и загрузить получившуюся программу в память контроллера робота. Сделать это можно непосредственно с пульта удаленного управления робота, не прибегая к подключению к внешнему компьютеру. Это позволяет студенту быстро и легко освоить основы интерфейса управления промышленными роботами, однако такой метод приемлем только для достаточно простых задач.

Следующим уровнем обучения является освоение специализированных интегрированных средств, реализующих визуальное программирование робота. Для описываемых комплексов на базе манипуляторов Kawasaki и Mitsubishi используются программы PC-ROSET и CIROS Automation Suite соответственно. Преимущество подобного программного обеспечения заключается в том, что студент может самостоятельно создать виртуальную графическую модель робота и его окружения и предварительно проверить работоспособность своей программы в симуляторе без риска повредить дорогостоящее оборудование.

Расширение возможностей использования таких робототехнических комплексов в учебных целях и прикладных исследованиях связано с внедрением открытых программных архитектур. Важно отметить, что навстречу разработчикам в этом направлении сейчас движутся крупнейшие производители [1, 2].

Типовые планы лабораторных занятий с использованием комплексов на базе роботов-манипуляторов могут быть следующими.

1. Прямое программирование робота (обучение) с использованием пульта

- Задать движение робота в ручном режиме:
  - независимое управление приводами сочленений;
  - линейное и дуговое перемещение и вращение по осям декартовых координат в системах отсчета относительно базы робота и центральной инструментальной точки;
  - выставление робота в исходную позицию для калибровки кинестетических сенсоров.
- Разработать программу перемещения по фиксированной траектории при свободном рабочем пространстве:
  - от начальной точки к конечной;
  - циклическое перемещение по замкнутым контурам, соответствующим геометрическим примитивам, арабским цифрам или печатным буквам.
- Оценить точность отработки заданной траектории при вариации скорости перемещения.

2. Программирование робота в визуальной среде

- Создать виртуальную модель рабочей зоны манипулятора с учетом имеющихся ограничений.
- Разработать программу управления манипулятором для следующих сценариев:
  - перемещение объектов с одной стороны препятствия на другую с оптимизацией траекторий подхода и переноса;
  - имитация сварочных операций – проход контрольным щупом по заданному контуру объемной детали с сохранением контакта;
  - сортировка деталей одинаковой формы по цветам с использованием внешней камеры;
  - сборка механических узлов в заданной последовательности из нескольких деталей разной формы, в том числе упругих;
  - поиск с помощью технического зрения и сбор по рабочему пространству за минимальное время деталей заданной геометрической формы и (или) цвета, изначально расположенных в неизвестных позициях.

– Провести моделирование программы в графическом симуляторе, включая проверку возможных столкновений и достижимость последовательности конфигураций манипулятора.
– Провести апробацию программы на робототехническом комплексе.

Массогабаритные параметры и эксплуатационные характеристики подобного оборудования позволяют проводить его установку в помещениях без специальной подготовки. В то же время до начала работы студенты должны быть ознакомлены с правилами работы и техникой безопасности, включая возможные риски, связанные с поражением электрическим током и травмами от подвижных элементов робота.

Таким образом, использование в курсах лабораторных работ современных комплексов на базе промышленных роботов позволяет, во-первых, формировать у студентов знания и умения, которые будут востребованы в дальнейшем на практике, а во-вторых, делает сам процесс обучения более интересным и плодотворным.

Литература

1. Schreiber G., Stemmer A., Bischoff R. The Fast Research Interface for the KUKA Lightweight Robot // Proc. of ICRA 2010 Workshop on Innovative Robot Control Architectures for Demanding (Research) Applications. – Anchorage, 2010. – P. 15–21.
2. Blomdell A., Dressler I., Nilsson K., Robertsson A. Flexible Application Development and High-Performance Motion Control Based on External Sensing and Reconfiguration of ABB Industrial Robot Controllers // Proc. of ICRA 2010 Workshop on Innovative Robot Control Architectures for Demanding (Research) Applications. – Anchorage, 2010. – P. 62–66.

Бобцов Алексей Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, bobtsov@mail.ifmo.ru

Колюбин Сергей Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, s.kolyubin@gmail.com

Пыркин Антон Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, a.pyrkin@gmail.com

УДК 528.851; 519.7

ВОПРОСЫ ОПТИМИЗАЦИИ ПРИМЕНЕНИЯ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ ДЛЯ ПОИСКА И СЛЕЖЕНИЯ ОБЪЕКТОВ

Р.Н. Абдулов, Н.А. Абдуллаев, Х.Г. Асадов

Проведен критический анализ известного метода разницы времени прибытия сигнала, применяемого для поиска и слежения за объектами с использованием беспилотного летательного аппарата. Результаты проведенных исследований показали возможности его оптимизации и дальнейшего синтеза качественно нового метода. Проведенная оптимизация известного метода поиска и слежения за объектом на базе группы беспилотных летательных аппаратов позволила выработать практические рекомендации по использованию синтезированного оптимального метода.

Ключевые слова: оптимизация, БПЛА, поиск, локация, измерения, корреляция.

Введение

Хорошо известно, что операции геолокации с применением беспилотного летательного аппарата (БПЛА) выполняются при решении многих комплексных задач как мирного, так и военного характера. Например, как указывается в работе [1], общая концепция возможных задач, выполняемых БПЛА, создает высокий практический и технологический потенциал для усиления возможностей военно-воздушных сил. Так, согласно прогнозу, данному в [1], в близком будущем БПЛА будут применяться в разведывательных целях, и полноценное оснащение БПЛА вооружением является задачей отдаленного будущего, когда БПЛА будут выполнять задачи геолокации, а также уничтожения цели без точного взаимодействия с пилотируемыми воздушными средствами. При этом, как указывается в [1], операции геолокации могут быть реализованы с помощью БПЛА при решении таких задач, как исследование мест применения оружия массового поражения; подготовка вооруженного нападения на стационарные или вражеские движущиеся объекты; операции идентификации, распознавания, разведки, поиска, спасения и т.д. Например, согласно [2], БПЛА, ранее использованные в Ираке и в Афганистане, в дальнейшем нашли применение в Новом Орлеане для поиска людей, оставшихся в наводненных кварталах этого города во время сильного урагана Катрина. В этой же работе изложены основы построения системы поиска и спасения «CLOSE-SEARCH», представляющей собой интеграцию БПЛА, устройства термовидения, мультисенсорной навигационной системы для проведения поисковых работ.

Вместе с тем существуют алгоритмы, позволяющие выполнить задачу позиционирования БПЛА без средств визуального наблюдения. Проведенный анализ известного метода размещения БПЛА вокруг цели и оценки разности времен прихода сигналов с цели показал возможность оптимизации этого метода и синтеза наилучшего варианта. Даны практические рекомендации по практическому использованию синтезированного оптимального метода.

Критический обзор существующих методов

В качестве примера реализации группового поиска и слежения движущихся целей с помощью группы БПЛА можно привести работу [3], где один оператор осуществляет контроль над многими БПЛА, которые, в свою очередь, реализуют следующие операции:

– автономное слежение за двигающимися и удаляющимися объектами;
– выдача отчета в централизованную базу данных.

Эта система осуществляет измерения экранных координат объекта в реальном времени и совместно с данными высоты, позиции и инерциальной навигационной системы формирует статистически точную оценку местонахождения объекта. Основным элементов в системе является бортовая передающая камера, управляемая многоступенчатой системой управления. Система имеет три контура обратной связи:

1. контур обратной связи в цепи гироскопа;
2. контур обратной связи GPS;
3. контур обратной связи видеокамеры.

Кроме этого, в системе предусматривается реализация алгоритма оптимизации траектории полета по критерию достижения максимального количества информации, полученной от датчиков. При этом алгоритм управления полетом должен одновременно удовлетворять двум условиям геолокации: во-первых, цель должна находиться в пределах поля зрения камеры, и, во-вторых, заданный часовой стрелочный угол между двумя БПЛА должен быть соблюден для наилучшей реализации геолокации. Что касается геолокации неподвижных объектов, то, как отмечается в работе [4], здесь возможно использование как георeференцированных спутниковых изображений объекта, так и изображений, полученных от бортовых датчиков, расположенных на малой высоте. В первом случае важен учет фактора изменения окружающей среды, чтобы исключить возможные ошибки геолокации.

Как было сказано выше, одним из методов рационального позиционирования БПЛА в окрестности объектов слежения является использование телевизионных камер. В то же время предложены эффективные алгоритмы, позволяющие выполнить эту задачу без использования телевизионных камер. Например, в работе [5] предлагается оптимизировать позиционирование БПЛА вокруг цели по разности времен прихода сигналов от цели. В этом случае задача формулируется следующим образом. Рассматриваются два БПЛА с приемниками сигналов. Один из приемников, например, первый БПЛА, принимает прямой сигнал $x_1(t)$ без эффекта многопутности, а второй приемник принимает суммарный сигнал $x_2(t)$, составленный из отдельных сигналов, прибывших по N различными путями распространения. Тогда

$$x_1(t) = s(t) + m_1(t), \tag{1}$$

$$x_2(t) = \sum_{n=1}^{N} a_n s(t - D_n) + m_2(t), \tag{2}$$

где $s(t)$ – интересующий нас переданный сигнал; $m_1(t)$ и $m_2(t)$ – аддитивные шумовые сигналы; a_n – ослабление сигнала при распространении по пути n; D_n – относительная временная задержка сигнала по пути n. При этом временная разница прибытия сигнала вычисляется путем нахождения такой величины задержки, при которой перекрестная корреляционная функция, вычисляемая по формуле

$$R_{x_1x_2}(\tau) = E\left[x_1(t) \cdot x_2(t+\tau)\right], \tag{3}$$

достигает относительного максимума. Как указывается в работе [5], многопутность прохождения сигналов приводит к искажениям вычисленного среднего значения временной задержки $\hat{D}$. Чтобы показать это, учтем (1) и (2) в (3). Получим

$$R_{x_1x_2} = \sum_{n=1}^{N} a_n R_{ss}(\tau - D_n) + R_{n_1n_2}(\tau). \tag{4}$$

Если шумы не коррелированы и не имеют нулевую составляющую, то, согласно [5], получаем:

$$R_{x_1x_2} = R_{ss}(\tau) \otimes \sum_{n=1}^{N} a_n \delta(\tau - D_n), \tag{5}$$

где $\otimes$ означает операцию свертки. В работе [4] предложено использование коэффициента корреляции, вычисленного на основе функции перекрестной корреляции с аргументом, равным вычисленной величине временной задержки, т.е.

$$\hat{\rho} = \frac{R_{x_1 x_2}\left(\hat{D}\right)}{\sqrt{R_{x_1 x_1}(0) \cdot R_{x_2 x_2}(0)}} . \tag{6}$$

Согласно [4], для узкополосных сигналов функция автокорреляции является более широкой, чем многопутевые временные задержки. Исходя из этого, при вычисленной временной задержке $\hat{D}$ уравнение (5) может быть аппроксимировано как

$$R_{x_1 x_2}(D) \cong R_{ss}(0) \sum_{n=1}^{N} a_n . \tag{7}$$

Подставляя (7) в (6), а также учитывая наличие шумов, получим

$$\hat{\rho} = \frac{R_{ss}(0) \sum_{n=1}^{N} a_n}{\sqrt{\left[R_{ss}(0) + \sigma_{11}^2\right]\left[R_{ss}(0) \sum_{n=1}^{N} a_n + \sigma_{22}^2\right]}} . \tag{8}$$

Оптимизация метода позиционирования БПЛА вокруг объекта по разности времени прихода сигналов от цели

В работе [5] выработаны различные практические рекомендации по идентификации условий распространения сигналов до приемников для дальнейшей выработки сигналов управления позиционированием БПЛА вокруг объекта. Однако в [5] не полностью раскрыты потенциальные возможности метода позиционирования БПЛА вокруг объекта по разности времени прихода сигналов от цели. Рассмотрим возможность оптимизации и обобщения данного метода.

Оптимизацию рассматриваемого метода осуществим путем нахождения такой оптимальной величины $R_{ss}(0) \sum_{n=1}^{N} a_n$, при которой $\hat{\rho}$ достигла бы экстремального значения. При этом следует учесть, что функция автокорреляции может достичь двух экстремальных значений, т.е. +1 и –1. Так как для аппаратурной реализации использование обоих этих экстремумов не представляет особого труда, то для поиска экстремума выражения (8) достаточно вычислить первую производную этого выражения и приравнять ее к нулю. В результате выполнения этих операций получим

$$\frac{\sqrt{\left[R_{ss}(0) + \sigma_{11}^2\right] \cdot \left[R_{ss}(0) \cdot \sum_{n=1}^{N} a_n + \sigma_{22}^2\right]} - \dfrac{\frac{1}{2} \cdot R_{ss}(0) \cdot \sum_{n=1}^{N} a_n \left[R_{ss}(0) + \sigma_{11}^2\right]}{\sqrt{\left[R_{ss}(0) + \sigma_{11}^2\right] \cdot \left[R_{ss}(0) \cdot \sum_{n=1}^{N} a_n + \sigma_{22}^2\right]}}}{\left[R_{ss}(0) + \sigma_{11}^2\right] \cdot \left[R_{ss}(0) \cdot \sum_{n=1}^{N} a_n + \sigma_{22}^2\right]} = 0 . \tag{9}$$

В выражении (9) примем следующие условные обозначения:

$$d_1 = R_{ss}(0) + \sigma_{11}^2 ; \tag{10}$$

$$z = R_{ss}(0) \cdot \sum_{n=1}^{N} a_n ; \tag{11}$$

$$d_2 = \sigma_{22}^2 . \tag{12}$$

С учетом (10)–(12) представим (9) в виде

$$d_1 \left(z + d_2\right) - \frac{z \cdot d_2}{2} = 0 . \tag{13}$$

Из выражения (13) находим

$$z = -2 d_2 . \tag{14}$$

Таким образом, с учетом (11), (12) и (14) функция автокорреляции $\hat{\rho}$ достигает экстремального значения при выполнении следующего равенства:

$$R_{ss}(0) \cdot \sum_{n=1}^{N} a_n = -2 \sigma_{22}^2 .$$

Очевидно, что $\sum_{n=1}^{N} a_n$ является положительной величиной, следовательно, оптимальным следует считать такой режим, при котором $R_{ss}(0)$ оказывается отрицательной величиной, равной $-2\sigma_{22}^2$.

Таким образом, полученный результат позволяет синтезировать качественно новый метод позиционирования, заключающийся в вычислении перекрестной корреляционной функции, в следующем виде:

$$R_{x_1\bar{x}_2} = \sum_{n=1}^{N} a_n R_{s\bar{s}}\left(\tau - D_n\right) + R_{n_1 n_2}\left(\tau\right).$$

Подобно выражению (4) выражения (5)–(8) в оптимизированном методе будут иметь соответственно форму записи (15)–(18):

$$R_{x_1\bar{x}_2}\left(\tau\right) = R_{s\bar{s}}\left(\tau\right) \otimes \sum_{n=1}^{N} a_n \delta\left(t - D_n\right), \tag{15}$$

$$\hat{\rho} = \frac{R_{x_1\bar{x}_2}\left(\hat{D}\right)}{\sqrt{R_{x_1\bar{x}_1}\left(0\right) \cdot R_{x_2\bar{x}_2}\left(0\right)}}, \tag{16}$$

$$R_{x_1\bar{x}_2}\left(D\right) \cong -2\sigma_{22}^2, \tag{17}$$

$$\hat{\rho} = -\frac{2\sigma_{22}^2}{\sqrt{\left[R_{s\bar{s}}\left(0\right) + \sigma_{11}^2\right] \cdot \left[R_{s\bar{s}}\left(0\right) \cdot \sum_{n=1}^{N} a_n + \sigma_{22}^2\right]}}. \tag{18}$$

Рассмотрим практические рекомендации, изложенные в [5], применительно к предложенному выше оптимальному методу. Для этого, подобно [5], проанализируем выражение (18).

Если отношение сигнал/шум имеет высокое значение, то сигнал функции автокорреляции при нулевой задержке намного превысит величину шумов приемника, и в этом случае получим

$$\hat{\rho} \cong \frac{\sum_{n=1}^{N} a_n}{\sqrt{\left[\sum_{n=1}^{N} a_n + \frac{\sigma_{22}^2}{R_{s\bar{s}}^2\left(0\right)}\right]}}.$$

Если сумма коэффициентов ослабления имеет намного большее значение, чем квадрат отношения шум/сигнал, то получим

$$\hat{\rho} \cong \left\{\sum_{n=1}^{N} a_n\right\}^{-\frac{1}{2}}.$$

Если ослабление сигналов настолько велико, что во втором приемнике доминируют шумы, то $R_{s\bar{s}}\left(0\right) << \sigma_{22}$.

В этом случае имеем

$$\hat{\rho} \cong \frac{R_{s\bar{s}}\left(0\right)}{\sigma_{22}} \sum_{n=1}^{N} a_n.$$

Если отношение сигнал/шум на обоих приемниках мало, то получим

$$\hat{\rho} \cong \frac{R_{s\bar{s}}\left(0\right)^2}{\sigma_{11} \cdot \sigma_{22}} \cdot \sum_{n=1}^{N} a_n.$$

Аналогично [5], эти выражения показывают, что при высоких значениях отношения сигнал/шум в коэффициенте корреляции доминирует сумма коэффициентов ослабления. При низких значениях отношения сигнал/шум коэффициент корреляции зависит как от величины этого отношения, так и от значения суммы коэффициентов ослабления. Следовательно, как и в [5], малая величина коэффициента корреляции будет означать, что либо приемник находится вдали от искомого объекта, либо на пути электромагнитного сигнала до приемника имеются препятствия. Таким образом, оператор БПЛА или бортовой вычислитель получает полезную информацию для осуществления соответствующей коррекции траектории полета БПЛА в целях слежения за объектом.

Заключение

В заключение сформулируем основные выводы и положения проведенного исследования.

1. Проведенный критический анализ известного метода разницы времени прибытия сигнала, применяемого для поиска и слежения за объектами, используя беспилотный летательный аппарат, показал наличие возможности его оптимизации и дальнейшего синтеза качественно нового метода.

2. В результате оптимизации известного метода поиска и слежения за объектом, используя группу беспилотных летательных аппаратов, выработаны практические рекомендации по использованию синтезированного оптимального метода.

Литература

1. Sarris Z. Survey of UAV applications in civil markets. – June 2001 [Электронный ресурс]. – Режим доступа: http://med.ee.nd.edu/MED9/Papers/Aerial_vehicles/med01-164.pdf, свободный. Яз. англ. (дата обращения 30.09.2012).
2. Molina P., Colomina I., Vitoria T., Silva P.F., Stebler Y., Skaloud J., Kornus W., Prades R. Egnos-based multi-sensor accurate and reliable navigation in search and rescue missions with UAVs // International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g). – Zurich, Switzerland, 2011. – V. XXXVIII-1/C22. – P. 87–93.
3. Wheeler M., Schrick B.,Whutacre W., Campell M., Rusdyk R., Wise R. Cooperative tracking of moving targets by a team of autonomous UAVs // Authorized licensed use limited to: Uppsala Universitetbibliotek. Downloaded on January 8. 2010 at 03:11 from IEE Xplore. Restrictions apply. – P. 5C2-1-5C2-9.
4. Nuske S., Dille M., Grocholsky B., Singh S. Representing substantial heading uncertainty for accurate geolocation by small UAVs // AIAA Guidance, Navigation, and Control Conference. – Toronto, Ontario, Canada, 2010. – 2–5 August [Электронный ресурс]. – Режим доступа: http://www.ri.cmu.edu/publication_view.html?pub_id=6677, свободный. Яз. англ. (дата обращения 25.08.2012).
5. Van Rheeden D.R., Brown B.C., Price J.C., Abbott B.A., Willden G.C., Chhokka K., Scott J., Barty T. Automatic positioning of UAVs to optimize TDOA geolocation performance / 0-7803-8539-X/04/$20.00© 2004 IEEE [Электронный ресурс]. – Режим доступа: http://www.ece.lsv.edu/pub/aravena/GoGS_Literature/Rheeden2004AutomaticPositioning21mproveGeolorat ion.pdf. свободный. Яз. англ. (дата обращения 20.09.2012).

Абдулов Рауф Нусрат оглы – Азербайджан. НИИ Министерства оборонной промышленности, заместитель главного инженера, Rabdulov2009@rambler.ru

Абдуллаев Новруз Алмамед оглы – Азербайджан. НИИ Министерства оборонной промышленности, кандидат технических наук, ст. научный сотрудник, начальник отдела, Naabdullayev@mail.ru

Асадов Хикмет Гамид оглы – Азербайджан. НИИ аэрокосмической информатики. доктор технических наук, доцент; начальник отдела, asadzade@rambler.ru

2 ОПТИЧЕСКИЕ И ОПТИКО-ЭЛЕКТРОННЫЕ СИСТЕМЫ. ОПТИЧЕСКИЕ ТЕХНОЛОГИИ

УДК 535.31; 681.7; 53.082.5

ГОМОГЕНИЗАТОРЫ ЛАЗЕРНОГО ИЗЛУЧЕНИЯ, ИЗГОТОВЛЕННЫЕ МЕТОДОМ ЛАЗЕРНОЙ ГРАВИРОВКИ ПОЛИМЕРНЫХ ПЛЕНОК

В.К. Баля, И.Ю. Денисюк, Е.В. Бараусова

Получены гомогенизаторы лазерного излучения на базе перекрещивающихся микропризменных структур и на базе нерегулярных диффузоров методом лазерной гравировки термочувствительной пленки и последующего наноимпринт-копирования.

Ключевые слова: гомогенизатор, лазерная абляция, микропризмы, нерегулярные диффузоры.

Введение

Гомогенизация лазерного излучения и формирование пучка являются ключевыми технологиями для многих современных приложений. Прежде всего, это те лазерные технологии, у которых изображение образца маски требует равномерного распределения интенсивности излучения по всей площади маски и, следовательно, на всей плоскости обработки. Другие приложения требуют однородной тонкой линии лазера; гомогенизируют только одно направление пучка.

Формирование луча с дифракционными оптическими элементами представляет очень изящный и мощный метод для генерации произвольных моделей облучения. Эти элементы, как правило, рассчитаны на определенную длину волны и фазовую функцию. Для достижения высокой производительности, т.е. однородности луча и эффективности, необходимы многоуровневые элементы. Другая концепция для генерации произвольных моделей облучения использует элементы мультиапертуры, которые делят поступающий луч на ряд элементарных лучей. Элементарные лучи накладываются друг на друга с помощью дополнительного объектива. Преимуществами этих элементов являются независимость от входного профиля интенсивности и широкий спектр длин волн. Хотя периодические структуры создают интерференционные эффекты, особенно с использованием высококогерентного света, тем не менее, успешная гомогенизация с этими элементами может быть достигнута с учетом физической оптики [1] и в некоторых случаях с использованием дополнительных элементов, таких как случайные диффузоры.

Интегратор Келера

Гомогенизация достигается с помощью интегратора Келера, как показано на рис. 1. Интегратор Келера состоит из двух массивов линз и конденсора, формирующего встречно направленные многократные системы освещения Келера [2]. Первый массив линз (LA_1) разделяет падающий свет и создает несколько изображений источников света в плоскости отверстия. Массив линз LA_1 также служит множеством диафрагм поля, определяющих освещенную область в плоскости объекта.

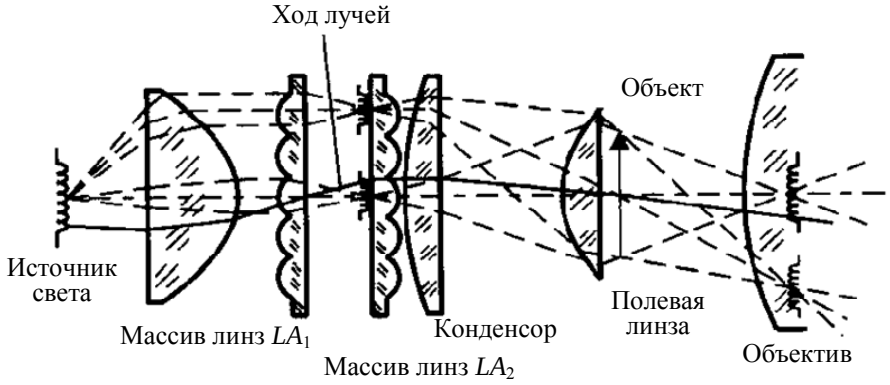


Рис. 1. Интегратор Келера для проецирования изображения

Второй массив линз (LA_2) находится в плоскости апертуры и представляет собой множество диафрагм. Линзы массива LA_2 и конденсор создают изображение отдельных диафрагм поля в плоскости объекта. Четкие изображения нити накала появляются в плоскости зрачка объектива. Это обеспечивает равномерное освещение как объекта, так и плоскости изображения. В плоскости объекта реальные изображения первого массива линз LA_1 накладываются. Если предположить, что освещенность в пределах каждого суботверстия LA_1 примерно равномерна или что излучение падающего света является симметрич-

ным, суперпозиция всех изображений обеспечивает равномерное распределение интенсивности в плоскости объекта.

Увеличение числа линз позволяет улучшить однородность распределения интенсивности. Тем не менее, если линзы слишком малы, дифракционные эффекты значительно исказят однородность. Большее количество линз позволит улучшить смешивание света, однако, если линзы становятся малыми, дифракционные эффекты будут искажать однородность.

Нерегулярные диффузоры

Одномерные нерегулярные диффузоры состоят из произвольных и статистически размещенных диффундирующих элементов, как показано на рис. 2. Эти диффузоры представляют собой маску щелей фиксированной ширины, которые статистически расположены на подложке [3]. Такие диффузоры эффективны для улучшения однородности линии. Гомогенизатор и сформированные линии лазерного излучения показаны на рис. 2.

Рис. 2. Схема гомогенизатора (а) и линий лазерного излучения, сформированных интегратором Келера (б) и одномерными нерегулярными диффузорами (в) [1]

Условия эксперимента

Подробно технология лазерной гравировки, которая была использована в настоящей работе, описана в работе авторов [4]. Лазерная гравировка выполнялась с использованием лазерного гравера Laser Graver LG 10F15, обеспечивающего получение структур с разрешением до 2,5 мкм (10160 dpi). Решетки формировались методом обработки черной полимерной или металлизированной пленки при абляции поверхностного слоя лучом непрерывного Nd:YAG-лазера с длиной волны 1,06 мкм. В данной работе эксперименты проводились с использованием полиэстеровой пленки с термочувствительным слоем на базе полиметилметакрилата (ПММА) с наночастицами сажи, обеспечивающими поглощение лазерного излучения. Толщина испаряемого слоя 13–15 мкм.

После изготовления заданной структуры она копировалась на акриловую фотоотверждаемую композицию методом наноимпринта. Метод описан в работе авторов [5].

Экспериментальные результаты

Решетки перекрещивающихся микропризменных структур. Решетки микропризменных структур, наложенные друг на друга и имеющие соответствующий период, могут быть использованы как элементы, обеспечивающие получение множественных изображений при наблюдении яркого источника через такую прозрачную структуру. На рис. 3 приведена фотография лампочки накаливания, полученная при наблюдении через подобную структуру. В данном случае (рис. 3) приведена фотография, полученная при наблюдении через решетку с углом между микропризменными растрами 90º. При изменении угла и наложении большего числа решеток могут быть получены и другие подобные структуры.

На рис. 4 приведена микрофотография решетки из пересекающихся микропризм. Решетка была получена путем нарезки полимерной пленки на лазерном гравере в двух пересекающихся направлениях. В соответствии с информацией из раздела «Условия эксперимента» эта структура является гомогенизатором лазерного излучения по типу Келера.

Лазерный генератор линии на базе нерегулярного диффузора. Рассмотренный элемент формирует пучок в виде узкой линии, в вертикальной плоскости которой высота определяется использованной отрицательной линзой и естественной расходимостью лазерного луча. Длина линии в горизонтальной плоскости определяется углами дифракции решетки и составляет ±20° по полуширине. Элемент универсален и имеет близкие характеристики для зеленого, желтого и красного диапазонов спектра, поскольку подъем интенсивности в области углов ±20° определяется преломлением света на треугольном штрихе.

Недостатком простой конструкции на базе одиночной дифракционной решетки постоянного периода являются четко видимые дифракционные максимумы в линии (рис. 5), что создает значительную неравномерность распределения света по длине линии.

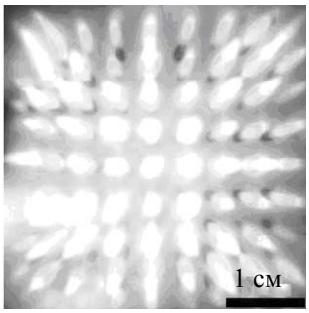


Рис. 3. Фотография лампы накаливания при наблюдении через структуру наложенных решеток микропризм

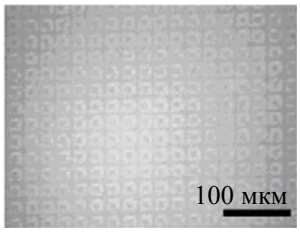


Рис. 4. Микрофотография полимерной структуры решетки

Наличие четких дифракционных порядков является принципиальным недостатком расщепителей, основанных на дифракционных решетках, и может быть преодолено только при изготовлении решетки со сбитым периодом (нерегулярный диффузор). В работе была получена такая решетка с чередованием областей с различным периодом штрихов. Решетка образована периодически повторяющимися элементами, в каждом из которых полоски идут с переменным периодом, например, 10; 12,5; 15 мкм, далее период повторяется. При этом решетка со сбитым периодом по причине непостоянства шага нарезки не дает четко видимых дифракционных порядков, и линия становится более однородной по длине.

На рис. 6 приведена схема дифракционной решетки со сбитым периодом. Параметры штрихов с размерами приведены в таблице. На рис. 6, в, хорошо видны группы штрихов одинакового периода, совокупность которых и образует необходимый оптический элемент.

Рис. 5. Фотография линии, сформированной периодической решеткой

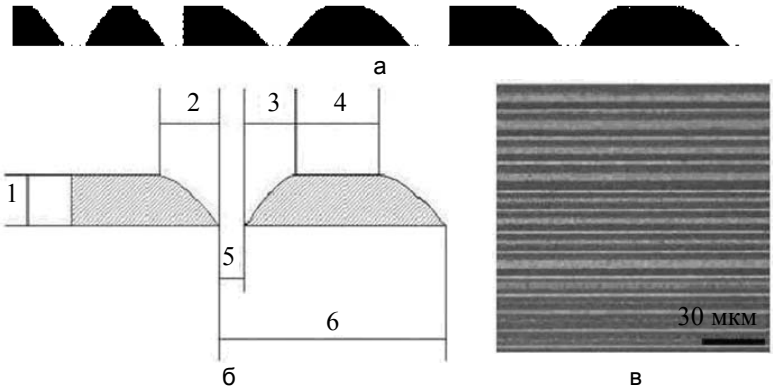


Рис. 6. Схема дифракционной решетки – схема расположения штрихов (а); геометрия одного штриха: 1–6 – параметры штрихов (б); микрофотография решетки (в)

Период полос	1	2	3	4	5	6
10 мкм	3 мкм	2,7 мкм	2,2 мкм	1,5 мкм	1,5 мкм	10 мкм
12,5 мкм	3 мкм	4 мкм	3 мкм	3 мкм	1,5 мкм	12,5 мкм
15 мкм	3 мкм	4 мкм	3 мкм	5 мкм	1,5 мкм	15 мкм

Таблица. Параметры штрихов дифракционной решетки со сбитым периодом

На рис. 7 представлено распределение энергии в линии, сформированной периодической решеткой (рис. 7, а) и решеткой со сбитым периодом (рис. 7, б).

Фотография лазерного генератора линии представлена на рис. 8. Генератор дает узкую линию света с заданными параметрами распределения интенсивности в телесном угле. В соответствии с информацией, указанной выше, эта структура является гомогенизатором лазерного излучения на базе нерегулярных диффузоров.

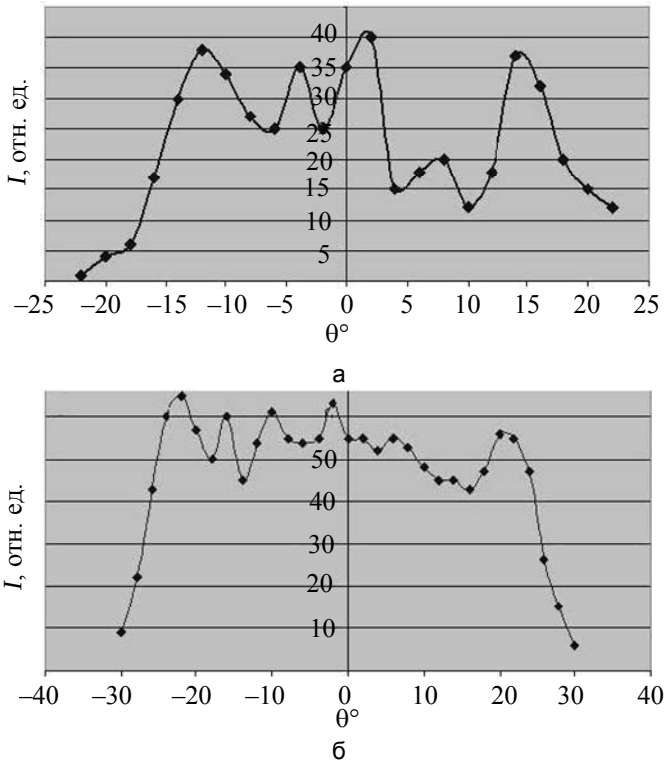


Рис. 7. Распределение энергии в линии, сформированной решетками: с постоянным периодом (а); со сбитым периодом (б)

Рис. 8. Фотография генератора линии, сформированной решеткой со сбитым периодом

Заключение

Показана возможность изготовления двумерных гомогенизаторов путем нарезки перпендикулярных линий на полимерной пленке.

Изготовлен лазерный генератор линии на базе нерегулярного диффузора методом лазерной гравировки термочувствительной пленки и последующего наноимпринт-копирования. Изготовление несложно, не требует прецизионного оборудования, а тиражирование возможно по методу hot embossing, когда получение элементов с нанометровым разрешением не является обязательным.

Литература

1. Harder I., Lano M., Lindlein N., Schwider J. Homogenizations and beam shaping with microlens arrays // Photon Management, Proceeding of SPIE. – 2004. – V. 5456. – P. 99–107.
2. Räntsch K., Bertele L., Sauer H., Merz A. Illuminating System. – US Patent 2186123. 1940.
3. Voelkel R., Weible K.J. Laser beam homogenizing: limitations and constraints // Proc. SPIE. – 2008. – V. 7102. – P. 71020J–71020J-12.
4. Баля В.К., Денисюк И.Ю. Формирование микрооптических элементов для защиты с использованием лазерного гравера // Изв. вузов. Приборостроение. – 2012. – Т. 55. – № 3. – С. 51–56.
5. Арефьева Н.Н., Денисюк И.Ю. Применение метода наноимпринта для единичного копирования полимерной френелевской и микрооптики // Оптический журнал. – 2008. – Т. 75. – № 7. – С. 71–74.

Баля Вера Константиновна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, Vera_Balya@mail.ru

Денисюк Игорь Юрьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор физ.-мат. наук, профессор, зав. кафедрой, denisiuk@mail.ifmo.ru

Бараусова Екатерина Васильевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, barausik@yandex.ru

3 ФОТОНИКА И ОПТОИНФОРМАТИКА

УДК 537.862

ХАРАКТЕРИСТИКИ МЕТАПЛЕНОК ДЛЯ ТЕРАГЕРЦОВОГО ДИАПАЗОНА ЧАСТОТ ПРИ МАСШТАБИРОВАНИИ ГЕОМЕТРИЧЕСКИХ ПАРАМЕТРОВ

Ю.Е. Терехов, М.К. Ходзицкий, Г.В. Белокопытов

Методом численного конечно-элементного моделирования получены коэффициенты отражения и прохождения метапленок, составленных из П-образных резонаторов, в частотном диапазоне 0,1–1 ТГц. Пропорционально масштабировались длина стороны, глубина зазора и период резонаторов. Получены экспериментальные спектры коэффициентов прохождения метапленок с П-образными резонаторами. Показана перестройка резонансных мод с 0,12 до 0,51 ТГц при масштабировании геометрических параметров П-образных резонаторов.

Ключевые слова: метаматериалы, П-образный резонатор, электродинамическое подобие, терагерцовый диапазон, плазмонный резонанс, метапленка.

Введение

В настоящее время активно исследуются метаматериалы – искусственные композитные среды с включениями, размер которых и расстояния между которыми много меньше длины волны внешнего излучения [1]. Ключевой структурной единицей («атомом фотоники») композитной среды является металлический П-образный резонатор микронных или субмикронных размеров [2, 3], представляющий собой на рабочих частотах аналог сосредоточенного LC-контура, образованного индуктивной петлей с емкостным зазором. Периодические структуры на основе таких субволновых резонаторов позволяют получать заданный отклик на внешнее электромагнитное излучение. Это свойство актуально при создании маскирующих покрытий [4] на основе резонаторов и устройств на метаматериалах, работающих в терагерцовом (ТГц) диапазоне частот [5]. Важным обстоятельством, которое облегчает создание метаматериалов ТГц диапазона, является свойство электродинамического подобия, которое проявляется в том, что собственные частоты субволновых резонансных элементов обратно пропорциональны их геометрическим размерам (при сохранении геометрических пропорций) [6, 7]. Вместе с тем, электродинамическое подобие всегда является неполным из-за дисперсии материальных параметров металлических и диэлектрических сред, причем именно в ТГц диапазоне частот дисперсия многих веществ особенно существенна. В связи с этим целью данной работы явилась проверка выполнения закономерностей электродинамического подобия в двумерных метаматериалах (метапленках) ТГц диапазона частот, которые представляли собой совокупность П-образных металлических резонаторов, расположенных в узлах квадратной решетки, а также моделирование и разработка механически перестраиваемого фильтра на метаматериалах (П-образных резонаторах) для ТГц диапазона частот.

Описание структуры исследуемых метаматериальных образцов

Исследуемая структура (метапленка) представляет собой массив периодически расположенных медных П-образных резонаторов высотой 15 мкм на ситалловой подложке толщиной 0,56 мм. Для адекватного численного и экспериментального моделирования метапленки была исследована частотная дисперсия показателя преломления ситалловой подложки в диапазоне частот 0,1–1 ТГц. Показатель преломления ситалла был измерен двумя методами:

1. по положениям резонансов Фабри–Перо в спектре прохождения через ситалловую пластину;
2. по разности фазовых задержек сигнала в свободном пространстве и пространстве с подложкой.

Спектры прохождения и отражения подложки были получены на ТГц фотометре [8] и рефлектометре [9] соответственно и представлены на рис. 1.

В первом методе показатель преломления ситалла определялся по формулам [10]

$$n = \frac{c \cdot k}{2d \cdot \left(\nu_{m+k} - \nu_m\right)},$$

где c – скорость света; d – толщина ситалловой подложки; ν_{m+k} и ν_m – частоты $m+k$ и m-го резонансов Фабри–Перо соответственно. Полученный этим методом средний показатель преломления равен $\overline{n} = 2,9$. Во втором методе показатель преломления определялся по формуле [11]

$$n(\nu) = 1 + \frac{c}{2\pi\nu d}\left[\varphi_s(\nu) - \varphi_r(\nu)\right],$$

где $\varphi_s(\nu)$ и $\varphi_r(\nu)$ – фазы комплексных амплитуд коэффициентов прохождения для подложки и свободного пространства соответственно. Зависимость фазы от частоты для подложки и свободного

пространства, а также частотная дисперсия показателя преломления ситалла, рассчитанная вторым методом, представлены на рис. 2.

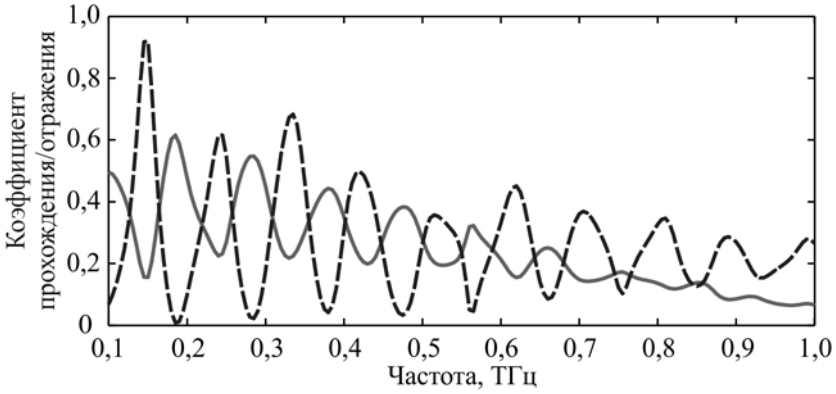


Рис. 1. Экспериментальные спектры прохождения (сплошная линия) и отражения (пунктирная линия) по мощности для ситалловой подложки толщиной 0,56 мм

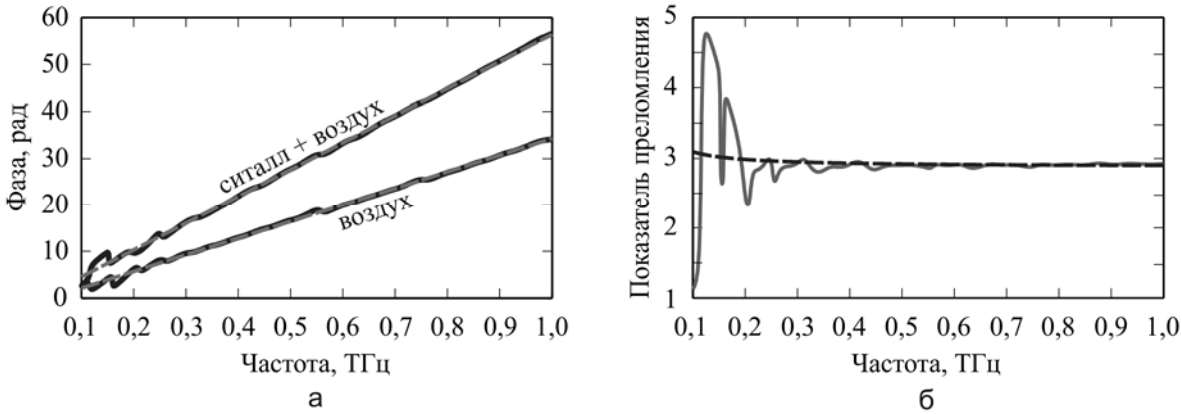


Рис. 2. Зависимость фазы от частоты (а) и дисперсия показателя преломления, определенная фазовым методом (б). Сплошные линии – экспериментально полученные фазы и показатель преломления, рассчитанный из разности экспериментальных фаз; пунктирные – фазы, полученные линейной аппроксимацией экспериментальных фаз и показатель преломления, рассчитанный из разности линейно-аппроксимированных экспериментальных фаз

Средний показатель преломления, определенный фазовым методом, в диапазоне частот 0,1–1 ТГц равен $\overline{n} = 2,9$. Разброс значений показателя преломления в области низких частот обусловлен погрешностями эксперимента.

Расчет коэффициентов прохождения и отражения метапленки

Численный расчет параметров прохождения и отражения исследуемой структуры производился методом конечно-элементного моделирования в среде COMSOL Multiphysics. Геометрия П-образного резонатора и области для численного расчета представлены на рис. 3.

В расчетах длина стороны резонатора l изменялась от 100 до 350 мкм с шагом 50 мкм и были выбраны следующие геометрические пропорции: $u/l = 0,6$; $g/l = 0,3$; $p/l = 1,5$. Для учета дисперсии материальных параметров резонаторов использовалась модель Друде–Лоренца

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega(\omega + i\omega_r)},$$

где плазменная частота ω_p = 8,76 эВ, частота релаксации ω_r = 0,0955 эВ (данные для меди в соответствии с [12]). Для моделирования в областях 1–3 показатель преломления принят равным n = 1, а в областях 5 и 6 показатель преломления подложки принят равным $n_{sub} = 2,9$ (средний показатель преломления ситалла в диапазоне 0,1–1 ТГц). На границе металлического резонатора для учета скин-слоя были заданы импедансные граничные условия, на внешних границах областей 1–6 (точечные линии на рис. 3, б), перпендикулярных осям x и z, установлены периодические граничные условия, а на внешних границах, перпендикулярных оси y, заданы граничные условия рассеяния. Таким образом, электродинамический

расчет производился (подобно [6]) для периодического массива частиц, при этом учитывалось взаимодействие между резонаторами. Ширина области моделирования p определяла период решетки резонаторов.

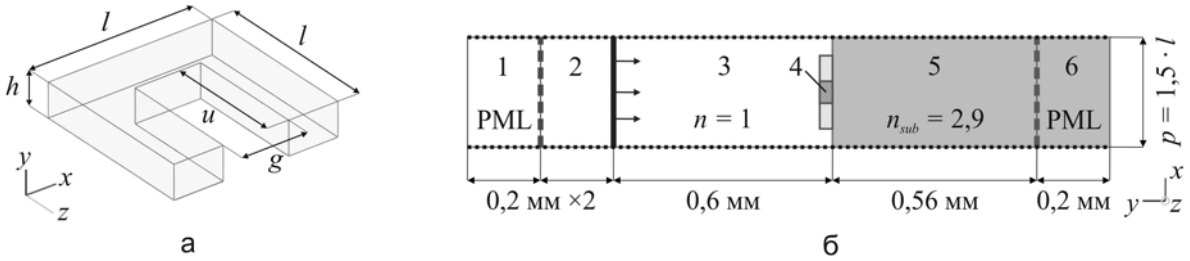


Рис. 3. Геометрия резонатора: l – сторона резонатора; h – высота; u – глубина зазора; g – ширина зазора (а). Вид области моделирования в плоскости *xy*: 1 – верхний поглощающий слой (PML); 2, 3 – области свободного пространства; 4 – резонатор; 5 – подложка; 6 – нижний поглощающий слой (PML) (б)

Плоская волна падает по нормали к поверхности резонатора (источник отмечен сплошной линией со стрелками на рис. 3, б), причем электрическая компонента поля поляризована перпендикулярно емкостному зазору резонатора (вдоль оси x). Интегрирование отраженной и прошедшей мощности производится по границам между областями 1/2 и 5/6 (пунктирные линии на рис. 3, б). Спектры коэффициентов прохождения и отражения метапленок рассчитывались в диапазоне длин волн 0,3–3 мм. Шаг по длинам волн составлял 30 мкм, соответственно каждый рассчитанный спектр состоял из 91 отсчета.

Результаты расчета приведены на рис. 4. Можно выделить две резонансные моды. Первая мода резонансов, LC-мода, соответствует круговому движению зарядов в плоскости резонатора, эффективно представляющему собой виток индуктивности с емкостным зазором. На рис. 4 эта низкодобротная мода наблюдается в диапазоне длин волн 0,6–3,0 мм и характеризуется резонансным увеличением коэффициента отражения и резонансным падением коэффициента прохождения. Вторая, коротковолновая, высокодобротная мода располагается в диапазоне 0,3–1,8 мм. Такие резонансы представляют собой согласованные локализованные колебания зарядов вдоль оснований резонаторов (далее – плазмонные резонансы). Для получения плазмонных резонансов на длине стороны резонатора, с учетом влияния подложки, должно укладываться целое число полуволн. В этом случае отклик П-образных резонаторов аналогичен отклику полуволновых антенн.

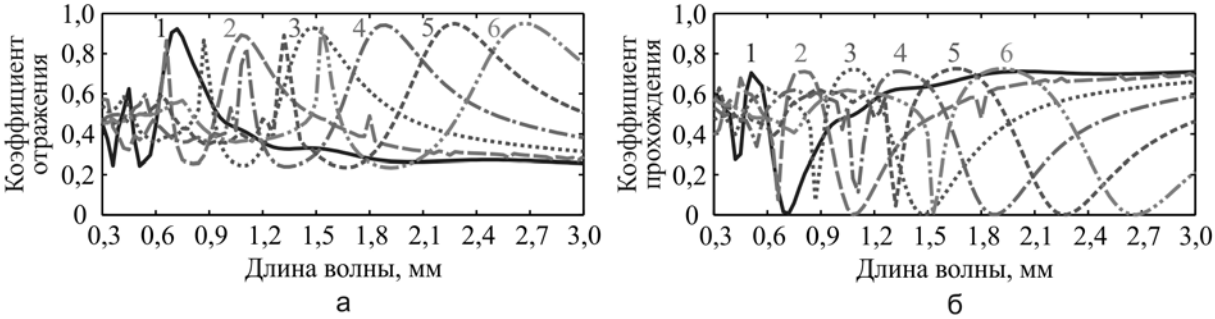


Рис. 4. Спектры коэффициентов отражения (а) и прохождения (б). Числами обозначены спектры метапленок, состоящие из П-образных резонаторов с различной длиной основания l : 1 – 100 мкм; 2 – 150 мкм; 3 – 200 мкм; 4 – 250 мкм; 5 – 300 мкм; 6 – 350 мкм

Зависимости резонансных длин волн и добротностей от длины П-образного резонатора представлены на рис. 5. Видно, что значения резонансных длин волн для обеих мод линейно зависят от длины стороны резонатора (рис. 5, а). Поскольку окружающая резонатор среда практически не имеет дисперсии, то при пропорциональном изменении геометрических параметров метапленки потери на излучение остаются неизменными, как следствие, не изменяется и добротность (рис. 5, б). Сильный разброс в добротностях плазмонных резонансов по сравнению с LC-резонансами связан с большими погрешностями моделирования на коротких длинах волн.

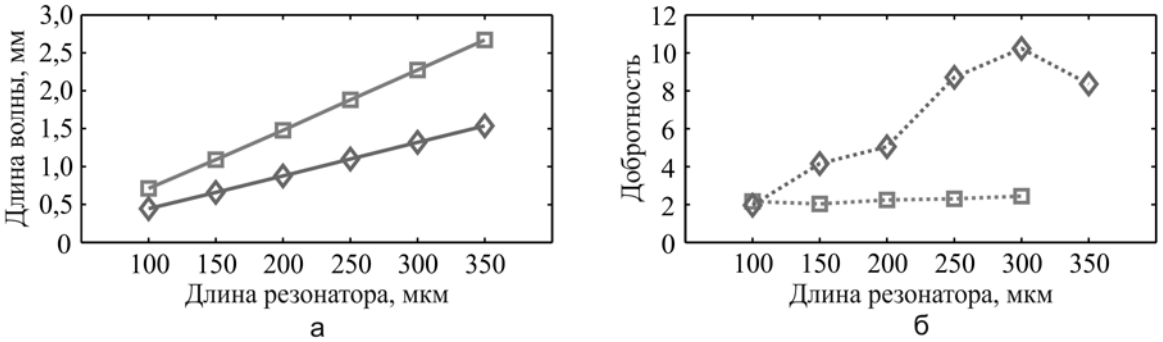


Рис. 5. Зависимость длины волны резонанса отражения (а) и его добротности (б) от длины резонатора. Квадратами отмечены значения для LC-резонансов, ромбами – для плазмонных резонансов

Эксперимент

По результатам расчета был изготовлен экспериментальный образец метапленки с 9 областями (рис. 6) с различными параметрами медных П-образных резонаторов (таблица), так чтобы каждая область имела свою LC-резонансную частоту. Метапленка представляет собой рисунок из П-образных резонаторов, выгравированных с точностью 5 мкм в слое меди на ситалле марки СТ-50-1. Медный слой толщиной 15 мкм наносился на ситалл методом ионно-вакуумного напыления. Для лучшей адгезии меди на ситалл предварительно напылялся слой хрома толщиной 10 нм.

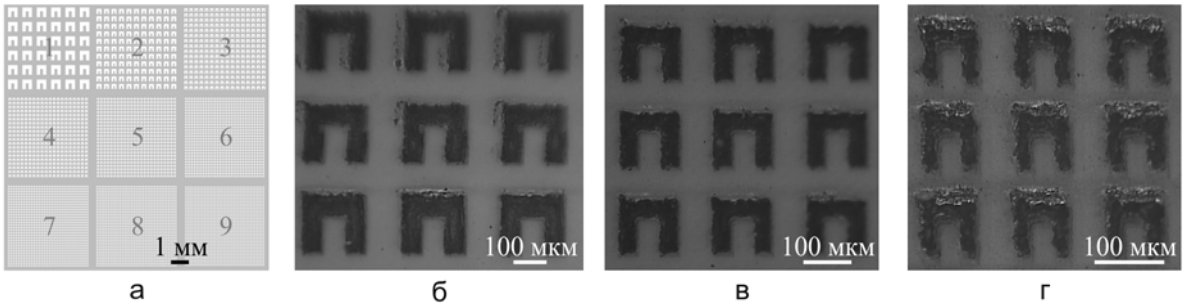


Рис. 6. Шаблон для изготовления экспериментального образца (а), цифрами отмечены номера областей. Фотографии резонаторов в областях № 3 (б), № 5 (в) и № 9 (г) изготовленного образца

Номер	l, мкм	u, мкм	p, мкм
1	542	325	813
2	282	169	423
3	195	117	293
4	152	91	228
5	126	76	189
6	108	65	162
7	96	58	144
8	87	52	131
9	79	47	119

Таблица. Параметры областей экспериментального образца

Размеры резонаторов для различных частот были определены с помощью зависимости на рис. 5. ТГц излучение по очереди фокусировалось в каждую из 9 областей размером 5×5 мм^2. Спектры прохождения и отражения для каждой из 9 областей метапленки были получены на ТГц фотометре ($P_{ср} = 48 \pm 4$ мкВт; $\tau_{имп} = 3$ пс; $\nu = 0{,}1–1{,}0$ ТГц) [8] и рефлектометре ($P_{ср} = 30 \pm 3$ мкВт; $\tau_{имп} = 3$ пс; $\nu = 0{,}1–1{,}0$ ТГц) [9] и представлены на рис. 7 в сравнении с рассчитанными спектрами. Как видно из рис. 7, в эксперименте наблюдается смещение частоты LC-резонанса в высокочастотную область при уменьшении геометрических размеров П-образного резонатора (механической перестройки сфокусированного ТГц излучения из одной области в другую) и хорошо согласуется с численным моделированием. Кроме того, следует отметить наличие дополнительных периодически расположенных резонансных особенностей, которые есть не что иное, как резонансы Фабри–Перо в толщине ситалловой подложки.

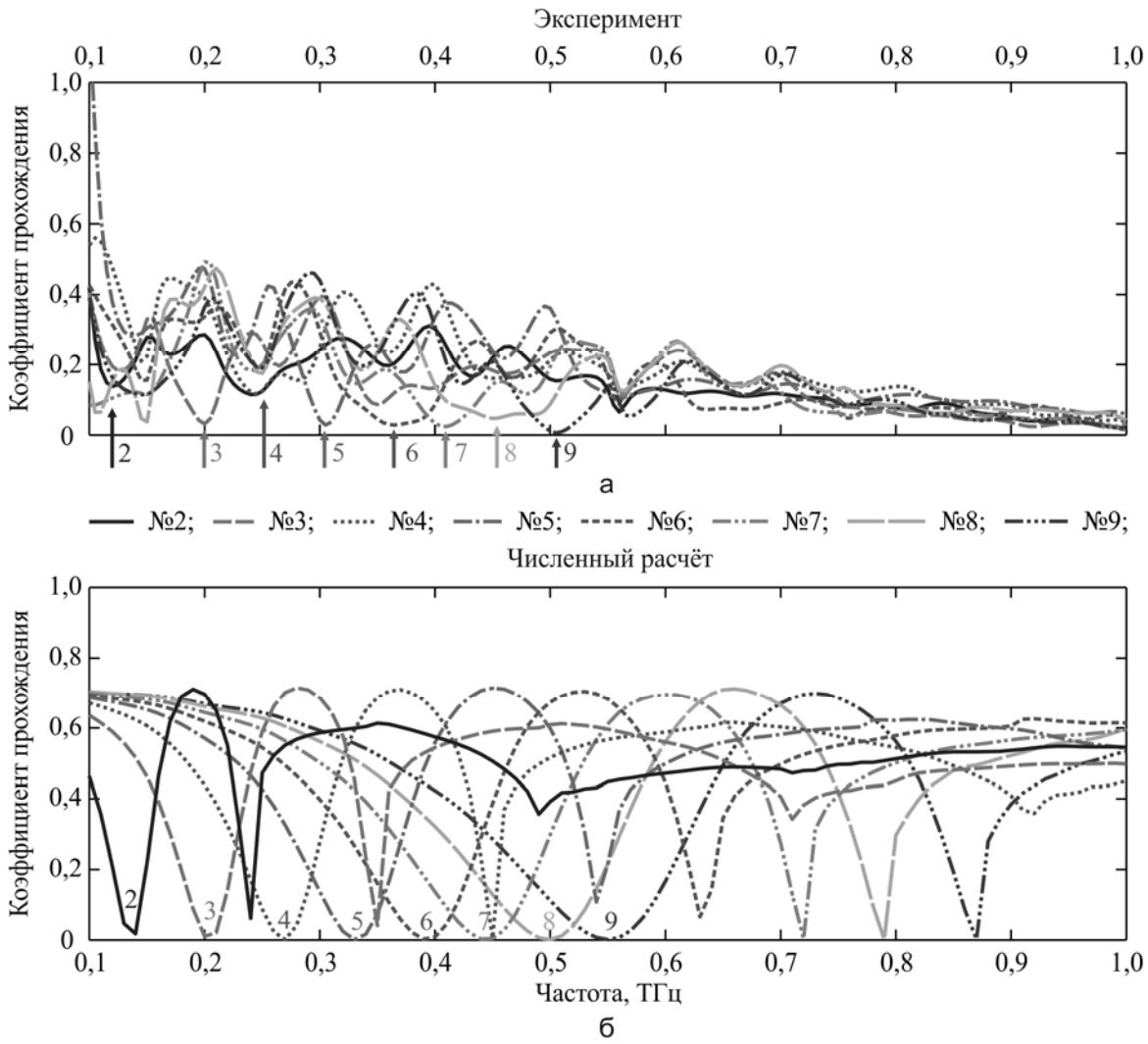


Рис. 7. Экспериментальные (а) и рассчитанные численно (б) спектры коэффициентов прохождения метапленок на ситалловой подложке. Числами обозначены кривые для резонаторов с различными геометрическими параметрами (сторона / глубина зазора / период решетки, мкм): 2 – 282/169/423 мкм; 3 – 195/117/293 мкм; 4 – 152/91/228 мкм; 5 – 126/76/189 мкм; 6 – 108/65/162 мкм; 7 – 96/58/144 мкм; 8 – 87/52/131 мкм; 9 – 79/47/119 мкм. Стрелками на экспериментальных графиках отмечены LC-резонансы

Заключение

В численных расчетах показана возможность перестройки частоты LC-резонанса как с помощью пропорционального изменения размеров резонатора, так и с помощью только изменения глубины зазора резонатора при постоянстве остальных геометрических параметров. Получены экспериментальные спектры прохождения метапленки с 9 областями, в которых перестройка частоты LC-резонанса с 0,12 ТГц до 0,51 ТГц осуществляется пропорциональным изменением размеров резонаторов. Продемонстрировано хорошее соответствие результатов расчета и эксперимента.

Работа выполнена в соответствии с программой Министерства образования и науки РФ «Поддержка развития внутрироссийской мобильности научных и научно-педагогических кадров путем выполнения научных исследований молодыми учеными и преподавателями в научно-образовательных центрах» (ГК № 14.B37.21.0248) и программой «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (ГК № 14.132.21.1421). Выражаем благодарность Я.В. Грачеву за помощь в проведении измерений.

Литература

1. Кившарь Ю.С., Орлов А.А. Перестраиваемые и нелинейные метаматериалы // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 3 (79). – С. 1–10.
2. Pendry J.B., Holden A.J. et al. Magnetism from conductors and enhanced nonlinear phenomena // IEEE Trans. Microw. Theory Tech. – 1999. – V. 47. – P. 2075–2084.

3. Smith D.R., Padilla W.J. et al. Composite Medium with Simultaneously Negative Permeability and Permittivity // Phys. Rev. Lett. – 2000. – V. 84. – P. 4184–4187.
4. Возианова А.В., Ходзицкий М.К. Маскирующее покрытие на основе спиральных резонаторов // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 4 (80). – С. 28–34.
5. Tao H., Padilla W.J., Zhang X., Averitt R.D. Recent Progress in Electromagnetic Metamaterial Devices for Terahertz Applications // IEEE Journal of Selected topics in Quantum Electronics. – 2011. – V. 17. – № 1. – P. 92–101.
6. Klein M.W., Enkrich C., Wegener M. et al. Single-slit split-ring resonators at optical frequencies: limits of size scaling // Optics Letters. – 2006. – V. 31. – № 9. – P. 1259–1261.
7. Белокопытов Г.В., Журавлев А.В., Терехов Ю.Е. Размерная зависимость поляризуемости металлических частиц // Вестник Московского университета. Серия 3. Физика. Астрономия. – 2012. – № 3. – С. 17–24.
8. Грачев Я.В., Куклин И.А., Смолянская О.А. Козлов С.А., Беспалов В.Г. и др. Исследование воздействия излучения 0,05–2 ТГц на биоткани разной толщины в медицинской диагностике // Оптический журнал. – 2010. – Т. 77. – № 11. – С. 92–94.
9. Bespalov V.G., Gorodetsky A.A. et al. Influence of THz broadband pulse radiation on some biotissues // Proc. of SPIE. – 2010. – V. 7547. – P. 754707-1.
10. Ахманов С.А., Никитин С.Ю. Физическая оптика. – М.: Изд-во МГУ, Наука. – 2004. – 656 с.
11. Zhang X.-C., Xu J. Introduction to THz Wave Photonics. – NY: Springer, 2010. – 246 p.
12. Zeman E.J., Schatz G.C. An Accurate Electromagnetic Theory Study of Surface Enhancement Factors for Ag, Au, Cu, Li, Na, AI, Ga, In, Zn, and Cd // J. Phys. Chem. – 1987. – V. 91. – P. 634–643.

Терехов Юрий Евгеньевич – Московский государственный университет им. М.В. Ломоносова, аспирант, yury.terekhov@phys.msu.ru

Ходзицкий Михаил Константинович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, ассистент, khodzitskiy@yandex.ru

Белокопытов Геннадий Васильевич – Московский государственный университет им. М.В. Ломоносова, кандидат физ.-мат. наук, доцент, gvb@phys.msu.ru

УДК 535.14, 57.043

ВОЗДЕЙСТВИЕ ТЕРАГЕРЦОВОГО ИЗЛУЧЕНИЯ НА РОСТ НЕЙРИТОВ

М.В. Цуркан, О.А. Смолянская, Н.Г. Брянцева

Исследовано воздействие широкополосного импульсного терагерцового излучения в диапазоне частот 0,05–2 ТГц на рост нейритов сенсорных ганглиев 10–12 дневных куриных эмбрионов. Описан дозозависимый эффект стимуляции роста нервных волокон.

Ключевые слова: терагерцовое излучение, терапия, медицина, клеточный рост, стимуляция, органотипическая культура ткани, нейрит.

Введение

Разработка новых источников терагерцового (ТГц) излучения с высокой спектральной чувствительностью и высоким разрешением позволила создать новые технологии в военной отрасли и системах безопасности, а также в биологии, химии и медицине. Уникальные свойства данного излучения заключаются в том, что в этом диапазоне находится значительная часть колебательно-вращательного спектра воды и многих органических молекул, в том числе биологически активных макромолекул (белков и нуклеиновых кислот), а так же частоты межмолекулярных взаимодействий. Однако энергии данного излучения недостаточно для тепловой вибрации, диссоциации межатомных связей и химических преобразований сложных молекул [1]. В связи с такими особенностями ТГц излучения публикации, посвященные его применению, в основном связаны с разработкой методов спектральных исследований молекул в дальнем ИК диапазоне спектра, где наблюдается большое число линий поглощения. Было показано, что сложные биологические молекулы, в том числе ДНК и РНК, также имеют поглощение в ТГц диапазоне частот [2, 3].

Применение источников ТГц излучения в различных областях, в том числе биологии и медицине, с каждым годом увеличивается. Вместе с этим увеличивается и потребность в более глубоком понимании механизмов этих взаимодействий и их возможных последствий, как положительных, так и отрицательных. Это отражено в увеличивающемся в последние годы количестве работ, связанных с исследованием возможных эффектов, которые вызваны поглощением этого излучения разнообразными биологическими объектами на различных уровнях биологической организации. Многие исследования проведены в

рамках масштабного европейского проекта THz-BRIDGE, который направлен на фундаментальное изучение воздействия ТГц излучения на биологические объекты [4].

Число работ, посвященных изучению влияния взаимодействия электромагнитного излучения ТГц частоты с собственно биологическими объектами относительно не велико. Однако в последние пять лет наблюдается рост количества работ в этом направлении, что связано с распространением источников ТГц излучения и возможностью их применения для спектроскопии в целях медицинской диагностики. Исследуются вопросы безопасности применения ТГц излучения.

Ряд работ, в которых применяется ТГц излучение различной длины волны и плотности мощности, посвящен изменению морфологии и жизнеспособности клеток [5–10], а также способности сохранять мембранный потенциал [11–16]. Авторами этих исследований отмечается как стимулирующий, так и угнетающий эффекты ТГц излучения на клеточные системы. Представлены разнонаправленные данные, которые не могут быть однозначно интерпретированы по отношению к влиянию ТГц излучения на биологические объекты.

Несмотря на возросшее количество научных направлений, посвященных практическому применению ТГц излучения, вопрос о фундаментальном обосновании возможности использования ТГц излучения для воздействия на биологические системы, в особенности на нервную систему, все еще открыт. Механизм данного воздействия до сих пор не изучен, в том числе не определены дозы и время облучения. В рамках данного вопроса на первом этапе необходимо выявить эффекты, возникающие в сложно организованных молекулярных системах – нервных клетках – при действии такого излучения.

Экспериментальная установка

В работе авторами применена специально разработанная в Санкт-Петербургском национальном исследовательском университете информационных технологий, механики и оптики (НИУ ИТМО) методика [17]. Оптическая схема эксперимента (рис. 1) основана на облучении кристалла InAs фемтосекундным волоконным лазером EFOA-SH.

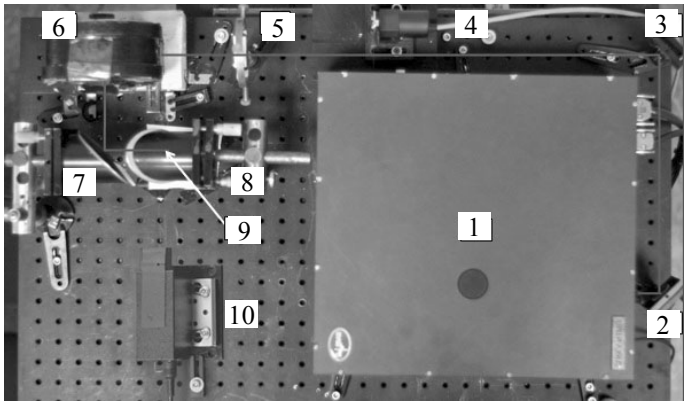


Рис. 1. Фотография оптической схемы установки: 1 – лазер; 2, 3 – система зеркал; 4 – механический модулятор; 5 – линза; 6 – магнит; 7, 8 – параболические зеркала; 9 – объект; 10 – оптико-акустический приемник

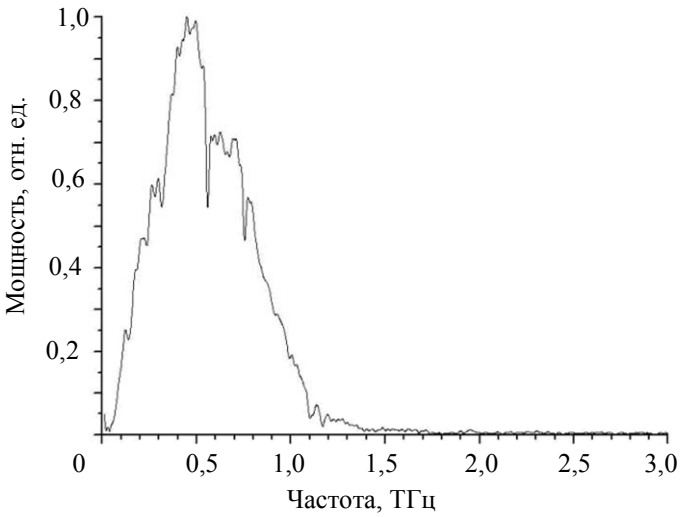


Рис. 2. Распределение мощности в импульсе

Лазер работает в импульсно-периодическом режиме с частотой повторения 50 МГц на длине волны $\lambda = 1{,}56$ мкм. Типичная длительность одного импульса составляет 125 фс, мощность излучения – 120 мВт. Излучение лазера 1 проходит систему зеркал 2, 3. Перед тем как попасть на кристалл, оно стробируется механическим модулятором 4 с частотой 13 Гц, а затем фокусируется линзой 5 на кристалл InAs, помещенный в магнит 6 для увеличения эффективности генерации. Мощность фемтосекундного излучения на входе в кристалл InAs составляла в эксперименте 120 мВт. Лазерное излучение фемтосекундной длительности создает свободные носители заряда, движение которых в магнитном поле генерирует ТГц излучение. Расходящееся ТГц излучение селектируется от излучения накачки с помощью фильтра из фторопласта, собирается внеосевыми параболическими зеркалами 7, 8 и направляется на чашку Петри с объектом 9. Генерируемое ТГц излучение имело полосу частот 0,05–2 ТГц. Мощность ТГц излучения измерялась с помощью оптико-акустического приемника 10 и варьировалась в эксперименте с помощью фильтров, составляя 10,84; 1,97 и 0,91 мкВт. Длительность импульса – 2,5 пс. Время облучения составило 3 мин. Площадь облучения – 0,18 см^2. Эксперименты проводились при комнатной температуре (24ºС). Контроль изменения температуры образца осуществлялся с помощью инфракрасного термодатчика. Распределение мощности в импульсе приведено на рис. 2.

Описание подготовки образцов и методики измерений

В качестве объекта исследования авторами использовались культивированные спинальные ганглии 10–12-дневных куриных эмбрионов, препарированные на уровне пояснично-крестцового отдела спинного мозга. После изъятия ганглии помещали в стерильную чашку Петри на покровные стекла с коллагеновой подложкой. Капли питательной среды наносятся на стенки чашки Петри, что создает подходящие условия для поддержания жизнедеятельности ганглия за время облучения. Питательная среда содержала 35% раствора Хенкса, 35% среды Игла с добавлением инсулина (0,5 ед./мл), глюкозы (0,6%), глютамина (2 мМ), антибиотиков, 5% куриного эмбрионального экстракта и 15% фетальной сыворотки коровы. Эксперимент проводился в закрытой полистирольной чашке Петри. Данный материал довольно хорошо пропускает ТГц излучение – измеренные нами потери составили 16%. Проведение облучения в таком режиме позволяет минимизировать влияние состава окружающей среды, что является важным фактором для экспериментов на культуре ткани.

Таким образом, во время облучения экспериментального эксплантата в чашке Петри отсутствовала питательная среда, температура окружающей среды изменялась не более чем на 1ºС. Образцы, не подвергшиеся влиянию ТГц излучения, считались контрольными.

Для прикрепления эксплантатов после облучения исследуемый объект помещался в CO_2-инкубатор при температуре 36,8ºС с содержанием 5% CO_2. Через трое суток культивирования в контрольных эксплантатах спинномозговых ганглиев формируются две отчетливо выраженные зоны: центральная, состоящая из немигрирующих дифференцирующихся нейробластов, и периферическая. Периферическая зона, представляющая собой характерный ореол вокруг ганглия, образуется растущими нейритами, фибробластоподобными клетками и глией. Рост нервных волокон в культуре ткани исследовался прижизненно с помощью светового микроскопа. Для количественной оценки влияния ТГц излучения на рост нервных волокон спинальных ганглиев применялся морфометрический метод. С целью унификации конечных показателей степени роста нервных волокон для ее оценки использовался относительный критерий – индекс площади (ИП). Интенсивность роста эксплантатов оценивалась по величине ИП, которая рассчитывалась как отношение площади всего эксплантата, включая периферическую зону роста, к исходной площади ганглия. Полученные данные соотносились с данными контрольных эксплантатов, которые не подвергались облучению и культивировались при тех же условиях. Для каждого значения плотности мощности ТГц излучения анализировали 10 экспериментальных эксплантатов. Полученные данные соотносились с данными контрольных эксплантатов, которые не подвергались облучению и культивировались при тех же условиях. Достоверность различий в ИП контрольных и экспериментальных эксплантатов оценивали с помощью *t*-критерия Стьюдента. Статистическая обработка результатов проведена в предположении нормальности распределения с доверительной вероятностью 0,95. Значения ИП выражали в процентах, причем за 100% принимали контрольные значения ИП.

Результаты

В данной работе исследовалось изменение роста нейритов сенсорных ганглиев в органотипической культуре ткани под действием ТГц излучения различной плотности мощности.

При плотностях мощности облучения 60 мкВт/см^2 и 11 мкВт/см^2 изменение ИП относительно контрольного образца составило 103% и 107% соответственно. Эти результаты не являются статистически достоверными и могут рассматриваться как погрешности на уровне контроля. Увеличение индекса площади до 124% ($p<0{,}05$) было получено при плотности мощности облучения 5 мкВт/см^2 (рис. 3), что позволяет говорить нам об эффекте стимуляции клеточного роста.

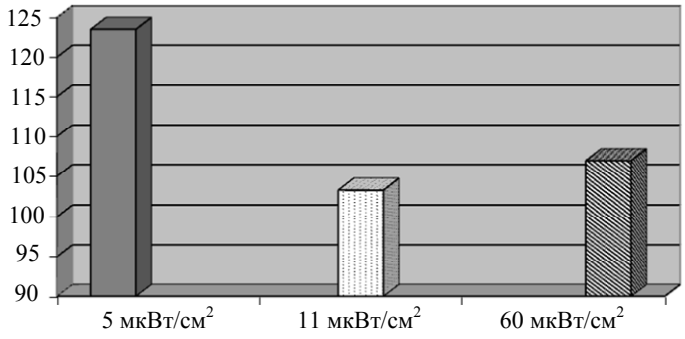


Рис. 3. Влияние ТГц излучения на рост нейритов спинномозговых ганглиев 10–12-дневных куриных эмбрионов. По вертикали – индекс площади эксплантата в % к контрольным значениям. Для плотности мощности 5 мкВт/см^2 p<0,05

Обсуждение результатов и сравнение с литературными данными

В эксперименте мы не наблюдали изменения температуры образца во время воздействия ТГц излучения. Сравнивая полученные нами результаты с литературными данными, можем предположить возникновение стимуляции роста нервных волокон через нетепловые механизмы воздействия, которые обеспечиваются при самой низкой мощности, на которой был достигнут наибольший эффект (5 мкВт/см^2), и непродолжительной экспозиции (3 мин). Однако с учетом того, что пики поглощения воды находятся в ТГц диапазоне, а клетка на 80% состоит из воды, возможно наличие микротермальных эффектов, выражаемых в локальном нагреве элементов клетки, что влияет на ее метаболизм и изменяет функциональное регулирование.

В большинстве исследований, авторы которых также предполагают наличие микротермальных эффектов, использовалось излучение, по мощности и длительности облучения значительно превышающее примененное нами, а также выполнялись требования текущего американского правительственного стандарта по максимально допустимой мощности облучения (MPE) –1 мВт/см^2 – и длительности – 6 мин. В этих пределах мощности облучения также были получены как стимулирующие, так и угнетающие эффекты на биологических объектах [18–22]. Кроме того, во всех известных нам работах, кроме [7, 12, 15], длительность облучения была равной или свыше 15 мин. Работа [15] может быть сравнима с нашей по используемой плотности мощности и длительности воздействия: в приведенной работе при воздействии ТГц излучения на частоте 0,06 ТГц наблюдалось ингибирование и стимуляция потенциала действия при плотностях мощности (0,28, 0,74 и 0,56 мкВт/см^2). Это входит в диапазон частот и мощностей, использованный в эксперименте нами.

Следует, однако, подчеркнуть, что почти во всех перечисленных нами работах в экспериментах использовался источник непрерывного излучения. Нам известна лишь одна печатная работа [11], авторы которой достигли увеличения проницаемости мембраны при использовании импульсного излучения и показали преимущества этого вида облучения по сравнению с непрерывным режимом. Фактически в своей работе мы впервые получили дозозависимый эффект импульсного широкополосного ТГц облучения стимуляции клеточного роста нервных клеток.

Заключение

Получены качественные характеристики и количественные данные, свидетельствующие о стимулирующем воздействии ТГц излучения малой интенсивности на нервные клетки. Обнаружено, что стимулирующий эффект максимально проявляется при сравнительно малой плотности мощности излучения (5 мкВт/см^2), а сам эффект является дозозависимым. Это исследование является первой ступенью масштабной работы. Результаты исследования имеют большое значение для обеспечения безопасности сотрудников при работе с электромагнитным излучением ТГц диапазона. С другой стороны, изменение проницаемости мембраны влечет за собой изменение ионных токов, что можно использовать для усиления клеточного метаболизма. Посредством неразрушающего процесса при подборе правильного режима облучения появляется возможность потенциального контроля нейронного ответа, что в дальнейшем может привести к созданию новых методов терапии.

Работа поддержана Министерством образования и науки РФ (ГК №14.B37.21.0248), ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (ГК № 16.513.11.3070).

Литература

1. Zhang X.-C., Xu J. Introduction to THz wave photonics. – NY: Springer Science+Business Media, 2010. – 248 p.

2. Globus T.R., Woolard D.L., Khromova T., Crowe T.W., Bykhovskaia M., Gelmont B.L., Hesler J. and Samuels A.C. THz-spectroscopy of biological molecules // J. Bio. Phys. – 2003. – V. 29. – № 2. – P. 89–100.
3. Цуркан М.В., Собакинская Е.А., Смолянская О.А., Беспалов В.Г., Вакс В.Л., Балбекин Н.С. Исследование спектра молекулы ДНК в терагерцовой области частот // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 1 (77). – С. 15–19.
4. Gian Piero Gallerano. Tera-Hertz radiation in Biological Research, Investigation on Diagnostics and study of potential Genotoxic Effects. Проектная документация [Электронный ресурс]. – Режим доступа: www.frascati.enea.it/thz-bridge/index.html, свободный. Яз. англ. (дата обращения 13.12.2012).
5. Webb S.J., Dodds D.D. Inhibition of bacterial cell growth by 136 gc microwaves // Nature. – 1968. – V. 218. – № 5139. – P. 374–375.
6. Blackman C.F., Benane S.G., Weil C.M., Ali J.C. Effects of nonionizing electromagnetic radiation on single cell biologic systems // Annals of the New York Academy of Sciences. – 1975. – V. 247. – № 1. – P. 352–366.
7. Berns M.W., Bewley W. Inhibition of nucleic acid synthesis in cells exposed to 200 micrometer radiation from the Free electron laser // Photochemistry and Photobiology. – 1987. – V. 46. – № 2. – P. 165–167.
8. Berns M.W., Bewley W., Sun C.H., Templin P. Free electron laser irradiation at 200 microns affects DNA synthesis in living cells // Proc. Natl. Acad. Sci. – 1990. – V. 87. – № 7. – P. 2810–2812.
9. Hadjiloucas S., Chahal M., Bowen J. Preliminary results on the non-thermal effects of 200–350 GHz radiation on the growth rate of S. cerevisiae cells in microcolonies // Phys. in Med. and Biol. – 2002. – V. 47. – № 21. – P. 3831–3839.
10. Bock J., Fukuyo Y., Kang S., Phipps M.L., Alexandrov L.B., Rasmussen K. Ø., Bishop A.R., Rosen E.D., Martinez J.S., Chen H.-T., Rodriguez G., Alexandrov B.S., Usheva A. Mammalian stem cells reprogramming in response to terahertz radiation // PLoS ONE. – 2010. – V. 5. – № 12. – e15806.
11. Ramundo-Orlando A., Gallerano G.P., Stano P., Doria A., Giovenale E., Messina G., Cappelli M., Arienzo M.D., Spassovsky I. Permeability changes induced by 130 GHz pulsed radiation on cationic liposomes loaded with carbonic anhydrase // Bioelectromagnetics. – 2007. – V. 28. – № 8. – P. 587–598.
12. Olshevskaya J.S., Ratushnyak A.S., Petrov A.K., Kozlov A.S., Zapara T.A. Effect of terahertz electromagnetic waves on neurons systems // Computational Technologies in Electrical and Electronics Engineering. 2008. SIBIRCON 2008. IEEE Region 8 International Conference. – 2008. – P. 210–211.
13. Olshevskaya J.S., Kozlov A.S., Petrov A.K., Zapara T.A., Ratushnyak A.S. Cell membrane permeability under the influence of terahertz (submillimeter) laser radiation // Vestnik Novosibirsk State University. – 2010. – № 5 (4). – P. 177–181.
14. Alekseev S.I., Ziskin M.C. Effects of millimeter waves on ionic currents of Lymnaea neurons // Bioelectromagnetics. – 1999. – V. 20. – № 1. – P. 24–33.
15. Siegel P.H., Pikov V. Can neurons sense millimeter waves? // SPIE Photonics West, BiOS. – 2010. – P. 7562-17.
16. Siegel P.H., Pikov V., Impact of low intensity millimeter-waves on cell membrane permeability // Infrared, Millimeter, and Terahertz Waves, IRMMW-THz (2009), 34th International Conference on 1(1). – 2009. – P. 21–25.
17. Tsurkan M.V., Penniyainen V.A. et all. Changing growth of neurites of sensory ganglions by terahertz radiation // Proc. of SPIE. – 2012. – V. 8261. – P. 82610S-1.
18. Kiselev V.K. Use of submillimeter laser technology in immunological studies // Radiotekhnicheskie Systemy Millimetrovogo i Submillimetrovogo Diapazonov. – 1991. – P. 176–181.
19. Makolinets V.I. Influence of submillimeter wave radiation on some manifestation of immunoresponsives Medithiskaya Reabilitologia // Kurortologiya i Fizioterapiya. – 1996. – V. 1. – P. 33–35.
20. Kirichuck V.F., Ivanov A.N., Antipova O.N., Krenickiy A.P., Mayborodin A.V., Tupikin V.D. Sex-specific differences in changes of disturbed functional activity of platelets in albino rats under the effect of terahertz electromagnetic radiation at nitric oxide frequencies // Bulletin of Experimental Biology and Medicine. – 2008. – V. 145. – № 1. – P. 75–77.
21. Kirichuk V., Tsymbal A. Effects of Terahertz Irradiation at Nitric Oxide Frequencies on Intensity of Lipoperoxidation and Antioxidant Properties of the Blood under Stress Conditions // Bulletin of Experimental Biology and Medicine. – 2009. – V. 148. – № 2. – P. 200–203.
22. Kirichuk V., Tsymbal A. Use of Terahertz Electromagnetic Waves for Correcting Hemostasis Functions // Biomedical Engineering. – 2010. – V. 44. – № 1. – P. 11–14.

Цуркан Мария Валериевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, tsurkan.maria@yandex.ru

Смолянская Ольга Алексеевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, доцент, denisiuk@mail.ifmo.ru

Брянцева Наталья Геннадьевна – Национальный исследовательский Томский государственный университет, кандидат физ.-мат. наук, мл. научный сотрудник, brjantseva@rambler.ru

УДК 535.015

ФЕМТОСЕКУНДНОЕ ФОТОННОЕ ЭХО НА НЕОДНОРОДНЫХ ПО РАЗМЕРУ КВАНТОВЫХ ТОЧКАХ ОКСИДА ЦИНКА ПРИ КОМНАТНОЙ ТЕМПЕРАТУРЕ

Н.С. Вашурин, И.И. Попов, С.Э. Путилин, В.Т. Сидорова, С.А. Степанов, Н.И. Сушенцов

Сообщается об обнаружении в пленке оксида цинка толщиной 100 нм, состоящей из различных по размеру квантовых точек, фемтосекундных сигналов самодифракции и фотонного эха при комнатной температуре. Получены экспериментальные результаты, подтверждающие проявление эффекта Аутлера–Таунса на экситонных переходах квантовых точек.

Ключевые слова: фемтосекундное фотонное эхо, самодифракция, пленка оксида цинка, магнетронное распыление, квантовые точки, эффект Аутлера–Таунса.

Введение

Материалы, содержащие квантовые точки, являются перспективными в области построения оптоэлектронных приборов, в медицине – для визуализации молекулярных объектов, а в биологии – для люминесценции [1]. Интенсивная работа ученых ведущих стран с этими объектами [2] определяется широким спектром их физических свойств и применений [3]. Оптические свойства рассматриваемых материалов интересны тем, что в квантовых точках возникает эффект размерного квантования их энергетического спектра, в то время как в объемных материалах энергетические спектры не зависят от размеров образца резонансной среды. Эти отличия оптических свойств материалов, содержащих квантовые точки, вносят особенность во взаимодействие таких материалов с внешними возмущениями, что приводит к изменению скорости релаксации возбужденных состояний резонансной среды. Применение фотонного эха как высокоразрешающего спектроскопического инструмента позволяет проводить исследования быстропротекающих релаксационных процессов, измерения внутри неоднородно-уширенного контура резонансной спектральной линии. При этом исследуемые релаксационные процессы не подвержены возмущающему действию интенсивного лазерного излучения. Фотонное эхо позволяет не только измерять интегральные релаксационные характеристики уровней и переходов, но и определять вклады в эти характеристики различных механизмов релаксации. В силу изложенного изучение наноразмерных сред, включающих квантовые точки на основе фотонного эха, является весьма перспективным направлением исследований, позволяющим решать актуальные задачи современной науки и техники.

Известны работы по регистрации сигналов аккумулированного фотонного эха при гелиевых температурах на квантовых точках CuBr [4] и сферических квантовых точках, внедренных в стекло CuCl [5]. Эксперименты по наблюдению сигналов фотонного эха в материалах, включающих квантовые точки, при комнатной температуре не проводились. Подобные эксперименты способны значительно повысить практическую значимость подобных работ. Авторами впервые проведено экспериментальное наблюдение фемтосекундных сигналов фотонного эха в пленках оксида цинка (ZnO), включающих квантовые точки, при комнатной температуре. Эксперимент является продолжением ранее выполненной работы [6], где обнаружены фемтосекундные сигналы самодифракции и фотонного эха в пленках ZnO наноразмерной толщины.

Экспериментальная установка

Возбуждение фемтосекундных сигналов фотонного эха проводилось в пленках ZnO толщиной 100 нм, полученных методом вакуумного магнетронного распыления на кремниевой подложке. Это позволило сформировать пленки, включающие квантовые точки, методами Фольмера–Вебера или Странски–Крастанова. Из-за высоких механических напряжений в тонком слое полупроводника, возникающих благодаря рассогласованию постоянных решетки подложки и наносимого материала, имеет место компенсация напряжения, приводящая к образованию трехмерных островков – квантовых точек.

Для возбуждения и регистрации фемтосекундных когерентных сигналов [7] была создана установка на основе титан-сапфировой лазерной системы с усилением чирпованного импульса. Она позволяет обеспечить условия формирования сигнала светового эха в твердом теле и его регистрацию после поступления возбуждающих импульсов, превышающих по интенсивности эхо-сигнал более чем на шесть порядков; дискретно изменять интервал между возбуждающими импульсами с помощью оптической линии задержки на базе линейного транслятора с шаговым двигателем и минимальным шагом 2,5 мкм. Оптическая схема экспериментальной установки для регистрации фемтосекундного фотонного эха приведена на рис. 1.

Фемтосекундная лазерная система 1 на кристаллах титан-сапфира состоит из задающего генератора, лазера накачки задающего генератора, стретчера, многопроходового усилителя, блока синхронизации и управления электрооптическим затвором, лазера накачки многопроходового усилителя и компрессора. Параметры фемтосекундной системы: длительность импульса 40 фс, частота повторения импульса 50 Гц, энергия импульса 1 мДж.

Фемтосекундные импульсы, выходящие из компрессора фемтосекундной лазерной системы, с помощью зеркал 2 и 3 попадают на светоделитель 7, где разделяются на два луча. Оба луча направляются

на фокусирующую линзу 9: один – через зеркала 4–6, второй – через линию задержки 8 с шагом 2,5 мкм (16,5 фс). Линза 9 с фокусным расстоянием 15 см увеличивает плотность мощности обоих пучков на твердотельном резонансном образце 10, находящемся на расстоянии 7,5 см от нее. Часть возбуждающих импульсов, прошедших через резонансную среду, и сформированный в ней сигнал первичного фотонного эха падают на фотодиод, подключенный к персональному компьютеру через аналогово-цифровой преобразователь.

Возбуждающие импульсы направлялись на твердотельный образец резонансной среды под углом $\varphi = 8°$. Фотонное эхо регистрировалось в направлении $k_э$ под углом φ ко второму возбуждающему импульсу, соответственно под углом 2φ к первому возбуждающему импульсу. Подбором зеркал и светоделителя соотношение энергии в направлении k_1 и направлении k_2 было близко к 1:2. Средняя энергия импульсов, зарегистрированных измерителем мощности Nova 2 Ophir в направлении k_1, составляла 7 мВт, а в направлении k_2 – 14,7 мВт. Интенсивность первого возбуждающего импульса с учетом потерь на оптических элементах составляла 0,1 ТВт/см^2, второго возбуждающего импульса – 0,2 ТВт/см^2.

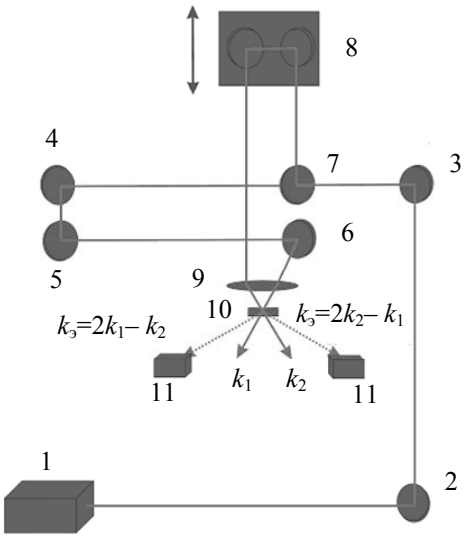


Рис. 1. Оптическая схема фемтосекундного эхо-спектрометра: 1 – фемтосекундная лазерная система; 2–6 – зеркала; 7 – диэлектрический светоделитель; 8 – оптическая линия задержки; 9 – линза; 10 – объект исследования; 11 – фотодетектор, спектрометр. $\boldsymbol{k}_1$ – волновой вектор первого возбуждающего импульса; $\boldsymbol{k}_2$ – волновой вектор второго возбуждающего импульса

Для регистрации спада интенсивности самодифракции и эхо-сигнала во временном интервале между возбуждающими импульсами обеспечено плавное изменение длительности этого интервала от нулевого значения до 800 фс. Это достигнуто за счет изменения длины хода второго луча с помощью оптической линии задержки, перестраиваемой с помощью шагового электропривода в сторону запаздывания ($+\tau_{12}$) или в сторону опережения ($-\tau_{12}$) второго импульса относительно первого. По этой причине одной из задач юстировки эхо-спектрометра является получение нулевой разности хода возбуждающих фемтосекундных импульсов в резонансной среде. Для определения нулевой задержки между импульсами используется неколлинеарная генерация второй оптической гармоники. На место объекта 9 ставится нелинейный кристалл BBO (β-Barium borate), на который под некоторым углом направляются два лазерных пучка в направлении k_1 и k_2 на частоте первой гармоники. При нулевой разности хода возникает неколлинеарная генерация второй гармоники с волновым вектором, направленным между волновыми векторами возбуждающих лазерных импульсов.

Для определения достоверности регистрации сигнала фотонного эха поочередно перекрывались возбуждающие лазерные пучки в направлении k_1 и k_2 или изымалась резонансная среда. Во всех этих случаях сигнал фотонного эха исчезал. Это подтверждает, что появление третьего (правого) импульса фотонного эха связано с наличием резонансного взаимодействия двух возбуждающих импульсов с резонансной средой. При этом эхо-сигнал не является переотражением (бликом) на оптических элементах схемы одного из возбуждающих импульсов.

Экспериментальные результаты

В эксперименте, выполненном при комнатной температуре методом фотонного эха, авторами проведено спектроскопическое исследование экситонных квантовых уровней, возбуждаемых в ансамбле квантовых точек пленки ZnO толщиной 100 нм.

На рис. 2 приведен результат обнаружения фемтосекундных сигналов самодифракции и фотонного эха в зависимости от разности хода между возбуждающими импульсами τ_{12} .

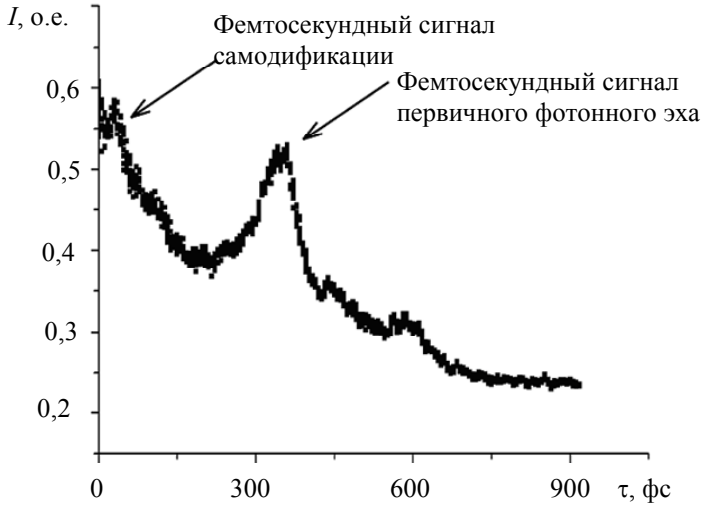


Рис. 2. Интенсивность спада фемтосекундного сигнала самодифракции и фотонного эха при различной разности хода между возбуждающими импульсами

На рис. 3 показаны спектры регистрируемых сигналов самодифракции и фотонного эха при разности хода между возбуждающими импульсами 0,240 фс, 480 фс и 720 фс.

Обсуждение результатов

Как видно из рис. 3, имеет место явное «красное» (в сторону больших значений длин волн) смещение и сужение спектра фемтосекундных сигналов самодифракции и фотонного эха. Это явление объясняется эффектом Аутлера–Таунса (или оптическим штарк-эффектом), при котором под действием приложенного нерезонансного, но мощного лазерного импульса наблюдается расщепление квантовых резонансных уровней [8]. В данном случае варьированием средней мощности и изменением разности хода между возбуждающими импульсами падающего на ансамбль квантовых точек пленки ZnO излучения меняется энергия перехода между основным состоянием материала ZnO и сверхтонкими структурами (верхним и нижним) расщепленного биэкситонного уровня (рис. 4). Непропорциональное сужение высокочастотной и низкочастотной части спектра эхо-сигнала (вносящее свой вклад в его смещение в длинноволновую часть) связано с несимметричным расщеплением резонансного уровня: верхняя компонента сверхтонкой структуры уровня имеет больший энергетический сдвиг, чем нижняя. Сканирование энергии запрещенного перехода между основным состоянием и биэкситонным уровнем происходит в пределах спектральной ширины лазерного импульса [9]. Второй причиной этого смещения является энергетическая релаксация экситонов в хаотическом потенциале в плоскости квантовых ям, когда экситоны мигрируют в их плоскости в поисках более низкоэнергетических локальных минимумов хаотического потенциала с испусканием акустических фононов. Такой механизм энергетической релаксации экситонов характеризуется широким распределением времен релаксации [10], а высокоэнергетичные экситоны имеют большую скорость рекомбинации вследствие их большего радиуса локализации.

Данное значение расщепления спектра резонансного биэкситонного уровня позволило определить момент дипольного перехода по формуле

$$p = \frac{1}{4}\hbar\Omega\sqrt{\frac{c}{\pi J_0}},$$

где $\hbar\Omega$ – разность между крайними значениями зарегистрированного расщепления резонансного перехода; J_0 – среднее значение падающей мощности возбуждения.

Для нашего случая, при максимальном расщеплении 17 мэВ, дипольный момент перехода составил p = 1,34 Д ($4{,}47\cdot10^{-30}$ Кл·м).

Аналогичный эксперимент проводился на пленке CuCl толщиной 10 мкм при гелиевых температурах [11]. При интенсивности возбуждения 18 кВт/см^2 расщепление Аутлера–Таунса оказалось равным 93 мкэВ. Измеренный в этом случае дипольный момент перехода составил p = 30 Д.

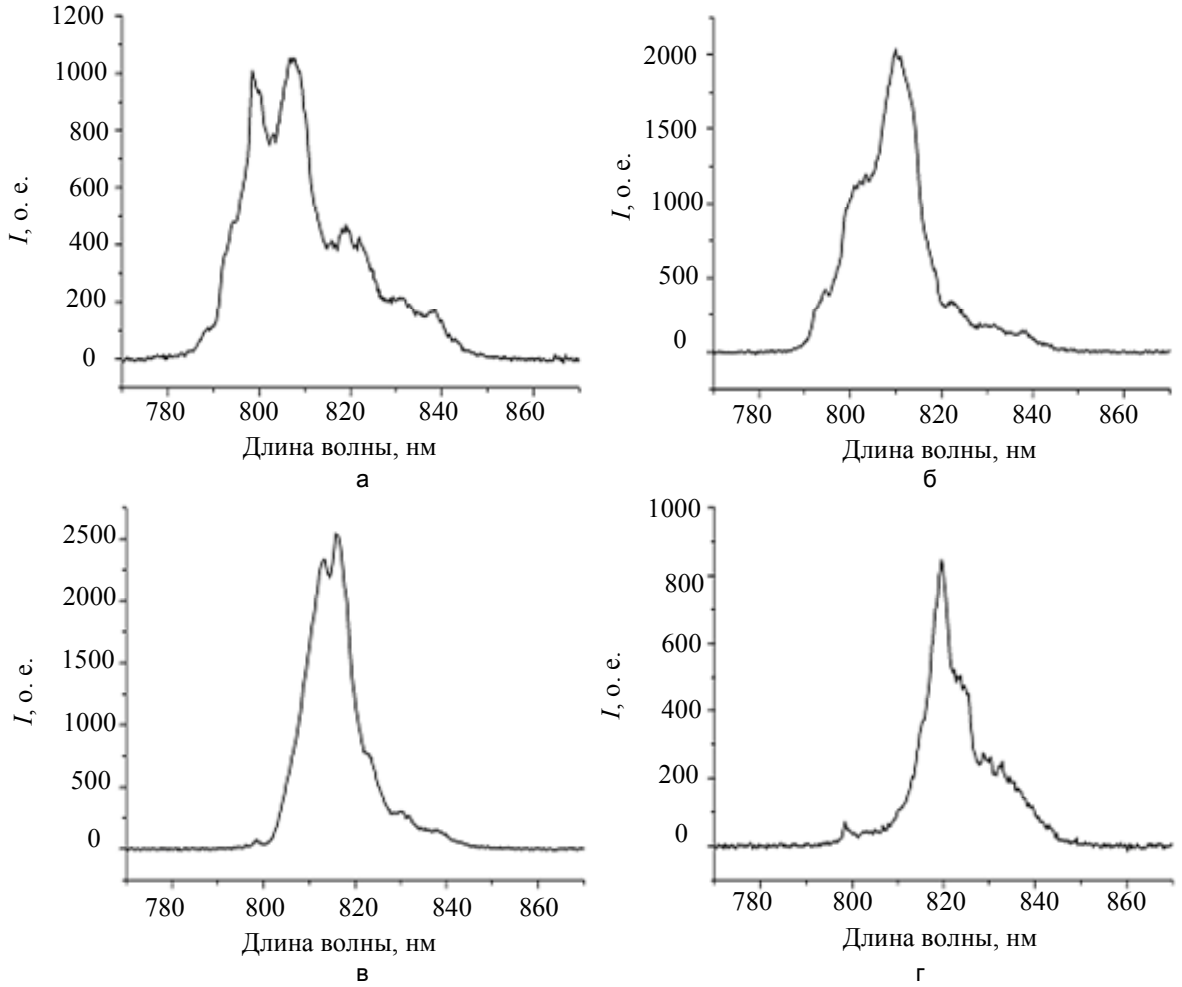


Рис. 3. Спектры фемтосекундных сигналов фотонного эха и самодифракции при различных значениях временного интервала между возбуждающими лазерными импульсами при различных значениях средней мощности возбуждающего излучения: средняя плотность мощности возбуждения P = 0,3 ТВт/см^2, τ_{12} = 0 фс (а); P = 0,0375 ТВт/см^2, τ_{12} = 240 фс (б); P = 0,0214 ТВт/см^2, τ_{12} = 480 фс (в); P = 0,0115 ТВт/см^2, τ_{12} = 720 фс (г)

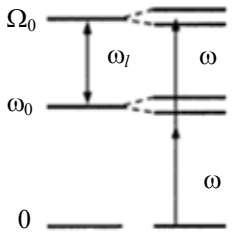


Рис. 4. Схема энергетических уровней (0 – основное состояние кристалла; ω_0 и Ω_0 – экситонное и биэкситонное состояние соответственно) и квантовых уровней под действием поля накачки с частотой ω_l

Заключение

При комнатной температуре наблюдался сигнал первичного фотонного эха, формирующийся в пленке оксида цинка 100 нм, включающей различные по величине квантовые точки. Установлен факт сужения и сдвига в длинноволновую часть спектра регистрируемых сигналов самодифракции и фотонного эха при уменьшении значения средней суммарной мощности лазерного импульса, падающего на поверхность пленки ZnO. Произведена оценка величины дипольного момента биэкситонного перехода, составившая p = 30 Д.

Работа поддержана Министерством образования и науки РФ (ГК №14.B37.21.0248) и грантом РФФИ 12-02-00736-а.

Литература

1. Medintz I.L., Uyeda H.N., Goldman E.R., Mattuosi Y. Quantum dot bioconjugates for imaging, labelling and sensing // Nature Materials. – 2005. – Vol. 4. – P. 435–446.
2. Пул Ч., Оуэнс Ф. Нанотехнологии. – М.: Техносфера, 2006. – 336 с.
3. Суздалев И.П. Нанотехнология: физико-химия нанокластеров, наноструктур и наноматериалов. – М.: КомКнига, 2006. – 592 с.
4. Kuroda T., Minami F., Inoue K., Baranov A.V. Accumulated photon echo in semiconductor microcrystalline quantum dots // Phys. Rev. B. – 1998. – V. 57. – № 4. – P. R2077–R2080.
5. Kuribyashi R., Inoue K., Sakoda K. et al. Long phase-relaxation time in CuCl quanum dots: Four-wave-mixing signals analogous to dye molecules in polymers // Phys. Rev. B. – 1998. – V. 57. – № 24. – P. R15084– R15087.
6. Вашурин Н.С., Попов И.И., Газизов К.Ш., Путилин С.Э., Степанов С.А., Сушенцов Н.И., Сидорова В.Т. Фотонное эхо в парах молекулярного йода и нанопленках как метод оптической обработки информации // Известия РАН. Серия физическая. – 2012. – Т. 76. – № 3. – С. 362–365.
7. Козлов С.А., Самарцев В.В. Основы фемтосекундной оптики. – М.: ФИЗМАТЛИТ, 2009. – 292 с.
8. Хаджи П.И., Коровай А.В., Ткаченко Д.В. Оптические свойства полупроводников в условиях действия мощной накачки в области М-полосы и двухфотонного зондирования биэкситонного уровня // Физика твердого тела. – 2002. – Т. 44. – Вып. 5. – С. 774–779.
9. Вашурин Н.С., Попов И.И., Путилин С.Э., Сушенцов Н.И., Степанов С.А. Обнаружение и исследование сигналов первичного и обращенного стимулированного фотонного эха в неорганических нанопленках // Материалы Шестой международной научной школы «Наука и инновации – 2011». – Йошкар-Ола: МарГУ, 2011. – С. 58–62.
10. Takagahara T. Localization and energy transfer of quasi-twodimensional excitons in GaAs-AlAs quantum-well heterostructures // Phys. Rev. B. – 1985. – V. 31. – P. 6552.
11. Shimano R., Kuwata-Gonokami M. Observation of Autler-Townes Splitting of Biexctons in CuCl // Phys. Rev. Lett. – 1994. – V. 72. – P. 530–533.

Вашурин Никита Сергеевич – Поволжский государственный технологический университет, аспирант, Nickita_Vashurin@mail.ru

Попов Иван Иванович – Поволжский государственный технологический университет, доктор физ.-мат. наук, профессор, popov@volgatech.net

Путилин Сергей Эдуардович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат физ.-мат. наук, ст. научный сотрудник, seputilin@yandex.ru

Сидорова Вера Тагировна Марийский государственный университет, кандидат физ.-мат. наук, доцент, veranig@yandex.ru

Степанов Сергей Александрович Марийский государственный университет, аспирант, stepan_mail@mial.ru

Сушенцов Николай Иванович Поволжский государственный технологический университет, кандидат технических наук, доцент, зав. кафедрой, sni@mari-el.ru

4 КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 001.891.573; 519.872

ПОГРЕШНОСТИ МОДЕЛИРОВАНИЯ ВЫСОКОНАГРУЖЕННЫХ СИСТЕМ В GPSS WORLD

Т.И. Алиев

Излагаются результаты исследования влияния программных генераторов псевдослучайных величин на точность моделирования высоконагруженных систем массового обслуживания в системе имитационного моделирования GPSS World. Показано, что при некоторых сочетаниях генераторов, называемых проблемными, погрешности определения среднего времени ожидания заявок в высоконагруженной системе могут достигать десятков процентов. Выполненный анализ проблемных сочетаний позволил выявить факторы, обусловливающие большие погрешности результатов моделирования.

Ключевые слова: имитационное моделирование, высоконагруженная система, генераторы псевдослучайных величин, погрешность моделирования, GPSS World.

Введение

Одной из наиболее апробированных и широко распространенных систем имитационного моделирования (СИМ) является GPSS (General Purpose Simulation System) и, в частности, реализованная в среде Microsoft Windows версия GPSS World [1], в которой используются программные генераторы равномерно распределенных в интервале (0;1) псевдослучайных величин. Последовательность формируемых в процессе моделирования значений псевдослучайных величин зависит от задаваемого в качестве параметра целочисленного начального значения, что позволяет говорить о множестве используемых в GPSS World генераторов случайных величин, которые будем обозначать как 1, 2, 3, … . Кроме того, в GPSS World имеется более 20 библиотечных процедур для формирования случайных величин с заданными законами распределений, базирующихся на встроенных генераторах равномерно распределенных величин.

Полувековая история существования СИМ GPSS и ее широкое применение в различных областях сформировали мнение, что при достаточно большой и обоснованно выбранной длительности моделирования погрешности результатов имитационного моделирования обычно не превышают 3–5%. Однако при моделировании высоконагруженных систем массового обслуживания (СМО), загрузка которых $\rho \geq 0{,}8$, было выявлено, что при некоторых сочетаниях генераторов случайных величин, задаваемых в виде {N1; N2} (здесь N1 и N2 – номера генераторов, используемых для формирования значений интервалов между поступающими в систему заявками и длительностей обслуживания заявок соответственно), погрешность результатов моделирования может составлять 30% и более. При этом для других сочетаний генераторов эта погрешность при той же длительности моделирования, задаваемой числом прошедших через систему заявок, оказывается менее 3%.

Для выяснения обстоятельств, связанных с появлением такого эффекта, было проведено исследование генераторов псевдослучайных величин с целью оценки качества используемых в СИМ GPSS World генераторов и выявления причин, обусловливающих большую погрешность результатов моделирования при использовании некоторых сочетаний. Исследование проводилось в несколько этапов.

На первом этапе были выявлены так называемые проблемные сочетания генераторов, при использовании которых погрешность результатов моделирования оказывается значительной.

Для объяснения возникающей погрешности были протестированы проблемные встроенные генераторы равномерно распределенных псевдослучайных величин и библиотечные генераторы на соответствие декларируемым законам распределения [2].

Поскольку большая погрешность моделирования возникает только при использовании некоторых конкретных сочетаний генераторов, необходимо было выполнить корреляционный анализ соответствующих последовательностей случайных значений, рассчитать коэффициенты корреляции и сравнить эти коэффициенты для проблемных и непроблемных сочетаний генераторов случайных величин.

На последнем этапе была проанализирована зависимость результатов моделирования, а именно, среднего времени ожидания заявок, от длительности моделирования, задаваемого количеством заявок, прошедших через модель.

Описание экспериментов

Для исследования погрешностей результатов имитационного моделирования и выявления проблемных сочетаний генераторов использовались GPSS-модели одно- и многоканальных СМО типа M/M/1, M/G/1 и M/M/2, поддающиеся точному аналитическому расчету [2]. Варьируя загрузку систем от 0,1 до 0,9 и пропуская через каждую модель по 1 млн. заявок, было выявлено, что погрешность результа-

тов моделирования лежит в приемлемых пределах (1–5%) для значений загрузки примерно до 0,7. Начиная со значения ρ≥0,8, при некоторых сочетаниях генераторов случайных величин погрешность начинала расти и составляла 10% и более, а при ρ=0,95 достигала 40%. При исследовании реальных систем (маршрутизаторов, коммутаторов, серверов в компьютерных сетях) наибольший интерес обычно представляют сведения о качестве функционирования системы при большой загрузке, что делает задачу обеспечения приемлемой точности в процессе имитационного моделирования весьма актуальной. В связи с этим все последующие эксперименты выполнялись для систем, загрузка которых составляла 0,9 и 0,95. Для указанных значений загрузки системы были выполнены более 10 тыс. экспериментов при различных сочетаниях генераторов с номерами от 1 до 100 и выявлены сочетания генераторов, при которых достигается наибольшая погрешность результатов моделирования. В процессе экспериментов было установлено, что погрешности моделирования зависят только от загрузки системы $\rho=\lambda b$ и длительности моделирования, задаваемого количеством заявок, прошедших через систему, и не зависят от значений параметров загрузки – интенсивности поступления заявок λ и средней длительности обслуживания заявок b. Кроме того, было установлено, что использование проблемных сочетаний генераторов случайных величин приводит к значительным погрешностям результатов моделирования для других, более сложных моделей, в частности, для многоканальных СМО и двухузловых сетей массового обслуживания (СеМО), при других законах распределений, использующих эти генераторы.

Проблемные сочетания генераторов

На модели M/M/1 с однородным потоком заявок и загрузкой ρ=0,9 было проведено 10 000 экспериментов с целью выявления всех проблемных сочетаний генераторов с номерами от 1 до 100. В каждом эксперименте через модель пропускался 1 млн. заявок. Относительная погрешность моделирования рассчитывалась для среднего времени ожидания заявок как $\delta w = \frac{w - w^*}{w^*} \cdot 100\%$, где w – значение, полученное в процессе имитационного эксперимента; w^* – точное значение, рассчитанное для СМО M/M/1. Для большинства сочетаний погрешность не превышала 5%. В то же время, как видно из табл. 1, для 40 сочетаний погрешность превысила 20% и для 128 сочетаний – 10%.

Погрешность	Проблемные сочетания генераторов								
10–20%	{1;2}	{3;6}	{3;9}	{4;6}	{4;12}	{5;15}	{6;3}	{6;18}	{7;14}
	{7;21}	{8;16}	{8;32}	{8;48}	{9;3}	{10;30}	{11;77}	{12;4}	{12;6}
	{13;26}	{13;65}	{14;7}	{14;42}	{15;3}	{15;5}	{15;10}	{15;45}	{16;8}
	{16;32}	{16;48}	{16;64}	{17;51}	{17;68}	{18;6}	{19;57}	{20;5}	{20;30}
	{20;60}	{20;100}	{21;7}	{21;14}	{21;63}	{22;44}	{22;66}	{24;48}	{24;72}
	{24;96}	{25;50}	{25;75}	{26;13}	{27;9}	{27;54}	{28;56}	{29;87}	{30;6}
	{30;10}	{30;20}	{30;60}	{32;8}	{32;48}	{33;22}	{33;66}	{33;99}	{36;12}
	{38;19}	{38;76}	{39;78}	{40;8}	{40;10}	{40;80}	{41;82}	{42;6}	{42;14}
	{42;28}	{44;22}	{44;88}	{46;69}	{46;92}	{47;94}	{48;8}	{48;16}	{48;24}
	{48;96}	{50;25}	{51;17}	{60;15}	{60;20}	{60;30}	{63;9}	{63;21}	{64;16}
	{65;13}	{66;22}	{68;17}	{69;46}	{70;35}	{72;24}	{75;15}	{76;38}	{80;40}
	{82;41}	{84;12}	{84;56}	{87;29}	{88;44}	{92;46}	{93;31}	{94;47}	{96;24}
	{96;48}	{99;33}	{100;20}	{100;50}					
>20%	{2;1}	{4;8}	{5;10}	{8;4}	{9;18}	{10;5}	{10;20}	{12;24}	{14;28}
	{15;30}	{18;9}	{20;10}	{20;40}	{21;42}	{24;12}	{26;52}	{28;14}	{30;15}
	{31;62}	{32;16}	{32;64}	{36;72}	{37;74}	{40;20}	{42;21}	{43;86}	{45;90}
	{49;98}	{52;26}	{54;27}	{57;19}	{62;31}	{64;32}	{66;33}	{72;36}	{74;37}
	{78;39}	{86;43}	{90;45}	{98;49}					

Таблица 1. Проблемные сочетания генераторов

Анализ представленных результатов показывает, что наиболее проблемными являются сочетания генераторов, номера которых соотносятся как 1:2, 1:3, 2:1 или 3:1.

Большинство проблемных сочетаний генераторов были проверены также для СМО общего вида G/G/1, в частности, U/U/1 и многоканальных СМО. Для сравнения в табл. 2 представлены результаты имитационного моделирования при использовании различных сочетаний генераторов в СМО M/M/1, загрузка которой ρ=0,95, а длительность обслуживания b=9,5 единиц времени. В этом случае точное значение среднего времени ожидания w=180,5 единиц.

Генераторы	Время ожидания, ед.	Погрешность, %
{2; 1}	114,4	–36,6
{2; 2}	173,6	–3,8
{2; 4}	155,8	–13,7
{20; 40}	111,8	–38,1
{40; 20}	113,2	–37,3
{110; 40}	179,9	–0,3
{200; 400}	136,5	–24,4
{999; 40}	177,1	–1,9
{2000; 4000}	119,3	–33,9

Таблица 2. Погрешности моделирования СМО M/M/1

Как видно из табл. 2, для проблемных сочетаний генераторов случайных величин погрешность результатов моделирования достигала 40%, в то время как у непроблемных сочетаний {2; 2}, {110; 40} и {999; 40} при той же длительности моделирования погрешность не превышала 5%. Более того, использование одного и того же генератора с номером 2 {2; 2} для формирования значений интервалов между поступающими заявками и длительности обслуживания позволяет получить существенно более точный результат, чем использование этого же генератора в сочетании с другим генератором. Перестановка генераторов в сочетании с {10; 20} на {20; 10} незначительно изменяет погрешность моделирования.

Аналогичные эксперименты были выполнены для двухузловой разомкнутой экспоненциальной СеМО, в которой заявки последовательно проходят через узлы, загрузка которых $\rho_1=\rho_2=0{,}95$. Расчетные значения среднего времени ожидания составили в узлах $w_1=w_2=180{,}5$ и в сети $W=w_1+w_2=361$ единиц. В табл. 3 представлены результаты имитационного моделирования при различных сочетаниях генераторов, задаваемых в виде {N1; N2; N3}, где N1, N2 и N3 – номера генераторов, используемых для формирования значений интервалов между поступающими в сеть заявками и длительностей обслуживания заявок в узлах 1 и 2 соответственно.

Сочетание генераторов	Узел 1		Узел 2		Сеть	
	w_1, ед.	δw_1, %	w_2, ед.	δw_2, %	W, ед.	δW, %
{10; 20; 40}	121,1	–32,9	134,7	–25,39	255,72	–29,16
{20; 10; 40}	106,6	–41,0	122,0	–32,39	228,61	–36,67
{40; 20; 10}	113,2	–37,3	128,0	–29,11	241,13	–33,20
{21; 15; 719}	169,0	–6,4	170,7	–5,43	339,68	–5,91
{20; 10; 20}	172,7	–4,3	185,4	2,69	358,04	–0,82
{14; 28; 56}	110,4	–38,9	157,2	–12,90	267,58	–25,88
{28; 56; 14}	138,4	–23,3	134,2	–25,65	272,62	–24,48
{140; 280; 560}	124,4	–31,1	133,5	–26,06	257,88	–28,57
{280; 560; 140}	116,1	–35,7	128,7	–28,71	244,80	–32,19

Таблица 3. Погрешности моделирования двухузловой СеМО

Как видно из табл. 3, относительная погрешность (разброс) результатов моделирования для проблемных сочетаний генераторов составляет более 25% как для узловых, так и для сетевых характеристик. Большой разброс наблюдается также для такой важной характеристики, как максимальная длина очереди заявок. Так, например, максимальная длина очереди заявок в узлах 1 и 2 для сочетания {20; 10; 20} составляет 153 и 186 соответственно, а для проблемного сочетания {20; 10; 40} – 79 и 95, т.е. имеем почти двукратное отличие.

Тестирование генераторов

Генератор равномерно распределенных случайных величин вырабатывает базовую последовательность значений в интервале (0; 1), которая используется для формирования случайных величин, распределенных по другим законам и в других интервалах. Поскольку погрешность может появиться как на этапе генерации базовой последовательности, так и на этапе формирования экспоненциально распределенной величины, было проведено тестирование базовых последовательностей и последовательностей экспоненциально распределенных случайных величин.

Относительная погрешность математического ожидания случайных величин для генераторов, входящих в состав проблемных сочетаний, в большинстве случаев составляла более 0,1%, в то время как

для других генераторов – менее 0,1%. Этот результат справедлив как для базовых последовательностей, так и для последовательностей экспоненциально распределенных случайных величин.

Для базовой последовательности относительная погрешность среднеквадратического отклонения для всех исследованных генераторов, включая генераторы, входящие в состав проблемных сочетаний, составила менее 0,01%. Для последовательностей экспоненциально распределенных случайных величин эта погрешность оказалась больше 0,01%, а для генераторов, входящих в состав проблемных сочетаний, достигала значений 0,2% и более.

Результаты корреляционного анализа

Корреляционный анализ базовых последовательностей и последовательностей, экспоненциально распределенных случайных величин на выборке по 1 млн. значений каждого генератора, показал, что коэффициенты корреляций проблемных сочетаний генераторов случайных величин в 5–10 раз превышают значения коэффициентов корреляции непроблемных сочетаний, причем они принимают положительные значения, т.е. большему (меньшему) значению первого генератора соответствуют большие (меньшие) значения второго генератора. Это означает, что с большой вероятностью заявке, поступившей в систему с большим интервалом (сформированным первым генератором) от момента поступления предыдущей заявки, будет соответствовать (сформировано вторым генератором) большее время обслуживания, и, наоборот, заявке, поступившей в систему с небольшим интервалом от момента поступления предыдущей заявки, будет соответствовать маленькое время обслуживания. Вероятность же ситуации, когда заявке, поступившей в систему с маленьким интервалом, будет соответствовать большое время обслуживания, оказывается незначительной. Все это приводит к тому, что результаты имитационного моделирования, в частности, среднее время ожидания, оказываются существенно заниженными даже при достаточно длительном моделировании.

Анализ переходных процессов в модели

Для проблемных сочетаний генераторов с целью выяснения характера изменения среднего времени ожидания от длительности имитационного моделирования, задаваемого числом заявок, проходящих через моделируемую систему, были выполнены эксперименты, при которых через систему пропускалось до 100 млн. заявок, а в некоторых моделях до 200 млн. заявок. Изучение переходного процесса в СМО позволило установить, что даже при таком большом количестве заявок, пропущенных через систему, среднее время ожидания не достигало расчетного значения, а погрешность могла составлять до 10%. На рис. 1 представлены зависимости среднего времени ожидания заявок и относительной погрешности результатов имитационного моделирования от числа заявок, прошедших через имитационную модель СМО M/M/1 с загрузкой 0,95, для сочетания генераторов {20;10}.

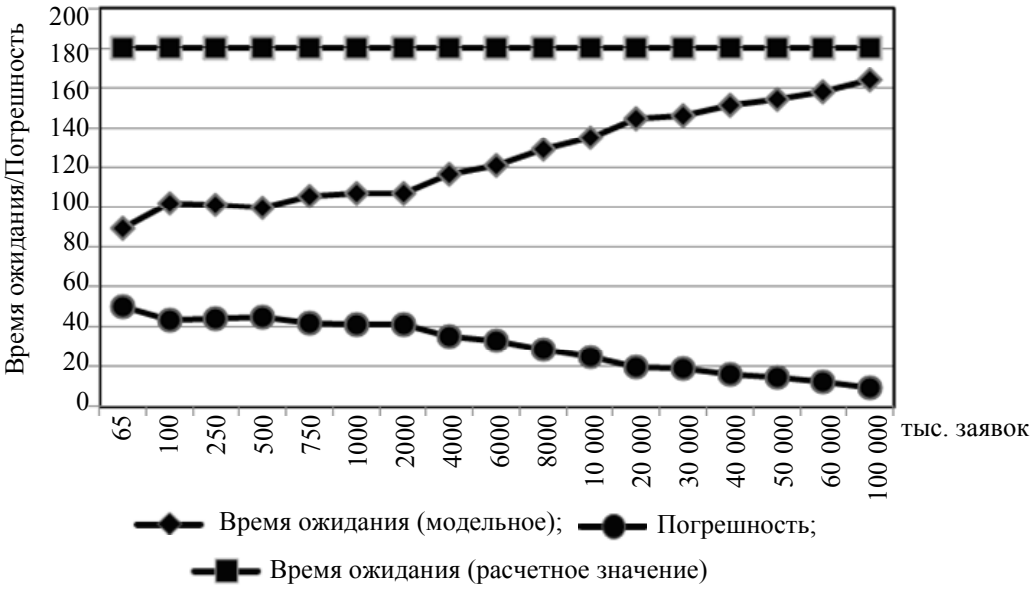


Рис. 1. Проблемное сочетание генераторов {20; 10}

Из графика видно, что даже после прохождения через модель 100 млн заявок относительная погрешность составляет около 10%. Отметим, что коэффициент корреляции генераторов 20 и 10 положителен и равен 0,00247.

Заметим также, что использование широко применяемого на практике метода «срезов» для определения вхождения системы в стационарный режим может привести к ошибочным результатам. Например, как видно из графика, среднее время ожидания при пропускании 100 тыс., 250 тыс. и 500 тыс. заявок остается практически неизменным (отличие менее чем на 1%), откуда можно сделать вывод, что система вошла в стационарный (установившийся) режим и, следовательно, процесс моделирования может быть завершен. При этом погрешность результата моделирования оказывается более 40%.

Для сравнения на рис. 2 показана аналогичная зависимость для той же модели при использовании непроблемного сочетания генераторов {110; 40}, откуда видно, что установившийся режим в системе существует уже после пропускания 500 тыс. заявок, и погрешность результатов моделирования в дальнейшем не превышает 4%. Следует обратить внимание, что такой характер зависимости модельного времени ожидания по отношению к расчетному может быть объяснен тем, что коэффициент корреляции генераторов 110 и 40 – отрицательный и равен –0,00176.

Рис. 2. Непроблемное сочетание генераторов {110;40}

Для двухузловой разомкнутой СеМО на рис. 3 представлены зависимости относительных погрешностей среднего времени ожидания заявок в узлах и в сети в целом от количества заявок, прошедших через имитационную модель при использовании проблемного сочетания генераторов {20; 10; 40}. Из графика видно, что даже после пропускания через модель 200 млн заявок относительная погрешность результатов моделирования составляет около 10%.

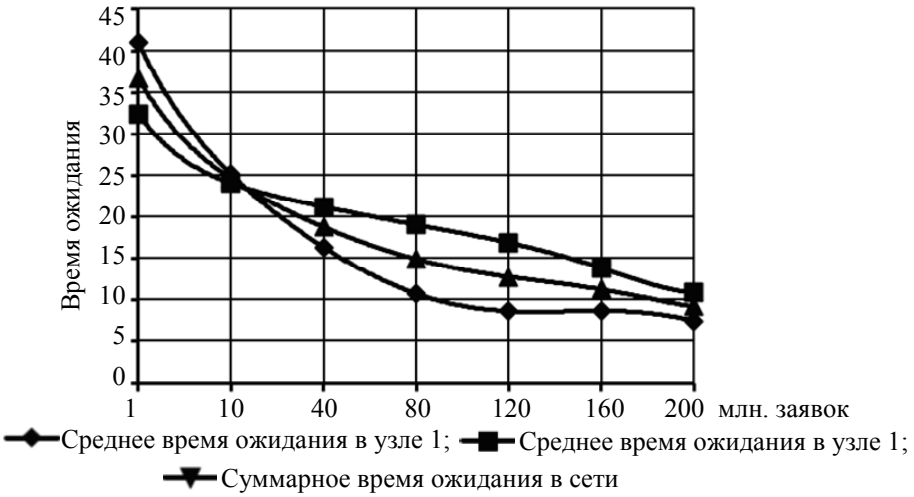


Рис. 3. Относительные погрешности для сочетания генераторов {20; 10; 40}

Заключение

Погрешность результатов моделирования высоконагруженных систем ($\rho \geq 0{,}8$) в системе имитационного моделирования на GPSS World при некоторых сочетаниях генераторов случайных величин, используемых для формирования значений интервалов между поступающими в систему заявками и длительностей обслуживания заявок, может оказаться значительной и достигать десятков процентов.

Результаты многочисленных экспериментов позволили выявить, что наибольшая погрешность достигается для сочетаний генераторов, номера которых соотносятся как 1:2, 1:3, 2:1 и 3:1.

Выполненный корреляционный анализ показал, что коэффициенты корреляций проблемных сочетаний генераторов принимают положительные значения и по абсолютной величине в 5–10 раз превышают значения коэффициентов корреляции непроблемных сочетаний, что обусловливает значительные погрешности результатов моделирования высоконагруженных систем.

Анализ переходных процессов в моделях высоконагруженных систем при использовании проблемных сочетаний генераторов случайных величин показывает, что среднее время ожидания заявок медленно увеличивается с увеличением длительности моделирования (числа заявок, пропускаемых через модель), но так и не достигает точного значения, полученного аналитически, в некоторых случаях даже после прохождения через моделируемую систему более 100 млн. заявок.

Литература

1. Бражник А.Н. Имитационное моделирование: возможности GPSS World. – СПб: Реноме, 2006. – 439 с.
2. Алиев Т.И. Основы моделирования дискретных систем. – СПб: СПбГУ ИТМО, 2009. – 363 с.

Алиев Тауфик Измайлович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, aliev@d1.ifmo.ru

УДК 004.8

ИЕРАРХИЯ ГЛОБАЛЬНЫХ СТРУКТУР АЛГЕБРАИЧЕСКОЙ БАЙЕСОВСКОЙ СЕТИ КАК СИСТЕМА ГРАФОВ И ГИПЕРГРАФОВ[1]

А.А. Фильченков

Алгебраические байесовские сети относятся к классу логико-вероятностных графических моделей и позволяют осуществлять логико-вероятностный вывод, в том числе и в отношении знаний с неопределенностью, формализуемых через скалярные и интервальные оценки истинности пропозициональных формул. В работе рассмотрены глобальные структуры алгебраической байесовской сети, предложена их систематизация через гиперграфовое представление и выявлена их функциональная иерархия. Предложенная систематизация позволяет приложить методы теории гиперграфов к решению ряда задач анализа глобальных структур алгебраической байесовской сети, в частности, предложить и обосновать критерий для выявления ацикличности ее первичной структуры, а также формирует теоретическую основу алгоритмизации автоматического обучения указанных сетей.

Ключевые слова: алгебраическая байесовская сеть, графы смежности, машинное обучение, глобальная структура, вероятностные графические модели.

Введение

Алгебраические байесовские сети (АБС) лежат на стыке двух классов основных подходов, применяемых к построению математических моделей баз знаний с неопределенностью. Первый из этих классов – подходы, использующие оценки истинности над фрагментами знаний (ФЗ). Ко второму классу относятся так называемые «сетевые» подходы, заключающиеся в использовании графов для моделирования связей (причинно-следственных, логических, реляционных) между элементами или наборами элементов базы знаний [1, 2].

Так, АБС являются математической моделью базы ФЗ с неопределенностью, в которых, в свою очередь, математической моделью ФЗ выступает идеал конъюнктов с заданными над ними оценками вероятности их истинности [1]. Декомпозиция предметной области на базу (набор) таких моделей ФЗ, с прикладной точки зрения, позволяет снизить требования к вычислительным ресурсам (памяти и времени), необходимым для их хранения и обработки, а с теоретической – служит причиной для выделения двух структур АБС: первичной и вторичной.

Первичная структура АБС тесно связана с подходами первого из рассмотренных классов, поскольку представляет собой набор ФЗ. Вторичная же структура АБС – граф [3, 4], построенный над первичной структурой, – согласуется с «сетевыми» подходами. Эти две структурные особенности позволяют отнести АБС к классу логико-вероятностных графических моделей [5], из которого АБС выделяются возможностью использовать интервальные оценки вероятности истинности для представления неопределенности [1, 2]. Помимо первичной и вторичной структур у АБС выделяют также ряд других глобальных структур [6], тесно связанных друг с другом и играющих ключевую роль при решении задачи глобально-

[1] Работа выполнена при финансовой поддержке РФФИ, гранты №№ 12-01-00945-а и 12-01-31202-мол_а.

го обучения АБС. Глобальное обучение АБС – это один из видов машинного (автоматического) обучения, развитие методов, моделей и алгоритмов которого является одной из самых актуальных задач в искусственном интеллекте [2, 5, 7, 8]. Цель данной работы – выявить отношения между указанными глобальными структурами АБС, т.е. описать их иерархию и систематизировать составляющие ее объекты на языке теории гиперграфов. Формально каждое подобное сопоставление является теоремой о представлении, однако и формулировки, и доказательства подобных теорем однотипны, поэтому для краткости мы не будем их воспроизводить в полном масштабе, ограничиваясь лишь содержательным обоснованием и заключением. Выявленные представления позволяют применить известные результаты теории графов (в частности, гиперграфов) в рамках задач теории АБС, в особенности в отношении вопросов построения ациклических и связных вторичных структур.

Будем везде далее под структурой понимать структуру АБС.

Фрагмент знаний и первичная структура как набор фрагментов знаний

Далее во всей работе будем следовать системе терминов теории АБС, основанной на работах [4, 6, 9, 10] и теории графов по [11]. Алфавит $A = \{x_1, \ldots, x_n\}$ – множество атомарных пропозициональных формул (атомов). Фрагмент знаний – пара $C_{A'}, \boldsymbol{p}_{A'}$, где $C_{A'}$ – идеал конъюнктов над подалфавитом $A' \subset A$, не содержащий пустой конъюнкт:

$$C_{A'} = \left\{ \bigwedge_{x_i \in K} x_i | K \subseteq A', K \neq \varnothing \right\},$$

а $\boldsymbol{p}_{A'}$ – функция, заданная на элементах $C_{A'}$, значениями которой являются либо значения из отрезка $[0;1]$, (тогда говорят о скалярных оценках), либо отрезок $[a;b]: 0 \le a \le b \le 1$ (тогда говорят об интервальных оценках). Если ФЗ непротиворечив, то $\boldsymbol{p}_{A'}$ в скалярном случае является вероятностью, а в интервальном – ассоциирована с заданием внешней и внутренней меры. Фрагмент знаний, построенный над A' – подалфавитом A – будем обозначать как $\mathrm{KP}_{A'}$.

Первичной структурой MKP_A, заданной над алфавитом A, называется набор максимальных по включению ФЗ, построенных над подалфавитами A, образующими покрытие A:

$$\mathrm{MKP}_A = \{\mathrm{KP}_{A_i}\}_{i=1..N} : \forall 1 \le i, j \le N C_{A_i} \subsetneq C_{A_j}, A_i \subset A, \text{ причем } \bigcup_{i=1..N} A_i = A.$$

Первичная структура как гиперграф и протоструктура

В данной работе нас будут интересовать лишь структуры АБС (ее «топология»), а не вероятностная семантика, поэтому оказывается достаточным считать первичной структурой набор соответствующих ФЗ подалфавитов [4]. Все рассуждения о подалфавитах, связанные с отношением включения, операциями объединения и пересечения, оказываются справедливы и для построенных над ними идеалов.

Таким образом, в рамках рассмотрения глобальных структур, первичной структурой над алфавитом A называется набор максимальных по включению подалфавитов A, образующих покрытие A: $\mathrm{PS}_A = \{A_i\}_{i=1..N}, \forall 1 \le i, j \le N A_i \subsetneq A_j, A_i \subset A,$ причем $\bigcup_{i=1..N} A_i = A.$

С точки зрения классической системы определений теории гиперграфов [11], которой мы будем придерживаться на протяжении всей работы, первичная структура является множеством ребер гиперграфа, вершинами которого выступают атомы A. Будем обозначать такой гиперграф как H_A: $H_{\mathrm{PS}_A} = \langle A, \mathrm{PS}_A \rangle.$

Теорема. Первичная структура является минимальным гиперграфом, т.е. она не содержит голых элементов, вложенных ребер и кратных ребер.

Доказательство. Благодаря максимальности подалфавитов первичная структура не содержит вложенных ребер, кратных ребер, а также голых ребер – т.е. ребер, степень которых равна нулю, поскольку любое неголое ребро будет содержать голое ребро. Благодаря тому, что подалфавиты образуют покрытие алфавита A, в гиперграфе нет голых вершин.

Протоструктурой *АБС* будем называть первичный граф первичной структуры, т.е. ненаправленный граф, построенный над тем же алфавитом, в котором две вершины смежны тогда и только тогда, когда они смежны в первичной структуре, т.е. входят в один и тот же подалфавит:

$$\mathrm{pS}(\mathrm{PS}_A) = L^{(2)}(H_{\mathrm{PS}_A}) = \langle A, \{(x_i, x_j) \mid \exists A_i \in \mathrm{PS}_A : x_i, x_j \in A_i\} \rangle .$$

Далее для краткости будем под протоструктурой понимать протоструктуру АБС.

Клики в протоструктуре будут совпадать с кликами в H_A, однако максимальные клики в протоструктуре вовсе не обязательно будут совпадать с подалфавитами, входящими в соответствующую первичную структуру.

Магистральная связность и вторичная структура

Если в предыдущем разделе мы строили графы над алфавитом, то теперь будем строить их над первичной структурой. В подобных графах введем функцию нагрузки (веса) вершины: нагрузкой (весом) вершины будем называть соответствующий ей подалфавит. Будем обозначать нагрузку вершины v как $W(v)$.

Сепаратор двух вершин v и u – пересечение нагрузок их вершин: $\mathrm{Sep}(v,u) = W(v) \cap W(u)$. Две вершины будем называть сочлененными, если их сепаратор непуст. Нагрузкой ребра в подобных графах будем называть сепаратор его концов: $W((v,u)) = \mathrm{Sep}(v,u)$.

Графы максимальных фрагментов знаний (граф МФЗ) над данной первичной структурой PS_A – графы, построенные над подалфавитами, входящими в PS_A, и в которых ребра возможны только сочлененными вершинами. Множество графов МФЗ над первичной структурой PS_A обозначим как $\mathbf{MKPG}(\mathrm{PS}_A)$:

$$G \in \mathbf{MKPG}(\mathrm{PS}_A) \Leftrightarrow \left(G = \langle \mathrm{PS}_A, E\rangle, (u,v) \in E \Rightarrow \mathrm{Sep}(v,u) \neq \varnothing\right).$$

В графе МФЗ нагрузка каждого ребра непуста. С точки зрения структуры гиперграфа H_{PS_A}, это означает, что для любых двух несмежных гиперребер соответствующие им вершины в графе МФЗ не могут быть связны.

Две сочлененные вершины в графе называются магистрально связными, если между ними существует магистральный путь, т.е. такой путь, в котором нагрузка каждой вершины содержит сепаратор концов этого пути. Множество пар магистрально связных сочлененных вершин в графе G обозначим как BBC_G:

$$\mathrm{BBC}_G = \{(u,v) \mid \exists \{x_i\}_{i=0}^{n} : u = x_0, v = x_n,$$

и $\forall i \in \overline{1,n}\ \ (x_{i-1}, x_i) \in E(G), W(x_i) \supset W(u) \cap W(v)\}$.

Магистрально связный граф – граф, в котором каждая пара сочлененных вершин магистрально связна. Множество магистрально связных графов над первичной структурой PS_A обозначим как $\mathbf{BBcon}(\mathrm{PS}_A)$:

$$G \in \mathbf{BBcon}(\mathrm{PS}_A) \Leftrightarrow \left(G = \langle \mathrm{PS}_A, E\rangle,\ u,v \in \mathrm{PS}_A, \mathrm{Sep}(v,u) \neq \varnothing \Rightarrow (v,\ u) \in \mathrm{BBC}_G\right).$$

С точки зрения структуры гиперграфа H_{PS_A} это означает, что для каждой пары его смежных гиперребер можно указать последовательность гиперребер, содержащих пересечение их вершин, такую, что в магистрально связном графе соответствующая последовательность вершин образует путь между вершинами, соответствующими исходным гиперребрам (следует отметить, что последовательность может быть и пустой, тогда две соответствующие исходным гиперребрам вершины должны быть соединены ребром).

Граф смежности – магистрально связный граф МФЗ. Множество графов смежности над первичной структурой PS_A обозначим как $\mathbf{JG}(\mathrm{PS}_A)$: $\mathbf{JG}(\mathrm{PS}_A) = \mathbf{MKPG}(\mathrm{PS}_A) \cap \mathbf{BBcon}(\mathrm{PS}_A)$.

Максимальный граф смежности $G_{\max}(\mathrm{PS}_A)$ – граф смежности, число ребер которого максимально: $G_{\max}(\mathrm{PS}_A) = \mathrm{argmax}_{G \in \mathbf{JG}(\mathrm{PS}_A)} |E(G)|$. Заметим, что определение корректно, поскольку над заданной первичной структурой существует единственный максимальный граф смежности [4]. Он обладает рядом интересных особенностей. Во-первых, множество нагрузок его ребер совпадает с множеством непустых сепараторов соответствующей первичной структуры. Во-вторых, максимальный граф смежности является для соответствующего первичной структуре гиперграфа реберным графом (или, иначе, дуальным графом), т.е. графом над ребрами гиперграфа, в котором вершины соединены ребром, если соответствующие им гиперребра смежны:

$$G_{\max}(\mathrm{PS}_A) = L(H_{\mathrm{PS}_A}) = \left\langle E(H_{\mathrm{PS}_A}), E\right\rangle, (v,u) \in E \Leftrightarrow W(v) \cap W(u).$$

Минимальным графом смежности называется граф смежности, число ребер которого минимально. Множество минимальных графов смежности над первичной структурой PS_A будем обозначать как $\mathbf{MJG}(\mathrm{PS}_A)$: $\mathbf{MJG}(\mathrm{PS}_A) = \{G' \mid G' = \mathrm{argmin}_{G \in \mathbf{JG}(\mathrm{PS}_A)} |E(G)|\}$. В общем случае минимальных графов смежности может быть несколько [4].

Вторичной структурой может выступать какой-либо граф смежности. Обычно в качестве вторичной структуры рассматривают минимальный граф смежности. Задача обучения вторичной структуры

сводится к построению вторичной структуры, оптимальной (или хотя бы приемлемой в определенном смысле) с точки зрения применения алгоритмов глобального логико-вероятностного вывода в АБС [12].

Связность и ацикличность первичной структуры

Вторичная структура, построенная над первичной структурой, должна быть связна; для работы известных алгоритмов апостериорного логико-вероятностного вывода также требуется, чтобы она была и ациклической [12]. Однако не над любой первичной структурой можно построить связную вторичную структуру, так же как и не над любой первичной структурой можно построить ациклическую вторичную структуру [13]. Первичная структура называется связной, если над ней возможно построить граф смежности, который бы являлся связным. Первичная структура называется ациклической, если над ней возможно построить граф смежности, который является ациклическим. Известно [13], что все графы смежности над первичной структурой связны или несвязны одновременно, а также то, что все минимальные по числу ребер графы смежности будут циклическими или ациклическими одновременно.

На основе выявленных в предыдущих разделах соответствий, позволяющих систематизировать объекты теории АБС через графовое и гиперграфовое представление, можно предложить короткое, но строгое и ясное доказательство для следующей теоремы.

Теорема. Первичная структура PS связна тогда и только тогда, когда связен соответствующий ей гиперграф H_{PS}.

Доказательство. Это следует непосредственно из того, что все графы смежности связны или несвязны одновременно, и из того, что максимальный граф смежности как дуальный граф гиперграфа связен или несвязен одновременно с протоструктурой как первичным графом гиперграфа [11].

Доказанная теорема дает простой и проверяемый за $O(n^2)$ критерий цикличности первичной структуры, где n – мощность алфавита A. Последнее важно в контексте синтеза первичной структуры АБС: поскольку известные алгоритмы логико-вероятностного вывода требуют, чтобы первичная структура АБС была ациклической, важно устанавливать это свойство на этапе синтеза именно первичной, а не вторичной структуры [13].

Третичная полиструктура и четвертичная структура

Сужением $G \downarrow U$ графа G (заданного над первичной структурой с введенной на вершинах и ребрах функцией нагрузки) на подалфавит U называется подграф G, состоящий только из тех вершин и ребер G, нагрузка которых содержит U: $G \downarrow U = \{V_i \mid V_i \in V(G), U \subseteq W(V_i)\}, \{E_i \mid E_i \in E(G), U \subseteq W(E_i)\}$.

Множество сужений графа G на произвольные нагрузки будем обозначать как $\mathbf{Nar}(G)$. Для краткости $\mathbf{Nar}(G_{\max}(\mathrm{PS}_A))$ будем обозначать как $\mathbf{Nar}(\mathrm{PS}_A)$. $\mathbf{Nar}(G)$ является частично-упорядоченным множеством, в котором отношение порядка задается включением сужений друг в друга: $(G \downarrow U) \prec (G \downarrow V) \Leftrightarrow (G \downarrow \mathrm{U}) \supset (G \downarrow V) \Leftrightarrow U \subset W$.

Значимой кликой с нагрузкой U для первичной структуры PS_A называется сужение графа $G_{\max}(\mathrm{PS}_A)$ на нагрузку, являющуюся непустым сепаратором (т.е. совпадающим с каким-либо ребром графа $G_{\max}(\mathrm{PS}_A)$). Множество значимых клик для первичной структуры PS_A обозначим как $\mathbf{Clique}(\mathrm{PS}_A)$. Оно как подмножество частично упорядоченного множества также является частично упорядоченным множеством.

Множество значимых клик образует набор гиперребер над МФЗ. Однако соответствующий гиперграф не является в общем случае минимальным, поскольку гиперребра содержатся в других гиперребрах (меньших их по определенному на множестве сужений порядку). В соответствии с этим над множеством $\mathbf{Clique}(\mathrm{PS}_A)$ можно построить направленный граф, в котором ребро от значимой клики P к значимой клике Q проведено тогда и только тогда, когда P содержит Q. Построенный таким образом граф называется родовым графом над множеством $\mathbf{Clique}(\mathrm{PS}_A)$. Родовой граф можно строить и над произвольным подмножеством $\mathbf{Nar}(\mathrm{PS}_A)$.

Родительский граф – это направленный граф над некоторым подмножеством $\mathbf{Nar}(\mathrm{PS}_A)$, в котором ребро проведено от P к Q тогда и только тогда, когда такое ребро содержится в родовом графе и одновременно не существует иного пути от P к Q. Определенный подобным образом родительский граф является транзитивной редукцией родового графа и выступает диаграммой Хассе для подмножества

$\mathbf{Nar}(\mathrm{PS}_A)$, над которым он построен, с порядком, индуцированным порядком над множеством $\mathbf{Nar}(\mathrm{PS}_A)$.

Третичной полиструктурой *АБС* (далее третичной полиструктурой) для данной первичной структуры PS_A называется множество графов (направленных, ненаправленных и гибридных), построенных над подмножествами $\mathbf{Nar}(\mathrm{PS}_A)$.

Замкнутым сверху множеством значимых клик $\mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)$ для данной первичной структуры PS_A называется множество $\mathbf{Clique}(\mathrm{PS}_A)$, к которому добавили праклику – сужение $G_{\max}(\mathrm{PS}_A)\downarrow\varnothing$. Праклика совпадает с $G_{\max}(\mathrm{PS}_A)$. Нагрузкой праклики является пустое множество. Благодаря этому множество нагрузок элементов $\mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)$ совпадает с множеством сепараторов PS_A.

Рассмотрим произвольный элемент $\mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)$, назовем его C_U, где U – его нагрузка. Через $\mathrm{Son}(C_U)$ будем обозначать подмножество $\mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)$, соответствующее сыновьям C_U в родительском графе над $\mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)$. Два элемента $\mathrm{Son}(C_U)$ называются собратьями, если они пересекаются как сужения ($\mathrm{Son}(C_U)$ – подмножество $\mathbf{Nar}(\mathrm{PS}_A)$, более того, даже $\mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)$):

$$(S_i, S_j) \in \mathrm{Fel} \Leftrightarrow (S_i, S_j) \in \mathrm{Son}(C_U) \wedge S_i \cap S_j \neq \varnothing .$$

Полусиблинговый граф HS_U для элемента $\mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)$ с нагрузкой U (назовем его C_U) – это граф, построенный над множеством сыновей C_U, ребра в котором проведены между собратьями: $\mathrm{HS}_U = \mathrm{Son}(C_U), \{(S_i, S_j) | S_i, S_j \in \mathrm{Fel}\}$.

Четвертичная структура – семейство полусиблинговых графов, построенных для каждого сепаратора: $\{\mathrm{HS}_U \mid U \in \mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)\}$.

Родственный граф – гибридный граф, получаемый дополнением направленного родительского графа над $\mathbf{Clique}^{\uparrow}(\mathrm{PS}_A)$ ненаправленными ребрами между собратьями. Согласно определению, родственный граф является элементом третичной полиструктуры, при этом множество его ненаправленных ребер совпадает с множеством ребер в четвертичной структуре.

Иерархия глобальных структур

Отношения, возникающие между введенными выше графовыми и гиперграфовыми объектами, представим явно – в виде единой иерархии.

Основной структурой на данный момент выступает первичная структура, с которой начинается построение АБС экспертом или программой, выявляющей в имеющихся данных скрытые закономерности, отражающие стохастические зависимости, независимости и условные независимости [1, 2].

Для данной первичной структуры PS однозначно восстанавливается ее протоструктура $\mathrm{pS}(\mathrm{PS})$. Протоструктура открывает один из возможных подходов осуществить глобальное автоматическое обучение первичной структуры (или первичной и вторичной структур одновременно) [12]. Кроме того, протоструктура может использоваться в анализе особенностей первичной структуры, в частности – для выявления ее связности и ацикличности [14].

Для данной первичной структуры однозначно определяется множество сужений $\mathbf{Nar}(\mathrm{PS})$, а также его подмножество – множество значимых клик $\mathbf{Clique}(\mathrm{PS})$. Последнее множество играет значительную роль в синтезе вторичной структуры, поскольку используется во всех известных алгоритмах ее построения [9, 14].

Множество $\mathbf{Nar}(\mathrm{PS})$, а также его подмножества, в особенности $\mathbf{Clique}(\mathrm{PS})$ выступают в роли множества вершин для графов (в том числе пустого графа), совокупность которых образует третичную полиструктуру. Элементы третичной полиструктуры используются для синтеза вторичной структуры, в том числе для синтеза всех возможных вариантов такой структуры [14]. Родительский граф над расширенным вверх множеством значимых клик $\mathbf{Clique}^{\uparrow}(\mathrm{PS})$ используется для построения четвертичной структуры. Последняя тесно связана с третичной полиструктурой: в частности, она индуцирует ненаправленные ребра семейного графа над $\mathbf{Clique}^{\uparrow}(\mathrm{PS})$. Кроме того, наблюдается также и функциональная связь: и элементы третичной полиструктуры, и четвертичная структура используются для выявления ацикличности первичной структуры [15]. Следует отметить, что метод выявления ацикличности первич-

ной структуры на основе четвертичной структуры дает сильный аналитический критерий, лежащий в основе алгоритмов устранения циклов первичной структуры [16].

Необходимо также указать, что родительский граф может использоваться для осуществления алгоритмов апостериорного логико-вероятностного вывода, причем предполагается, что по скорости работы и минимизации ошибок в вычислении оценок такой вывод окажется лучше, чем над любой вторичной структурой, построенной над той же первичной структурой.

Заключение

В рамках данной работы были рассмотрены и систематизированы глобальные структуры алгебраической байесовской сети, в частности, введено понятие ее протоструктуры. Протоструктура, элементы третичной полиструктуры и четвертичная структура не используются непосредственно для осуществления алгоритмов логико-вероятностного вывода или для обеспечения представления знаний, хранящихся в сети, однако играют важную роль в решении задач глобального обучения АБС (являющегося подвидом машинного обучения вероятностных графических моделей), в частности, непосредственно для построения вторичной структуры.

Благодаря представлению первичной структуры и элементов третичной полиструктуры как гиперграфов открывается возможность для использования существующих в этой теории результатов для решения задач в теории АБС; в частности, в работе был предложен критерий ацикличности первичной структуры, опирающийся на свойства протоструктуры.

Выделение объекта как структуры АБС основано на соображениях функциональности: каждая глобальная структура выполняет одну или несколько функций в рамках теории АБС. Иерархия структур строится на основе отношений зависимости при синтезе указанных структур. Так, на вершине иерархии структур, описанной в настоящей работе, оказывается первичная структура, тогда как вторичная структура оказывается в самом низу.

Литература

1. Тулупьев А.Л., Николенко С.И., Сироткин А.В. Байесовские сети: логико-вероятностный подход. – СПб: Наука, 2006. – 607 с.
2. Тулупьев А.Л., Сироткин А.В., Николенко С.И. Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. – СПб: Изд-во СПбГУ, 2009. – 400 с.
3. Тулупьев А.Л., Столяров Д.М., Ментюков М.В. Представление локальной и глобальной структуры алгебраической байесовской сети в Java-приложениях // Труды СПИИРАН. – 2007. – Вып. 5. – С. 71–99.
4. Фильченков А.А., Тулупьев А.Л. Структурный анализ систем минимальных графов смежности // Труды СПИИРАН. – 2009. – Вып. 11. – С. 104–127.
5. Alpaydin E. Introduction to Machine Learning. – 2nd ed. – Cambridge, Mass. MIT Press, 2010. – 581 p.
6. Фильченков А.А., Тулупьев А.Л. Третичная структура алгебраической байесовской сети // Труды СПИИРАН. – 2011. – Вып. 18. – С. 164–187.
7. Егоров К.В., Царев Ф.Н., Шалыто А.А. Применение генетического программирования для построения автоматов управления системами со сложным поведением на основе обучающих примеров и спецификации // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 5 (69). – С. 81–89.
8. Тихомиров А.В., Шалыто А.А. Применение генетического подхода для генерации клеточных автоматов // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 2 (72). – С. 62–66.
9. Опарин В.В., Фильченков А.А., Тулупьев А.Л., Сироткин А.В. Матроидное представление семейства графов смежности над набором фрагментов знаний // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 4 (68). – С. 73–76.
10. Фильченков А.А., Тулупьев А.Л. Совпадение множеств минимальных и нередуцируемых графов смежности над первичной структурой алгебраической байесовской сети // Вестник СПбГУ. Серия 1. Математика. Механика. Астрономия. – 2012. – Вып. 2. – С. 65–74.
11. Зыков А.А. Основы теории графов. – М.: Наука, 1987. – 384 с.
12. Тулупьев А.Л., Фильченков А.А., Вальтман Н.А. Алгебраические байесовские сети: задачи автоматического обучения // Информационно-измерительные и управляющие системы. – 2011. – Т. 9. – № 11. – С. 57–61.
13. Фильченков А.А., Тулупьев А.Л., Сироткин А.В. Управление глобальной структурой знаний в интеллектуальных системах, основанных на алгебраических байесовских сетях // Материалы конференции «Информационные технологии в управлении» (ИТУ–2012). – СПб: ОАО «Концерн «ЦНИИ «Электроприбор». – 2012. – С. 25–33.
14. Фильченков А.А. Алгоритмы построения элементов третичной полиструктуры алгебраической байесовской сети // Труды СПИИРАН. – 2011. – Вып. 3 (18). – С. 237–266.

15. Фильченков А.А., Тулупьев А.Л. Косвенные признаки цикличности вторичной структуры алгебраической байесовской сети // «Гибридные и синергетические интеллектуальные системы: теория и практика». Материалы 1-го международного симпозиума. Т. 2. – Калининград: БФУ им. Канта, 2012. – С. 9–18.
16. Фильченков А.А., Фроленков К.В., Тулупьев А.Л. Устранение циклов во вторичной структуре алгебраической байесовской сети на основе анализа ее четвертичной структуры // Труды СПИИРАН. – 2012. – Вып. 21. – С. 143–156.

Фильченков Андрей Александрович – СПИИРАН,. мл. научный сотрудник, Санкт-Петербургский государственный университет, аспирант, aaafil@mail.ru

УДК 004.912

ИЗВЛЕЧЕНИЕ И РАНЖИРОВАНИЕ КЛЮЧЕВЫХ ФРАЗ В ЗАДАЧЕ АННОТИРОВАНИЯ

С.В. Попова, И.А. Ходырев

Для решения задачи аннотирования проводится сравнительный анализ двух подходов ранжирования ключевых фраз. Первый основан на оценке веса извлекаемых фраз с помощью TextRank, второй основан на использовании *tf-idf* оценки. Исследование проведено на базе коллекции INSPEC dataset. Представлены описание экспериментов и сравнительные результаты. Экспериментально показано, что подход, основанный на использовании *tf-idf*, дает лучший результат.

Ключевые слова: аннотирование, извлечение и ранжирование ключевых фраз, оценка качества аннотаций.

Введение

Тенденция к распространению электронных форматов представления научной информации стимулирует активное развитие научного сектора Интернета. Выражено это появлением огромного числа электронных публикаций и каталогов цитирования, доступных через сеть интернет, что, в свою очередь, способствует развитию и научных электронных библиотек. Очевидно, что комфортная работа пользователя с таким большим объемом информации невозможна без быстрого автоматического поиска нужных материалов. Для решения этой задачи необходимы данные о смысловом содержании документа, представленного в виде короткой аннотации. В работе под аннотацией понимается список ключевых слов/словосочетаний (фраз), характеризующих электронный документ. Наборы ключевых фраз или слов могут быть также использованы в задачах кластеризации и классификации, в задаче автоматического построения/пополнения онтологий, в задаче определения основных трендов, в задаче поиска новой информации и т.д. Под аннотированием в работе будем иметь в виду автоматическое извлечение из текста ключевых слов/словосочетаний (фраз).

Для решения задачи аннотирования выделяют два подхода. Первый использует обучающую выборку, второй – нет.

В первом подходе задача сводится к разработке классификатора, определяющего для поступившего на вход текста, какие из его частей являются ключевыми фразами, а какие нет [1, 2]. В работе [2] предложен генетический алгоритм и параметризованная система по извлечению ключевых фраз Extractor. Генетический алгоритм позволяет определить оптимальные значения параметров. В [1] использован наивный байесовский классификатор. В [3] выполнена интеграция лингвистических данных в машинное обучение, показано преимущество использования информации о частях речи.

В рамках второго подхода наиболее популярным является метод, основанный на представлении текста в виде графа, предложенный в работе [4]. Вершины графа – целостные части текста (отдельные слова, *n*-граммы, предложения). Веса дуг графа характеризуют тип связи между вершинами по выбранному принципу (например, встречаться вместе в окне размера *n*, т.е. на расстоянии не более *n* слов друг от друга). В [4] в качестве вершин графа рассматриваются отдельные слова текста; вес дуги, соединяющей две вершины-слова, показывает, сколько раз эти два слова встретились в тексте в окне *n*. Для оценки веса каждой вершины-слова в [4] используется величина, основанная на модификации формулы PageRank [5]:

$$S(v_i) = (1-d) + d \sum_{V_j \in In(V_i)} \frac{1}{|Out(v_j)|} S(v_j),$$

где $In(v_i)$ – дуги, входящие в вершину v_i; $Out(v_j)$ – дуги, исходящие из вершины v_j. Представленная выше формула была изменена [4] с учетом того, что каждая дуга имеет вес w:

$$S(v_i) = (1-d) + d \sum_{v_j \in In(V_i)} \frac{w_{ij}}{\sum_{v_k \in Out(v_j)} w_{jk}}. \quad (1)$$

Формула (1) получила название TextRank. Данная формула в [4] используется для расчета весов вершин-слов. Вершины-слова ранжируются по значению их веса, после чего отбирается только часть вершин с наибольшим весом. В [4] отобранные таким образом слова склеивались в фразы. Два слова объединяются в одну фразу, если в оригинальном тексте они непосредственно следуют друг за другом. В работе [8], помимо слов, рассмотрены многословные части текста – тогда вершинами графа текста являются группы слов. В качестве таких групп слов, например, взяты *n*-граммы (последовательность из *n* слов, следующих в тексте друг за другом), существительные фразы [8]. Для полученных многословных вершин рассчитывался вес на основе (1), лучшие вершины отбирались как ключевые фразы. После этого две последовательности слов склеивались в одну фразу, если они непосредственно следовали друг за другом в оригинальном тексте. В противном случае группа слов, образующих вершину, рассматривалась как самостоятельная фраза. Использование многословных частей текста при описанном выше подходе не дает лучших результатов в сравнении с использованием слов [8].

Оценка веса части текста с помощью TextRank в задаче аннотирования получила дальнейшее развитие. В работе [6] при построении графа учитывается содержание *k* ближайших документов. В [7] учитывается информация о семантической близости между построенными вершинами, дополнительно используются WordNet и Wikipedia. Полученные результаты являются одними из лучших в предметной области.

В данной работе задача извлечения ключевых фраз разделена на два этапа: 1) построение фраз-претендентов; 2) оценка веса каждой фразы-претендента и выбор лучших *k* из них как ключевых фраз. Также как и в работе [4], вес слов рассчитывается с помощью (1). Но, в отличие от [4], здесь не отбираются слова с наивысшей величиной TextRank, которые затем склеиваются в фразы. Вместо этого все слова после предобработки склеиваются в фразы, после чего оценивается вес каждой фразы, и лучшие фразы отбираются в качестве ключевых. Вес фразы считается равным весу лучшего слова (лучшее слово – слово в фразе с набольшим весом).

В работе поставлена задача проверки адекватности оценки веса фразы на основе информации о весах входящих в фразу слов, оцененных с помощью (1). Для сравнительного анализа в работе использован второй способ оценки весов слов и фраз соответственно [8]. Для этого использована популярная оценка веса слова *tf-idf* [11], которая рассчитывается для каждого конкретного слова в каждом документе как произведение частоты слова в данном документе tf на инвертированную частоту документов *idf*,

$$idf = \log \frac{|N|}{df},$$

где N – множество документов, df – число документов, в которых хотя бы раз встретилось слово. Далее будем обозначать вес слова v_i в документе как $(tf\text{-}idf)(v_i)$. С помощью $tf\text{-}idf$ оценивается вес каждого слова в документе. Вес фразы, построенной для текста, считается равным весу слова, входящего в фрау с максимальным значением *tf-idf*.

Для работы выбрана коллекция INSPEC dataset – одна из самых популярных в исследованиях по аннотированию текстов ключевыми фразами [3, 4, 7, 8, 10]. Коллекция содержит англоязычные аннотации к научным публикациям (abstracts, далее в тексте использован термин «абстракт», под которым понимается аннотация к научной публикации. Заметим, что термин «абстракт» использован во избежание путаницы, так как термин «аннотация» в работе используется для обозначения набора ключевых фраз документа). Коллекция состоит из трех подколлекций: training dataset (1000 документов), evaluation dataset (500 документов), testing dataset (500 документов). Каждый текст имеет «золотой стандарт», состоящий из фраз, отобранных для документа экспертом. Золотой стандарт включает в себя два подмножества аннотаций: contr set и unconter set. Аналогично [3, 4, 7, 8, 10], в работе использовано множество uncontr set. Подробное описание коллекции приведено в [3].

Оценка качества

Изначально для оценки качества извлеченных ключевых фраз использовалась оценка, основанная на F-score [3, 11], интегрирующая информацию о точности и полноте извлеченных фраз (Precision и Recall [11]). Однако данный способ оценки не накладывает четких ограничений на число извлеченных фраз. В работе [8] предложен подход, основанный на использовании R-Precision вместо F-score. R-Precision – значение Precision при условии, что число извлеченных ключевых фраз в точности совпадает с числом фраз в золотом стандарте. Для расчета R-Precision требуется информация о числе правильно извлеченных фраз. Пусть G_t – множество автоматически извлеченных фраз из текста t, C_t – множество ключевых фраз золотого стандарта для текста t (в золотой стандарт для каждого текста входят идеальные аннотации, вручную построенные экспертом). Возникает вопрос, какие из фраз G_t принадлежат $G_t \cap C_t$. В большинстве работ по извлечению ключевых фраз используется точное совпадение. Тогда фраза k («продвинутый автоматический перевод») и фраза золотого стандарта g («автоматический перевод»)

распознаются как различные. Аналогично ситуация возникает, например, и для фраз «хороший автомобиль» и «хорошая машина». В [8] сравниваются способы определения правильности извлеченной фразы:

– точное совпадение двух фраз (exact);
– совпадение двух фраз при наличии только морфологического различия (morph);
– совпадение двух фраз, если фраза k содержит в себе фразу g (include);
– совпадение двух фраз, если фраза g содержит в себе фразу k (part-of).

Показано, что использование part-of является неудачным. Для оценки совпадения двух фраз в [8] использовались exact и include+morph. В работе [9] расширено число способов оценки извлеченной ключевой фразы: BLEU, METEOR, NIST, ROUGE. Определено отношение, лучше отражающее подход экспертов к определению корректности фразы:

$$rp = \frac{|\,слова_фразы \in G_t \cap слова_фразы \in C_t\,|}{\max\{|\,слова_фразы \in G_t\,|, |\,слова_фразы \in C_t\,|\}}.$$

В данной работе для оценки качества результатов аннотирования используются метод и мера, предложенные в работах [8, 9], что позволяет в дальнейшем использовать результаты, полученные в данной работе, для сравнительной оценки качества новых алгоритмов, также опирающихся на [8, 9].

Для оценки качества аннотирования используется R-Precision [8]. Для оценки качества каждой конкретной извлеченной фразы рассматриваются три способа.

1. Случай exact. Здесь
$$|\,G_t \cap C_t\,| = |\,\{k \in G_t : \exists g \in C_t \wedge exact(k,g) = 1\}\,|;$$
где $exact(k,g) = 1$, если фразы k и g совпадают, $exact(k,g) = 0$ – в противном случае.
2. Случай include. Здесь
$$|\,G_t \cap C_t\,| = |\,\{k \in G_t : \exists g \in C_t \wedge include(k,g) = 1\}\,;$$
где $include(k,g) = 1$, если фраза k содержит в себе фразу g, иначе $include(k,g) = 0$.
3. Случай rp [9]. Здесь
$$|\,G_t \cap C_t\,| = \frac{1}{|\,G_t\,|}\sum_{k \in G_t}\max_{g \in C_t}\left(\frac{|\,k \cap g\,|}{\max(|\,k\,|, |\,g\,|)}\right).$$

Описание эксперимента

Для формирования исходных данных выполнена предобработка текстов. Каждый текст t представлен набором слов, из которого изъяты стоп-слова и все другие слова, кроме существительных и прилагательных [4]. Для определения части речи слов использован Stanford POS tagger tool [12]. Для каждого текста t построены фразы-претенденты: несколько слов из множества b_t объединялись в одну фразу, если в t они непосредственно следовали друг за другом. В построенной фразе-претенденте не могло быть больше четырех слов [8]. В работе поставлена следующая задача: проанализировать способы ранжирования полученных фраз-претендентов, для отбора лучших *n* из них как ключевых фраз, где $n = |\,C_t\,|$.

Примем следующие обозначения: $\mu(k)$ – вес фразы k, где $k = \{s_1, s_2, ...s_n\}$; s_i или s_i^k – слова фразы k; $TR(s_i^k)$ – вес слова s_i^k, рассчитанный с использованием формулы (1). Для ранжирования фраз на основе TextRank и *tf-idf* вес фразы рассчитывался по формулам

$$\mu(k) = \max_{s_i \in k} TR(s_i), \quad \mu(k) = \max_{s_i \in k}\left(tf - ifd\right)(s_i).$$

Для расчета $TR(s_i^k)$ требуется определить способ оценки веса дуг w_{ij}, соединяющий вершины-слова v_i и v_j. Рассмотрены варианты:

1. $w_{ij} = occur(v_i, v_j)$, где $occur(v_i, v_j)$ – число раз, когда слова v_i и v_j встретились вместе в тексте t в окне $n = 2$ (в работе [4] было показано, что такой размер окна является наилучшим);
2. $w_{ij} = mi(v_i, v_j)$, где $mi(v_i, v_j) = \log\dfrac{p(v_i, v_j)\cdot N}{p(v_i)\cdot p(v_j)}$ есть взаимная информация между v_i и v_j, $p(v_i, v_j)$ – число раз, когда слова v_i и v_j встретились вместе в текстах коллекции в окне 2, $p(v_i)$, $p(v_j)$ – число вхождений слов v_i и v_j в текстах коллекции, N – общее число словоформ в коллекции;
3. $w_{ij} = \left(tf - idf\right)(v_i)\cdot\left(tf - idf\right)(v_j)$.

Также поставлен дополнительный эксперимент: для ранжирования на основе TextRank строился граф, вершинами которого являлись все слова коллекции после предобработки, а $w_{ij} = p(v_i, v_j)$. Отличие от рассмотренного выше способа расчета весов вершин на основе формулы (1) – в том, что используется один граф для всей коллекции, а не отдельные графы для каждого текста.

Фразы ранжировались по значению весов. Лучшие n фраз с наибольшими значениями весов отбирались как ключевые фразы. На завершающей стадии эксперимента проводилась постобработка. Если для одного текста в качестве претендентов были получены две фразы, такие, что одна является частью другой, то оставлялась одна фраза наибольшего размера.

Результаты экспериментов и обсуждение

Результаты экспериментов представлены для трех подколлекций INSPEC dataset: test dataset, evaluate dataset, trial dataset (табл. 1–3). Приняты следующие обозначения:

1. ранжирование фраз с помощью TextRank:
   TR for text – граф строился для каждого отдельного текста, $w_{ij} = occur(v_i, v_j)$;
   TR for text*mi – граф строился для каждого отдельного текста, $w_{ij} = mi(v_i, v_j)$;
   TR for text***tf-idf* – граф строился для каждого отдельного текста, $w_{ij} = (tf - idf)(v_i) \cdot (tf - idf)(v_j)$;
   TR collection – граф строился для всей коллекции, $w_{ij} = p(v_i, v_j)$;
2. результаты ранжирования с помощью *tf-idf*.

INSPEC dataset	TR for text	TR for text*mi	TR for text**(*tf-idf*)	TR collection	*tf-idf*
test dataset	0,22	0,21	0,23	0,19	0,28
evaluate dataset	0,21	0,20	0,22	0,17	0,26
trial dataset	0,21	0,19	0,22	0,16	0,26

Таблица 1. R-Precision exact

INSPEC dataset	TR for text	TR for text*	TR for text**	TR collection	*tf-idf*
test dataset	0,29	0,27	0,31	0,25	**0,37**
evaluate dataset	0,27	0,26	0,29	0,23	0.34
trial dataset	0,27	0,26	0,29	0,23	0.34

Таблица 2. R-Precision include

INSPEC dataset	TR for text	TR for text*	TR for text**	TR collection	*tf-idf*
test dataset	0,42	0,41	0,43	0,37	0,48
evaluate dataset	0,40	0,39	0,42	0,36	0.46
trial dataset	0,40	0,40	0,42	0,35	0.46

Таблица 3. R-Precision rp

Результаты показали, что ранжирование фраз, основанное на TextRank, уступает ранжированию на основе *tf-idf*. TextRank использует данные, собранные из одного текста. Такая ограниченность плохо сказывается на качестве аннотирования. *Tf-idf* интегрирует информацию о важности слов в отдельном тексте и информацию о распространенности слов в коллекции, позволяя отсеять общеупотребительные для коллекции слова, не потеряв важных для текста терминов. Интересно, что *tf-idf* хорошо работает для коллекций абстрактов, несмотря на то, что документы коллекции является короткими текстами, а используемая коллекция – сравнительно небольшая. Формула *tf-idf* требует сбора статистики по встречаемости слов, для чего обычно требуется большой объем текстовых данных. Вывод: важные слова часто повторяются в конкретной аннотации к научной статье и редки в других аннотациях (по другим темам). Такая закономерность позволяет собрать нужную статистику на небольшом наборе данных и обосновывает использование *tf-idf*.

Использование $w_{ij} = mi(v_i, v_j)$ для TextRank не улучшает качества ранжирования по сравнению с классическим вариантом $w_{ij} = occur(v_i, v_j)$. Некоторое улучшение достигается за счет использования для оценки дуг $w_{ij} = (tf - idf)(v_i) \cdot (tf - idf)(v_j)$. Использование взаимной информации между словами: $mi(v_i, v_j)$, не улучшает качества ранжирования в сравнении с *tf-idf*, несмотря на то, что данные собираются по всей коллекции. При использовании mi оценивается не значимость отдельных слов, а значимость связей между словами. Эта мера имеет тенденцию к завышению значимости связей с/между редкими (случайными) словами. В работе при расчете mi не ставилось ограничение, которое позволило бы рассматривать связи только со словами, встретившимися в коллекции больше, чем l раз (в противном

случае вес связи был бы равен нулю). Полученный результат служит индикатором того, что в аннотациях к публикациям ключевые слова не бывают случайными (одиночными относительно коллекции).

Результаты, полученные в дополнительном эксперименте, могут считаться неудовлетворительными. Вероятно, это связано с тем, что теряется информация об особенностях отдельных текстов.

Еще одно интересное наблюдение касается способа оценки каждой автоматически извлеченной фразы. Рассматривалось три способа: exact, include, rp. Существует высокая корреляция между результатами, полученными для каждого из этих способов. Коэффициент корреляции Спирмена рассчитан для пар exact-include, include-rp, exact-rp и составил 0,98–0,99, для расчета коэффициентов использовано 15 наблюдений (Табл. 1 для exact, Табл. 2 для include, Табл. 3 для rp), корреляция значима на уровне 0,001. Была проверена корреляция между результатами, полученными в [8], где использовались такие способы оценки, как exact и include+morph. Получено значение 0,91, корреляция значима на уровне 0,001. Планируется провести дополнительные исследования, отвечающие на вопрос: «насколько высока вероятность того, что рассматриваемые способы могут быть взаимозаменяемыми», что позволит использовать единственный способ оценки качества.

Заключение

В работе представлены результаты серии экспериментов по извлечению ключевых слов на базе коллекции INSPEC datset с использованием Stanford POS tagger tool. Показано, что для данной коллекции *tf-i*df предложенный авторами способ ранжирования дает лучший результат, чем TextRank. Использование *tf-idf* требует сбора статистики по встречаемости слов, для чего обычно требуется большой объем данных. Экспериментально полученный результат показывает пригодность *tf-idf* для обработки небольших коллекций абстрактов. Это означает, что важные для текста слова часто повторяются в конкретном абстракте и редки в других абстрактах, что можно считать обоснованием использования *tf-idf* для небольших коллекций абстрактов. По результатам экспериментов сделан вывод, что в абстрактах ключевые слова имеют вхождение в коллекцию большее, чем 1–2 раза. Показано, что используемые в работе три способа оценки качества аннотирования могут оказаться взаимозаменяемыми.

Работа выполнена в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» поисковые научно-исследовательские работы по лоту шифр «2011-1.2.1-302-031», государственный контракт № 16.740.11.0751.

Литература

1. Frank E., Paynter G.W., Witten I.H., Gutwin C., Nevill-Manning C.G. Domain-specific keyphrase extraction // Proc. of IJCAI. – 1999. – P. 688–673.
2. Turney P. Learning to Extract Keyphrases from Text. – NRC/ERB-1057. – 1999. – February 17. – 43 p.
3. Hulth A. Improved automatic keyword extraction given more linguistic knowledge // Proc. of the Conference on Empirical Methods in Natural Language Processing. – 2003. – P. 216–223.
4. Mihalcea R., Tarau P. TextRank: Bringing order into texts // Proc. of the Conference on Empirical Methods in Natural Language Processing. – 2004. – P. 404–411.
5. Brin S. and Page L. The anatomy of a large-scale hypertextual web search engine // Computer Networks and ISDN Systems. – 1998. – V. 30. – № 1–7. – P. 107–117.
6. Wan Xiaojun and Jianguo Xiao Single document keyphrase extraction using neighborhood knowledge // Proceedings of the 23rd AAAI Conferenceon Artificial Intelligence. – 2008. – P. 855–860.
7. Tsatsaronis G., Varlamis I., Norvag K. SemanticRank: Ranking Keywords and Sentences Using Semantic Graphs // Proc. of the 23rd International Conference on Computational Linguistics. – 2010. – P. 1074–1082.
8. Zesch T., Gurevych I. Approximate Matching for Evaluating Keyphrase Extraction // International Conference RANLP 2009. – Borovets, Bulgaria, 2009. – P. 484–489.
9. Su Nam Kim, Baldwin T., Min-Yen Kan. Evaluating N-gram based Evaluation Metrics for Automatic Keyphrase Extraction // Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). – Beijing, 2010. – P. 572–580.
10. Kazi Saidul Hasan, Vincent Ng. Conundrums in Unsupervised Keyphrase Extraction: Making Sense of the State-of-the-Art // Proceedings of Coling 2010. – Poster Volume, Beijing, 2010. – P. 365–373.
11. Manning C., Raghavan P., Schütze H. Introduction to Information Retrieval // Cambridge University Press. – 2009. – 581 p.
12. Интернет страница инструмента Standford POS tagging tool [Электронный ресурс]. – Режим доступа: http://nlp.stanford.edu/software/tagger.shtml, свободный. Яз. рус. (дата обращения 09.11.2012).

Попова Светлана Владимировна – Санкт-Петербургский государственный университет, Санкт-Петербургский государственный политехнический университет, ст. преподаватель, svp@list.ru

Ходырев Иван Александрович – ООО «Висмарт», зам. генерального директора по науке, kivan.mih@gmail.com

УДК 004.89

ВЕРИФИКАЦИЯ ДАННЫХ В СИСТЕМАХ ОТСЛЕЖИВАНИЯ ЗАДАЧ С ПОМОЩЬЮ ПРОДУКЦИОННЫХ ПРАВИЛ

Р.С. Катериненко, И.А. Бессмертный

Предлагается методика верификации данных в системах отслеживания задач с помощью модели продукционных правил. Данная модель позволяет в декларативном стиле формулировать условия, которым должно соответствовать информационное наполнение, и использовать логический вывод. В качестве практических результатов описывается применение разработанной системы верификации для реального проекта разработки программного обеспечения.
Ключевые слова: продукционные правила, логический вывод, системы отслеживания задач, верификация.

Введение

Система отслеживания задач (СОЗ) – прикладная программа, разработанная с целью помочь контролировать ход отдельных задач и проекта в целом. Основные компоненты такой системы – это база данных задач и графический интерфейс с возможностью редактирования задач и построения аналитических отчетов. Информация в подобной системе меняется и используется большим количеством специалистов из разных областей, что является одним из источников ошибок. Какой бы хорошей ни была СОЗ, со временем эксплуатации в содержащейся информации накапливаются неточности, противоречия, ошибки. Этому способствуют человеческий фактор и неточность естественного языка. В свою очередь, от адекватного состояния информационного наполнения зависят продуктивность работы участников проекта и его успех. Другой проблемой является миграция неточностей из базы СОЗ в автоматически генерируемые отчеты. Неправильные данные в таких отчетах могут привести к неверно подсчитанным трудозатратам, финансовым затратам и ошибочному планированию.

В основном СОЗ создаются под каждый отдельно взятый проект на базе общей платформы СОЗ. Платформа для создания СОЗ имеет язык высокого уровня абстракций, подходящий для описания широкого круга проектов в терминах рабочего процесса, артефактов рабочего процесса, связей между артефактами, состояний артефактов и т.д. Из-за высокой универсальности этот язык позволяет реализовать только основные проверки целостности информации на уровне правильности заполнения полей артефактов. Его возможности ограничиваются синтаксическими проверками.

Авторы предлагают использовать для целей верификации информационного наполнения СОЗ логические правила – продукции. Преимущества продукционного языка – декларативность и язык запросов, основанный на логическом выводе. Продукционный язык обладает большими выразительными возможностями, позволяющими формулировать сложные ограничения на семантику информации в СОЗ, в том числе рекурсивные правила.

Модель продукционных правил

Модель продукционных правил относится к классу дедуктивных логических систем, в основе которых лежит принцип дедукции – порождение истинных фактов из заданных аксиом и правил. В общем случае дедуктивная система состоит из набора аксиом, правил вывода и изначально заданных формул. В модели продукционных правил с целью повышения эффективности логического вывода введены ограничения:

– правила вывода (продукционные правила) являются хорновскими выражениями;
– факты представляют собой атомарные формулы.

Хорновское выражение (horn clause) – это дизъюнкция литералов с одним единственным положительным литералом. Литералом называется атомарная формула, а атомарная формула с отрицанием называется отрицательным литералом. Если применить эквивалентность $\neg A \vee B \equiv A \rightarrow B$, то хорновский дизъюнкт превращается в привычный вид импликации или продукции:

$$\neg A_1 \vee \neg A_2 \vee \ldots \vee \neg A_{n-1} \vee \mathrm{A}_n \equiv A_1 \wedge A_2 \wedge \ldots \wedge A_{n-1} \rightarrow A_n .$$

В некоторых продукционных системах поддерживаются рекурсивные правила, отрицание по «принципу закрытого мира» и термы-функции. Согласно «принципу закрытого мира» (closed world assumption) все те факты, которые не заложены в систему и не могут быть получены логическим выводом, объявляются ложными. Другими словами, все, о чем нет информации в системе, объявляется ложным.

Преимуществами указанного типа моделей являются:

– простота и ясность основной единицы – продукционного правила;
– независимость продукций и легкость модификации баз знаний;
– строгость, простота и изученность механизма логического вывода;
– эффективность логического вывода.

К недостаткам можно отнести:

– малую степень структурированности базы знаний;
– неуниверсальность.

Применение модели продукционных правил

Как было показано выше, для построения логической модели необходимы изначально заданные факты и правила. Часть информации СОЗ, подлежащая верификации, должна быть представлена в виде фактов для формирования предикатов. Правилами здесь являются записанные на продукционном языке свойства, которым должны удовлетворять данные. Проверка того, что данные СОЗ удовлетворяют заданным правилам, называется верификацией СОЗ.

Рассмотрим предлагаемую методику на примере СОЗ, эксплуатируемой в проекте разработки программного обеспечения длительностью 10 лет. Основными артефактами являются спецификации, тикеты, проекты. Тикетом будем называть представление задачи в СОЗ. Тикет является минимальной атомарной единицей информации о задаче. В данном примере одной задаче соответствует один тикет. В проекте может быть произвольное число тикетов и спецификаций, связанных между собой различными связями. Тикет бывает следующих видов:

- описание дефекта;
- описание задачи для разработки;
- описание исследовательской задачи;
- описание дизайнерской задачи;

и т.д.

У тикета имеется около 30 обязательных и необязательных для заполнения полей. Помимо полей, у тикета есть именованные связи с другими тикетами. Связи представляют собой ссылки от одного тикета к другому и могут иметь следующие типы: «has clone», «duplicates», «introduces», «incorporates», «relates». Симметричными им являются соответственно: «clone of», «is duplicated by», «introduced by», «is incorporated by», «is related to». Семантика связей следует из англоязычных названий связей и не является существенной для раскрытия содержания темы, поэтому выходит за рамки настоящей работы.

Существенная для верификации информация экспортируется из системы отслеживания задач (СОЗ) в предикаты следующим образом.

- Каждому полю тикета сопоставляется двухместный предикат тикет–значение. Например, если тикет № 8520 имеет поле «Номер спецификации» («requirement»), равное 328, то соответствующий предикат будет иметь вид *hasRequirement(T1,T2): hasRequirement(8520, 328) = true*.
- Каждому типу связи ставится в соответствие двухместный предикат тикет–тикет. Например, двум тикетам, связанным связью вида «introduces», будет соответствовать предикат вида *introduces(T1, T2)*. Свойства, которым должно удовлетворять информационное наполнение системы, формулируются в виде продукционных правил. В нашей системе одним из правил было наличие у каждого тикета ссылки (связи) на спецификацию, под которой по нему велись работы. Если тикет был причиной других тикетов, то спецификация должна была наследоваться. При составлении биллинговых отчетов часы, затраченные на выполнение спецификации, брались как сумма часов по связанным с ней тикетам. Нарушение указанного правила приводило к большим расходам. Сформулируем это ограничение в виде правил – продукций:

introducesClosure(T1, T2) :– introduces(T1, T2) (1)

introducesClosure(T1, T3) :– introduces(T1, T2), introducesClosure(T2, T3) (2)

requirementForClosure(R, T2) :– requirementFor(R, T1), introducesClosure(T1, T2) (3)

Правила (1) и (2) определяют предикат *introducesClosure* как транзитивное замыкание по свойству «introduces». После выполнения логического вывода он должен содержать все пары тикетов, первый из которых является причиной возникновения второго. Правило (2) задает предикат *requirementClosure*, который должен содержать все пары вида спецификация–тикет, учитывая наследование спецификаций. Эти три правила ((1)–(3)) задают условие, которому должно соответствовать правильное информационное наполнение. Но в реальной системе всегда содержатся ошибки. В связи с этим следующим шагом является написание диагностического правила:

failRequirement(R, T) :– requirementForClosure(R, T), not requirementFor(R, T) (4)

Предикат *failRequirement* должен содержать те пары вида спецификация-тикет, которых не было в системе, но которые должны в ней быть, следуя описанным ограничениям. Исходя из этого, срабатывание данного правила соответствует ошибочной ситуации, когда в СОЗ у тикета *T* не проставлена спецификация *R*. После верификации информационное наполнение может быть скорректировано администратором СОЗ таким образом, чтобы диагностические предикаты стали «пустыми» после следующей итерации.

Общий алгоритм верификации

На рис. 1 в схематическом виде приведены основные этапы алгоритма верификации и данные, передаваемые между ними.

1. Выгрузка данных. На этом этапе происходит выгрузка данных из СОЗ в виде отчетов, результатов выполнения SQL-запросов или html-страниц.

2. Преобразование. Из выгруженных данных формируется логическая модель: предикаты, домены (области определения аргументов предиката).
3. Логический вывод. К логической модели добавляются продукционные правила для верификации (1), (2), (3) и вспомогательные диагностические продукционные правила (4) для локализации ошибочной информации.
4. Коррекция данных. На данном шаге аналитик или администратор системы должен исправить ошибочную информацию, указанную в диагностических предикатах.

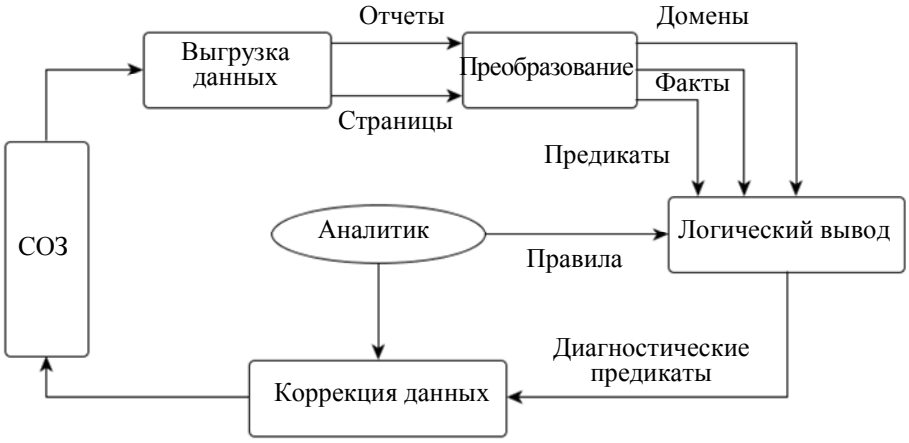


Рис. 1. Основные этапы алгоритма верификации

Описанная итерация может быть проделана несколько раз по мере добавления диагностических правил и коррекции информации. Информация считается верифицированной, когда все диагностические предикаты пусты.

Описание программной реализации

На рис. 2 приведена схема модулей программной реализации.

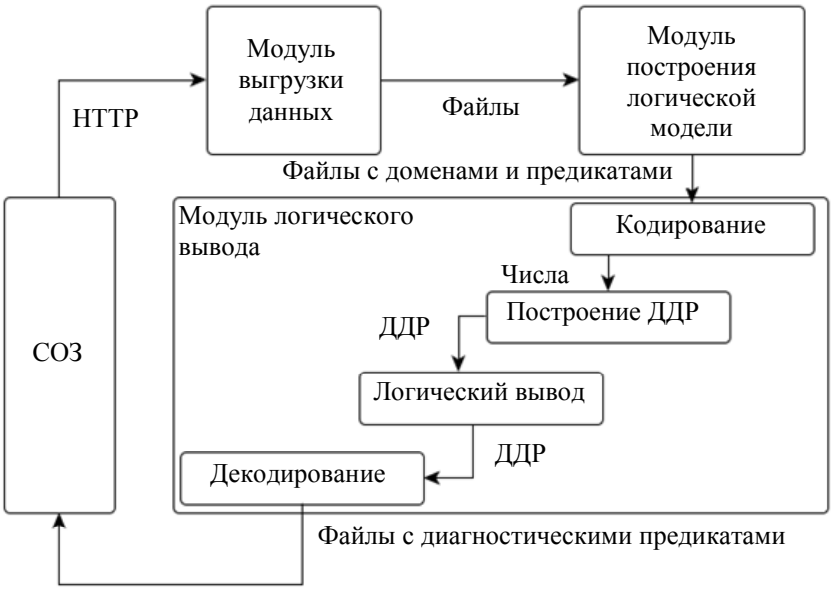


Рис. 2. Схема модулей программной реализации

Основными модулями являются: модуль выгрузки данных, модуль построения логической модели и модуль логического вывода. Функционал модуля выгрузки данных заключается в извлечении верифицируемых данных из СОЗ. В данном случае взаимодействие с СОЗ, построенной на программной платформе JIRA, осуществляется по протоколу HTTP, но может быть реализовано непосредственно через базу данных СОЗ и программное API. Выгрузка данных может быть полной, когда каждый раз выгружаются все необходимые для верификации данные, или инкрементальной, когда добавляются только новые данные.

Полученные данные в виде файлов поступают на вход модуля построения логической модели, где формируются домены и предикаты. В нашем приложении этот модуль позволяет декларативно задать правила для преобразования данных в предикаты. При этом домены создаются автоматически.

Построенная логическая модель и логическая программа, состоящая из продукционных правил, поступают на вход модуля логического вывода. Существуют разные подходы для осуществления логического вывода на продукционных правилах. Авторами был выбран подход на основе двоичных диаграмм решений (ДДР, Binary Decision Diagram), поскольку авторы ведут исследования в области применения ДДР для логического вывода [1–3]. ДДР – это структура данных, хорошо зарекомендовавшая себя на практике в области верификации аппаратных средств, позволяющая выполнять манипуляции над булевыми формулами. Компактность и эффективность операций над ДДР может быть использована для представления отношений и быстрого логического вывода [4–6]. В этом модуле последовательно выполняются следующие шаги:

1. кодирование доменов и предикатов числами;
2. построение ДДР;
3. логический вывод над ДДР;
4. декодирование ДДР.

Заключительным шагом является осуществление коррекции информации в СОЗ.

Основные результаты

Мы провели сравнительный эксперимент на реальной СОЗ для нашей задачи, описанной в разделе «Применение модели продукционных правил». К моменту начала логического вывода в базе данных содержалось 4667 факта и 6 продукционных правил. В качестве системы для сравнения была выбрана система DES (Datalog Educational System), не использующая ДДР. Результаты эксперимента приведены в таблице. Время указано в миллисекундах. T_1 – время загрузки фактов в систему DES. T_2 – время логического вывода в системе DES. T_3 – время загрузки фактов в нашу систему. По сравнению с DES, в нашей системе можно получить отдельно время логического вывода и отдельно время записи его результата в файл или выдачи на экран. Этими временами являются T_4 и T_5 соответственно. Для сравнения производительности логического вывода берутся времена получения ответа от систем с уже загруженными фактами – T_2 и T_6. Как видно, T_2 превосходит T_6 более чем в 200 раз. Таким образом, наша система на основе ДДР осуществляет логический вывод в 200 раз быстрее DES. Загрузка фактов осуществляется быстрее в 1,4 раза. Система реализована на языке Java. Верифицируемая СОЗ создана с помощью системы JIRA. Эксперимент проводился на платформе со следующими характеристиками: Intel Core I5 2.3 ГГц, 4 Гб ОЗУ, Windows 7 64-bit, Java 1.6.0_25.

Количество фактов	Время логического вывода, в мс					
	DES T_1	DES T_2	ДДР T_3	ДДР T_4	ДДР T_5	ДДР $T_6 = T_4 + T_5$
4667	1001	130000	721	311	233	544

Таблица. Времена логического вывода

В данной работе получен опыт применения продукционной модели и разрабатываемого механизма логического вывода на реальной прикладной задаче.

Авторы считают, что создание базы знаний не является очень трудоемкой задачей, потому что может быть легко автоматизировано. Как правило, все СОЗ представляют исходные данные, необходимые для верификации, некоторым регулярным для всех проектов образом. Разработанный единожды модуль выгрузки данных под конкретную СОЗ может быть использован на протяжении всей эксплуатации системы без особых изменений. Построение логической модели и дедуктивный вывод осуществляются автоматически разработанной системой. Участие аналитика требуется лишь при коррекции исходных данных по результатам вывода и формулировки правил для верификации.

Выразительность продукционной модели варьируется в зависимости от ее свойств. Решаемая задача дала возможность выделить свойства продукционного языка, которые должны поддерживаться механизмом вывода для широкого применения на практике в подобного рода прикладных задачах: поддержка отрицания, рекурсии, временных предикатов, функций для работы со строками и числами.

Заключение

Предложенный подход может быть реализован в виде отдельного приложения для верификации систем отслеживания задач или в качестве плагина к системе отслеживания задач. В результате применения были обнаружены и скорректированы ошибки в информационном наполнении в эксплуатируемой системе. Выбранный метод логического вывода на основе двоичных диаграмм решений показал хорошую производительность и имеет большой запас по производительности.

На взгляд авторов, методику имеет смысл применять в проектах при выполнении двух условий:

1. над проектом работает достаточно большое количество специалистов разных областей, что способствует быстрому накоплению ошибок в системе отслеживания задач;
2. актуальность и корректность данных в системе отслеживания задач напрямую влияет на ключевые характеристики проекта, такие как трудозатраты, или влияет на качество и работоспособность производимого продукта.

Работа выполнена при финансовой поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (соглашение № 14.B37.21.0406).

Литература

1. Bessmertny I.A., Katerinenko R.S. The method of acceleration of logical inference in production knowledge model // Programming and Computer Software. – Springer, 2011. – V. 37. – № 4. – P. 197–199.
2. Бессмертный И.А. Теоретико-множественный подход к логическому выводу в базах знаний // Научно-технический вестник СПбГУ ИТМО. 2010. – № 2 (66). – С. 43–48.
3. Бессмертный И.А. Быстрый логический вывод в среде программирования Visual Prolog // Научно-технический вестник СПбГУ ИТМО. – 2010. – № 3 (67). – С. 50–56.
4. Iwaihara M., Inoue Y. Bottom-up Evaluation of Logic Programs Using Binary Decision Diagrams // Proc. 11th Int. Conf. Data Engineering. – Taipei, Taiwan, Mar. 1995. – P. 467–474.
5. Bryant R. Graph-based Algorithms for Boolean Function Manipulation // IEEE Transactions on Computers. – August 1986. – V. 35. – № 8. – P. 677–691.
6. Andersen H.R. An Introduction to Binary Decision Diagrams // Lecture notes for Advanced Algorithms. – Lyngby, Denmark, 1997. – P. 1–37.

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

Бессмертный Игорь Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, igor_bessmertny@hotmail.com

УДК 004.93+57.087.1

АГЛОМЕРАТИВНАЯ КЛАСТЕРИЗАЦИЯ РЕЧЕВЫХ СЕГМЕНТОВ ФОНОГРАММЫ НА ОСНОВЕ БАЙЕСОВСКОГО ИНФОРМАЦИОННОГО КРИТЕРИЯ

О.Ю. Кудашев

Дано описание реализации системы агломеративной кластеризации речевых сегментов фонограммы на основе байесовского информационного критерия. Приведены результаты численных экспериментов с применением различных акустических признаков, а также с использованием полной и диагональной матриц ковариации. Для аудиозаписей радио «Свобода» на разработанной системе был достигнут уровень ошибки DER 6,4%.

Ключевые слова: кластеризация речевых сегментов, вариационный байесовский анализ, речевые технологии.

Введение

В последнее время наблюдается значительный рост интереса к системам автоматической сегментации фонограмм. Подобный рост вызван, в первую очередь, значительным увеличением объема звуковых данных, а также быстрым развитием технологий обработки речи. В значительной степени интерес научного сообщества в этой области поддерживается Национальным институтом стандартов и технологий (National Institute of Standards and Technology, NIST), разработавшим методику оценки качества систем автоматической обработки речи (Rich Transcription Evaluation Project, RTE) [1]. Одной из подзадач RTE является задача разделения дикторов (MDE (Metadata Extraction) Speaker Diarization), в рамках которой необходимо произвести кластеризацию (объединение) речевых сегментов фонограммы, принадлежащих одному диктору.

Задачи разделения дикторов и методы их решения можно классифицировать в соответствии с областью их применения. Так, для кластеризации речевых сегментов аудиозаписей диалогов наиболее эффективным методом является вариационный байесовский анализ в пространстве собственных голосов [2, 3]. Интересом данной работы являются аудиозаписи радиовещаний. Особенностью таких аудиозаписей является относительно редкая смена дикторов, а также отсутствие ограничения на их количество. Общепризнанным решением в этом случае является метод агломеративной кластеризации речевых сегментов на основе байесовского информационного критерия (Bayesian Information Criterion, BIC). Системы разделения дикторов, основанные на этом методе, зарекомендовали себя с точки зрения оптимального соотношения эффективности и производительности [4, 5].

Целью настоящей работы является реализация системы агломеративной кластеризации речевых сегментов на основе BIC. Кроме того, в данной работе будут представлены результаты применения разработанной системы к русскоязычному корпусу, в частности, к аудиозаписям радио «Свобода».

Применение BIC для кластеризации речевых сегментов

BIC является широко распространенным статистическим критерием, на основе которого производится выбор модели. В соответствии с этим критерием качество модели M, описывающей распределение данных $X=\{x_1, \ldots, x_N\}$, $x_i \in R^d$, оценивается по формуле

$$BIC(M) = \log L(x_1,...,x_N \mid M) - \frac{\lambda}{2}\nu(M)\log N, \quad (1)$$

где $L(x_1,...,x_N \mid M)$ – функция правдоподобия; $\nu(M)$ – число степеней свободы модели M (число свободных параметров); λ – настраиваемое пороговое значение, теоретически равное 1.

Пусть даны два независимых набора данных X_1, X_2. Задача кластеризации наборов данных X_1, X_2 может быть сведена к задаче выбора модели. Рассмотрим две альтернативные модели:

1. модель M_1 – данные X_1 и X_2 подчиняются одному гауссову распределению, $\{X_1, X_2\} \sim N(\mu, \Sigma)$.
2. модель M_2 – данные X_1 и X_2 подчиняются двум различным гауссовым распределениям, $\{X_1\} \sim N(\mu_1, \Sigma_1)$, $\{X_2\} \sim N(\mu_2, \Sigma_2)$.

Тогда, в соответствии с формулой (1),

$$BIC(M_1) = \log L(X_1, X_2 \mid \mu, \Sigma) - \frac{\lambda}{2}\nu(M_1)\log(N_1 + N_2),$$

$$BIC(M_2) = \log L(X_1 \mid \mu_1, \Sigma_1) + \log L(X_2 \mid \mu_2, \Sigma_2) - \frac{\lambda}{2}\nu(M_2)\log(N_1 + N_2),$$

где N_1, N_2 – количество данных в наборах X_1 и X_2 соответственно.

Разницу между этими двумя значениями обозначим ΔBIC:

$$\Delta BIC = BIC(M_1) - BIC(M_2) = \log\frac{L(X_1, X_2 \mid \mu, \Sigma)}{L(X_1 \mid \mu_1, \Sigma_1)L(X_2 \mid \mu_2, \Sigma_2)} + \frac{\lambda}{2}\left(\nu(M_2) - \nu(M_1)\right)\log N. \quad (2)$$

Положительное значение величины ΔBIC свидетельствует о том, что модель M_1 является наиболее предпочтительной, следовательно, наборы данных X_1 и X_2 следует отнести к одному кластеру.

Формулу (2) можно переписать в следующем виде:

$$\Delta BIC = \frac{1}{2}\left(N_1 \log|\Sigma_1| + N_2 \log|\Sigma_2| - (N_1 + N_2)\log|\Sigma_{1,2}|\right) + \frac{\lambda}{2}\beta\log(N_1 + N_2), \quad (3)$$

где Σ_1, Σ_2 – ковариационные матрицы данных X_1 и X_2 соответственно; $\Sigma_{1,2}$ – ковариационная матрица объединенных данных $\{X_1, X_2\}$; β – число свободных параметров, для диагональной ковариационной матрицы $\beta = d$, для полной ковариационной матрицы $\beta = \frac{d(d+1)}{2}$.

Алгоритм агломеративной кластеризации речевых сегментов

Алгоритм агломеративной кластеризации заключается в последовательном иерархическом попарном объединении (агломерации) кластеров (наборов речевых сегментов), принадлежащих одному диктору. При этом в качестве начальных кластеров используются речевые сегменты фонограммы, ограниченные двумя точками смены дикторов. Как для поиска точек смены дикторов, так и для определения принадлежности двух кластеров одному диктору используется BIC.

Разработанный алгоритм агломеративной кластеризации состоит из 4 этапов.

1. Выделение речевых сегментов фонограммы.
2. Расчет акустических признаков по всей фонограмме.
3. Поиск точек смены дикторов.
    1. Акустические признаки, лежащие внутри выделенных речевых сегментов, последовательно объединяются в непрерывный массив данных. В дальнейшем работа идет только в рамках этого массива.
    2. По всему полученному массиву перемещается окно фиксированной длины $2w$ с фиксированным шагом h. Окно разбивается на две равные части, левой части окна соответствуют данные X_1, правой – X_2 (формула (3)). Для точки $h \cdot i + w$, являющейся серединой окна, рассчитывается величина ΔBIC_i при $\lambda = \lambda_{CPD}$.
    3. Среди всех полученных таким образом значений выбираются локальные минимумы $M = \{m : \Delta BIC_{m-1} > \Delta BIC_m < \Delta BIC_{m+1}, \Delta BIC_m < 0\}$.

4. Среди всех полученных локальных минимумов M выбирается значение $m' = \arg\min_{m \in M}(\Delta BIC_m)$. При этом локальные минимумы из множества M, располагающиеся к выбранному значению ближе, чем w, удаляются: $M = M \setminus \{m : |m - m'| < w\}$.
5. Пункт 4 повторяется до тех пор, пока множество M не пусто.
6. Точки $\{h \cdot m' + w\}$ берутся в качестве точек смены дикторов.

4. Агломеративная кластеризация.
   1. Формируется набор кластеров $C = \{c_k\}$, каждый элемент которого состоит из множества речевых сегментов, ограниченных двумя соседними точками смены дикторов.
   2. Подсчитываются попарные значения $\{\Delta BIC_{k,l}\}_{k,l=1}^{|C|}$ по формуле (3) между всеми кластерами при $\lambda = \lambda_{AC}$.
   3. Если существует пара кластеров $c_{k*}, c_{l*} : \{k*, l*\} = \arg\max_{k,l}(\Delta BIC_{k,l}), \Delta BIC_{k*,l*} > 0$, то они объединяются в один кластер $c_{k*} = c_{k*} \bigcup c_{l*}$, при этом кластер c_{l*} удаляется: $C = C \setminus \{c_{l*}\}$.
   4. Пункты 2–3 повторяются до тех пор, пока происходят объединения.
   5. Получившийся набор кластеров C и будет являться решением задачи разделения дикторов.

Важно отметить необходимость использования различных пороговых значений при подсчете величины ΔBIC на этапе определения точек смены дикторов (λ_{CPD}) и на этапе агломеративной кластеризации (λ_{AC}). Эта необходимость обусловлена тем, что при поиске точек смены дикторов ключевым требованием является низкий уровень ошибки пропуска. Возникающий в этом случае высокий уровень ошибки ложного срабатывания компенсируется на этапе агломеративной кластеризации.

Результаты численных экспериментов

В качестве тестовой базы использовались 20 60-минутных аудиозаписей радио «Свобода» [6]. Все аудиозаписи имеют один канал с частотой дискретизации 16000 Гц. Для всей базы тестирования была создана эталонная разметка с указанием имен дикторов и принадлежащих им речевых сегментов. Речевые сегменты каждого диктора были дополнительно разделены на 4 категории:

1. чистая речь (71,86%);
2. речь на фоне шума (8,66%);
3. речь на фоне речи (11,20%);
4. речь на фоне музыки (8,28%).

Среднее количество дикторов в одном файле базы тестирования составило 25.

Для численных экспериментов речевые сегменты брались из эталонной разметки без использования какой-либо дополнительной обработки.

В качестве акустических признаков использовались:

– мел-частотный банк фильтров (MBF);
– мел-частотные кепстральные коэффициенты (MFCC).

Для построения акустических признаков были использованы следующие характеристики:

– окно быстрого преобразования Фурье (БПФ) – 16 мс;
– шаг окна БПФ – 10 мс;
– частотный диапазон для банка фильтров – от 0 до 8000 Гц.

Дополнительно к акустическим признакам были добавлены логарифм энергии (E), производные первого (Δ) и второго (Δ^2) порядков.

Для оценки качества системы использовался стандартный показатель вероятности ошибки разделения дикторов (Diarization Error Rate, DER), используемый в NIST RTE 2006 [1]:

$$DER = \frac{\sum_S \left\{ dur(S) \cdot \left(\max\left[N_{ref}(S), N_{sys}(S) \right] - N_{correct}(S) \right) \right\}}{\sum_S \left\{ dur(S) \cdot N_{ref}(S) \right\}} \cdot 100\%,$$

где S – непрерывный речевой сегмент фонограммы; $dur(S)$ – длина этого сегмента; $N_{ref}(S)$ – число дикторов, которым принадлежит речевой сегмент S, в соответствии с эталонной разметкой; $N_{sys}(S)$ – число дикторов, которым принадлежит речевой сегмент S, в соответствии с полученной разметкой; $N_{correct}(S)$ – число корректно определенных дикторов на речевом сегменте S.

Размер окна w для поиска точек смены дикторов брался равным 3 с, шаг h брался равным 0,2 с.

Результаты численных экспериментов, представляющих зависимость величины DER от используемых акустических признаков и типа ковариационной матрицы, представлены в табл. 1.

Акустические признаки	Размерность признаков	Полная ковариационная матрица			Диагональная ковариационная матрица		
		λ_{CPD}	λ_{AC}	DER (%)	λ_{CPD}	λ_{AC}	DER (%)
MBF	20	1,2	3,2	**8,0**	5	30	21,1
MFCC	20	1,0	3,5	**7,7**	4,8	27	8,5
MFCC + E	20	1,0	3,3	**7,8**	4,8	27	8,3
MFCC + $E + \Delta$	40	0,6	1,3	7,3	3	18	**6,4**
MFCC + $E + \Delta + \Delta^2$	60	0,55	1,1	7,2	2,3	14	**6,6**

Таблица 1. Зависимость DER от акустических признаков и типа ковариационной матрицы

Результаты численных экспериментов, представляющих зависимость относительной производительности системы от типа ковариационной матрицы и размера акустических признаков, представлены в табл. 2, где относительная производительность измеряется как отношение продолжительности всех речевых сегментов фонограммы ко времени работы системы. Все измерения были проведены при работе алгоритма на одном ядре процессора Intel Core i5 760 2.8 GHz.

Размерность акустических признаков	Относительная производительность	
	Полная ковариационная матрица	Диагональная ковариационная матрица
20	210	**3600**
40	56	**1900**
60	23,5	**1280**

Таблица 2. Зависимость относительной производительности системы от типа ковариационной матрицы и размера акустических признаков

Заключение

Разработанная система агломеративной кластеризации речевых сегментов фонограммы представляет собой эффективное решение с точки зрения соотношения эффективности и производительности. Как показывают численные эксперименты, данная система способна показать значение ошибки DER равной 6,4% на русскоязычных аудиозаписях радиовещания, где происходит редкая смена дикторов. Как показано в табл. 2, относительная производительность системы сильно зависит от применения полной или диагональной ковариационной матрицы. Интересным представляется тот факт, что в случае использования слабо коррелирующих акустических признаков мел-частотных кепстральных коэффициентов применение диагональной ковариационной матрицы способно привести к уменьшению ошибки. При использовании полной ковариационной матрицы увеличение размерности акустических признаков приводит к сильному падению производительности прямо пропорционально квадрату размерности признаков, что еще раз демонстрирует целесообразность использования диагональной ковариационной матрицы.

Данная система разработана на кафедре «Речевые информационные системы», являющейся базовой кафедрой компании ООО «Центр речевых технологий». Она успешно применяется в системах автоматической обработки и распознавания речи.

Литература

1. Rich Transcription Evaluation Project [Электронный ресурс]. – URL: http://www.itl.nist.gov/iad/mig/tests/rt/, свободный. Яз. англ. (дата обращения 20.09.2012).
2. Kenny P. Bayesian Analysis of Speaker Diarization with Eigenvoice Priors // Technical report, Centre de recherche informatique de Montreal (CRIM). – Montreal, Canada. – May 2008. – 17 p.
3. Кудашев О.Ю., Пеховский Т.С. Проблема инициализации систем сегментации дикторов на основе вариационного байесовского анализа // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 3 (79). – С. 83–87.
4. Reynolds D., Kenny P., Castaldo F. A Study of New Approaches to Speaker Diarization // Proc. Interspeech – 2009. – P. 1047–1050.
5. Jin Q., Laskowski K., Schultz T., Alex Waibel A. Speaker segmentation and Clustering in Meetings // Proc. ICASSP-2004 Meeting Recognition Workshop. – Montreal, Canada. – May 2004. – P. 112–117.
6. Радио Свобода [Электронный ресурс]. – URL: http://www.svobodanews.ru/, свободный. Яз. рус. (дата обращения 20.09.2012).

Кудашев Олег Юрьевич – ООО «ЦРТ-инновации», программист, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, kudashev@speechpro.com

УДК 004.04 + 004.22 + 004.415.2 + 004.658.2

ПРОГРАММНАЯ СИСТЕМА И ИНФОРМАЦИОННЫЙ БАНК МЕТАДАННЫХ ТРЕТИЧНЫХ СТРУКТУР БЕЛКОВ

Т.А. Никитин, Ю.Б. Порозов

Приведена архитектура модели базы хранения метаданных о результатах проверок трехмерных структур белков. Построена концептуальная модель базы данных. Представлен сервис и процедура обновления базы данных, а также алгоритмы преобразования данных о записях структур белков и их качестве. Приведены наиболее важные сведения о записях и форме их представления для хранения, выборки и предоставления пользователям. Разработан комплекс программного обеспечения для реализации функциональных задач с использованием языка программирования Java в среде разработки NetBeans v.7.0, а также языка JQL для формирования запросов и взаимодействия с базой данных JavaDB. Проведено тестирование сервиса. Результаты работы системы показали ее эффективность при фильтрации записей структур белков PDB.

Ключевые слова: программная система, информационный банк, структура белка, PDB, качество структуры, сервер проверок, метаданные.

Введение

В настоящее время идет бурное развитие многих областей современной науки вследствие совершенствования техник и технологий, проведения экспериментов, требующих хранения огромного массива получаемых данных. Одним из таких направлений является биоинформатика – наука, занимающаяся разработкой методов анализа, управления и обработки данных о биологических последовательностях и их взаимосвязях. Эти данные хранятся в специально созданных информационных банках (ИБ), функционирующих на базе научно-исследовательских институтов или университетов, ведущих экспериментальные работы в области биохимии и молекулярной биологии [1].

В последние годы наблюдается взрывной рост количества записей в ведущих базах данных (БД) нуклеотидных последовательностей и структур белков. Как следствие, остро встает вопрос об администрировании, упорядочивании, фильтрации и периодической ревизии общедоступных некоммерческих БД, которые часто хранят неполные сведения, результаты, полученные на устаревшем оборудовании или имеющие ошибки. Наличие таких записей в результатах поиска приводит к неверным расчетам и выводам, ошибкам в постановке экспериментов и излишним временным и материальным затратам на ручную сортировку.

Важно отметить, что задача состоит не только в сборе информации и ее хранении, но и в предоставлении сервисов для выборки и анализа данных, выявлении в них закономерностей и взаимосвязей, построении регуляторных моделей и моделей для предсказания структуры, функции и клеточного расположения [2].

В связи с этим разработка нового программного обеспечения и совершенствование способов описания, хранения, простого и быстрого поиска информации о последовательностях и структурах становятся чрезвычайно актуальными [3].

Описываемая разработка имеет целью существенно облегчить работу специалистов в области структурной биологии белков, предоставляя только заранее проверенные и ранжированные записи из общедоступных БД, отсеивая или предупреждая о сомнительных, неполных или ошибочных данных.

Обзор существующих программных систем и БД хранения информации о последовательностях и структурах

В настоящее время существуют сотни веб-сайтов, предоставляющих данные по различным аспектам геномики, протеомики, системной биологии и средства их анализа. Многие из них имеют свой формат хранения данных, степень избыточности и связи с родственными или аналогичными БД. Каждый ресурс имеет свои средства работы с информацией – различные поисковые программы, программные средства визуализации, пополнения базы. Крупнейшие хранилища нуклеотидных и аминокислотных последовательностей пополняются аннотированными записями непосредственно исследователями и лабораториями. Однако в курируемых БД новая информация проверяется обслуживающим ее персоналом [4].

Необходимо подчеркнуть, что в отличие от традиционной библиографической научно-технической информации, собираемой и распространяемой на печатных носителях и в электронной форме, данные в биоинформатике являются фактографическими и гораздо более тесно привязаны к источникам их происхождения. В связи с этим большинство биоинформационных БД созданы и поддерживаются крупными центрами, например, European Molecular Biology Laboratories [5], European Bioinformatics Institute [6], National Center for Biotechnological Information [7], DNA DataBank of Japan. В то же время научные группы и лаборатории, ведущие работы в области биохимии, молекулярной и структурной биологии, также поддерживают свои БД. Как правило, в первом случае обеспечивается функционирование общих банков данных, содержащих информацию о последовательностях белков и нуклеиновых кислот, а во втором – специализированных банков данных.

В зависимости от содержания различают несколько типов биологических БД. Первый тип – архивные: GenBank [8], EMBL, Protein Data Bank (PDB) [9]. Второй тип – курируемые, к которым относится Swiss-Prot – наиболее качественная БД, содержащая аминокислотные последовательности белков. Третий тип – производные БД, получающиеся в результате обработки данных из архивных и курируемых баз: SCOP (структурной классификации белков), PFAM (семейств белков), Gene Ontology, ProDom (белковых доменов). Благодаря такому разделению обеспечивается высокий уровень экспертной оценки поступающих и сохраняемых данных.

Вместе с тем в настоящее время остро стоит проблема интеграции и фильтрации информации, содержащейся в различных базах, так как большое количество записей содержит фрагменты белков, устаревшие данные, дублирующие записи (структура одного и того же белка в комбинации с разными лигандами или фрагментами других молекул), записи низкого разрешения с ошибками структур различного рода или не соответствующие спецификации PDB, что затрудняет работу с ними.

Материалы и методы

Разработанное программное обеспечение собирает данные о результатах проверок записей белковых структур и поддерживает в актуальном состоянии БД с ними. Созданный сервис позволяет предоставить отфильтрованные по критерию качества данные, соответствующие заданным параметрам поиска. Потенциал сервиса проявляется в комплексном взаимодействии с банками данных, хранящими информацию о трехмерных структурах белков, сервисами их проверок и учете запросов в соответствии с индивидуальными потребностями пользователей сервиса.

Выбранная для работы сервиса мировая база PDB предоставляет информацию о структурах в форматах PDB и mmCIF – координаты атомов и дополнительную информацию (ротамеры, элементы вторичной структуры, ссылки на статьи и т.д.) [10].

Чтобы сравнить записи конкретного белка и организовать работу системы, необходимо иметь все индексы записей PDB. Актуальные индексы представлены в файле на специализированном сервисе WHY_NOT [11], который индексирует все записи в банке PDB и составляет списки доступных записей, предоставляя тем самым удобный способ получения информации и статистики. Сами записи белковых структур хранятся в различных форматах, например, в сжатом виде (*.gz), и доступны для загрузки [11].

Проверки на сервере What If. Анализ качества белковых структур

Для качественного сравнения трехмерных структур белков был использован онлайн-сервер What If [12], предоставляющий результаты следующих проверок:

– общих сведений в заголовке записи (имя, номенклатура, потерянные атомы);
– сведений о симметрии (количество молекул и состав объекта);
– параметров геометрии (межатомные расстояния, длины ковалентных связей, параметры планарных и торсионных углов);
– структурных параметров (ошибки вторичной структуры в соответствии с DSSP, качество упаковки, положение атомов, углов и аномальных связей, анализ *B*-фактора).

Результаты проведенных проверок находятся на сервере What If в открытом доступе. Для каждой белковой записи хранятся результаты трех проверок: оригинальной записи из PDB, записи после консервативной и полной оптимизации. Авторами были отобраны 10 оценок, отражающих качество анализируемых структур: разрешение структуры (Å); общая оценка качества; качество структуры 1 и 2 уровней; распределение по диаграмме Рамачандрана; характеристики углов χ^1 и χ^2; конформация основной цепи; длины ковалентных связей; планарные углы; число столкновений атомов; распределение структуры.

Информация об используемых данных, доступных в явном виде, помогающая находить объект, называется метаданными, которые особо важны, поскольку их использование позволяет повысить качество поиска. При этом хранимые данные не дублируют исходные ресурсы, а производится сбор и хранение лишь выборочной информации для получения ускоренного доступа к оригиналу [13]. Сохраняемые метаданные содержат название исходной записи, ее уникальный идентификатор PDB, дополнительную информацию и результаты проверок. Главный параметр записи – ее идентификатор, представляющий собой четырехсимвольную последовательность, первый символ которой является цифрой, а последующие могут быть буквами или цифрами. Названием белковой записи обычно является одно или несколько слов, определяющих название структуры. Дополнительная информация берется из раздела заголовка PDB-файла TITLE. Ссылки на адреса самой структуры и дополнительной информации о ней, серверы проверок и основных используемых ресурсов в сети Интернет хранятся в виде URL-ссылок.

База данных

Разработанная БД представляет собой ИБ метаданных, организованных в логическую структуру формального вида для эффективного поиска и обработки данных. В данном случае целесообразно использование централизованной БД для минимизации затрат ресурсов и времени на получение данных [14].

Для обеспечения минимальной логической избыточности БД при проектировании отношений учитывалась совокупность требований – нормальные формы. БД создана в соответствии с тремя базовыми требованиями классической реляционной модели данных, которая является фактическим стандартом для современных систем управления базами данных (СУБД) [15]. Она была взята за основу благодаря своим достоинствам, таким как упрощение схемы БД (таблицы), логическая и физическая независимость, доступ к БД на языках высокого уровня, улучшение целостности и защиты данных. Поскольку реляционная структура концептуально проста, она позволяет реализовывать небольшие и простые БД [16]. Концептуальная модель хранилища данных представлена главными сущностями и отношениями между ними, используемыми при разработке модели (рис. 1).

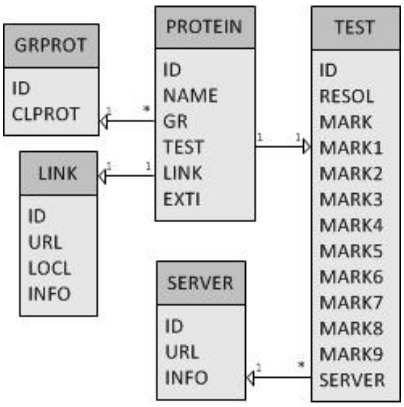


Рис. 1. Концептуальная модель

Модель содержит пять основных сущностей, используемых для обеспечения минимальных требований, предъявляемых к работе системы и ее функциям: PROTEIN, GRPROT, LINK, TEST, SERVER, главной из которых является таблица PROTEIN, содержащая записи метаданных белковых структур: идентификатор записи, имя и дополнительную информацию о ней. Связь каждой записи с ее группой по первичному ключу, находящемуся в таблице GRPROT, является связью «многие-к-одной», поскольку записи входят в состав группы (принадлежат определенной группе). Для каждой записи белка в таблице PROTEIN имеется связь с записями хранения ссылок по первичному ключу в таблице LINK, являющаяся отношением «один-к-одному», поскольку в ней хранятся уникальные ссылки.

Основная таблица PROTEIN содержит шесть полей (уникальный идентификатор PDB, имя белковой записи, идентификаторы для связи с таблицами хранения группы, тестов, ссылок и дополнительную информацию). Таблица LINK содержит ссылки на локальные и внешние источники, а также дополнительные сведения о них. Таблица GRPROT содержит информацию о группах записей белковых структур и их названия.

В таблице TEST хранятся результаты проверок, связанных первичным ключом с соответствующей записью таблицы PROTEIN (отношение «один-к-одному»). Записи о проверках качества на различных серверах связаны с записями о самом сервере по ключу таблицы SERVER (отношение «многие-к-одному»), хранящей адреса серверов проверок и дополнительную информацию о них.

Типы полей разработанной БД обусловлены хранимой в них информацией.

Доступ к базе данных

При интеграции приложения с БД решалась задача организации максимально прозрачного доступа к ней. Один из подходов к разрешению этой задачи заключается в использовании так называемых тонких клиентов. В этом случае клиентское приложение обеспечивает реализацию презентационной логики, а сервер объединяет логику доступа к ресурсам и бизнес-логику. Интерфейс пользователя разрабатывался при помощи языка гипертекстовой разметки HTML с использованием HTML-форм, которые являются наиболее удобным механизмом передачи запросов.

Доступ к БД был обеспечен на стороне веб-сервера, который динамически формирует представление, используя классы Java. Существует универсальный способ доступа к реляционным данным – разработанный компанией Sun ряд стандартных Java-классов [17]. Он позволяет выполнять запросы к БД и ее модификацию непосредственно из Java-программ: классов связанности баз данных – пакетов JDBC (Java Database Connectivity), которые имеют описание всех свойств СУБД, доступных с помощью драйверов JDBC, и обеспечивают абсолютную независимость приложения от БД благодаря инструментам администрирования СУБД в Java [18].

Преимущество использованного J2EE (Java 2 Platform, Enterprise Edition) состоит в том, что модель приложения заключает уровни функциональных возможностей в определенные типы компонентов [19]. Взаимодействие с клиентом осуществляется посредством обычных веб-страниц с использованием

апплетов на основе Java-технологии – Java Server Pages (JSP), или посредством автономных Java-приложений. Веб-приложение Java создает интерактивные веб-страницы, содержащие различные типы языков разметки, например HTML, а динамическое содержимое формируется веб-компонентами – страницами JSP [20], сервлетами и средствами JavaBeans, которые позволяют изменять данные, осуществлять их временное хранение, взаимодействовать с БД и веб-службами.

Для предоставления качественного сервиса пользователям, содержимое БД должно быть актуальным и своевременно обновляться, а также иметь удобный интерфейс и продуманную форму представления.

Сервис актуализации состояния базы данных

Раз в неделю по четвергам на сервере PDB обновляются базы хранения белковых структур. По этой причине целесообразно также раз в неделю запускать сервис обновления. Работа данной функции начинается со сканирования файлов на сервере PDB, содержащих информацию о новых записях. Сервис обращается к европейскому зеркалу мировой базы белковых структур (наиболее близкому, а, следовательно, и быстрому по отношению к нашему серверу) и сканирует ее на предмет наличия новых записей. После этого происходит обращение к серверу проверок What If и просмотр информации о наличии результатов проверок новых записей белковых структур. В случае обнаружения новых записей сервис считывает необходимую информацию о новой белковой структуре, результатах ее проверок и записывает их в БД (рис. 2). После успешного обновления и вывода идентификаторов новых записей, пользователь может найти их по поиску или просмотреть подробную информацию о них (информация о записи 2EGX приведена на рис. 3).

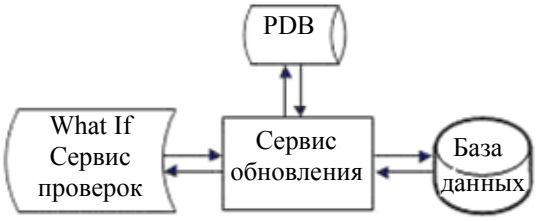


Рис. 2. Процедура обновления состояния базы данных

ID	null
Name	CRYSTAL STRUCTURE OF THE PUTATIVE ACETYLGLUTAMATE KINASE FROM THERMUS THERMOPHILUS
Classification	STRUCTURAL GENOMICS, UNKNOWN FUNCTION
SpaceGroup	P 65
Test service	PDBe
Resolution	1.919
R-value	0.2249
1st generation packing quality	0.823
2nd generation packing quality	-0.0689
Ramachandran plot appearance	-0.725
Chi-1/Chi-2 rotamer normality	-0.4029
Backbone conformation	0.2249
Bond length RMS Z-score	0.2569
Bond angle RMS Z-score	0.5849
Total number of bumps	0.2311
Inside/Outside distribution	70

Рис. 3. Подробная информация о белке 2EXG

Сервис обновления не дублирует записи, а сохраняет их в виде метаданных в локальной БД и использует их, ссылаясь на расширенные отчеты результатов проверок [21] и сами записи в PDB. Ссылки формируются на страницах представления благодаря использованию шаблонов адресов.

Тестирование системы и результаты

Для подтверждения работоспособности разработанной программной системы была проведена проверка обновления БД с последующим контролем появления новых записей в ней и их соответствия реальным данным. Проверка работоспособности системы включала в себя тесты на корректность представления метаданных и соответствующей информации из БД на всех страницах веб-модуля и тест работоспособности ее EJB-модуля.

Была выполнена выборка данных, удовлетворяющих критериям поиска по определенным параметрам на поисковой странице системы. В качестве теста был произведен поиск в локальной базе по записям, содержащим в своем имени строку «calmodulin». Поисковый запрос по ключевому слову «calmodulin» в базе PDB находил 377 записей. В отличие от нее, в разработанной и обновленной базе данных программной системы поиск выдает 90 записей, поскольку, как уже говорилось ранее, не все записи проходят проверку качества и имеют оценки, многие из этих записей устарели или имеют неправильный формат данных. Результатом поиска являются представленные в табличной форме отфильтрованные записи, содержащие поля «идентификатор PDB», «разрешение», «z-score» (кумулятивная оценка качества) и ссылки на внешние источники (рис. 4). Для таблиц вывода записей в системе существует возможность их сортировки по полям и ограничению количества записей на одной странице с указанием максимального количества записей по конкретному запросу в целом и смещения этого интервала относительно найденных записей.

ID	Name	Res	Z-score	SpaceGroup	Links
1exr	THE 1.0 ANGSTROM CRYSTAL STRUCTURE OF CA+2 BOUND CALMODULIN	1.0	0.134	P 1	
2o5g	CALMODULIN-SMOOTH MUSCLE LIGHT CHAIN KINASE PEPTIDE COMPLEX	1.08	0.142	P 1 21 1	
2pq3	N-TERMINAL CALMODULIN ZN-TRAPPED INTERMEDIATE	1.299	0.143	P 43 21 2	
3gof	CALMODULIN BOUND TO PEPTIDE FROM MACROPHAGE NITRIC OXIDE SYNTHASE	1.45	0.1449	P 1 21 1	
2jam	CRYSTAL STRUCTURE OF HUMAN CALMODULIN-DEPENDENT PROTEIN KINASE I G	1.7	0.155	P 1	
3gp2	CALMODULIN BOUND TO PEPTIDE FROM CALMODULIN KINASE II (CAMKII)	1.46	0.1599	C 1 2 1	
2wel	CRYSTAL STRUCTURE OF SU6656-BOUND CALCIUM/CALMODULIN- DEPENDENT PROTEIN KINASE II DELTA IN COMPLEX WITH CALMODULIN	1.899	0.1606	P 21 21 21	
2vz6	STRUCTURE OF HUMAN CALCIUM CALMODULIN DEPENDENT PROTEIN KINASE TYPE II ALPHA (CAMK2A) IN COMPLEX WITH INDIRUBIN E804	2.299	0.1633	P 21 21 21	
1vrk	THE 1.9 ANGSTROM STRUCTURE OF E84K-CALMODULIN RS20 PEPTIDE COMPLEX	1.899	0.171	P 1 21 1	
3gwr	CRYSTAL STRUCTURE OF PUTATIVE CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE TYPE II ASSOCIATION DOMAIN (YP_315894.1) FROM THIOBACILLUS DENITRIFIC	2.009	0.1729	P 21 21 21	

Рис. 4. Вывод ограниченного количества записей, отсортированных по оценке качества структур z-score

Заключение

Основной целью проекта являлась разработка программного обеспечения и базы данных, предоставляющих отсортированные по критерию качества структуры записи PDB. Известно, что при поиске по ключевым словам в PDB пользователь обычно получает несколько записей (зачастую несколько десятков), выбрать лучшую из которых непросто. Разработанная система актуализации представления метаданных и сервис динамического пополнения новыми записями и данными о результатах проверки качества структур помогает пользователю быстро находить несколько наиболее качественных структур среди всех записей белка (и его частей) в PDB. Была разработана архитектура и создана база данных и служба ее актуализации. В качестве фундамента была выбрана база данных JavaDB. Связь с ней была организована посредством JDBC, а языком программной реализации был выбран язык высокого уровня Java для обеспечения совместимости приложения с любыми операционными системами на уровне исполняемого кода.

Проект выгодно отличается по простоте использования от аналогичных сервисов – в нем нет сложной структурной организации возможностей использования и поиска по многочисленным критериям выборки. Разработанная программная система использует результаты работы сервиса онлайн-проверки белковых структур, что обеспечивает точность оценки качества. В то же время она достаточно просто организована для облегчения понимания результатов и ускорения принятия решений о выборе той или иной структуры для работы. Созданная система является своего рода «фильтром», помогающим в поиске наиболее качественных структур среди многих, зачастую одинаковых, неполных, устаревших или ошибочных данных. Интуитивно понятный графический и простой текстовый интерфейсы позволяют пользователю быстро разобраться с основными принципами работы сервиса и начать его эксплуатацию в повседневной работе совместно с сервисами PDB, что способствует сокращению временных затрат на поиск и выборку необходимой информации из PDB.

Дальнейшая работа над комплексом предполагает решение таких задач, как обеспечение взаимодействия с другими сервисами и расширение набора хранимых данных для более гибкого задания параметров выборки посредством подключения иных сервисов онлайн-проверок, информационных банков хранения информации о белковых структурах и увеличения количества атрибутов записей, расширив их представление в базе метаданных.

Исследование поддержано грантом Министерства образования и науки Российской Федерации (ГК № 07.514.11.4163).

Литература

1. Леск А. Введение в биоинформатику. – М.: Бином. Лаборатория знаний, 2009. – 324 с.
2. Игнасимуту С. Основы биоинформатики. – М.: Ижевск: НИЦ «Регулярная и хаотическая динамика». Институт компьютерных исследований, 2007. – 320 с.
3. Афанасьева Г. Биоинформатика: виртуальный эксперимент в шаге от реальности // Наука и жизнь. – 2004. – № 11. – С. 20–24.

4. Молекулярно-биологические базы данных // Объединенный центр вычислительной биологии и биоинформатики [Электронный ресурс]. – Режим доступа: http://jcbi.ru/baza/index.shtml, свободный. Яз. рус. (дата обращения 16.05.2012).
5. Stoesser G., Baker W., van den Broek A. et al. The EMBL Nucleotide Sequence Database // Nucleic acids research. – 2002. – V. 30. – № 1. – P. 21–26.
6. Emmert D.B. The European Bioinformatics Institute (EBI) databases // Nucleic acids research. – 1994. – V. 22. – № 17. – P. 3445–3449.
7. Geer L.Y., Marchler-Bauer A., Geer R.C. et al. The NCBI BioSystems database // Nucleic acids research. – 2010. – V. 38 (Database issue). – D492–496.
8. Benson D.A., Karsch-Mizrachi I., Lipman D.J. et al. GenBank // Nucleic acids research. – 2005. – V. 33. – suppl. 1. – D34–38.
9. Bernstein F.C. The Protein Data Bank: a computer-based archival file for macromolecular structures // Journal of molecular biology. – 1977. – V. 112. – № 3. – P. 535–542.
10. Henrick K., Feng Z., Bluhm W. et al. Remediation of the Protein Data Bank Archive // Nucleic acids research. – 2008. – V. 36. – Suppl. 1. – D426–D433.
11. Joosten R.P., Beek T.A.H., Krieger E. et al. A series of PDB related databases for everyday needs // Nucleic acids research. – 2011. – V. 39. – Suppl. 1. – D411–D419.
12. Vriend G. WHAT IF: a molecular modeling and drug design program // Journal of Molecular Graphics. – 1990. – V. 8. – Is. 1. – P. 52–56.
13. Башмаков А.И., Старых В.А. Систематизация информационных ресурсов для сферы образования: классификация и метаданные. – М.: Европейский центр по качеству, 2003. – 384 с.
14. Дейт К.Дж. Введение в системы баз данных. – 8-е изд. – М.: Вильямс, 2005. – 1328 с.
15. Горев А., Ахаян Р., Макашарипов С. Эффективная работа с СУБД. – СПб: Питер, 2006. – 704 с.
16. Кузнецов С.Д. Базы данных: языки и модели: Учебник. – М.: Бином. Лаборатория знаний, Интернет-университет информационных технологий, 2008. – 720 с.
17. Монахов В. Язык программирования Java и среда NetBeans. – 2-е изд. – СПб: БХВ-Петербург, 2009. – 720 с.
18. Хейк Б. JDBC: Java и базы данных. – М.: Лори, 1999. – 320 с.
19. Ноутон П., Шилдт Г. Java 2. – СПб: БХВ-Петербург, 2008. – 1072 с.
20. Steelman A., Murach J. Murach's Java Servlets and JSP. – 2nd ed. – Fresno, CA, USA: Mike Murach & Associates, 2008. – 729 p.
21. Hooft R. A WHAT IF check report: what does it mean. – 2007 [Электронный ресурс]. – Режим доступа: http://swift.cmbi.ru.nl/gv/pdbreport/checkhelp/explain.html, свободный. Яз. рус. (дата обращения 16.05.2012).

Никитин Тимофей Александрович – Санкт-Петербургский государственный политехнический университет, студент, tim04k@gmail.com

Порозов Юрий Борисович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат медицинских наук, зав. лабораторией, porozov@ifc.cnr.it

УДК 621.391

АНАЛИЗ ЧАСТОТЫ ПОВТОРЕНИЙ RLE-БЛОКОВ В СЕМЕЙСТВАХ БИНАРНЫХ КОДОВ, НАИЛУЧШИХ ПО МИНИМАКСНОМУ КРИТЕРИЮ АВТОКОРРЕЛЯЦИОННОЙ ФУНКЦИИ

А.А. Ковылин, Д.В. Злобин, А.Ю. Родионов

Рассматриваются вопросы отыскания двоичных псевдослучайных последовательностей с автокорреляционной функцией, близкой к идеальной, предназначенных для использования в современных системах передачи информации, в том числе в мобильной связи и интерфейсах беспроводной передачи данных. При синтезе наборов двоичных последовательностей поставлена задача комплектования их на основе минимаксного критерия, по которому последовательность считается оптимальной в соответствии с предполагаемой областью применения. Получены оптимальные последовательности размерностью до 52, для них проведен анализ кодирования длин серий. Выявлены закономерности в распределении количества серий различной длины в кодах, оптимальных по выбранному критерию, что в дальнейшем позволит оптимизировать процесс поиска таких кодов.

Ключевые слова: псевдослучайные последовательности, хаотические сигналы, коды Баркера, минимаксный критерий, автокорреляционная функция, кодирование длин серий.

Введение

В связи с ростом популярности систем, использующих хаотические сигналы, актуально изучение функций, имеющих с ними интегральное соответствие, на основе псевдослучайных последовательностей

с наилучшими автокорреляционными свойствами. В соответствии с теоремой Винера–Хинчина сигналы с идеальной автокорреляционной функцией (АКФ) имеют наилучшие стохастические свойства. Подобные последовательности имеют широкое применение в системах радиолокации, синхронизации, расширения спектра и т.п. Отмечено использование шумоподобных последовательностей при анализе цепочек дезоксирибонуклеиновой кислоты (ДНК).

Учитывая большой интерес к представленной научной теме, встает вопрос о синтезе более длинных последовательностей, нежели те, что используются на сегодняшний день. Поиск оптимальных кодов большой длины методом простого перебора является весьма ресурсозатратной задачей. Все это подталкивает к созданию других методов поиска, обеспечивающих заданный критерий оптимальности, но при этом содержащих меньшее количество вычислительных операций, а следовательно, обладающих меньшим временем расчета. В работе авторами рассматривается один из способов оптимизации такого поиска.

Оптимальные дискретные сигналы

Для количественного определения степени отличия сигнала $u(t)$ и его смещенной во времени копии $u(t-\tau)$ принято вводить АКФ сигнала $u(t)$:

$$B_u(\tau) = \int_{-\infty}^{\infty} u(t)\, u(t-\tau)\, dt \ .$$

Для дискретного сигнала $u = \{u_0, u_1, \ldots, u_{M-1}\}$ АКФ имеет следующий вид:

$$B_u(n) = \sum_{j=-\infty}^{\infty} u_j\, u_{j-n} \ .$$

При синтезе оптимальных дискретных сигналов принято использовать минимаксный критерий: оптимальным считается сигнал с наименьшим уровнем наибольшего из боковых лепестков АКФ. Такой критерий отвечает существу проблемы [1].

Дискретные сигналы с наилучшей структурой АКФ являются объектом интенсивных исследований. Среди них большую известность получили сигналы Баркера. Эти сигналы обладают уникальным свойством: при всех $n \neq 0$ значения их АКФ не превышают единицы. Установлено, что не существует сигналов Баркера с числом элементов, большим 13. Однако в [2] говорится о том, что если последовательность Баркера длиной более 13 существует, то $n = 189\,260\,468\,001\,034\,441\,522\,766\,781\,604$ (т.е. более $2 \cdot 10^{30}$).

Свойства полученных кодов

В результате поиска всех существующих последовательностей по заданным критериям посредством разработанного алгоритма было обнаружено семейство кодов, с учетом зеркальных и обратных последовательностей. В табл. 1 приведено сравнение рассчитанных значений максимального уровня бокового лепестка (УБЛ) АКФ и значений, приведенных в [3]. Следует отметить, что параметры некоторых обнаруженных последовательностей превосходят результаты, приведенные в [3]. Такие параметры отмечены в табл. 1 символом ($^+$).

Длина кода	Максимальный УБЛ АКФ		Количество кодов
	расчет	литература [3]	
14	2	2	72
15	2	2	104
16	2	2	80
17	2	2	32
18	2	2	16
19	2	2	8
20	2	2	24
21	2	2	24
22	3	3	3024
23	3	3	4084
24	3	3	6864
25	**2**$^+$	3	8

Длина кода	Максимальный УБЛ АКФ		Количество кодов
	расчет	литература [3]	
33	3	–	1132
34	3	–	408
35	**3**$^+$	4	888
36	**3**$^+$	4	1288
37	**3**$^+$	4	440
38	3	–	136
39	3	–	240
40	**3**$^+$	4	456
41	**3**$^+$	4	120
42	**3**$^+$	5	32
43	**3**$^+$	4	96
44	**3**$^+$	4	120

Длина кода	Максимальный УБЛ АКФ		Количество кодов
	расчет	литература [3]	
26	3	3	1936
27	3	3	3096
28	2	2	16
29	3	3	2244
30	3	3	688
31	3	3	2008
32	3	–	3376

Длина кода	Максимальный УБЛ АКФ		Количество кодов
	расчет	литература [3]	
45	3	–	32
46	$\mathbf{3}^+$	5	8
47	$\mathbf{3}^+$	4	8
48	$\mathbf{3}^+$	5	32
49	4	–	392704
50	4	–	201352
51	3	–	8
52	4	–	264464

Таблица 1. Сводная таблица кодов

К примеру, один из кодов длиной 52 будет иметь вид (0100101001000001011101000001001100111110011111001110).

Можно отметить, что отношение главного пика АКФ к максимальному УБЛ для кодов длиной 51 и 28 элементов больше, чем у 13-элементного кода Баркера (рис. 1). Это говорит о существовании последовательностей длиной, превосходящей коды Баркера, обладающих лучшим отношением максимума АКФ к боковому лепестку.

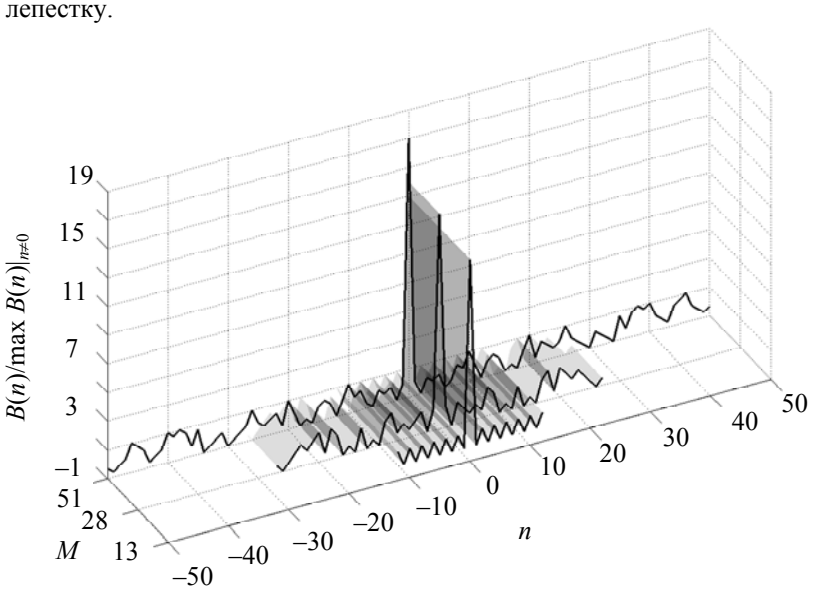


Рис. 1. АКФ оптимальных кодов длиной 13, 28 и 51 соответственно с нормой на максимальный уровень бокового лепестка

В современных системах радиоэлектроники широко применяются псевдослучайные последовательности, сформированные по определенным правилам, например, М-последовательности, последовательности Де Брейна, последовательности Гордона–Милса–Велча (GMW) [4]. Однако их длина кратна степени 2 ± 1, что существенно ограничивает выбор длины кода. При этом значения УБЛ АКФ с ростом длины последовательности уступают аналогичным показателям предложенных кодов (табл. 2).

Последовательности	Максимальное значение УБЛ АКФ при длине последовательности					
	15	16	17	31	32	33
М-последовательности	3	–	–	4	–	–
Последовательности Де Брейна	–	3	–	–	4	–
GMW-последовательности	–	–	3	–	–	6
Найденные коды	2	2	2	3	3	3

Таблица 2. Выборочное сравнение последовательностей по максимальному УБЛ АКФ

Анализ кодирования длин серий

Поиск кодов большой длины методом простого перебора вариантов оказывается весьма громоздким и является проблемой даже для современных вычислительных мощностей: время поиска всех бинарных последовательностей (по заданным критериям) длиной 52 элемента составляло около 180 дней на вычислительном сервере (4×4 ядра с тактовой частотой 3,2 ГГц) [5]. Между тем имеется явная тенденция применять сигналы с все большей размерностью, и это оправдывает поиск других методов синтеза, не связанных с подобными трудностями.

В настоящей работе проводится исследование частоты повторений RLE-блоков (Run-Length Encoding, кодирование длин серий) бинарных последовательностей, полученных с помощью разработанного программного приложения. При кодировании длин серий кодовая последовательность разбивается на блоки, состоящие из идущих подряд одинаковых элементов кода. Код при этом записывается как последовательность длин этих блоков. Таким образом, если рассмотреть известную бинарную последовательность Баркера длиной 13 – (1111100110101), то запись ее в формате RLE выглядит как (5221111), количество анализируемых RLE-блоков составит 7, а число уникальных блоков – 3.

Анализ сигнатур кодов показал характерное количество RLE-блоков, из которых были рассчитаны доли от общей длины кода (рис. 2). Это хорошо согласуется с теоремой Винера–Хинчина, так как конкретные RLE-блоки отвечают за нахождение определенных частот внутри кода, а неравномерность количества элементов в разных блоках показывает равномерность спектральной плотности кода.

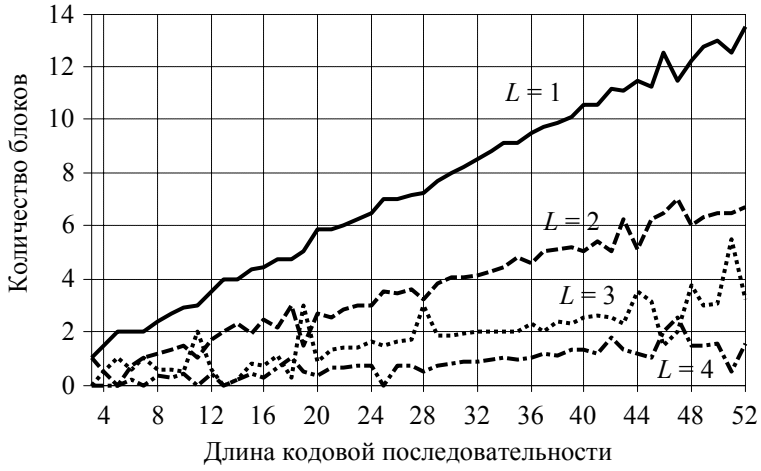


Рис. 2. Количество RLE-блоков длиной 1, 2, 3 и 4 элемента в зависимости от длины кода

Существуют три основных свойства любой двоичной псевдослучайной последовательности, которые могут быть использованы в качестве проверки на случайность. Это сбалансированность, цикличность и корреляция. Циклом (группой) называется непрерывная последовательность одинаковых двоичных чисел. Появление одной двоичной цифры автоматически начинает новый цикл. Длина группы равна количеству цифр в нем. Желательно, чтобы в каждом фрагменте последовательности приблизительно половину составляли циклы обоих типов длиной 1, приблизительно одну четверть – длиной 2, приблизительно одну восьмую – длиной 3 и т.д. [6]. Таким образом, доля содержания конкретной группы в рассматриваемом коде определяется выражением

$$p_i = 2^{-i},$$

где i – длина группы. В результате проведенного авторами в рамках данной работы исследования синтезированных сигнатур кодов вышеупомянутое распределение оказалось несколько иным:

$$p_i = 2^{-i-1}. \quad (1)$$

Варианты такого распределения представлены на рис. 3 (утолщенная линия). Распределение долей RLE-блоков в коде имеет медианный характер, так как сумма RLE-блоков в заданной пропорции не равна длине кода. Необходимую длину кода дополняют отклоненные от медианного значения RLE-блоки большой длины. В действительности эти отклонения незначительно будут влиять на время расчета кода необходимой длины. Алгоритм расчета подобных кодов предполагает рекуррентный анализ вновь найденных кодов для обновления значений распределения RLE-элементов в кодах. Количество RLE-элементов большой длины для найденных кодов имеет хаотический характер в пределах определенных чисел, в то время как число RLE-блоков малой длины имеют характерную линейную зависимость от длины кода, с распределением (1). С увеличением длины кода RLE-элементы большой длины также начинают подчиняться распределению (1).

Это предполагает сокращение времени расчета кодов, в отличие от метода простого перебора, в котором время расчета имеет степенную зависимость от длины кода.

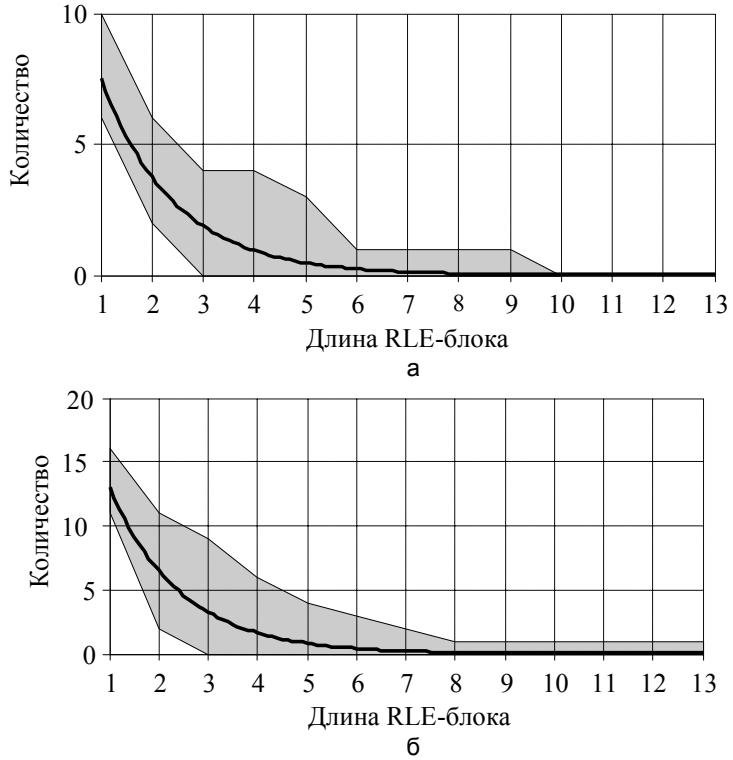


Рис. 3. Диапазон значений количества RLE-блоков в кодах длиной 30 (а) и 52 (б) (утолщенная линия – аппроксимирующая кривая)

Равномерность спектральной плотности кода – главное условие для получения минимального значения УБЛ его АКФ. График зависимости доли блока в коде от длины блока (рис. 3) хорошо апроксимируется экспоненциальной функцией.

Заключение

В результате поиска оптимальных двоичных последовательностей с размерностью, большей 13, было обнаружено существование кодов, превосходящих по своим автокорреляционным свойствам все известные. Выявлена тенденция к периодическому повышению «качества» кода с увеличением его длины, что говорит об актуальности дальнейшего поиска более длинных кодов.

Анализ частоты повторений RLE-блоков в найденных семействах бинарных кодов позволил выявить закономерности распределения количества блоков в зависимости от длины кода (1). Данные закономерности в дальнейшем позволят сузить область поиска оптимальных кодов при решении задач по синтезу и оптимизации структуры сложных сигналов.

Литература

1. Гантмахер В.Е., Быстров Н.Е., Чеботарев Д.В. Шумоподобные сигналы. Анализ, синтез, обработка. – СПб: Наука и техника, 2005. – 400 с.
2. Mossinghoff M.J. Wieferich pairs and Barker sequences // Designs, Codes and Cryptography. – 2009. – V. 53. – № 3. – P. 149–163.
3. Свердлик М.Б. Оптимальные дискретные сигналы. – М.: Советское радио, 1975. – 200 с.
4. Golomb S.W., Gong G. Signal design for good correlation for wireless communication, cryptography and radar. – US: Cambridge University Press, 2005. – 438 p.
5. Чусов А.А., Ковылин А.А., Стаценко Л.Г., Миргородская Ю.В. Параллельный поиск сигналов с заданными взаимно и автокорреляционными свойствами на многопроцессорных платформах // Известия вузов. Радиоэлектроника. – 2010. – Т. 54. – № 8. – С. 29–35.
6. Скляр Б. Теоретические основы и практическое применение: Пер. с англ. – 2-е изд., испр. – М.: Вильямс, 2004. – 1104 с.

Ковылин Александр Александрович – Дальневосточный федеральный университет, инженер, kalexer@hotmail.com

Злобин Дмитрий Владимирович – Дальневосточный федеральный университет, инженер, memrbomel@mail.ru

Родионов Александр Юрьевич – Дальневосточный федеральный университет, кандидат физ.-мат. наук, доцент, deodar1618@mail.ru

УДК 004.4'242

МИНИМАЛЬНАЯ МОДИФИКАЦИЯ АВТОМАТНЫХ ПРОГРАММ ПРИ ИЗМЕНЕНИИ СЦЕНАРИЕВ ИХ РАБОТЫ

А.В. Шестаков

Предлагается метод модификации автоматных программ при изменении сценариев их работы. Метод, в отличие от разработанных ранее, не использует эволюционные алгоритмы. Его важной особенностью является требование минимальности изменения программы. Метод может быть применен для построения автоматных программ на основе набора сценариев. Его суть заключается в сведении исходной задачи к задаче поиска кратчайшего пути в ациклическом взвешенном графе, построенном по исходному автомату и добавляемому сценарию работы.

Ключевые слова: автоматное программирование, сценарий работы, автомат Мура.

Введение

Одним из подходов к программированию объектов со сложным поведением является использование парадигмы автоматного программирования [1] или «программирования с явным выделением состояний» [2]. При использовании этого подхода поведение программы задается в виде управляющих конечных автоматов. Состояния каждого автомата соответствуют состояниям управляемого объекта, а переходы между состояниями осуществляются при выполнении заданных условий.

Автоматное программирование хорошо подходит для создания событийно-ориентированных приложений [3, 4]. По каждому событию автомат в зависимости от состояния генерирует определенное выходное воздействие, а последовательность событий порождает последовательность выходных воздействий. Такая цепочка пар «событие – воздействие» называется сценарием работы автоматной программы.

Изменение сценариев работы требует модификации автоматной программы. Метод, предложенный в настоящей работе, позволяет изменить программу в соответствии с новыми сценариями. Использование этого метода также гарантирует минимальность изменений. Существуют методы, позволяющие построить автомат на основе определенного множества сценариев, например, с помощью генетических алгоритмов [5]. Предлагаемый метод, в отличие от разработанных ранее, не использует эволюционные алгоритмы.

Таким образом, постановка задачи работы состоит в следующем. Требуется модифицировать автоматную программу таким образом, чтобы полученная программа удовлетворяла измененному набору сценариев работы. Дополнительным требованием является минимальность отличия модифицированной программы от исходной. Способ определения количества изменений приведен в разделе «Нахождение кратчайшего пути».

Предметная область

Логика автоматной программы задается с помощью детерминированного конечного автомата. В работе в качестве автоматной модели используется автомат Мура второго рода [2]. Используемую модель автомата можно представить в виде графа переходов, в котором при входе в каждое состояние генерируется выходное воздействие, переходы осуществляются по событиям и определено стартовое состояние, а также одно или несколько конечных.

Сценарием работы в настоящей работе называется список пар $<e, z>$, где e – событие, z – выходное воздействие. Сценарий работы определяет поведение программы при появлении определенной последовательности событий.

Под рабочим множеством в работе понимается множество сценариев, определенное пользователем, причем автомат должен удовлетворять сценариям из этого множества. Сценариям, не принадлежащим рабочему множеству, автомат может, как удовлетворять, так и не удовлетворять.

Рабочее множество может быть изменено несколькими способами:

- добавление сценария;
- удаление сценария;
- изменение существующего сценария.

При удалении сценарий убирается из рабочего множества, а автомат можно не изменять. При добавлении сценария может возникнуть ситуация конфликта. Конфликтом называется ситуация, когда требуется добавить переход из заданного состояния по событию e в состояние с выходным воздействием на входе z_1, но уже существует переход в состояние с воздействием на входе z_2.

Добавить сценарий, конфликтующий с хотя бы одним сценарием из рабочего множества, невозможно. Однако, если добавляемый сценарий конфликтует с автоматом, но не конфликтует со сценариями рабочего множества, то его добавление возможно с помощью предварительного изменения автомата при помощи рефакторинга [6]. Добавление сценария при условии конфликта с автоматом подробно рассмотрено в работе [7]. В случае отсутствия конфликтов сценарий следует добавить с минимальным изменением автоматной программы при помощи изложенного далее алгоритма.

Добавление сценария

Суть алгоритма заключается в сведении исходной задачи модификации автомата к хорошо изученной задаче поиска кратчайшего пути в графе. Перед началом работы алгоритма требуется убедиться, что сценарий не конфликтует с исходным автоматом. В случае обнаружения конфликта его следует устранить с помощью метода, описанного в работе [7]. Алгоритм выявления конфликта представлен ниже.

1. Пусть необходимо добавить сценарий $sc=\{e_i, z_i\}_{i=1..n}$. Текущее состояние *currentState=startState* (стартовое состояние), номер рассматриваемого элемента сценария *pairNum*=1.
2. Если *pairNum* совпадает с длиной сценария *n* и текущее состояние выделено как конечное, то автомат уже удовлетворяет сценарию, добавление сценария не требуется.
3. Если *pairNum* совпадает с длиной сценария *n* и текущее состояние не является конечным, то сценарий является префиксом другого сценария.
4. Если существует исходящий переход из *currentState* по событию $e_{pairNum}$ в состояние *state* с воздействием на входе $z_{pairNum}$, то *currentState=state*, *pairNum=pairNum*+1. Переход к шагу 3.
5. Если существует исходящий переход из *currentState* по событию $e_{pairNum}$ в состояние *state* с воздействием на входе $z \neq z_{pairNum}$, то обнаружен конфликт.
6. Если не существует исходящего перехода из *currentState* по событию $e_{pairNum}$, то конфликта нет, и можно добавлять сценарий.

Результатом проверки является либо сообщение о том, что сценарий в автомат добавить нельзя, либо хвост исходного сценария $subSc=\{e_i, z_i\}_{i=m..n}$ и состояние *S*, из которого требуется этот остаток добавить в автомат. Для удобства далее будем считать, что задача состоит в добавлении в автомат сценария *subSc*, пары которого перенумерованы, начиная с единицы. Следовательно, $subSc=\{e_i, z_i\}_{i=1..n-m}$, а стартовым состоянием считается состояние *S*.

Построение графа по автомату и сценарию

Для сведения исходной задачи к задаче нахождения кратчайшего пути строится взвешенный граф. Кратчайший путь в этом графе будет соответствовать такому минимальному изменению автомата, что полученный автомат будет содержать требуемый сценарий.

Для построения графа рассматривается список воздействий $\{z_i\}_{i=1..n}$ добавляемого сценария. Каждому воздействию z_i сопоставляется список состояний автомата, при входе в которые выполняется это воздействие: $z_i \rightarrow \{s_1^i, s_2^i, \ldots\}$. Далее список состояний, соответствующий *i*-му воздействию сценария, называется *i*-м слоем. Для того чтобы добавить сценарий, необходимо построить непротиворечивый путь через эти слои, причем разрешается использовать только одно состояние из слоя.

Для реализации каждого воздействия может использоваться не только состояние из автомата, но и новое состояние, добавленное в автомат,. Для того чтобы избежать ситуации, когда для реализации воздействия невозможно выбрать состояние из списка, в каждый список добавляется *k* новых состояний, где *k* – число раз, когда пара z_i, e_i встречалась в сценарии. Таким образом, если $<z_i,e_i>=<z_j, e_j>$, то:

– $z_i \rightarrow \{s_1^i, s_2^i, \ldots, ns_1^i\}$;
– $z_j \rightarrow \{s_1^j, s_2^j, \ldots, ns_1^j, ns_2^j\}$.

Здесь верхний индекс обозначает номер слоя, а нижний – номер состояния. При этом ns_1^j и ns_1^i указывают на одно и то же состояние, хотя находятся в разных слоях.

Построенные списки состояний используются для построения графа. Каждый список рассматривается как множество вершин, каждой из которых соответствует либо состояние автомата, либо новое состояние, которое, возможно, придется добавить в автомат. В граф добавляется вершина, соответствующая стартовому состоянию, полученному на этапе выявления конфликта.

Для добавления ребер в граф рассматривается каждая вершина *v*. Пусть этой вершине соответствует состояние *s* исходного автомата. Ребра добавляются по следующим правилам:

- если *s* – стартовое состояние, то надо добавить ребра из *v* во все вершины первого слоя. Вес каждого ребра устанавливается равным единице;
- если вершина *v* лежит в *i*-м слое и у состояния *s* существует исходящий переход $s \rightarrow e_{i+1} \rightarrow s_{zi+1}$, где s_{zi+1} – некоторое состояние с воздействием z_{i+1} на входе, то необходимо добавить ребро $v \rightarrow w$, где *w* – вершина из (*i*+1)-го слоя, соответствующая состоянию s_{zi+1}. Вес этого ребра устанавливается равным нулю;
- если вершина *v* лежит в *i*-м слое и у состояния *s* существует исходящий переход $s \rightarrow e_{i+1} \rightarrow s_{zx}$, где s_{zx} – некоторое состояние с воздействием $z \neq z_{i+1}$ на входе, то ребро не добавляется;
- если вершина *v* лежит в *i*-м слое и у состояния *s* нет исходящего перехода по событию e_{i+1}, то добавляются ребра из *v* во все вершины следующего слоя. Вес ребер устанавливается равным единице;
- если вершина *v* лежит в *i*-м слое, ей соответствует новое состояние *ns*, не представленное в исходном автомате, то добавляются ребра из *v* во все вершины следующего слоя. Вес ребер устанавливается в 1+*W*, где *W* – вес добавления нового состояния. Вес добавления нового ребра считается равным единице.

Так как возможно, что не из всех вершин графа были добавлены исходящие ребра, то необходимо удалить все вершины, из которых недостижима хотя бы одна из вершин, соответствующих конечным состояниям исходного автомата. Следует заметить, что, исходя из построения, в графе нет вершин, недостижимых из стартовой вершины (в каждом слое существует хотя бы одна вершина, из которой достижим весь следующий слой). Пример построенного графа приведен на рисунке.

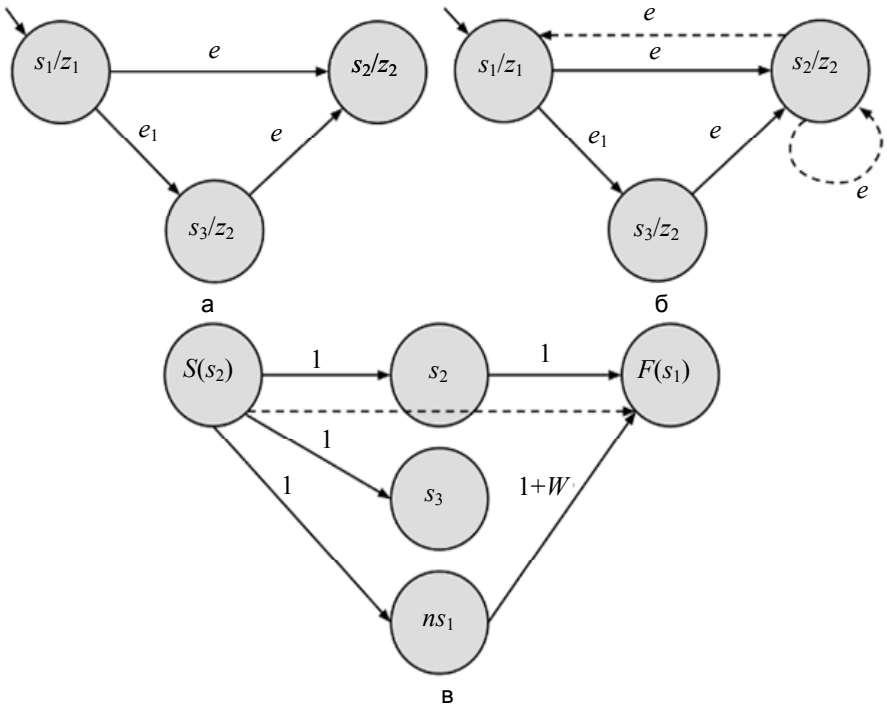


Рисунок. Добавление в автомат сценария $e\ z_2\ e\ z_2\ e\ z_1$: исходный автомат (а); автомат после добавления требуемого сценария (б); граф, построенный по исходному автомату и добавляемому сценарию (в)

Добавляемый сценарий $e\ z_2\ e\ z_2\ e\ z_1$ уже частично присутствует в автомате, поэтому добавляется его остаток $e\ z_2\ e\ z_1$. Стартовым состоянием S считается состояние s_2. Найденный в графе кратчайший путь $s_2\ s_2\ s_1$ (на рисунке отмечен пунктиром) имеет вес два. Соответствующие ему переходы добавляются в исходный автомат.

Нахождение кратчайшего пути

Задача добавления нового сценария свелась к задаче поиска кратчайшего пути в построенном графе. На рисунке такой путь помечен пунктиром. В настоящей работе метрика изменения автомата задается как $\mu=E+W\cdot S$, где E – число добавленных переходов; S – число добавленных состояний; W – вес добавления состояния. Если кратчайший путь найден, то остается добавить в автомат новые состояния, соответствующие вершинам, через которые проходит путь, и переходы, соответствующие ребрам пути ненулевого веса.

Проблема, из-за которой нельзя напрямую применять описанный в работе [8] метод поиска кратчайшего пути для ациклических взвешенных графов, состоит в том, что вес и существование ребра зависят от уже пройденного пути:

– если был совершен переход по ребру $u\to e\to v$, то следующий переход по ребру $u\to e\to v$ имеет вес ноль;
– если был совершен переход по ребру $u\to e\to v$, то дальнейший переход по ребру $u\to e\to w$ невозможен;
– пусть u соответствует новому состоянию, которого еще нет в автомате, тогда ребро $u\to e\to v$ имеет вес $1+W$, однако следующий переход $u\to e_1\to w$ будет иметь вес, равный единице, потому что состояние u уже создано.

Первые две ситуации возможны только в том случае, когда в добавляемом сценарии $sc=\{e_i, z_i\}_{i=1..n}$ существуют повторяющиеся пары $z_i\ e_{i+1}$. Ребра, прохождение по которым изменяет граф, будем называть спорными. Ситуация третьего типа возможна при повторении в сценарии отдельных воздействий и не зависит от событий сценария. В качестве примера рассмотрим граф на рисунке. В добавляемом сценарии $e\ z_2\ e\ z_2\ e\ z_1$ есть повторение пары $z_2\ e$. Путь из вершины S в вершину F не может проходить через вершины $s_2\ s_3\ s_1$, так как из-за наличия ребра $s_3\to e\to s_2$ ребро $s_3\to e\to s_1$ становится недопустимым.

Поиск кратчайшего пути в описанных условиях реализуется с помощью функции послойной релаксации ребер, которая рекурсивно вызывает себя при прохождении через спорное ребро. Следовательно, функция фиксирует спорное ребро и ищет кратчайший путь в графе с учетом пройденного пути.

Пусть в каждой вершине хранится информация о предке этой вершины в кратчайшем пути и ее расстоянии до стартовой вершины (до начала работы все расстояния считаются бесконечными). Между двумя слоями графа могут быть как спорные, так и обычные переходы, поэтому возникает проблема: функция может ослабить обычное ребро, обновив расстояние в вершине v, затем для следующего спорного ребра вызвать себя рекурсивно и снова обновить расстояние в вершине v, но уже с учетом спорного ребра, что не всегда приводит к корректному результату. Предлагается для каждого слоя сначала запускать рекурсивно функции для каждого спорного ребра, а затем продолжать релаксацию обычных ребер. Также каждая функция должна хранить, какие из вершин следующего слоя доступны из текущего, иначе возможно некорректное построение кратчайшего пути. Кроме того, корректная релаксация ребра $u \to v$ затруднена тем, что текущей функции неизвестно, кто задал вершине v ее текущее расстояние до стартовой вершины, так как это могла сделать и другая функция. Таким образом, у каждой функции должен быть свой идентификатор, который позволял бы определить, можно ли ослабить ребро.

Если в некоторый момент времени функция находится в вершине v, то из вершины v уже известен кратчайший путь до стартовой вершины. Эта информация позволяет определить, было ли уже создано состояние, соответствующее v.

Пусть у каждой вершины графа есть следующие поля:

1. *parent* – предок вершины в кратчайшем пути;
2. *distance* – расстояние до стартовой вершины;
3. *createdStates* – множество состояний, созданных при прохождении по кратчайшему пути к этой вершине;
4. *lastModified* – идентификатор функции, которая последней обновляла поля вершины.

Тогда псевдокод ослабления ребра $u \to v$ с весом w функцией с идентификатором *func* выглядит следующим образом:

```
IF(lastModified != func OR u.distance + w < v.distance)
      v.distance = u.distance + w
      v.parent = u
      v.lastModified = func
      v.createdStates = u.createdStates
      IF(u соответствует новому состоянию)
           v.createdStates.add(u)
```

Приведем псевдокод функции нахождения кратчайшего пути:

Вход: стартовая вершина *start*, множество сделанных спорных переходов E.

Выход: кратчайший путь P.

1. Пусть P – кратчайший путь из *start*, $P=\varnothing$.
2. Пусть Q – множество вершин текущего слоя, $Q=\{start\}$.
3. Пока $Q \neq \varnothing$:
    1. $Q'=\varnothing$ – вершины следующего слоя, доступные из текущего. Для всех спорных ребер $u \to v$, таких, что u принадлежит Q:
        1. Если $u \to v$ принадлежит E, то ослабляется ребро $u \to v$, вес считается равным нулю. $Q'=Q'$ в объединении с $\{v\}$.
        2. Если $u \to v$ не принадлежит E, то ослабляется ребро $u \to v$. Рекурсивно вызывается функция из вершины v с учетом сделанного спорного перехода. Обновляется путь P, если путь, возвращенный вызванной функцией, короче.
        3. Пропустить ребро, если переход по нему невозможен из-за совершенных ранее спорных переходов.
    2. Для всех обычных ребер $u \to v$, таких, что u принадлежит Q , ослабить ребро $u \to v$. $Q'=Q'$ в объединении с $\{v\}$.
    3. $Q=Q'$.
4. Если конечная вершина была достигнута по обычным ребрам и расстояние до нее меньше, чем длина P, то обновить P.

По найденному кратчайшему пути добавляются необходимые ребра и состояния в исходный автомат. В автомате появляется путь, соответствующий добавляемому сценарию.

Заключение

В работе исследована проблема изменения автоматной программы при модификации множества ее сценариев работы. Поставлена задача достижения минимального изменения автоматной программы при добавлении сценария в рабочее множество. Для решения предложен метод, который также может быть использован для построения автомата по набору сценариев работы.

Литература

1. Шалыто А.А. Парадигма автоматного программирования // Научно-технический вестник СПбГУ ИТМО. – 2008. – № 8 (53). – С. 3–24.
2. Поликарпова Н.И., Шалыто А.А. Автоматное программирование. – СПб: Питер, 2010. – 176 с.
3. Шалыто А.А. SWITCH-технология. Алгоритмизация и программирование задач логического управления. – СПб: Наука, 1998. – 628 с.
4. Шалыто А.А. Алгоритмизация и программирование для систем логического управления и «реактивных» систем // Автоматика и телемеханика. – 2001. – № 1. – С. 3–29.
5. Царев Ф.Н. Метод построения управляющих конечных автоматов на основе тестовых примеров с помощью генетического программирования // Информационно-управляющие системы. – 2010. – № 5. – С. 31–36.
6. Федотов П.В., Степанов О.Г. Внесение изменений в автоматные программы // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 1 (71). – С. 77–83.
7. Шестаков А.В. Разработка методов модификации автоматных программ при изменении их сценариев работы. Бакалаврская работа. СПбГУ ИТМО, 2011 [Электронный ресурс]. – Режим доступа: http://is.ifmo.ru/papers/2011-bachelor-shestakov/, свободный. Яз. англ. (дата обращения 07.03.2012).
8. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. – М.: Вильямс, 2007. – 677 с.

Шестаков Алексей Викторович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, shestakov@rain.ifmo.ru

УДК 621.317.42

ФАЗОВО-ГРАДИЕНТНЫЙ МЕТОД ИЗМЕРЕНИЯ МАГНИТНОГО ПОЛЯ В ЭЛЕКТРОМОБИЛЕ[2]

Н.Г. Птицына, В.С. Исмагилов, Ю.А. Копытенко, А.Г. Коробейников

При эксплуатации электромобилей и гибридных автомобилей необходим контроль магнитного поля в реальном масштабе времени, например, для пожарной и электромагнитной безопасности. В работе предложен метод измерения магнитных полей на борту электрического автомобиля с учетом специфики этих полей. Метод опирается на дифференциальные методы измерения и минимизирует количество магнитометрических датчиков.

Ключевые слова: электромобиль, магнитное поле, электромагнитная безопасность, измерение магнитного поля, фазово-градиентный метод.

Введение

Электрификация дорожного транспорта является в настоящее время приоритетным направлением развития науки, технологий и техники. Это связано с такими факторами и рисками, как изменение климата, здоровье населения, зависимость от энергии, а также стоимость исходных материалов. Дорожный транспорт является основным источником загрязнения среды в месте нахождения автомобиля. В городах концентрация вредных выхлопов особенно велика. Растет число доказательств, связывающих автомобильное загрязнение городов с серьезными рисками для здоровья, как, например, с респираторными и кардио-респираторными заболеваниями, а также с раком легких. По сведениям Всемирной организации здравоохранения, вредные выхлопы автомобилей приводят к большему количеству смертей, чем автомобильные аварии. Исходя из этого, Европейское Сообщество в последнее время инвестирует в разработки экологичного автомобиля «green car» суммы порядка миллиарда евро. Замена обычных автомобилей, использующих двигатели внутреннего сгорания, на электрический автотранспорт – электромобили и их гибриды – представляется логичным решением и панацеей от экологических бед.

Преимуществом электромобилей по сравнению с автомобилями, оснащенными двигателями внутреннего сгорания, является отсутствие вредных выхлопов в месте нахождения автомобиля и экологичность, ввиду отсутствия нефтяного топлива, антифризов, трансмиссионных и моторных масел, а также фильтров для этих жидкостей. Дополнительным существенным преимуществом является низкая стоимость эксплуатации электромобиля, особенно в Российской Федерации, где стоимость электроэнергии сравнительно невелика, в первую очередь, в ночное время, когда, скорее всего, и будут заряжаться автомобили.

[2] Работа выполнена при поддержке Министерства образования и науки РФ, государственный контракт № 11.519.11.1011.

Необходимость анализа магнитных полей в электромобиле

С другой стороны, мировые производители электрических автомобилей столкнулись с серьезной проблемой обеспечения электромагнитной безопасности пользователей электрического автомобильного транспорта. Существует озабоченность населения и средств массовой информации по поводу возможных рисков для здоровья и безопасности движения из-за воздействия электромагнитных полей (ЭМП), которые будут генерироваться в электрическом автотранспорте сильными токами, текущими в электропроводах и кабелях. Эта озабоченность подкрепляется многочисленными работами, свидетельствующими о том, что ЭМП промышленной частоты (50–60 Гц) могут представлять угрозу здоровью, в частности, могут провоцировать канцерогенные заболевания, опухоли мозга, различные неврологические нарушения, депрессии [1–4]. Кроме того, магнитные поля (МП) сверхнизкой частоты замедляют реакцию человека, что может представлять серьезную угрозу безопасности движения, если такие явления наблюдаются у водителей [3]. При этом следует учесть два «отягчающих обстоятельства»: во-первых, водители и пассажиры электромобиля будут находиться в непосредственной близости к источникам МП, и, во-вторых, они будут подвергаться экспозиции полей в течение длительного времени. Мы пользуемся феном, электробритвой или микроволновой печью в течение нескольких минут или десятков минут, а в автомобиле люди часто проводят многие часы. Оба вышеперечисленных фактора увеличивают риск негативного воздействия МП.

В то время как во всех типах автомобилей присутствуют внешние ЭМП, включая МП Земли, а также ЭМП от различных бортовых электронных устройств, электрический и гибридный автомобили благодаря своим конструктивным особенностям генерируют существенные внутренние ЭМП в широком диапазоне частот. Электрооборудование в электромобилях является источником переменного ЭМП, имеющего сильную временную и пространственную неоднородность в диапазоне частот от долей герца до сотен мегагерц [5–7]. С экологической точки зрения наиболее важной является магнитная составляющая ЭМП, так как доказано, что именно она может приводить к негативным последствиям для здоровья. МП также могут представлять риск для электромагнитной совместимости различных электротехнических средств и электронных устройств электромобиля.

Таким образом, измерение и оценка МП, а также определение их топологии в электрическом автомобиле является актуальной задачей. До сих пор работ о детальных магнитных измерениях в гибридных автомобилях очень мало, а опубликованные статьи по тестированию МП в полностью электрических автомобилях практически отсутствуют [5, 8]. Однако суммирование известных к настоящему времени сведений и сравнение с данными, полученными для других видов электрического транспорта [9–13] позволило выявить характерные, общие для таких полей, черты. В отличие от синусоидальных полей от линий передач (на частоте 50 или 60 Гц), ЭМП в электрическом автомобиле являются мультичастотными полями, т.е. они являются суперпозицией полей, генерируемых множеством источников на борту автомобиля. Кроме того, выявлено, что МП в электромобиле являются иррегулярными, быстро меняющимися во времени и крайне неоднородными в пространстве салона.

Перечисленные характерные черты МП в электромобиле осложняют подробные и точные измерения этих полей в салоне движущегося автомобиля. Для корректных измерений полей и их градиентов в салоне и в непосредственной близости от электромобиля необходимо учитывать большую пространственную неоднородность поля в салоне автомобиля, аддитивную индустриальную помеху, а также естественное постоянное и переменное геомагнитное поле в том же диапазоне частот [3, 14].

Для мониторинга внутри салона автомобиля пространственно сильно неоднородных МП и построения 3D-топологии необходимо использование многочисленных равномерно распределенных датчиков МП. В настоящей работе для решения этой задачи, предлагается метод тестирования МП в электромобиле, базирующийся на дифференциальных методах измерения.

Фазово-градиентный метод

Одним из дифференциальных методов является фазово-градиентный метод, впервые предложенный в [15–20] для исследования усилителей ультранизких частот (F=0,001–100 Гц) геомагнитных вариаций. Этот метод позволяет по трем трехкомпонентным магнитным датчикам, расположенным треугольником на небольшом расстоянии друг от друга (магнитный градиентометр), построить вектора градиентов и фазовых скоростей магнитных вариаций для любой из трех компонент МП. Вектора градиентов геомагнитных волн, как правило, направлены к локальному источнику пульсаций, а вектора фазовых скоростей – от источника.

Градиенты и фазовые скорости магнитных вариаций можно определять по фазовым задержкам и разностям величин амплитуд вариаций между двумя любыми парами магнитного градиентометра, состоящего из трех разнесенных датчиков. Поскольку координаты магнитных датчиков и расстояние между ними известно, то можно определить градиенты и фазовые скорости для двух пар датчиков, выбранных из трех датчиков магнитного градиентометра, и затем построить вектор фазовой скорости и градиента вариаций следующим образом:

для вектора фазовой скорости:

$$\begin{cases} \alpha = \operatorname{arctg}\dfrac{V_{12}\cos(a_1) - V_{13}\cos(a_2)}{V_{12}\sin(a_1) + V_{13}\sin(a_2)} \\ V = V_{12}\cos\left(\alpha + a_1\right) \end{cases}; \quad (1)$$

для вектора градиента:

$$\begin{cases} \alpha = \operatorname{arctg}\dfrac{G_{13}\cos(a_1) - G_{12}\cos(a_2)}{G_{12}\sin(a_1) + G_{13}\sin(a_2)} \\ G = G_{12} / \sin\left(\alpha + a_1\right) \end{cases}. \quad (2)$$

В выражениях (1)–(2) $V_{12}, V_{13}, G_{12}, G_{13}$ – величины фазовых скоростей и градиентов в направлении между парами трехкомпонентных магнитных датчиков 1 и 2, 1 и 3; α – направление соответствующего вектора относительно заданной оси (V и G – величины фазовой скорости и градиента). Углы a_1 и a_2 определяются через координаты магнитных датчиков 2 и 3 (базовый датчик 1 расположен в начале системы координат): $a_1 = \operatorname{arctg}\left(\dfrac{x_2}{y_2}\right)$, $a_2 = \operatorname{arctg}\left(\dfrac{x_3}{y_3}\right)$.

Градиенты и фазовые скорости могут быть выражены и через другие тригонометрические функции [16–18]. Использование выражений (1)–(2) в этом методе предполагает достаточно хорошую синхронизацию работы всех трех магнитных датчиков (если данные не записываются в один многоканальный аналого-цифровой преобразователь) и высокую дискретность регистрируемых данных.

В рамках модели плоской электромагнитной волны величина фазовой скорости между двумя точками определяется через амплитуды соответствующих компонент вариаций МП с учетом фазовой задержки следующим образом [15, 19]:

$$V_{12} = \frac{2\pi d_{12}}{T\ln\left[B_1(t)/B_2(t+\tau)\right]}. \quad (3)$$

В выражении (3) для магнитных вариаций с периодом T величины B_1 и B_2 определяются в момент времени t на первой станции и в момент времени $t + \tau$ на второй станции (τ – фазовая задержка при прохождении магнитной волны расстояния d_{12} между двумя станциями). Определив по (3) фазовые скорости V_{12} и V_{13}, можно затем воспользоваться выражениями (1) для определения направления и величины вектора фазовой скорости электромагнитных волн. При применении этого метода в качестве величин B_1 и B_2 могут быть использованы среднеквадратичные значения амплитуд.

Применение фазово-градиентного метода для измерения магнитных полей в электромобиле

Обозначим величину и градиент магнитного поля, создаваемого электрооборудованием электромобиля в какой-то точке внутри салона, как B_1 и ΔB_1, а величину поля и градиент внешних источников (индустриальная помеха и естественное постоянное и переменное геомагнитное поле) как B_2 и ΔB_2. Пусть $B_2 \geq B_1$, а ΔB_1 много больше ΔB_2. Тогда разность полей в двух точках внутри салона, поскольку ΔB_1 много больше ΔB_2, с большой точностью будет выглядеть следующим образом:

$$\Delta B = (B_1 + \Delta B_1 + B_2 + \Delta B_2) - (B_1 + B_2) = \Delta B_1 + \Delta B_2 \approx \Delta B_1,$$

т.е. разность ΔB магнитных полей в двух точках внутри салона является градиентом внутренних источников электромобиля, а внешние поля исключаются.

Источники переменного ЭМП в электромобилях имеют небольшие размеры, и их можно рассматривать как излучающие точечные диполи, поля от которых затухают с расстоянием по кубическому закону, поэтому градиенты ЭМП внутри электромобиля на расстояниях до нескольких метров должны быть весьма велики. Действительно, как было показано прямыми измерениями в гибридном автомобиле «Крайслер» [21], пространственные градиенты доходили до 1 Гс/м (100 мкТл/м). В то же время градиенты внешних источников малы в связи с их удаленностью и большими размерами.

Для того чтобы выявить топологию МП в электромобиле с учетом его крайней пространственной неоднородности, датчики в салоне должны располагаться достаточно плотно, т.е. количество датчиков должно быть не менее 10–20 (равномерно на расстоянии десятка сантиметров на уровне пола, на уровне головы водителей и пассажиров и т.д.).

Для уменьшения количества датчиков предлагается использовать фазово-градиентный метод. Так как мы предполагаем, что МП от внутренних источников меняются с расстоянием как поле диполя, то мы можем использовать этот метод для построения трехмерного распределения МП внутри салона. В этом случае минимальное количество трехкомпонентных магнитных датчиков – четыре. Три из них образуют треугольник (например, два датчика на спинках сидений и один вблизи двигателя/батареи), необходимый для измерения полей и градиентов в плоскости треугольника, четвертый поднят вверх или опущен вниз для построения градиентов, направленных вертикально. В пирамиде можно выделить четыре тройки датчиков.

Используя эти тройки, можно построить четыре вектора градиентов МП. Величина и направление этих векторов позволяет интерполировать МП на весь салон электромобиля.

На рисунке представлена примерная схема расположения минимального количества датчиков (4 сенсора) в плоскости пола автомобиля (рисунок, а) и в продольном сечении, перпендикулярном полу (рисунок, б).

Практическую точность метода следует определить в эксперименте.

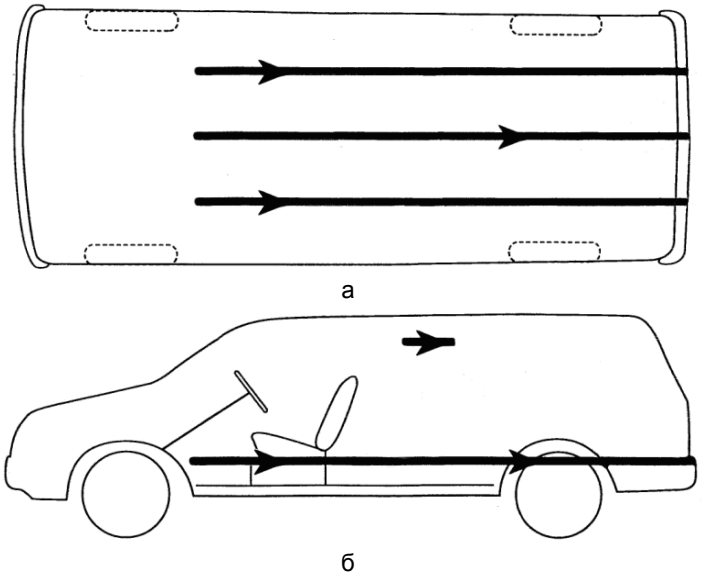


Рисунок. Схема расположения минимального количества датчиков для измерения магнитного поля и его пространственных градиентов в электрическом автомобиле: вид сверху (а); вид сбоку (б). Стрелки обозначают позицию датчиков

Заключение

Предложенный метод позволит разделить внешние и внутренние поля и определить источники увеличенных магнитных полей. В результате можно будет определить 3D-топологию магнитных полей в электрическом/гибридном автомобиле и выявить критические точки для электромагнитной совместимости, а также особенности поля и их источники, потенциально опасные для здоровья водителей и пассажиров.

Литература

1. Григорьев Ю.Г., Григорьев О.А., Степанов В.С., Пальцев Ю.П. Электромагнитное загрязнение окружающей среды и здоровье населения Россию. – М., 1997. – 91 с.
2. Portier C.J., Wolfe M.S. eds. Assessment of Health Effects from Exposure to Power-Line Frequency Electric and Magnetic Fields. – NIEHS Working Group Report NIH Publication. – 1998. – № 98-3981. – 508 p.
3. Птицына Н.Г., Виллорези Дж., Дорман Л.И., Юччи Н., Тясто М.И. Естественные и техногенные низкочастотные магнитные поля как факторы, потенциально опасные для здоровья // Успехи физических наук. – 1998. – Т. 168. – № 7. – С. 767–791.
4. Григорьев Ю.Г., Степанов В.С., Григорьев О.А., Меркулов А.В. Электромагнитная безопасность человека. Справочно-информационное пособие. – М.: Российский национальный комитет по защите от неионизирующих излучений, 1999. – 146 с.
5. Muc A.M. Electromagnetic Fields Associated with Transportation Systems, Radiation Health and Safety Consulting. – Contract Report 4500016448, Air Health Effects Division, Healthy Environments and Consumer Safety Branch Health, Canada, 2002. – 52 p.
6. Schick M. and Jakobus U. Advanced EMC modeling features in FEKO for Automotive Problems // Proceedings of the 22th International Conference on Electromagnetic Compatiability «EMC Europe 2012». – Rome, Italy, 2012. – 17–21 September. – 22 p.
7. Ptitsyna N., Ponzetto A. Magnetic Fields Encountered in Electric Transport: Rail Systems, Trolleybus and Cars // Proceedings of the 22th International Conference on Electromagnetic Compatiability «EMC Europe 2012». – Rome, Italy, 2012. – 17–21 September. – P. 1–6.
8. Swiss Federal Office of Public Health, Cars, hybrid cars. – Technical Report, 2009 [Электронный ресурс]. – Режим доступа: http://www.bag.admin.ch/themen/strahlung/00053/00673/02377/index.html?lang =en, свободный. Яз. англ. (дата обращения – 05.11.2012).

9. Vedholm K., Hamnerius Y.K. Personal Exposure Resulting from Low Level Low Frequency Electromagnetic Fields in Automobiles // Second World Congress for Electricity and Magnetism in Medicine and Biology. – Bologna, Italy, 1997. – June 8–13. – Abstract F-9. – 445 p.
10. Farag A.S., Hussain H., Said I., Abdel Kader M., Abdul Rahman N. Electromagnetic Fields Associated with Transportation Systems in Malaysia // Proceedings of the International Conference on Non-Ionizing Radiation at UNITEN (ICNIR 2003). – Universiti Tenaga NasionalSelangor, Malaysia, 2003. – 15 p.
11. Ptitsyna N.G., Villoresi G. and Kopytenko Yu.A. Magnetic fields from railway: environmental aspects // Railway Transportation: Policies, Technology and Perspectives. – NY, Nova publishers, 2009. – P. 87–140.
12. Птицына Н.Г., Виллорези Дж., Копытенко Ю.А. Тясто М.И. Магнитные поля на электротранспорте и экология человека. – СПб: Нестор-История, 2010. – 120 с.
13. Halgamuge M.N., Abeyrathne C.D., Mendis P. Measurement and Analysis of Electromagnetic Fields from Trams, Trains and Hybrid Cars // Radiation Protection Dosimetry. – 2010. – V. 141. – № 3. – P. 255–268.
14. Ptitsyna N.G., Kopytenko Y.A., Villoresi G., Pfluger D.H., Ismaguilov V., Iucci N., Kopytenko E.A., Zaitzev D.B., Voronov P.M., Tyasto M.I. Waveform Magnetic Field Survey in Russian DC- and Swiss AC-powered Trains: a Basis for Biologically Relevant Exposure Assessment // Bioelectromagnetics. – 2003. – V. 24. – P. 546–556.
15. Копытенко Ю.А., Исмагилов В.С., Копытенко Е.А., Воронов П.М., Зайцев Д.Б. Магнитная локация источников геомагнитных возмущений // ДАН. Сер. Геофизика. – 2000. – Т. 371. – № 5. – С. 685–687.
16. Ismaguilov V.S., Kopytenko Yu.A., Hattori K., Voronov P.M., Molchanov O.A. and Hayakawa M. ULF Magnetic Emissions Connected with Under Sea Bottom Earthquakes // Natural Hazards and Earth Sys. Sci. – 2001. – V. 1. – P. 1–9.
17. Ismaguilov V.S., Kopytenko Yu.A., Hattori K., Hayakawa M. Variations of phase velocity and gradient values of ULF geomagnetic disturbances connected with the Izu strong earthquakes // Natural Hazards and Earth Sys. Sci. – 2002. – V. 20. – P. 1–5.
18. Kopytenko Yu.A., Ismagilov V.S., Hattori K., Hayakawa M. Investigation of the ULF electromagnetic phenomena related to earthquakes: contemporary achievements and the perspectives // Annali di Geofisika. – 2001. – V. 44. – № 2. – P. 325–334.
19. Исмагилов В.С., Копытенко Ю.А., Хаттори К., Хаякава М. Использование градиентов и фазовых скоростей УНЧ геомагнитных возмущений для определения местоположения очага будущего сильного землетрясения // Геомагнетизм и аэрономия. – 2006. – Т. 46. – № 3. – С. 423–430.
20. Коробейников А.Г., Копытенко Ю.А., Исмагилов В.С. Интеллектуальные информационные системы магнитных измерений // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 1 (71). – С. 39–44.
21. Snyder M. Magnetic Shielding for Electric Vechicles. Program Review. Contract DAAE07-93-C-R107. Army TACOM, Chrysler Corp. 1995. – 56 p.

Птицына Наталья Григорьевна – Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова РАН, кандидат физ.-мат. наук, ст. научный сотрудник, Nataliaptitsyna@yandex.ru

Исмагилов Валерий Сарварович – Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова РАН, кандидат физ.-мат. наук, ст. научный сотрудник, ученый секретарь, IVS@izmiran.spb.ru

Копытенко Юрий Анатольевич – Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова РАН, доктор физ.-мат. наук, профессор, директор, office@izmiran.spb.ru

Коробейников Анатолий Григорьевич – Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова РАН, доктор технических наук, профессор, зам. директора, Korobeynikov_A_G@mail.ru

5 МАТЕРИАЛОВЕДЕНИЕ И НАНОТЕХНОЛОГИИ

УДК 539.21+539.216.2

ГАЛЬВАНОМАГНИТНЫЕ СВОЙСТВА ТОНКИХ ПЛЕНОК ВИСМУТА, ЛЕГИРОВАННОГО ТЕЛЛУРОМ, ПОЛУЧЕННЫХ МЕТОДОМ ТЕРМИЧЕСКОГО ИСПАРЕНИЯ В ВАКУУМЕ

В.А. Комаров, Д.Ю. Матвеев, Е.В. Демидов, А.Н. Крушельницкий

Проведено исследование влияния донорной примеси теллура в диапазоне концентраций 0,005–0,15 ат.% на гальваномагнитные свойства (сопротивление, магнетосопротивление и коэффициент Холла) тонких пленок висмута различных толщин. Исследование проведено в интервале температур 77–300 К. Установлено существенное проявление классического размерного эффекта, ослабевающее при увеличении степени легирования пленок. Проведен анализ экспериментальных результатов в приближении закона дисперсии Джонса–Шенберга. Выполнен расчет концентрации и подвижностей носителей заряда в исследуемых пленках.

Ключевые слова: висмут, легирование, тонкие пленки, удельное сопротивление, магнетосопротивление, коэффициент Холла, размерный эффект.

Введение

В связи с развитием технологии и физики наноструктур возросла актуальность исследования причин и закономерностей изменения физических свойств вещества при переходе от массивного монокристалла к низкоразмерным системам и наносистемам [1]. Дополнительную информацию для решения указанной задачи может дать исследование тонких пленок полуметаллов, легированных донорными примесями [2, 3]. Свойства таких систем напрямую зависят от их кристаллической структуры, поэтому это определяет необходимость получения пленок висмута, легированного теллуром, со структурой, приближающейся к структуре объемного монокристалла [4–6]. Процесс легирования является эффективным методом управления свойствами тонких пленок в широком интервале изменения концентрации и энергии носителей заряда, что позволяет получить дополнительную информацию о механизмах рассеяния электронов, об энергетическом спектре носителей заряда и проявлении классического размерного эффекта в электронной подсистеме, обусловленного ограничением длины свободного пробега носителей заряда толщиной пленки [3, 7].

Представленная работа посвящена изучению гальваномагнитных свойств тонких пленок висмута, легированного теллуром. В отличие от ранее выполненных работ [2, 3, 7–10], в которых рассматривались особенности структуры и явлений переноса в тонких пленках чистого и легированного висмута, в настоящей работе проведено целенаправленное комплексное исследование влияния донорной примеси теллура на гальваномагнитные свойства пленок висмута различной толщины.

Методика эксперимента

Изготовление тонких пленок висмута, легированного теллуром, производилось методом дискретного термического испарения в вакууме $1{,}5\cdot10^{-5}$ мм рт.ст. Использование дискретного метода обусловлено большим различием давления паров висмута и теллура при температуре испарения висмута. В качестве подложки для легированных пленок использовалась слюда мусковит. Выбор материала подложки обусловлен тем, что в лаборатории накоплен большой экспериментальный материал по исследованию пленок чистого висмута на подложке из слюды, а также тем, что пленки висмута на слюде имеют преимущественную текстуру с тригональной плоскостью параллельно подложке [7, 8]. Температура подложки в процессе напыления составляла 393 К [3]. Все пленки отжигались в вакууме непосредственно после осаждения при температуре 513 К в течение 30 мин. Скорость осаждения составляла около 10 нм/с. Выбор указанных режимов основан на результатах исследования пленок чистого висмута, полученных в лаборатории полуметаллов отдела физики конденсированного состояния на базе РГПУ им. А.И. Герцена. Измерение толщины пленок осуществлялось методом оптической интерферометрии [11].

Исследование структуры пленок производилось методами рентгеноструктурного анализа и атомно-силовой микроскопии (АСМ). Количество теллура в пленке задавалось путем выбора для распыления кристалла висмута с заранее известной концентрацией теллура. Кристаллы выращивались методом зонной плавки при двух встречных проходах зоны [12]. Такой метод обеспечивает равномерное распределение легирующей примеси по длине слитка. Дополнительный выборочный контроль содержания теллура производился с использованием времяпролетного масс-спектрометра «ЛЮМАС-30». Измерение удельного сопротивления, магнетосопротивления и коэффициента Холла проводилось в интервале температур 77–300 К и в магнитном поле до 0,65 Тл в стационарных условиях на постоянном токе классическими методами при ступенчатом изменении температуры.

Экспериментальные результаты и их обсуждение

В работе исследовались пленки висмута с содержанием теллура 0,005–0,15 ат.% и толщиной 270–670 нм.

Исследование структуры полученных пленок методами рентгеноструктурного анализа и АСМ показало, что они имеют блочную структуру с размерами блоков, примерно на порядок превышающими толщину пленки, и преимущественной ориентацией тригональной плоскости (111) кристаллитов параллельно подложке. Анализ АСМ-изображений показал существенное уменьшение размеров фигур роста [3] при увеличении концентрации теллура по сравнению с фигурами роста в пленках чистого висмута, что, возможно, обусловлено инородными атомами теллура, ухудшающими процесс роста кристалла висмута. Учитывая кристаллографическую ориентацию пленки, следует отметить, что экспериментально измеряемые гальваномагнитные коэффициенты для пленок висмута при индукции магнитного поля, параллельной нормали к плоскости пленки $(\mathbf{B} \parallel \mathbf{n})$, соответствуют следующим компонентам тензоров гальваномагнитных коэффициентов монокристалла висмута: $\rho=\rho_{11}$, $R=R_{12,3}$, $(\Delta\rho/B^2)=\rho_{11,33}$.

На рис. 1 представлены температурные зависимости удельного сопротивления (рис. 1, а) и относительного магнетосопротивления (рис. 1, б) пленок висмута различной толщины с содержанием теллура от 0,005 ат.% до 0,15 ат.%. Температурная зависимость удельного сопротивления пленок Bi (0,005 ат.% Te) имеет ярко выраженный полупроводниковый характер (рис. 1, а), свойственный тонким пленкам чистого висмута [7]. С увеличением концентрации теллура в висмуте до 0,15 ат.% наблюдается плавный переход температурной зависимости удельного сопротивления пленок к виду, характерному для металлов [3, 9]. Как видно из рисунка, при любом значении концентрации примеси теллура с уменьшением толщины пленки висмута ее удельное сопротивление при температуре 77 К увеличивается, причем, чем больше примеси теллура, тем меньше это различие. Наблюдаемый рост удельного сопротивления с уменьшением толщины пленок связан с проявлением классического размерного эффекта, обусловленного ограничением длины свободного пробега носителей заряда поверхностью пленки при уменьшении ее толщины [3, 7]. При увеличении концентрации легирующей примеси теллура в висмуте происходит существенное уменьшение длины свободного пробега электронов, в результате чего размерный эффект проявляется слабее при той же толщине пленки.

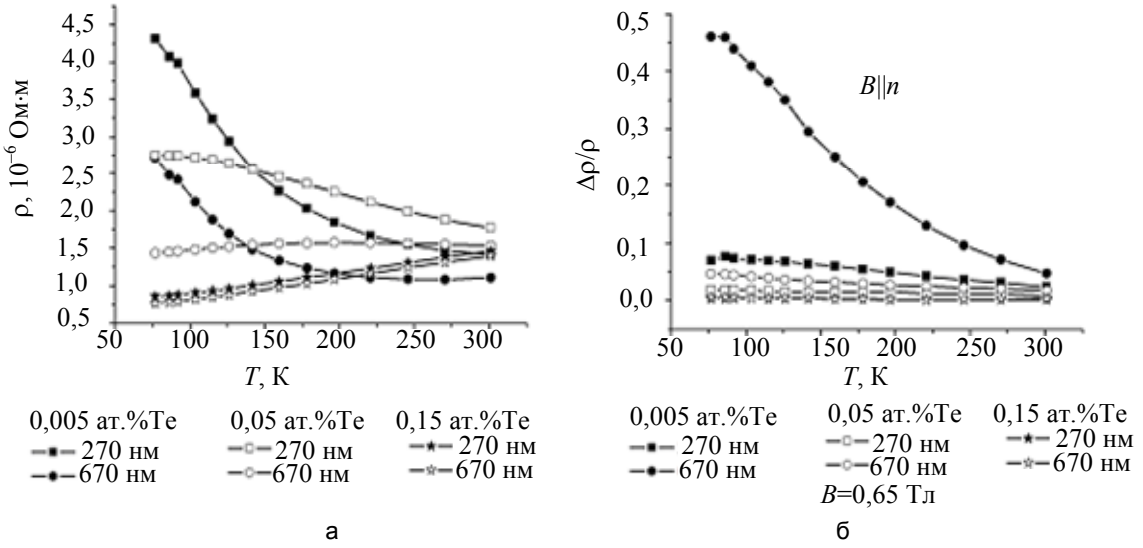


Рис. 1. Температурные зависимости характеристик пленок Bi (*x* ат.% Te) различной толщины: удельного сопротивления (а); относительного магнетосопротивления в магнитном поле *B*, равном 0,65 Тл (б). *x* – содержание теллура в атомных процентах

Зависимость от температуры относительного магнетосопротивления $\Delta\rho/\rho$ пленок висмута, легированного теллуром, в количестве 0,005 ат.% Te, 0,05 ат.% Te и 0,15 ат.% Te в магнитном поле *B*, равном 0,65 Тл, приведена на рис. 1, б. При понижении температуры относительное магнетосопротивление возрастает при всех уровнях легирования, но величина относительного магнетосопротивления значительно убывает с увеличением степени легирования. Магнетосопротивление пленок висмута, легированного теллуром, убывает и с уменьшением толщины, причем, чем тоньше пленка, тем слабее магнетосопротивление зависит от температуры [9]. В области температур ниже 120 К для всех представленных пленок наблюдается переход к насыщению. При увеличении толщины пленок, независимо от степени легирования, переход к насыщению смещается в область более низких температур.

Значительное влияние оказывает степень легирования и толщина пленки на коэффициент Холла [3]. На рис. 2 представлена температурная зависимость коэффициента Холла (*R*) для пленок

Bi (0,005 ат.% Te), Bi (0,05 ат.% Te) и Bi (0,15 ат.% Te) толщиной 270 нм и 670 нм в магнитном поле *B*, равном 0,65 Тл. Коэффициент Холла для пленок Bi (0,005 ат.% Te) при комнатной температуре имеет положительное значение. При понижении температуры происходит плавное уменьшение значения коэффициента Холла, а затем смена знака на отрицательный. В пленках Bi (0,05 ат.% Te) коэффициент Холла отрицателен во всем интервале температур, причем с уменьшением температуры его абсолютная величина увеличивается [3]. В области температур ниже температуры перехода к собственной проводимости (по данным для монокристаллов, в Bi (0,05 ат.% Te) переход к собственной проводимости происходит около 200 К) скорость изменения коэффициента Холла замедляется по сравнению с областью высоких температур. Это обусловлено постоянством концентрации носителей заряда в области низких температур. Увеличение в пленках висмута концентрации теллура до 0,15 ат.% приводит к практически не зависящей от температуры величине коэффициента Холла, что связано с постоянством концентрации носителей заряда в исследуемом интервале температур.

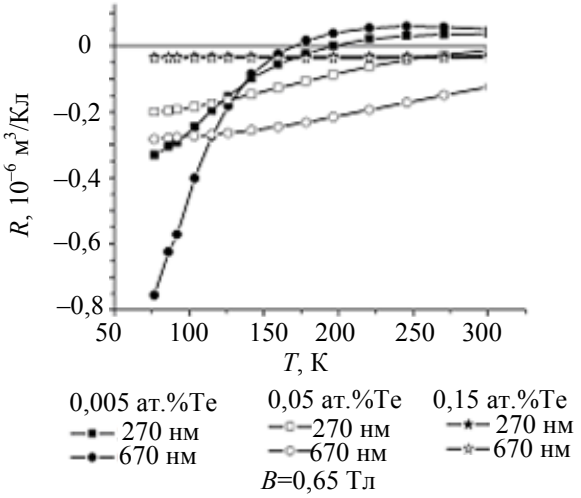


Рис. 2. Зависимость коэффициента Холла пленок Bi (*x* ат.%Te) различной толщины от температуры

Анализ концентрации и подвижности носителей заряда

Изменение гальваномагнитных коэффициентов пленок определяется температурной зависимостью концентрации и подвижности носителей заряда [2, 10]. Концентрация носителей заряда в пленках висмута, легированного теллуром, была рассчитана в рамках двух различных приближений. Первый вариант расчета проводился в рамках квадратичного закона дисперсии зоны проводимости и валентной зоны, который дает удовлетворительные результаты при расчете концентраций в монокристаллах висмута, легированного теллуром [13].

Выбор модели закона дисперсии обусловлен тем, что при увеличении уровня химического потенциала и концентрации носителей заряда данные по экспериментальному исследованию явлений переноса в кристаллах висмута, легированного теллуром, в области однозонности хорошо согласуются с квадратичной моделью [12, 13]. В работе [3] подтверждается правомерность использования данного подхода при анализе свойств пленок чистого висмута.

Концентрации электронов n_e и дырок n_p в пленках рассчитывались по формулам $n_e = C_L T^{3/2} F_{1/2}(\mu_e)$, $n_p = C_T T^{3/2} F_{1/2}(\mu_p)$, где C_L и C_T – коэффициенты, зависящие от эффективных масс электронов и дырок в направлениях главных осей симметрии поверхности Ферми в висмуте; $F_{1/2}(\mu_e)$, $F_{1/2}(\mu_p)$ – интегралы Ферми, зависящие от уровня химического потенциала μ электронов и дырок [10, 13]. Значения коэффициентов C_L и C_T были взяты из работы [13]. Суть расчета заключалась в следующем: варьируя значения уровня химического потенциала, определялось его значение, при котором концентрация электронов превышала концентрацию дырок на величину, равную концентрации донорных атомов с учетом их коэффициента отдачи. Далее, используя найденное значение уровня химического потенциала, рассчитывались значения концентрации электронов и дырок. Избыточная концентрация определялась по степени легирования:

$$n_e - n_p = \eta \cdot N_{\text{пр}} = \eta \cdot \frac{\rho_{\text{Bi}}}{M_{\text{Bi}}} \cdot N_a \nu_{\text{Te}}, \qquad (1)$$

где n_e, n_p – концентрация электронов и дырок соответственно; η – коэффициент отдачи (для теллура в висмуте η = 0,7 [10]), M_{Bi}, ρ_{Bi} – молярная масса и плотность висмута соответственно; N_a – постоянная Авогадро; ν_{Te} – процентное содержание атомов теллура в висмуте; $N_{\text{пр}}$ – значение концентрации введенной в образец примеси.

Второй вариант расчета основывался на экспериментальных результатах исследования гальваномагнитных свойств пленок висмута, легированного теллуром. Расчет производился в рамках однозонной модели, что справедливо для сильнолегированных пленок. Указанный метод был предложен Г.А. Ивановым [13] для обработки результатов, полученных при исследовании свойств монокристаллов висмута, легированного теллуром. Таким образом, пренебрегая вкладом дырок в явления переноса, из уравнений для гальваномагнитных коэффициентов, приведенных в работе [13], можно выразить концентрацию электронов n_e в виде

$$n_e = \frac{R}{\rho^2 \cdot \left(\frac{\Delta\rho_B}{\rho \cdot B^2} + \frac{R^2}{\rho^2} \right) \cdot e}, \qquad (2)$$

где R – постоянная Холла; ρ – удельное сопротивление; $\Delta\rho_B$ – изменение удельного сопротивления в магнитном поле; B – индукция магнитного поля; e – заряд электрона.

На рис. 3 приведены температурные зависимости концентрации носителей заряда в висмуте, легированном теллуром, полученные при реализации первого и второго вариантов расчета. На рис. 3, а, для сравнения представлены температурные зависимости концентрации электронов и дырок в Bi (0,005 ат.% Te) и чистом висмуте, рассчитанные первым методом. Для чистого висмута, так же как и для Bi (0,005 ат.% Te), наблюдается увеличение концентрации носителей заряда с повышением температуры, причем в Bi (0,005 ат.% Te) в интервале 77–150 К наблюдается почти постоянство концентрации. При увеличении степени легирования от 0,05 ат. % Te до 0,15 ат.% Te (рис. 3, б) концентрация электронов n_e перестает зависеть от температуры и становится равной концентрации примеси теллура с учетом коэффициента отдачи, что свидетельствует о переходе вещества в однозонное состояние. Концентрация дырок n_p с увеличением степени легирования заметно уменьшается во всем интервале температур.

Сравнение результатов двух вариантов расчета значений концентрации электронов, представленных на рис. 3, б, показывает хорошее согласие для пленок Bi (0,15 ат.% Te), что подтверждает правомерность использования для данного состава однозонного приближения [12, 13].

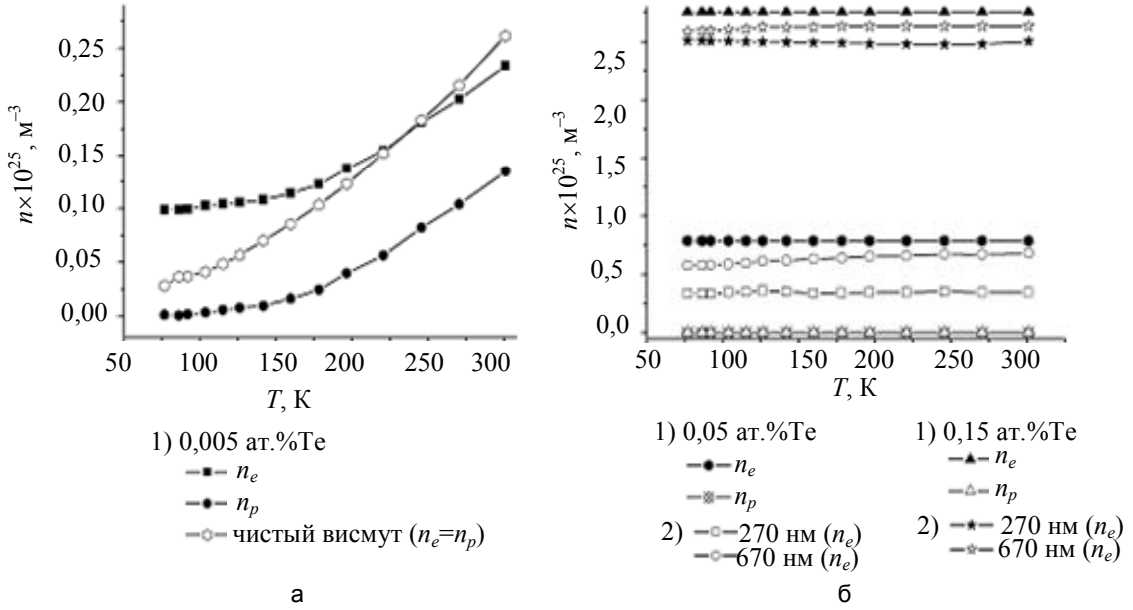


Рис. 3. Зависимости концентрации носителей заряда от температуры: в пленках Bi (0,005 ат.% Te) при реализации 1 варианта расчета (а); в пленках Bi (0,05 ат.% Te) и Bi (0,15 ат.% Te). Расчет выполнен в рамках двух приближений (б)

На основе экспериментальных результатов были рассчитаны подвижности свободных носителей заряда. Для расчета подвижности использовались уравнения для коэффициентов переноса в слабом магнитном поле. С учетом кристаллографической ориентации пленки они имеют вид [4]:

$$\sigma = e[\tfrac{1}{2} n_e (u_1^- + u_2^-) + n_p u^+]\,; \qquad (3)$$

$$R = \frac{e}{\sigma^2}[n_e B_n u_1^- u_2^- - n_p B_p (u^+)^2]\,; \qquad (4)$$

$$\frac{\Delta\rho_B}{B^2} = -\rho^2[e n_p C_p u^{+3} + \frac{1}{2} e n_e C_n u_1^- u_2^- (u_1^- + u_2^-)] - \sigma(R)^2\,, \qquad (5)$$

где σ – проводимость; R – постоянная Холла; n_e – концентрация электронов; n_p – концентрация дырок; $\Delta\rho_B$ – изменение удельного сопротивления в магнитном поле; B – индукция магнитного поля; $u^+ = u_1^+ = u_2^+$, u_1^-, u_2^- – подвижности дырок и электронов в тригональной плоскости; B_n, B_p – Холл-факторы и C_n, C_p – факторы магнетосопротивления, которые можно принять равными единице, ввиду того, что состояние носителей заряда в висмуте характеризуется большой степенью вырождения по сравнению с полупроводниками.

Для пленок Bi (0,15 ат.% Te), когда в явлениях переноса принимают участие только электроны L-экстремумов зоны проводимости из уравнений (3)–(5), можно определить компоненты подвижности L-электронов u_1^-, u_2^-:

$$u_1^- = \frac{\sigma - \sigma \cdot \sqrt{1 - en_e R}}{en_e}; \; u_2^- = \frac{\sigma + \sigma \cdot \sqrt{1 - en_e R}}{en_e}.$$

В пленках Bi (0,005 ат.% Te), Bi (0,05 ат.% Te) в приближении слабого магнитного поля для расчета усредненной подвижности электронов ($u^-_{\text{ср}}$) можно воспользоваться выражением $u^-_{\text{ср}} = \frac{1}{B}\sqrt{\frac{\Delta\rho}{\rho}}$.

На рис. 4 приведены результаты расчета температурной зависимости усредненной подвижности электронов ($u^-_{\text{ср}}$) для пленок Bi (0,005 ат.% Te), Bi (0,05 ат.% Te) и Bi (0,15 ат.% Te).

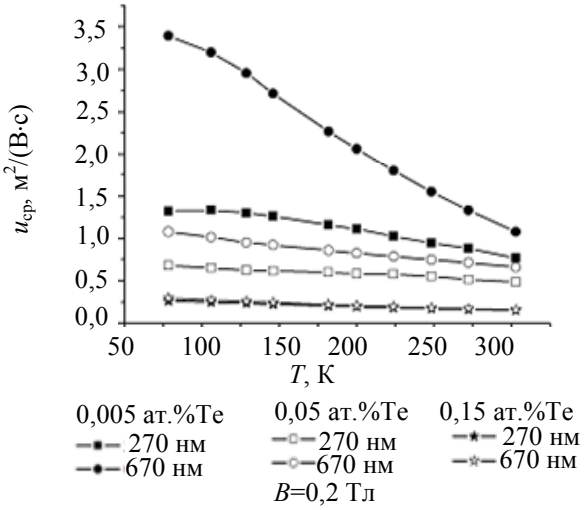


Рис. 4. Зависимость усредненной подвижности электронов пленок Bi (*x*. ат.% Te) от температуры в магнитном поле *B*, равном 0,2 Тл

Из рис. 4 следует, что при понижении температуры усредненная подвижность носителей заряда возрастает. Однако в пленках с концентрацией теллура 0,005 ат.%, наблюдается переход к насыщению, обусловленный, вероятно, рассеянием носителей заряда на несовершенствах кристаллической структуры образца (дефектах) и поверхности пленки. Уменьшение подвижности носителей заряда при уменьшении толщины пленок обусловлено проявлением классического размерного эффекта. Из рис. 4 видно, что в пленках легированного висмута, также как и в объемных монокристаллах висмута, легированного теллуром [12, 14], увеличение концентрации донорной примеси теллура приводит к значительному уменьшению подвижности электронов в исследуемом интервале температур [3]. Данная зависимость наблюдалась и другими исследователями в работе [2].

Заключение

В результате проведенных в работе исследований в пленочных образцах висмута, легированного теллуром, в температурном интервале 77–300 К установлено, что уменьшение толщины пленок висмута, легированного теллуром, приводит к переходу температурной зависимости удельного сопротивления от металлической к характерной для полупроводников, а также к уменьшению магнетосопротивления, что является проявлением классического размерного эффекта.

Увеличение концентрации легирующей примеси теллура в пленках висмута приводит к переходу температурной зависимости удельного сопротивления от полупроводниковой к металлической, уменьшению магнетосопротивления во всем температурном интервале, уменьшению коэффициента Холла по абсолютной величине и отсутствию зависимости его от температуры в пленках Bi (0,15 ат.% Te).

Проведенный в работе расчет концентраций носителей заряда в пленках висмута, легированного теллуром, в рамках двухзонной квадратичной модели и обработка экспериментальных данных в рамках

однозонного приближения показали хорошее совпадение результатов этих расчетов для сильнолегированных составов. В пленках висмута, легированного теллуром, при увеличении степени легирования наблюдается смена знака коэффициента Холла с положительного на отрицательный, что говорит об определяющем вкладе электронов в гальваномагнитные явления.

Полученные в настоящей работе результаты могут быть использованы при создании низкоразмерных структур на основе висмута с контролируемой концентрацией электронов, а также при изучении закономерностей изменения свойств вещества при переходе от массивного монокристалла к низкоразмерным и наноструктурам на основе полуметаллов.

Работа выполнена при финансовой поддержке Минобрнауки России в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы» (соглашение № 14.B37.21.0891).

Литература

1. Гусев А.И. Наноматериалы, наноструктуры, нанотехнологии. – М.: ФИЗМАТЛИТ, 2007. – 416 с.
2. Орлова Д.С., Рогачева Е.И. Гальваномагнитные свойства тонких пленок висмута, легированного теллуром // ІМФ: Наносистемы, наноматериалы, нанотехнологии. – Харьков. – 2009. – Т. 7. – № 2. – С. 487–493.
3. Комаров В.А., Матвеев Д.Ю., Худякова И.И., Крушельницкий А.Н. Явления переноса в тонких пленках висмута, легированного теллуром // Известия РГПУ им. А.И. Герцена. Естественные и точные науки: Научный журнал. – СПб, 2012. – № 147. – С. 50–63.
4. Грабов В.М., Комаров В.А., Демидов Е.В. Гальваномагнитные и термоэлектрические явления в тонких пленках висмута и сплавов висмут-сурьма. – СПб: Изд-во РГПУ им. А.И. Герцена, 2011. – 124 с.
5. Cho S., Kim Y., Freeman A.J., Wong G.K.L., Ketterson J.B., Olafsen L.J., Vurgaftman I., Meyer J.R. and Hoffman C.A. Large magnetoresistance in postannealed Bi thin films // Applied Physics Letters. – 2001. – V. 79. – № 22. – P. 3651–3653.
6. Yang F.Y., Liu Kai., Chien C.L. and Searson P.C. Large magnetoresistance and finite-size effects in electrodeposited single-crystal Bi thin films // Physical Review Letters. – 1999. – V. 82. – № 16. – P. 3328–3331.
7. Комаров В.А., Климантов М.М., Логунцова М.М., Пылина С.Н., Демидов Е.В. Кинетические явления и структура пленок висмута // Известия РГПУ им. А.И. Герцена. Физика: Научный журнал. – СПб, 2006. – № 6 (15). – С. 131–143.
8. Грабов В.М., Демидов Е.В., Комаров В.А. Атомно-силовая микроскопия пленок висмута // ФТТ. – 2008. – Т. 50. – С. 1312–1316.
9. Матвеев Д.Ю., Христич Е.Е. Электрические свойства пленок висмута и висмута, легированного теллуром // Сборник трудов VI Всероссийской конференции молодых ученых. – СПб: СПбГУ ИТМО, 2009. – Вып. 3. – С.75–81.
10. Грабов В.М., Комаров В.А., Худякова И.И., Яковлева Т.А. Физика полуметаллов и низкоразмерных структур на их основе. – СПб: Изд-во РГПУ им. А.И. Герцена, 2010. – 254 с.
11. Минайчев В.Е. Нанесение пленок в вакууме // Технология полупроводниковых приборов и изделий микроэлектроники. – Кн. 6. – М.: Высшая школа, 1989. – 111 с.
12. Иванов Г.А., Грабов В.М. Физические свойства кристаллов типа висмута // ФТП. – 1995. – Т. 29. – № 5, 6. – С. 1040–1050.
13. Грабов В.М., Лужковский А.В., Яковлева Т.А. Температурная зависимость энергетических параметров зон кристаллов висмута и сплавов Bij-x Sbx (0 < x < 0,22) в интервале температур 77–300 К // Доклады VIII Межгосударственного семинара «Термоэлектрики и их применения». – СПб: ФТИ им. А.Ф. Иоффе РАН, 2000. – С. 96–101.
14. Грабов В.М., Парахин А.С., Багулин Л.С., Урюпин О.Н. Общая закономерность, определяющая величину, температурную и концентрационную зависимость удельного сопротивления кристаллов типа висмута // Известия РГПУ им. А.И. Герцена. Естественные и точные науки: Научный журнал. – СПб, 2006. – № 6 (15). – С. 86–100.

Комаров Владимир Алексеевич – Российский государственный педагогический университет им. А.И. Герцена, кандидат физ.-мат. наук, доцент, va_komar@mail.ru

Матвеев Даниил Юрьевич – Российский государственный педагогический университет им. А.И. Герцена, аспирант, danila200586@mail.ru

Демидов Евгений Владимирович – Российский государственный педагогический университет им. А.И. Герцена, кандидат физ.-мат. наук, ст. научный сотрудник, demidov_evg@mail.ru

Крушельницкий Артемий Николаевич – Российский государственный педагогический университет им. А.И. Герцена, студент, ak.spb.ru@gmail.com

УДК 621.315.592; 621.382002

ВЛИЯНИЕ ПРОЦЕССА ЛАЗЕРНОГО МИКРОСТРУКТУРИРОВАНИЯ НА ЭЛЕКТРОФИЗИЧЕСКИЕ ПАРАМЕТРЫ СИСТЕМЫ SiO_2/Si

А.М. Скворцов, Р.А. Халецкий, Хуинь Конг Ту

Исследованы электрофизические параметры системы SiO_2/Si после облучения импульсным иттербиевым волоконным лазером типа ИЛИ-1-50. Показано, что при воздействии лазерного излучения в системе кремний–окисел происходит образование дефектов, ответственных за появление различного рода зарядов. Обнаружено, что лазерно-индуцированные дефекты появляются и на участках, удаленных от зоны облучения. Установлено, что скорость генерации дефектов на границе раздела кремний–окисел зависит от их энергетического расположения в запрещенной зоне кремния и достигает максимального значения вблизи уровня Ферми.

Ключевые слова: лазерное облучение, дислокации, линии скольжения,·микроструктурирование, вольт-фарадные характеристики.

Введение

В настоящее время лазерное излучение широко используется при исследовании различных свойств материалов в качестве инструмента, позволяющего гибко и точно варьировать режимы облучения [1–3]. Особое место при этом занимают исследования по влиянию лазерного излучения на свойства материалов, использующихся в микроэлектронной технологии. Одним из основных материалов, играющих важную роль в микроэлектронике, является тонкая пленка двуокиси кремния, выращенная методом термического окисления кремниевой подложки.

При исследовании влияния лазерного излучения на систему кремний–двуокись кремния (SiO_2/Si) обнаруживаются новые явления, стимулирующие повышенный интерес к этой системе. Авторами работы [4], например, отмечается, что при лазерном воздействии происходит изменение цвета и диэлектрической проницаемости пленки окисла. В наших работах [5, 6] сообщалось о том, что при облучении системы SiO_2/Si в окисле индуцируются новые заряды. Обнаружен эффект дальнодействия, при котором изменения электрофизических свойств системы кремний–окисел появляются в областях подложки, удаленных от зоны непосредственного воздействия лазерного луча. Однако до сих пор процессы, протекающие в структуре системы кремний–окисел при воздействии лазерного излучения, остаются плохо изученными. Это тем более актуально в связи с тем, что разработаны и выпускаются лазеры с новыми эксплуатационными характеристиками (большими энергиями и частотой следования импульсов). Целью настоящей работы является изучение влияния на электрофизические параметры системы SiO_2/Si облучения импульсным иттербиевым волоконным лазером типа ИЛИ-1-50, приводящего к образованию сеток линий скольжения.

Методика эксперимента

Экспериментальная структура SiO_2/Si изготавливалась высокотемпературным окислением кремниевой пластины электронной проводимости с удельным сопротивлением 4,5 Ом·см (КЭФ–4,5, $N_d = 9{,}3\cdot10^{14}$ см$^{-3}$) и кристаллографической ориентацией (100). Окисел толщиной 150 нм выращивался при температуре 1000ºС в среде влажного кислорода. Размер исследуемых чипов составлял 15×15 мм. Отсутствие повреждений и загрязнений поверхности исходных образцов проверялось при помощи оптического микроскопа. Образцы подвергались воздействию импульсного иттербиевого волоконного лазера с длиной волны излучения 1062 нм с частотой следования импульсов 50 кГц, длительностью импульсов 120 нс и энергией 1 мДж. Применена схема облучения пластины в сходящихся лучах (рис. 1). При этом образец располагался выше плоскости фокуса на $d = 44$ мм с целью уменьшения плотности мощности излучения, что, в свою очередь, обеспечивает лучший контроль над процессом микроструктурирования. Диаметр облученной области составлял около 700 мкм, плотность мощности излучения – $1{,}3\cdot10^4$ Вт/см^2.

Выбранные схема облучения и режимы исключают появление даже мельчайших точек плавления кремния и признаков разрушения пленки двуокиси кремния. Температура поверхности пластины и ее распределение по площади контролировались в реальном времени с помощью инфракрасной камеры FLIR серии SC7000 (тепловизора).

На рис. 2 приведена микрофотография центральной части облученной области (рис. 2, а) и экспериментальный график изменения во времени температуры области облучения (рис. 2, б). После облучения наблюдается четко выраженная сетка линий скольжения, образованная пересечением ортогональных полос линий скольжения, характерных для монокристаллов кремния, ориентированных в кристаллографических плоскостях {100}. При длительности облучения 10 с суммарная доза составила 56,4 Дж/см^2.

Исследование характера изменения электрофизических свойств изучаемой системы кремний–-окисел, обусловленных воздействием лазерного облучения, проводилось методом высокочастотных ($\omega = 1$ МГц) вольт-фарадных характеристик (ВФХ) структуры металл–окисел–полупроводник (МОП). Измерения производились на образцах, подвергнутых лазерному облучению различной длительности. В качестве металлического электрода, контактирующего с окисной пленкой, использовался эвтектический

сплав In–Ga (температура плавления $T_{пл}$ = 15,7ºС). На рис. 3 показаны полученные в ходе эксперимента типичные, нормированные по максимальной емкости, ВФХ структур, полученных после лазерного облучения различной длительности.

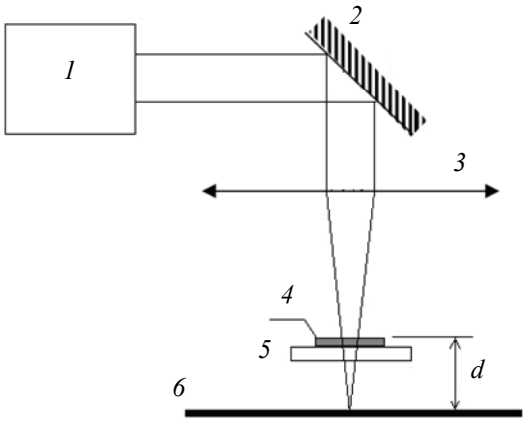


Рис. 1. Схема расположения структуры SiO_2/Si в сходящихся лучах лазера: 1 – лазер; 2 – поворотное зеркало; 3 – объектив; 4 – образец; 5 – металлическая подложка; 6 – фокальная плоскость объектива; *d* – расстояние между фокальной плоскостью объектива и поверхностью образца

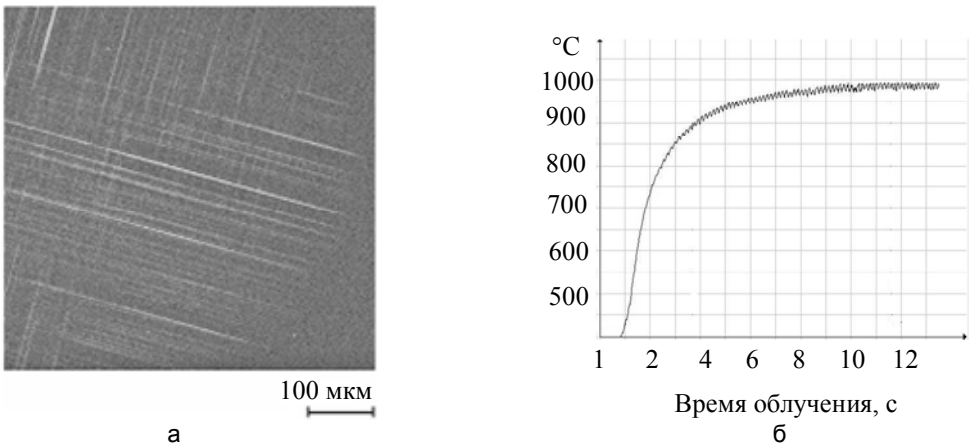


Рис. 2. Микрофотография центральной части облученной области (а) и временной график изменения температуры области облучения (б)

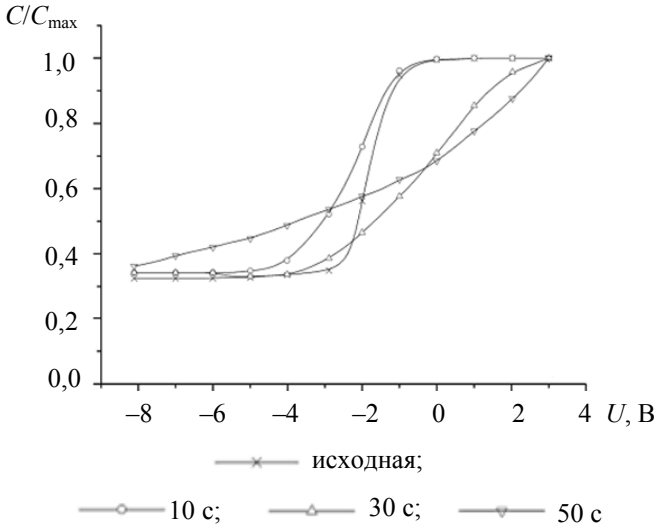


Рис. 3. ВФХ структур SiO_2/Si после лазерного облучения различной длительности

Обсуждение полученных результатов

Ранее А.Ф. Банишевым с сотрудниками в ряде работ [7–9] были рассмотрены вопросы генерации как обратимых, так и необратимых дефектов в монокристаллах кремния под действием лазерного облучения. В экспериментах использовался импульсный YAG:Nd лазер. Особое внимание уделялось собственным точечным дефектам (междоузельным атомам кремния и вакансиям), а также генерации дислокаций и периодических структур на поверхности кремния.

В настоящей работе, как было отмечено ранее, применен импульсный иттербиевый волоконный лазер с той же длиной волны (1062 нм), что и в вышеперечисленных работах А.Ф. Банишева. Однако авторами облучение производилось при более высоких частоте следования импульсов (50 кГц) и энергии импульса (1 мДж).

Следует отметить особенности системы SiO_2/Si. Контактирующие структурные фазы имеют общий уровень Ферми, что приводит к соответствующему перераспределению зарядов между ними. Одновременно разная структура кремния и диоксида кремния обуславливают наличие упруго-напряженного состояния системы SiO_2/Si, причем на поверхности кремния действуют максимальные касательные, растягивающие кристаллическую решетку напряжения.

При лазерном облучении системы SiO_2/Si фотоны беспрепятственно достигают поверхности кремния, так как пленка SiO_2 прозрачна для излучения этой длины волны. Здесь часть энергии квантов расходуется на разогрев кристаллической решетки, часть – на образование собственных точечных дефектов и дислокаций. В результате действия многих факторов – исходного напряженного состояния системы, повышения температуры в облученной области, ускоренной генерации структурных дефектов, размножения и скольжения дислокаций – происходит пластическая деформация, реализуемая в виде сетки линий скольжения (рис. 2, а). Более подробные сведения о микроструктурировании системы SiO_2/Si, при котором возникают сетки линий скольжения, приведены в работе [10].

В результате структурных изменений, связанных с образованием на облученной поверхности кремния сетки линий скольжения, происходит образование дефектных центров в приграничном к кремнию слое окисла, ответственных за встроенный заряд в пленке SiO_2. Из рис. 3 видно, что в процессе облучения происходит модификация ВФХ, которая свидетельствует об изменении электрофизических свойств оксида кремния и границы раздела кремний-окисел.

Смещение ВФХ по оси напряжений на уровне емкости C/C_{max}≈0,45, соответствующее режиму пересечения уровнем Ферми середины запрещенной зоны кремния, пропорционально изменению плотности встроенного в окисле заряда ΔQ_f. На рис. 4 показан график изменения плотности встроенного заряда в зависимости от продолжительности облучения.

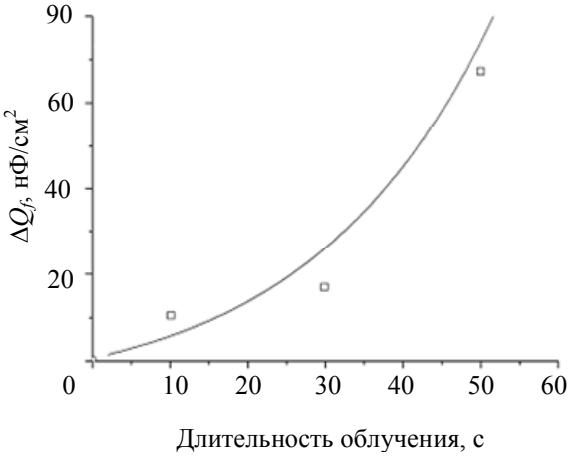


Рис. 4. Изменение плотности встроенного заряда в окисле в зависимости от длительности лазерного облучения

При длительности лазерного воздействия до 30 с наблюдаются незначительные изменения ΔQ_f. Однако при дальнейшем увеличении длительности облучения происходит резкое увеличение плотности встроенного положительного заряда. Это свидетельствует об интенсивной генерации структурных дефектов в слое окисла. При этом с высокой долей вероятности можно утверждать, что эти дефекты индуцируются в переходной области окисла, прилежащей к границе раздела и являющейся наиболее напряженной в системе SiO_2/Si.

Изменение наклона ВФХ в области напряжений, соответствующих переходу МОП-структуры из режима обогащения в режим инверсии (рис. 3), свидетельствует об изменении плотности заряда, пропорциональном изменению плотности поверхностных состояний ΔN_{SS} на границе раздела кремний–окисел. Определение плотности поверхностных состояний, генерируемых лазерным излучением, произ-

водилось по методу Термана, суть которого заключается в следующем. На основе предварительно рассчитанной теоретической ВФХ определяется сдвиг по оси напряжений $\Delta U(C)$ экспериментальной кривой относительно теоретической при различных значениях емкости МОП-структуры. Так как емкость МОП-структуры при высокой частоте измерения определяется только значением поверхностного потенциала кремниевой подложки Ψ_S, то на основе зависимости $\Delta U(C)$ формируется функция $\Delta U(\Psi_S)$. Далее графическим дифференцированием кривой $\Delta U(\Psi_S)$ рассчитывается ΔN_{SS} с использованием следующего выражения:

$$\Delta N_{SS} = \frac{C_0}{q}\frac{d(\Delta U)}{d\Psi_S},$$

где q – заряд электрона; C_0 – удельная емкость окисла.

На рис. 5 приведены графики, показывающие изменение плотности поверхностных состояний ΔN_{SS} в областях энергий, соответствующих запрещенной зоне кремния, при лазерном облучении различной длительности. Характер изменения ΔN_{SS} различается для разных энергетических диапазонов. Наибольшие изменения плотности поверхностных состояний наблюдаются в верхней половине запрещенной зоны вблизи уровня Ферми. В этом диапазоне энергий при увеличении времени облучения характер изменения ΔN_{SS} коррелирует с характером изменения плотности встроенного заряда (рис. 4), т.е. происходит резкое увеличение скорости генерации поверхностных состояний на границе раздела при продолжительности облучения более 30 с.

В областях, удаленных от зоны облучения, также наблюдались изменения ВФХ. Однако эти изменения были не такими выраженными по сравнению с изменениями ВФХ в зоне облучения. Таким образом, можно утверждать, что дефекты, ответственные за появление различного рода зарядов при лазерном воздействии, индуцируются и за пределами облученных областей, но их максимальная концентрация локализована вблизи линий скольжения.

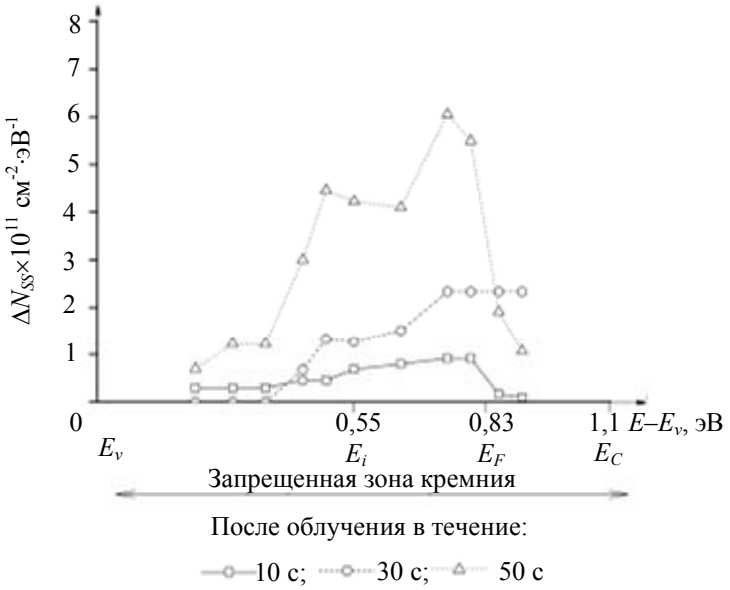


Рис. 5. Энергетическое распределение плотности поверхностных состояний в зависимости от длительности лазерного облучения

Заключение

Таким образом, при воздействии лазерного излучения в системе кремний–двуокись кремния происходят структурные изменения, которые обусловливают изменение электрофизических свойств системы SiO_2/Si. Изменения электрофизических свойств окисла и границы раздела в большей степени наблюдаются в зоне непосредственного воздействия лазерного луча на кремниевую подложку. Здесь происходят максимальные изменения структуры, проявляемые в виде сетки линий скольжения. Также авторами обнаружено появление лазерно-индуцированных дефектов и на участках, удаленных от зоны облучения.

Необходимо отметить, что при воздействии лазерного излучения наблюдается нелинейный рост плотности дефектов при увеличении дозы облучения. Причем было обнаружено, что при превышении определенной дозы происходит резкое увеличение скорости генерации лазерно-индуцированных дефектов. Для дефектов, расположенных на границе раздела кремний-двуокись кремния, скорость генерации определяется их энергетическим положением в запрещенной зоне кремния и максимальна вблизи уровня Ферми легированного полупроводника.

Литература

1. Вейко В.П., Дышловенко С.С., Скворцов А.М. Лазерное микроструктурирование поверхности кремния // Научно-технический сборник «Диагностика и функциональный контроль качества оптических материалов». Ч. 2. – СПб: СПбГУ ИТМО, 2004. – С. 138–153.
2. Zavestovskaya I.N., Eliseev P.G., Krokhin O.N. Nonlinear absorption mechanisms in ablation of transparent materials by high-intensity and ultrashort laser pulses // Applied Surface Science. – 2005. – V. 248. – P. 313–315.
3. Zavestovskaya I.N., Eliseev P.G., Krokhin O.N., Menkova N.A. Analysis of the nonlinear absorption mechanisms in ablation of transparent materials by high-intensity and ultrashort laser pulses // Appl. Phys. A. – 2008. – V. 92. – P. 903–906.
4. Medvid A., Onufrijevs P., Mellikov E., Kropman D., Muktepavela F., Bakradze G. Low-K factor of SiO_2 layer on Si irradiated by YAG:Nd laser // Journal of Non-Crystalline Solids. – 2007. – V. 353. – Is. 5–7. – P. 703–707.
5. Skvortsov A.M., Veiko V.P., Sokolov V.I., Pham Qung Tung, Khaletsky R.A. Laser modification of thermal oxide films on silicon // International conference «Fundamentals of laser assisted micro- and nanotechnologies» (FLAMN-10): Abstracts. – 2010. – P. 103.
6. Khaletsky R.A., Zamoraynskaya M.V., Kolesnikova E.V., Skvortsov A.M., Sokolov V.I., Pham Qung Tung, Veiko V.P. «Long-range action» effect under laser irradiation of SiO_2–Si system // International conference «Fundamentals of laser assisted micro- and nanotechnologies» (FLAMN-10): Abstracts. – 2010. – P. 105.
7. Банишев А.Ф., Володин Б.Л., Емельянов В.И., Мерзляков К.С. Образование периодических структур дислокаций при лазерном воздействии на поверхность полупроводников // ФТП. – 1990. – Т. 32. – Вып. 9. – С. 25–29.
8. Банишев А.Ф., Новикова Л.В. Образование обратимых и необратимых структурных дефектов на поверхности кремния под действием лазерного импульса // Физика и химия обработки материалов. – 1992. – № 4. – С. 55–59.
9. Banishev A.F., Golubev V.S., Kolmnev Y.U. Generation and accumulation of dislocations on the silicon surface under the action of pulse-periodic emission from a YAG:Nd laser // Technical Physics. – 2001. – V. 46. – № 8. – P. 962–967.
10. Скворцов А.М., Вейко В.П., Хуинь Конг Ту. Применение импульсного волоконного лазера для микроструктурирования системы SiO_2/Si // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 5 (81). – С. 128–133.

Скворцов Альберт Матвеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, a-skvortsov@yandex.ru

Халецкий Роман Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, доцент, halecky@yandex.ru

Хуинь Конг Ту – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, Picochip912@yahoo.com

6 СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

УДК.681.324

ПРИМЕНЕНИЕ СИМПЛЕКС-МЕТОДА И МЕТОДА ИСКУССТВЕННОГО БАЗИСА ПРИ ПРОЕКТИРОВАНИИ БОРТОВОГО ПРИБОРНОГО ОБОРУДОВАНИЯ

М.С. Дейко, И.О. Жаринов

Рассматривается применение методов оптимизации проектных решений (симплекс-метод и метод искусственного базиса) в приборостроении. Приводятся примеры их использования при проектировании бортового приборного оборудования.

Ключевые слова: оптимизация, приборостроение, симплекс-метод, метод искусственного базиса.

Введение

Методы оптимизации применяются в различных областях науки и техники, в том числе в авиационном приборостроении. Оптимизация является одной из важнейших процедур при проектировании. Под оптимизацией будем понимать процесс получения наиболее удачного расположения органов управления проектируемого прибора относительно друг друга по заданному критерию. Критерий исходит из требований стандартов (ГОСТ, ОСТ и др.), технологичности изготовления изделия, требований технического задания.

Лицевая панель бортового приборного оборудования состоит из множества функциональных элементов, которые находятся в непосредственной близости друг от друга. Имеются элементы, с которыми взаимодействует человек (индикационный экран, функциональные кнопки и др.), такие элементы входят в состав информационно-управляющего поля кабины пилота. Они должны быть расположены рационально, должно обеспечиваться удобство сборки прибора на заводе-изготовителе и удобство его эксплуатации.

Для определения параметров конструирования органов управления лицевой панели информационно-управляющего поля кабины пилота летательного аппарата необходимо решить общую оптимизационную задачу проектирования $extr\left\{f\left(x_1, x_2 \ldots x_n\right)\right\}$, где $x_1, x_2 \ldots x_n$ – параметры проектируемого устройства; n – количество этих параметров. Общая задача проектирования решается в два этапа. На первом этапе определяются внешние параметры (габариты) устройства, на втором определяются внутренние параметры (расстояния между функциональными элементами устройства). Первый этап разрешается симплекс-методом, второй этап – методом искусственного базиса. Целевая функция проектирования первого этапа, описывающая внешний параметр проектируемого устройства, имеет вид

$$f = c_1x_1 + c_2x_2 + \ldots + c_nx_n + C_x \to \max,$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m, \end{cases} \tag{1}$$

$$x_j \ge 0,\ j = 1, 2, \ldots, n;\ b_i \ge 0,\ i = 1, 2, \ldots, m,$$

где $x_1, x_2 \ldots x_n$ – параметры расстояния между функциональными элементами проектируемого прибора в горизонтальной плоскости; C_x – сумма горизонтальных размеров всех функциональных элементов лицевой панели; $c_1, c_2 \ldots c_n$ – весовые коэффициенты целевой функции проектирования; $a_{11}, a_{12} \ldots a_{mn}$ – коэффициенты системы ограничений; $b_1, b_2 \ldots b_m$ – свободные члены системы ограничений; m – количество неравенств в системе. Параметры a, b, c выбираются с использованием экспертных оценок [1].

Целевая функция проектирования второго этапа имеет вид

$$f = c_1x_1 + c_2x_2 + \ldots + c_nx_n + c_{n+1}y_1 + c_{n+2}y_2 + c_{n+k}y_k + C_x + C_y \to \max,$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m, \end{cases} \begin{cases} a_{(m+1)1}y_1 + a_{(m+1)2}y_2 + \ldots + a_{(m+1)k}y_k \le b_{m+1}, \\ a_{(m+2)1}y_1 + a_{(m+2)2}y_2 + \ldots + a_{(m+2)k}y_k \le b_{m+2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{(m+t)1}y_1 + a_{(m+t)2}y_2 + \ldots + a_{(m+t)k}y_k \le b_{m+t}, \end{cases} \tag{2}$$

$$x_j \ge 0,\ j = 1, 2, \ldots, n;\ y_h \ge 0,\ h = 1, 2, \ldots, k;\ b_i \ge 0,\ i = 1, 2, \ldots, m + t,$$

где $y_1, y_2, \ldots, y_k$ – параметры расстояния между функциональными элементами проектируемого прибора в вертикальной плоскости; C_y – сумма вертикальных размеров всех функциональных элементов лицевой панели; k – количество этих параметров; t – количество уравнений в системе граничных условий для параметров $y_1, y_2, \ldots, y_k$. Задача проектирования устройства в этом случае определяется в виде поиска значений параметров $x_1, x_2 \ldots x_n$, $y_1, y_2, \ldots, y_k$, удовлетворяющих системам граничных условий (1) и (2), при которых целевые функции $f = (x_1, x_2 \ldots x_n)$ и $f = (y_1, y_2, \ldots, y_k)$ будут принимать максимальные значения.

Применение симплекс-метода для определения внешних параметров проектируемого устройства в горизонтальной плоскости

В основу симплекс-метода [2] положена идея последовательного улучшения получаемого проектного решения. Опишем методику получения значений параметров $x_1, x_2 \ldots x_n$ проектируемого устройства.

1. Система уравнений (1) приводится к канонической форме по правилу

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \Rightarrow \sum_{j=1}^{n} a_{ij} x_j + s_i = b_i, \; i = 1, 2, \ldots, m, \tag{3}$$

где s – дополнительно вводимая переменная. Результат имеет следующий вид:

$$f = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n + C_x \to \max,$$

$$\begin{cases} a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n + s_1 = b_1, \\ a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n + s_2 = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n + s_m = b_m, \end{cases} \tag{4}$$

$$x_j \ge 0, \; j = 1, 2, \ldots, n; \; s_i \ge 0, \; b_i \ge 0, \; i = 1, 2, \ldots, m.$$

2. Составляется симплекс-таблица. Форма симплекс-таблицы представлена в табл. 1.

i	Базис	x_1	x_2	...	x_n	s_1	s_2	...	s_m	b_i
1	s_1	z_{11}	z_{12}	...	z_{1n}	$z_{1(n+1)}$	$z_{1(n+2)}$	...	$z_{1(n+m)}$	$z_{1(n+m+1)}$
2	s_2	z_{21}	z_{22}	...	z_{2n}	$z_{2(n+1)}$	$z_{2(n+2)}$	...	$z_{2(n+m)}$	$z_{2(n+m+1)}$
...	...	...	...	...	...	...	...	...	...	...
m	s_m	z_{m1}	z_{m2}	...	z_{mn}	$z_{m(n+1)}$	$z_{m(n+2)}$	...	$z_{m(n+m)}$	$z_{m(n+m+1)}$
	c_j	$z_{(m+1)1}$	$z_{(m+1)2}$	...	$z_{(m+1)n}$	$z_{(m+1)(n+1)}$	$z_{(m+1)(n+2)}$	...	$z_{(m+1)(n+m)}$	$z_{(m+1)(n+m+1)}$

Таблица 1. Форма симплекс-таблицы

Начальная симплекс-таблица составляется по следующим правилам:

$$\begin{cases} z_{ij} = a_{ij}, \; i = 1, 2, \ldots, m, \; j = 1, 2, \ldots, n; \\ z_{ij} = 1, \; i = 1, 2, \ldots, m, \; j = n+1, n+2, \ldots, n+m, \text{ при } i + n = j; \\ z_{ij} = 0, \; i = 1, 2, \ldots, m, \; j = n+1, n+2, \ldots, n+m, \text{ при } i + n \ne j; \\ z_{ij} = b_i, \; i = 1, 2, \ldots, m, \; j = n+m+1; \\ z_{ij} = -c_j, \; i = m+1, \; j = 1, 2, \ldots, n; \\ z_{ij} = 0, \; i = m+1, \; j = n+1, n+2, \ldots, n+m; \\ z_{ij} = L, \; i = m+1, \; j = n+m+1, \end{cases} \tag{5}$$

где L – значение целевой функции при текущем базисном проектном решении. Заполненная начальная симплекс-таблица, составленная по правилам (5), имеет вид, представленный в табл. 2.

3. Условие оптимальности проектного решения определяется по критерию

$$\forall -c_j \ge 0, \; j = 1, 2, \ldots, n. \tag{6}$$

4. Разрешающий столбец определяется по правилу

$$c_{\max} = \max\left(\left|-c_1, -c_2, \ldots, -c_n\right|\right). \tag{7}$$

Значение (7) достигается при $j = r$, где r – номер разрешающего столбца.

5. Достаточное условие получения проектного решения при условии выполнения (6) и (7) имеет вид

$$\exists a_{ir} \ge 0, \; i = 1, 2, \ldots, m. \tag{8}$$

i	Базис	x_1	x_2	...	x_n	s_1	s_2	...	s_m	b_i
1	s_1	a_{11}	a_{12}	...	a_{1n}	1	0	...	0	b_1
2	s_2	a_{21}	a_{22}	...	a_{2n}	0	1	...	0	b_2
...	...	...	...	...	...	...	...	...	...	...
m	s_m	a_{m1}	a_{m2}	...	a_{mn}	0	0	...	1	b_m
	c_j	c_1	c_2	...	c_n	0	0	...	0	L

Таблица 2. Общий вид начальной заполненной симплекс-таблицы

6. Для генерации кортежа проектных решений определяется разрешающая строка по правилу

$$\min\left(b_i / a_{ir}\right),\ a_{ir} \ge 0,\ i = 1,2,\ldots,m\ . \qquad (9)$$

Минимум (9) достигается при $i = p$, где p – номер разрешающей строки.

7. Переход к новому базисному решению (пересчет элементов симплекс-таблицы) осуществляется по следующим формальным правилам:

1. для элементов разрешающей строки используются зависимости

$$a'_{pj} = \frac{a_{pj}}{a_{pr}},\ b'_p = \frac{b_p}{a_{pr}},\quad j = 1,2,\ldots,n+m\ , \qquad (10)$$

где a'_{pj}, b'_p – новые значения пересчитываемых элементов; a_{pj}, b_p – старые значения пересчитываемых элементов; a_{pr} – значение разрешающего элемента;

2. элементы разрешающего столбца (кроме разрешающего элемента) обнуляются:

$$a_{pr} = 0\ ,\ c_r = 0\ ,\ i = 1,2,\ldots,m\ \text{, кроме } i = p\ ; \qquad (11)$$

3. элементы, не принадлежащие разрешающему столбцу и строке, пересчитываются по формулам

$$a'_{ij} = a_{ij} - \frac{a_{ir}a_{pj}}{a_{pr}}\ ,\ b'_i = b_i - \frac{a_{ir}b_p}{a_{pr}}\ ,\ c'_j = c_j - \frac{a_{pj}c_r}{a_{pr}}\ ,$$

$$L' = L - \frac{c_r b_p}{a_{pr}}\ ,\ i = 1,2,\ldots,m\ ,\ j = 1,2,\ldots,n+m\ , \qquad (12)$$

где a'_{ij}, b'_i, c'_j, L' – новые значения пересчитываемых элементов; a_{ij}, b_i, c_j, L – старые значения пересчитываемых элементов.

Блок-схема алгоритма определения параметров $x_1, x_2 \ldots x_n$ симплекс-методом приведена на рис. 1, а.

Применение метода искусственного базиса для определения внутренних параметров проектируемого устройства

Методика определения параметров $y_1, y_2, \ldots, y_k$ с помощью метода искусственного базиса [3, 4] представлена ниже.

1. Система уравнений (2) приводится к расширенной канонической форме:

$$f = c_1x_1 + c_2x_2 + \ldots + c_nx_n + c_{n+1}y_1 + c_{n+2}y_2 + c_{n+k}y_k - Ms_1 - Ms_2 - \ldots$$
$$\ldots - Ms_m - Ms_{m+1} - Ms_{m+2} \ldots - Ms_{m+t} \to \max,$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n - s_1 + l_1 = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n - s_2 + l_2 = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n - s_m + l_m = b_m, \end{cases}$$

$$\begin{cases} a_{(m+1)1}y_1 + a_{(m+1)2}y_2 + \ldots + a_{(m+1)k}y_k - s_{m+1} + l_{m+1} = b_{m+1}, \\ a_{(m+2)1}y_1 + a_{(m+2)2}y_2 + \ldots + a_{(m+2)k}y_k - s_{m+2} + l_{m+2} = b_{m+2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{(m+t)1}y_1 + a_{(m+t)2}y_2 + \ldots + a_{(m+t)k}y_k - s_{m+t} + l_{m+t} = b_{m+\mathrm{t}}, \end{cases} \qquad (13)$$

$x_j \ge 0,\ j = 1,2,\ldots,n$; $y_h \ge 0,\ h = 1,2,\ldots,k$; $s_i \ge 0,\ b_i \ge 0,\ l_i \ge 0,\ i = 1,2,\ldots,m+t$;

по правилам:

$$\begin{cases} \sum_{j=1}^{n} a_{ij}x_j \ge b_i \Rightarrow \sum_{j=1}^{n} a_{ij}x_j - s_i + l_i = b_i, \ \ (i = 1,2,\ldots,m), \\ \sum_{h=1}^{k} a_{ih}y_j \ge b_i \Rightarrow \sum_{j=1}^{n} a_{ij}y_j - s_i + l_i = b_i, \ \ (i = m+1, m+2,\ldots,m+t), \\ \sum_{j=1}^{n} a_{ij}x_j = b_i \Rightarrow \sum_{j=1}^{n} a_{ij}x_j + l_i = b_i, \ \ (i = 1,2,\ldots,m), \\ \sum_{h=1}^{k} a_{ih}y_j = b_i \Rightarrow \sum_{j=1}^{n} a_{ij}y_j + l_i = b_i, \ \ (i = m+1, m+2,\ldots,m+t), \\ f = \sum_{j=1}^{n} c_j x_j + \sum_{j=n+1,h=1}^{n+k,k} c_j y_h + C_x + C_y \to \max \Rightarrow \\ f = \sum_{j=1}^{n} c_j x_j + \sum_{j=n+1,h=1}^{n+k,k} c_j y_h + C_x + C_y - \sum_{i=1}^{m+t} Ms_i \to \max, \end{cases} \tag{14}$$

где M>>1; l – вспомогательные переменные.

2. Составляется симплекс-таблица. Правила составления симплекс-таблицы аналогичны (5), за исключением дополнительно вводимой (m+2)-ой строки. Правила заполнения (m+2)-ой строки симплекс-таблицы имеют вид

$$\begin{cases} z_{ij} = d_j, \ i = m+t+2, \ j = 1,2,\ldots,n+m, \\ z_{ij} = G, \ i = m+t+2, \ j = n+m+1, \end{cases} \tag{15}$$

где $d_j = \sum_{v=1}^{m+t} a_{vj}$, $j = 1,2,\ldots,n+m$ – сумма коэффициентов системы ограничений, содержащих искусственные переменные; $G = -\sum_{v=1}^{m+t} b_v$ – сумма всех свободных членов системы ограничений, содержащих искусственные переменные, взятая с обратным знаком.

3. Условия оптимальности проектного решения определяются по критериям

$$\forall d_j \ge 0, \ \ j = 1,2,\ldots,n+m, \tag{16}$$

$$\forall -c_j \ge 0, \ j = 1,2,\ldots,n+m. \tag{17}$$

4. Переменная, вводимая в базис (разрешающий столбец симплекс-таблицы) определяется как

$$\max\left(\left|d_j\right|\right), \ j = 1,2,\ldots,n+m. \tag{18}$$

Значение (18) достигается при $j = r$, где r – номер разрешающего столбца.

5. Достаточное условие получения проектного решения при условии выполнения критериев (16) и (17) имеет вид

$$\exists a_{ir} \ge 0, \ \ i = 1,2,\ldots,m+t. \tag{19}$$

6. Для генерации кортежа проектных решений определяется переменная, исключаемая из базиса (разрешающая строка симплекс-таблицы):

$$\min\left(b_i / a_{ir}\right), \ a_{ir} \ge 0, \ i = 1,2,\ldots,m+t. \tag{20}$$

Минимум (20) достигается при $i = p$, где p – номер разрешающей строки.

7. Переменная, вводимая в базис (разрешающий столбец симплекс-таблицы), определяется как

$$\max\left(\left|-c_j\right|\right), \ j = 1,2,\ldots,n+m. \tag{21}$$

Значение (21) достигается при $j = r$, где r – номер разрешающего столбца.

8. Достаточное условие получения проектного решения при условии выполнения критериев (4) и (5) имеет вид

$$\exists a_{ir} \ge 0, \ \ i = 1,2,\ldots,m+t. \tag{22}$$

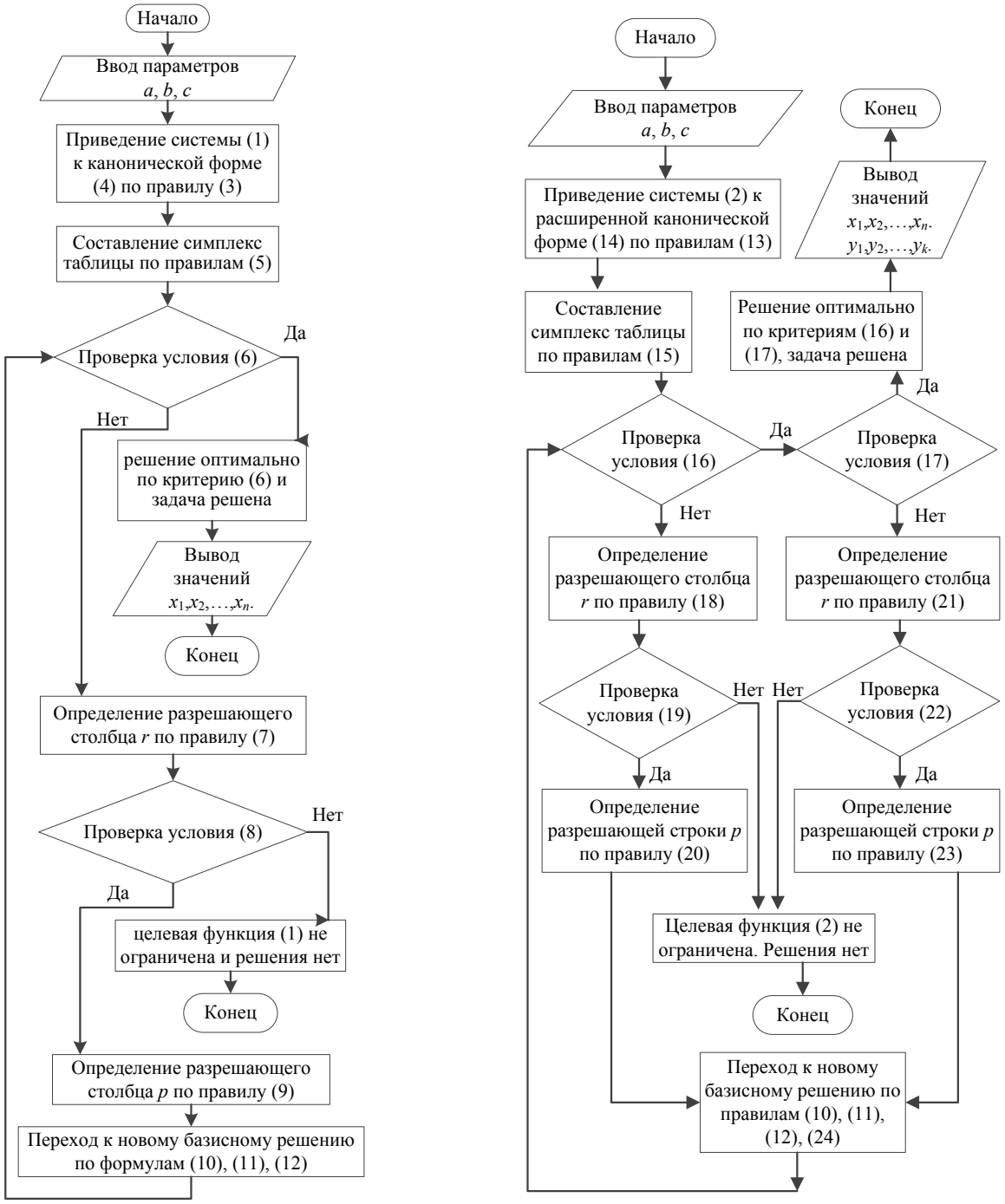
9. Для генерации кортежа проектных решений определяется переменная, исключаемая из базиса (разрешающая строка симплекс-таблицы):

$$\min\left(b_i / a_{ir}\right), \ a_{ir} \ge 0, \ i = 1,2,\ldots,m+t. \tag{23}$$

Минимум (23) достигается при $i = p$, где p – номер разрешающей строки.

$$d'_j = d_j - \frac{a_{pj}d_r}{a_{pr}}, \ G' = G - \frac{c_r b_p}{a_{pr}}, \ j = 1,2,\ldots,n+m, \tag{24}$$

где d'_j, G' – новые значения пересчитываемых элементов; d_j, G – старые значения пересчитываемых элементов. Блок-схема алгоритма определения параметров $y_1, y_2, \ldots, y_k$ проектируемого устройства методом искусственного базиса приведен на рис. 1, б.



а б

Рис. 1. Блок-схема алгоритмов для задачи проектирования бортового приборного оборудования: симплекс-метода (а); метода искусственного базиса (б)

10. Переход к новому базисному решению (пересчет элементов симплекс-таблицы) осуществляется последовательно по правилам (10)–(12) и по правилу:

$$d'_j = d_j - \frac{a_{pj} d_r}{a_{pr}}, \; G' = G - \frac{c_r b_p}{a_{pr}}, \; j = 1, 2, \ldots, n+m, \tag{24}$$

где d'_j, G' – новые значения пересчитываемых элементов; d_j, G – старые значения пересчитываемых элементов. Блок-схема алгоритма определения параметров $y_1, y_2, \ldots, y_k$ проектируемого устройства методом искусственного базиса приведен на рис. 1, б.

Заключение

Симплекс-метод и метод искусственного базиса были применены на практике в ФГУП «СПб ОКБ «Электроавтоматика» им П.А. Ефимова» в процессе проектирования бортового пульта управления и индикации, входящего в состав информационно-управляющего поля кабины пилота летательного аппарата. В результате решения системы уравнений (1) симплекс-методом и системы уравнений (2) методом искусственного базиса определены значения рабочих параметров проектирования прибора (рис. 2): x_1=5 мм; x_2=4 мм; y_1=8 мм; y_2=14 мм; y_3=12 мм.

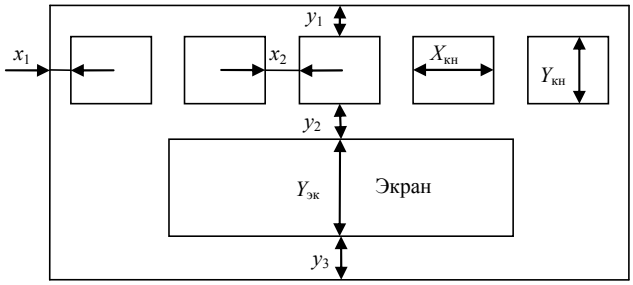


Рис. 2. Фрагмент лицевой панели пульта управления и индикации информационно-управляющего поля кабины пилота ($X_{кн}$ – горизонтальный размер кнопки; $Y_{кн}$ – вертикальный размер кнопки; $Y_{эк}$ – вертикальный размер экрана)

Проанализировав приведенные методы оптимизации и результаты расчетов на их основе, можно сказать, что они применяются при различных вариантах граничных условиях. Симплекс-метод применяется, когда в условиях проектной задачи в системе уравнений (1) содержится только отношение предпочтения вида «≤». Метод искусственного базиса используется, когда в условиях системы уравнений (2) присутствуют отношения предпочтения «≥» и «=» между левой и правой частями уравнений.

Литература

1. Гатчин Ю.А., Жаринов И.О. Основы проектирования вычислительных систем интегрированной модульной авионики: монография. – М.: Машиностроение, 2010. – 224 с.
2. Экономико-математические методы. Линейное программирование [Электронный ресурс]. – Режим доступа: http://emm.ostu.ru/lect/lect2_2.html, свободный. Яз. рус. (дата обращения 14.05.2012).
3. Акулич И.Л. Математическое программирование в примерах и задачах. – М.: Высшая школа, 1986. – 319 с.
4. Ермаков В.И. Общий курс высшей математики для экономистов. – М.: ИНФРА-М, 2007. – 656 с.

Дейко Михаил Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, магистрант, MadCat_MKII@mail.ru

Жаринов Игорь Олегович – ФГУП «СПб ОКБ «Электроавтоматика» имени П.А. Ефимова», доктор технических наук, доцент, руководитель учебно-научного центра, igor_rabota@pisem.net

УДК 004.78

АНАЛИЗ МЕТОДИК ПРОЕКТИРОВАНИЯ БЕСПРОВОДНЫХ СЕНСОРНО-АКТУАТОРНЫХ СЕТЕЙ

Д.А. Киров, А.А. Ожиганов

Рассмотрены существующие методики проектирования информационно-управляющих систем. С учетом специфики беспроводных сенсорно-актуаторных сетей выделены и проанализированы аспектный, актор-ориентированный и платформо-ориентированный подходы, в общем виде описаны процессы проектирования исследуемых систем на их основе. Предложен ряд направлений к улучшению данных методик применительно к беспроводным сенсорно-актуаторным сетям, произведен анализ перспектив.

Ключевые слова: беспроводные сенсорно-актуаторные сети, встроенные интеллектуальные системы, аспектное проектирование, акторное проектирование, платформо-ориентированное проектирование, WSN, WSAN.

Введение

Технология миниатюрных интеллектуальных сетей датчиков, известная как беспроводные сенсорные сети (БСС; Wireless Sensor Network – WSN), в настоящее время является уже довольно известной как

в академической, так и в производственной среде. Идеологически их можно назвать новым, стремительно развивающимся классом информационно-управляющих систем (ИУС). С учетом современных тенденций, связанных с внедрением автоматизации во всевозможные сферы человеческой деятельности, все более актуальным становится применение БСС в задачах управления. Задачей подобных сетей будет не только мониторинг и обработка получаемых данных, но и контроль объектов управления (ОУ), например, станков, роботов, и передача им необходимых команд. Подобному классу задач больше соответствует понятие беспроводных сенсорно-актуаторных сетей (БСАС; Wireless Sensor and Actuator Network – WSAN), активно используемое многими исследовательскими группами.

Говоря о подобном новом классе сетей, следует в первую очередь остановиться на проблемах их проектирования. Это достаточно сложный и трудоемкий процесс, поскольку разработчику необходимо продумывать организацию системы на всех уровнях, от аппаратного до прикладного, а также знать устройство ОУ. В связи с этим, адаптация существующих методик проектирования и разработка элементов автоматизированной системы проектирования (САПР) для БСАС являются актуальными научно-исследовательскими задачами и носят новый характер. Целью настоящей работы является анализ методик проектирования БСАС. На основе его результатов необходимо обозначить основные направления к доработке и улучшению этих методик, т.е. выполнить постановку задач для дальнейших исследований в изучаемой области.

Область применения технологии

БСАС находят применение в большом количестве разнообразных прикладных областей, основными из которых являются: автоматизация производственных процессов, мониторинг окружающей среды, «умные дома», охранные системы, медицина и т.д. В качестве конкретного примера можно привести производственную конвейерную линию, оборудованную набором роботов-манипуляторов, осуществляющих сборку изготавливаемых объектов (например, сверление, крепление и другие действия), как объект управления, и набор беспроводных датчиков, осуществляющих контроль состояния этих роботов (обнаружение чрезмерных вибраций и т.п.).

Наиболее перспективными для применения БСАС являются на сегодняшний день так называемые встроенные интеллектуальные системы (Physically Embedded Intelligent Systems – PEIS), описанные в [1, 2]. Они представляют собой набор механических устройств разной сложности (это может быть и обычная лампочка, и современный мобильный робот), коллективное взаимодействие которых осуществляется при помощи БСАС, выполняющей функции контроля и управления.

Проблемы проектирования БСАС

Процесс проектирования и разработки систем, подобных вышеописанной, обладает рядом особенностей, присущих не каждой ИУС. Основные из них перечислены ниже.

1. БСАС состоят, главным образом, из миниатюрных устройств, каждое из которых обладает автономным источником питания (батареей). Таким образом, для обеспечения надежности и длительности функционирования системы важнейшим требованием является минимизация энергопотребления каждым ее узлом. Наибольший расход батареи связан с вычислительными операциями на узле и, разумеется, с работой беспроводного трансивера (как в режиме ожидания, так и в режиме приема/передачи). Типовым решением описанной проблемы является применение алгоритмов и сетевых механизмов, сокращающих вычислительную нагрузку и обеспечивающих минимальное время нахождения трансивера в активном состоянии путем его периодического включения/выключения, а также промежуточной обработки данных на узле сети.
2. К БСАС часто предъявляют и другие требования, не свойственные ИУС. Хорошим примером здесь могут служить системы данного класса, решающие задачи позиционирования различных объектов. В последние годы построение таких систем на основе сенсорных сетей стало достаточно популярным решением, что описано, к примеру, в [3, 4]. Работоспособность применяемого метода позиционирования в различных условиях, точность, гранулярность измерений и другие требования играют здесь ключевую роль. Следует отметить, что процесс поиска оптимального метода на сегодняшний день недостаточно формализован, что и является основной трудностью для проектирования БСАС при наличии описанных требований.
3. Отдельно следует отметить сложность и трудоемкость процесса проектирования БСАС. Большое число узлов и гетерогенность оборудования являются для них типичными. В большинстве случаев от разработчика требуется детальное понимание предметной области и прикладной задачи, знание низкоуровневых функций конкретной аппаратуры, навыки в области теории автоматического управления и т.п. Основным способом снижения сложности разработки является использование промежуточного программного обеспечения (ПО), предоставляющего верхним уровням различные сервисы (отправки/приема данных, опроса, локализации и т.п.), скрывая при этом особенности их реализации на конкретном оборудовании. Примеры такого ПО могут быть найдены в [5] или [6].

Помимо указанных особенностей, к БСАС, разумеется, предъявляются те же требования, что и к большинству ИУС: работа в реальном масштабе времени (часто задаваемое ограничением WCET – Worst Case Execution Time), массогабаритные ограничения, высокая надежность, масштабируемость. Данный список можно расширить введением дополнительных требований, таких как шифрование передаваемых данных, динамическое реконфигурирование, самовосстановление и т.п.

Методики проектирования БСАС

Существующие методики проектирования ИУС (к которым относятся и БСАС) показаны на рис. 1.

Рис. 1. Методики проектирования ИУС

Методики во многом похожи на некоторых этапах, однако их основные принципы различны. Следует также отметить, что проектирование БСАС, как и некоторых других сложных распределенных систем, чаще всего принято отделять от проектирования и разработки компонент, из которых они состоят, т.е. аппаратуры. Это обусловлено следующими причинами. Во-первых, процесс проектирования и связывания между собой всех уровней таких систем (от верхнего прикладного до самого нижнего аппаратного) является невероятно трудоемкой задачей. Во-вторых, разработкой некоторых частей, рассматриваемых как часть сенсорно-актуаторной сети, таких как объекты наблюдения или управления, могут заниматься совершенно разные группы людей. Наконец, в-третьих, в настоящее время в области БСС все больше заметна тенденция к созданию так называемых аппаратно-программных платформ, т.е. многофункциональных контроллеров и приемопередатчиков, ориентированных на различные классы задач.

По указанным выше причинам из дальнейшего рассмотрения исключаются узконаправленные проприетарные решения, и делается предположение о том, что имеется возможность выбора различных вариантов готового оборудования, а также базового ПО (операционные системы, сетевые протоколы и т.п.). Следует отметить, что число существующих аппаратно-программных платформ и ПО достаточно велико, и это подразумевает достаточно серьезную проблему выбора. В рамках данного контекста наибольший интерес представляют аспектная, акторная и платформо-ориентированная методики.

Аспектное проектирование

В рамках аспектной методики предполагается детальное рассмотрение проектируемой системы в различных аспектах, каждый из которых учитывает определенный ряд взаимосвязанных свойств. Примерами аспектов могут служить конструктивный, надежностный, экономический, поведенческий и т.п. В процессе аспектного проектирования выделяются слабо связанные аспекты системы (ортогональные). Это существенно облегчает построение моделей системы, так как построение всеобъемлющей сложной модели заменяется возможностью описывать и исследовать систему с различных точек зрения на нее.

Типовой набор аспектов, рассматриваемых в ИУС, в случае БСАС может быть дополнен, главным образом, так называемым *энергосберегающим аспектом*, в рамках которого проектировщик стремится сформировать такое структурное и функциональное устройство системы, которое минимизировало бы энергопотребление ее узлов. Большое внимание здесь уделяется сетевым протоколам. К примеру, грамотный выбор размера временных слотов и их распределение в протоколах TDMA (Time Division Multiple Access – множественный доступ с разделением по времени) позволяют добиться существенной экономии энергии.

Аспектная методика проектирования во многом пересекается с другими методиками, что будет показано далее.

Акторное проектирование

Данный вид проектирования представляет собой использование формальных высокоуровневых моделей, представляющих вычислительный процесс как взаимодействие компонент, скрывающих внутреннюю реализацию и имеющих четко определенный внешний интерфейс. Эти компоненты называются

акторами. Для обработки сигналов акторы содержат порты, задающие направление (вход и (или) выход) и тип обрабатываемых данных. С этой точки зрения их можно сравнить с объектами ООП (объектно-ориентированного программирования). Процесс акторного проектирования во многом основан на принципе иерархического многоуровневого моделирования с постепенной детализацией моделей. В результате получается высокоуровневая платформа, позволяющая осуществить формальный анализ характеристик вычислительного процесса и его реализацию на конкретной аппаратуре с использованием конкретного языка программирования [7].

Начальный этап проектирования БСАС на основе данной методики представляет собой построение базовой акторной модели, описывающей взаимодействие крупных частей системы (можно назвать их «виртуальными объектами»). Это может быть, к примеру, набор датчиков и актуаторов, связанных с конкретным ОУ (количество и вид самих объектов, а также функции системы по отношению к ним, разумеется, определяются заранее). Уже на данном этапе задаются основные ограничения, осуществляется параметризация модели на основе известных характеристик оборудования, производится перебор и анализ разных вариантов высокоуровневой структурной организации системы.

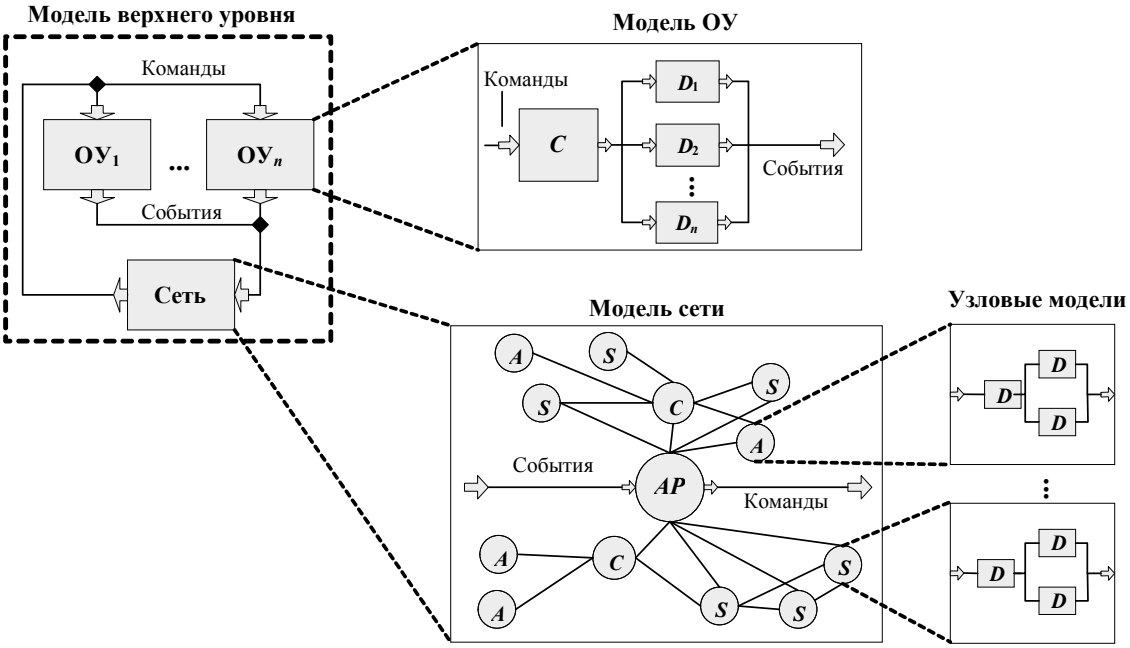


Рис. 2. Процесс построения акторных моделей

Уровни моделей	Виды моделей	Синтезируемые параметры
Системные	– начальная (общее устройство и связи)	– число ОУ – число сетей/подсетей
Сетевые	– каналы связи – топологическая модель (совокупность узлов)	– сетевая топология – взаимное расположение узлов, координаты – маршруты
Узловые	– сенсорного узла – актуатора – контроллера сети – комбинированные	– приоритеты задач – комплектация узлов (аппаратный состав)
Модели компонентов	– трансивера – микроконтроллера – другие элементы аппаратуры	– мощность сигнала – форматы команд, данных – набор сетевых протоколов
Модели отдельных механизмов	– прикладные алгоритмы – сетевые протоколы и механизмы	– активный интервал трансивера – длина и число временных слотов

Таблица. Основные уровни акторных моделей

Далее производится детализация базовой модели, т.е. введение дополнительных параметров, а также внутреннее описание элементов верхнего уровня. На каждом уровне (т.е. в базовой модели и во

вложенных) может использоваться наиболее удобная модель вычислений (дискретные события, конечный автомат, динамические потоки данных и т.п.) [7]. В принципе, детализация может продолжаться сколь угодно долго, однако сложность модели при этом стремительно растет. Этот факт следует выделить как основной недостаток акторного проектирования. В случае резкого увеличения сложности можно осуществлять выборочную детализацию с целью исследования какой-либо определенной характеристики, в чем проявляется связь данной методики с аспектным проектированием.

На рис. 2 в упрощенной форме описан процесс построения моделей БСАС в виде взаимосвязанных акторов. Видно, что из модели верхнего уровня, описывающей общее устройство системы (число ОУ, их связи друг с другом и с сетью), производится переход к сетевым моделям и моделям ОУ (на рисунке $ОУ_1$, ..., $ОУ_n$ – объекты управления; акторы D_1, ..., D_n – устройства ОУ и узлов сети; C – контроллеры сети или ОУ; S – сенсорные узлы; A – актуаторы; AP – модельная точка доступа в сеть).

В целом процесс функционирования системы можно представить как обнаружение и обработку *событий*, происходящих с ОУ, и генерацию соответствующих управляющих команд. Детализация моделей происходит и далее, и на каждом уровне осуществляется синтез различных параметров системы. Основные уровни моделей и примеры параметров представлены в таблице.

Многие компоненты и связи в приведенном выше описании изображены в упрощенном виде или вовсе не упоминаются, однако, с точки зрения авторов, это не мешает получению общего представления о процессе акторного проектирования БСАС.

Платформо-ориентированное проектирование

Данный вид проектирования подразумевает под собой последовательное создание ряда уровней представления системы, именуемых *платформами*, с последующим отображением их друг на друга и постепенным снижением абстрактности. Для этой методики характерно наличие нисходящей (top-down) и восходящей (bottom-up) фаз проектирования. Также часто выделяют финальную центральную фазу (meet in the middle), на которой происходит сопоставление абстрактного устройства системы, поступившего «сверху» (из пространства приложений), и более конкретной системной организации «снизу» (из архитектурного пространства).

Использование данного подхода для проектирования БСАС представляет собой создание нескольких платформ, каждая из которых скрывает особенности реализации платформы более низкого уровня. Начальным этапом нисходящей фазы является формирование набора сервисов, которые система будет предоставлять пользователю (например, сервис опроса, сервис управления и т.д.). На данном этапе система рассматривается как совокупность «виртуальных» узлов (это не физические сущности, а абстракции, описывающие наблюдаемые области и алгоритмы управления). Производится определение всех вариантов взаимодействия этих узлов (каждый из этих вариантов представляет собой сочетание запросов и команд) [8]. Это позволяет оценить некоторые характеристики системы (например, временные задержки передачи между конечными точками) и преобразовать их в требования, которым она гарантированно должна соответствовать.

На этапе восходящей фазы производится изучение доступного аппаратного обеспечения и выбор наиболее подходящего по ряду существенных критериев. Также на данном этапе должен быть специфицирован интерфейс, скрывающий особенности реализации аппаратных элементов. В основном это делается посредством установки и конфигурирования специализированной операционной системы (например, TinyOS – [9]).

При переходе к центральной фазе первым шагом является преобразование полученных оконечных задержек и энергозатрат в соответствующие задержки и энергозатраты при передаче данных между соседними узлами. Производится анализ и выбор протоколов, подходящих к применению в проектируемой системе (обычно они обладают уже известными характеристиками) и составляется некоторое число вариантов ее структурно-функциональной организации. Выбор оптимальной из них и вычисление соответствующих межузловых характеристик осуществляется при помощи формализованного математического аппарата (решение проблемы оптимизации с ограничениями). Обычно это делается при участии элементов САПР (пример подобного ПО описан в [8]). На основании вышеописанных действий, а также анализа предметной области описывается «виртуальная» архитектура системы, содержащая уже упоминавшийся набор сервисов, сетевую модель и ряд необходимых программных примитивов [10]. Завершающим этапом центральной фазы является отображение синтезированной «виртуальной» архитектуры на аппаратно-программную платформу и генерация кода программ, непосредственно устанавливаемых на узлы сети.

На рис. 3 в общем виде проиллюстрирован вышеописанный подход. Результатом нисходящей фазы проектирования является формирование платформы сервисов, результатом восходящей – *аппаратно*-программная платформа на базе выбранных контроллеров. Фазы сходятся в центральной области, где посредством анализа их результатов формируется оптимальный вариант организации системы, на основе которого производится описание «виртуальной» архитектуры системы с последующим преобразованием в реальную.

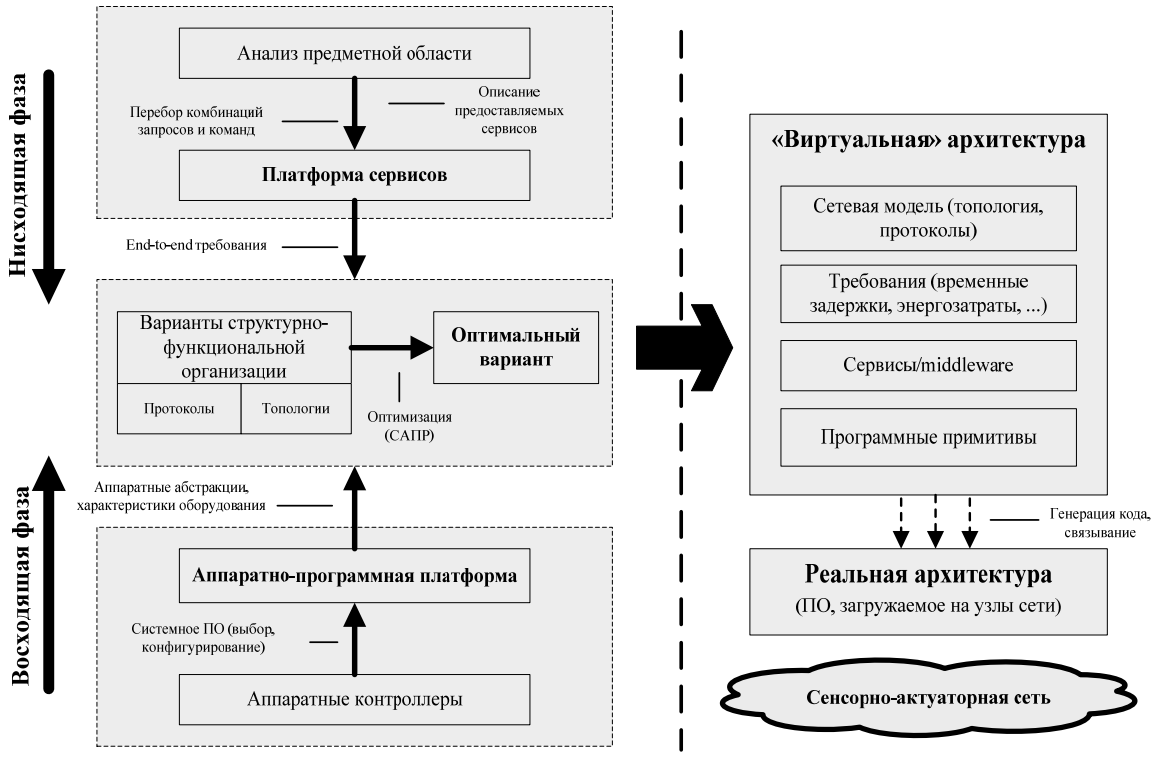


Рис. 3. Платформо-ориентированный подход к проектированию БСАС

Перспективы применения методик

В целом рассмотренные в данной работе методики проектирования ИУС в широком смысле являются вполне подходящими для БСАС. Однако, разумеется, каждая из них нуждается в пересмотре и доработке. Аспектный подход, к примеру, на первый взгляд вполне удобен, но, как было показано, требует рассмотрения вместе с типовыми дополнительных аспектов системы (таких как энергосберегающий), что существенно сказывается на сложности моделирования и «увязывания» полученных результатов друг с другом. Актуальным в рамках данного подхода является выявление аспектов, которые удобнее рассматривать совместно, а также поиск наиболее эффективных вариантов применения аспектной методики в сочетании с другими.

Безусловным достоинством акторного проектирования является визуальность моделей и удобство высокоуровневого (и не только) описания системы. Недостатком, разумеется, является резкое возрастание сложности акторных моделей с увеличением вложенности, что указывает на необходимость их разделения (связь с аспектной методикой). Также достаточно трудоемкой задачей является моделирование работы сетевых протоколов при помощи акторов, ввиду отсутствия их готовой инкапсулированной реализации. Интересной и важной задачей на сегодняшний день является создание готовых наработок (например, реализаций протоколов, типовых ОУ и т.п.) в какой-либо среде акторного моделирования (к примеру, в Ptolemy II).

С точки зрения авторов, наибольшей актуальностью обладает развитие платформо-ориентированной методики. Абстрагирование от конкретной аппаратно-программной платформы и переход к более высокому уровню представления системы является важнейшим шагом в развитии методов и средств проектирования вычислительных систем и БСАС, в частности. Создание высокоуровневых платформ в виде библиотек промежуточного ПО существенно облегчит процесс прототипирования БСАС и разработки конечных систем, решающих различные прикладные задачи. Примером может служить проект системы управления мобильными объектами, описанный в [4]. Проблемой же платформо-ориентированного подхода является, безусловно, недостаток средств САПР, применяемых на различных фазах проектирования. Процессы определения и перебора вариантов взаимодействия компонентов системы и решения оптимизационных задач, без сомнения, нуждаются в автоматизации. Требуется реализация методических заготовок в существующих математических пакетах (MATLAB, Mathcad и др.) и пакетах имитационного моделирования (например, Ptolemy II, SensorSim, OMNeT++, AnyLogic).

Общей задачей развития рассмотренных методик проектирования является формализация и приоритизация требований, дополнение и совершенствование процесса составления архитектурных и функциональных спецификаций.

Заключение

Технология беспроводных сенсорно-актуаторных сетей идеологически соответствует концепции «повсеместных» вычислений (ubiquitous computing), сформулированной Марком Уэйзером еще в начале 90-х г.г. прошлого века и получающей все большую популярность в современном мире. В настоящей работе в общем виде произведено рассмотрение основных проблем проектирования беспроводных сенсорно-актуаторных сетей, являющихся, по сути, камнем преткновения для их широкомасштабного распространения в различных прикладных областях. Проанализирован ряд методик, которые на сегодняшний день применяются для проектирования изучаемого класса систем. Проведенный анализ позволил сформулировать актуальные задачи по доработке и улучшению рассмотренных методик. Решение поставленных задач во многом определит как дальнейшее применение технологии, так и судьбу упомянутой концепции развития вычислительной техники.

Литература

1. Bordignon M., Pagello E., Saffiotti A. An Inexpensive, Off-The-Shelf Platform for Networked Embedded Robotics // International Conference on Robot Communication and Coordination (RoboComm). – Athens, Greece, 2007. – P. 1–4.
2. Bordignon M., Rashid J., Broxvall M., Saffiotti A. Seamless Integration of Robots and Tiny Embedded Devices in a PEIS-Ecology // International Conference on Intelligent Robots and Systems (IROS). – San-Diego, California, 2007. – P. 1–6.
3. Van de Goor M. Indoor Localization in Wireless Sensor Networks // Master's Thesis, Netherlands. – 2009. – P. 1–31.
4. Киров Д.А. Перспективы применения беспроводных сенсорных сетей в задачах управления мобильными роботами // Сборник трудов молодых ученых и сотрудников кафедры ВТ / Под. ред. д.т.н., проф. Т.И. Алиева. – СПб: СПбНИУ ИТМО, 2012. – С. 34–38.
5. Branch J.W., Davis II J.S., Sow D.M., Bisdikian C. Sentire: A Framework for Building Middleware for Sensor and Actuator Networks // Proc. of the 3rd International Conference on Pervasive Computing and Communications Workshops. – 2005. – P. 1–5.
6. Jimenez A., Martinez-de Dios J.R. et al. Towards an open testbed for the cooperation of robots and wireless sensor networks. – University of Sevilla, 41092 Sevilla, Spain. – 2010. – 6 p.
7. Lee E.A., Neuendorffer S., Wirthlin M.J. Actor-Oriented Design of Embedded Hardware and Software Systems // Journal of Circuits, Systems and Computers. – November 20, 2002. – P. 12–14.
8. Bonivento A., Carloni L.P., Sangiovanni-Vincentelli A. Platform-Based Design of Wireless Sensor Networks for Industrial Applications. – University of California at Berkeley. – 2005. – 6 p.
9. Handziski V., Polastre J., Hauer J-H., Sharp C., Wolisz A., Culler D., Gay D. Hardware Abstraction Architecture [Электронный ресурс]. – Режим доступа: http://www.tinyos.net/tinyos-2.1.0/doc/html/tep2.html, свободный. Яз. англ. (дата обращения 10.10.2012).
10. Bakshi A., Prasanna V.K. Algorithm Design and Synthesis for Wireless Sensor Networks. – Department of Electrical Engineering. University of Southern California, Los Angeles, CA 90089. – 8 p.

Киров Дмитрий Александрович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, kdedge@mail.ru

Ожиганов Александр Аркадьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, ojiganov@mail.ifmo.ru

УДК 681.31.00

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МОДЕЛИРОВАНИЯ ЭВОЛЮЦИИ ДЛЯ ОПТИМИЗАЦИИ ДОКУМЕНТООБОРОТА НА ПРЕДПРИЯТИИ

Ю.О. Чернышев, А.В. Басова, П.А. Панасенко, В.И. Поляков

Предлагается новый подход к созданию информационных систем управления на предприятии. Для решения актуальных задач нелинейной оптимизации разработан генетический алгоритм, позволяющий существенно повысить качество решения и быстродействие. Описана разработанная процедура инициализации начальной популяции. Предложены модифицированные операторы кроссинговера, мутации, обоснован новый подход к отбраковке нежизнеспособных решений. Предлагается введение в генетический алгоритм макроэволюции, т.е. развития нескольких субпопуляций. Отмечено, что работа не с одной, а с несколькими начальными популяциями позволяет получить большее разнообразие генетического материала и вести эволюцию в отдельных субпопуляциях.

Ключевые слова: оптимизация, генетический алгоритм, информационные системы, транспортная задача.

Введение

Современный подход к управлению на предприятии предполагает создание информационных систем управления, включающих механизмы сбора, обработки, передачи и использования информации.

Первым шагом в решении этой проблемы является создание адекватной модели, которая упростит и удешевит процесс исследования оригинала. В качестве модели обработки и передачи информации может быть использована графо-аналитическая модель вычислительного процесса [1]. Следующий шаг – оптимизация модели и использование полученных знаний для повышения эффективности реальной работы на предприятии. Создание информационных механизмов в настоящий момент становится все более актуальным, необходимым звеном процесса принятия решений во всех областях человеческой деятельности. В основе различных методов принятия решений лежит математический аппарат, позволяющий осуществить оптимизацию и найти лучшее решение из возможных. В этой связи необходимо отметить, что выбор метода оптимизации – сложнейшая задача, и если ее решить неверно, то это повлечет за собой существенные временные и материальные затраты. В работе предложен генетический алгоритм оптимизации документооборота на предприятии.

Постановка задачи

Организационная структура работает эффективно, если решены задачи документального (информационного) обеспечения. Оценку эффективности работы такой структуры можно свести к оценке информационной загрузки отделов предприятия.

В качестве основной информационной единицы можно выбрать как стандартную единицу измерения шенноновской информации, так и более наглядную – «документ» – основную информационную единицу внутри информационной системы (ИС). Типы документов и их характеристики могут быть получены путем анализа документооборота конкретного предприятия. Функции предприятия могут быть раздроблены на более примитивные составляющие – функциональные единицы, причем множество функциональных единиц имеет взаимно однозначное соответствие множеству документов.

Основным фактором, обеспечивающим эффективность любого предприятия, является оптимизация затрат на выполнение функциональных единиц и на обеспечение документооборота.

Задачу оптимизации информационных потоков предприятия можно свести к транспортной задаче (ТЗ) путем введения условной единицы передачи документов. Основное отличие предлагаемой постановки ТЗ от классической заключается в том, что передача документов ограничена заданной структурой информационных связей предприятия. Другое отличие состоит в том, что возможная исходная несбалансированность ликвидируется не путем введения фиктивных элементов, а путем изменения фактической структуры (введение реальных потребителей документов). В рассматриваемой трактовке ТЗ распределение транспортируемых единиц (документов) определяется заданной структурой системы.

Математическая модель организации документооборота на предприятии

Имеется n источников документов и m потребителей документов. Для каждого источника i задан объем производства информации $a[i] > 0$, а для каждого потребителя j – объем потребления информации $b[j] > 0$. Стоимость передачи условной единицы документооборота между i-м источником и j-м потребителем равна $c[i, j]$. Таблица издержек $C[A, B]$ задана. Мы стремимся к тому, чтобы суммарная стоимость издержек была бы минимальной.

В данном случае, как упоминалось выше, рассматривается закрытая задача, в которой отсутствуют нерасшиваемые очереди в структурных элементах:

$$\sum_{i=1}^{n} a[i] = \sum_{j=1}^{m} b[j].$$

На практике целевая функция зачастую является вогнутой, что связано прежде всего с так называемым законом экономии от масштаба, в соответствии с которым удельные затраты на передачу информации убывают с ростом объемов перевозок [2]. При этом целевая функция имеет вид $y(x) = A \cdot \alpha$, где A – константа, а $\alpha < 1$.

Другой причиной, обусловливающей вогнутость целевой функции, является наличие фиксированных доплат. Эта функция равна сумме постоянных и линейных затрат, зависящих от мощности для любого положительного ее аргумента, и нулевым затратам при нулевой мощности. К этому виду относятся обычно затраты на проектирование и подготовку технической документации по созданию объекта, изготовлению макетов и опытных образцов, затраты на подготовительные работы и т.д. В частности, можно оценить затраты на создание и функционирование ИС предприятия:

$$y(x) = \begin{cases} 0, & \text{если} \quad x = 0, \\ h1 + h2 \cdot x, & \text{если} \quad x > 0 \end{cases},$$

где $h1$ – фиксированные затраты (денежные издержки на создание ИС); $h2$ – текущие затраты.

Зачастую в реальных условиях может возникнуть комбинированная задача, объединяющая оба описанных случая. Во многих практических ситуациях нелинейность целевой функции существенна. В связи с этим аппроксимация ее линейной функцией приводит к решению, далекому от истинного оптимума, и является неприемлемой.

Таким образом, можно сделать вывод о необходимости решения нелинейной транспортной задачи

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}(x_{ij}) \longrightarrow \min,$$

где $c_{ij}(x_{ij})$ – вогнутая ценовая функция,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix},$$

x_{ij} – количество информации, передаваемой от i-го источника j-му потребителю. Каждая матрица решения $\mathbf{X}$ должна удовлетворять следующим ограничениям:

$$\sum_{i=1}^{n} x_{ij} = b_j, \quad j = 1,2,\dots,m,$$

$$\sum_{j=1}^{m} x_{ij} = a_i, \quad i = 1,2,\dots,n,$$

$$x_{ij} \geq 0, \quad i = 1,2,\dots,n, \quad j = 1,2,\dots,m.$$

Было предпринято множество попыток эффективно решить данную задачу, так как она и по сей день является актуальной с практической и интересной с теоретической точек зрения. Перспективным представляется применение для решения оптимизационных задач методов моделирования процессов, протекающих в живой природе [3–5].

Структура генетического алгоритма решения ТЗ

Опишем базисную структуру разработанного генетического алгоритма. В начале работы генерируется множество решений (популяция) случайным, направленным или случайно-направленным способом. Решения представляются в виде закодированных хромосом. Способ кодирования определяется спецификой задачи. Далее, для каждого решения с помощью моделирующей функции оценивается его качество. Эта функция также основана на знаниях из той области, где решается задача. Инструмент селекции выбирает индивидуальности для скрещивания (кроссинговера). Операция мутации применяется случайно с некоторой вероятностью P_k. Затем формируется новая популяция, и процесс повторяется.

Генетический алгоритм должен иметь критерий остановки. Можно завершить работу после фиксированного числа генераций и взять лучшее из имеющихся решений. Можно остановиться, когда все организмы в популяции тождественны, т.е. $f_i = f_j$ для всех i и j или $|f_i - f_j| < \varepsilon$, где ε – наперед заданная малая положительная величина. Выбор критерия зависит от целей и задач, стоящих перед исследователем.

Поясним приведенную схему, рассмотрев более подробно каждый ее блок.

При выборе стратегий формирования начальной популяции следует руководствоваться целями, которые обеспечили бы наилучшие условия для дальнейшего развития эволюции. Сгенерированная начальная популяция должна содержать индивиды, принадлежащие к множеству D допустимых решений задачи.

В разработанном алгоритме (рисунок) формирование стартовой популяции решений $X^0 = x_1^0 \left\{x_1^0, x_2^0, \dots, x_n^0\right\}$ осуществляется с помощью механизма последовательного насыщения строк и столбцов, используемом в методе северо-западного угла и других методах [6]. Предложенная процедура отличается случайным выбором номера первого элемента, с которого начинается очередной цикл алгоритма инициализации, не опираясь на его северо-западность.

Пусть заданы $a_1, a_2, \dots, a_n$ – объемы источников документов и $b_1, b_2, \dots, b_m$ – объемы потребления документов. Пусть множество E_0 состоит из элементов $\{1, 2, \dots, n\}$, а множество F_0 – из элементов $\{1, 2, \dots, m\}$.

Функция оценки годности (моделирующая функция) естественным образом записывается как целевая функция исходной задачи

$$F(x) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} \cdot x_{ij},$$

где $c_{ij} \in C(x)$, а $C(x)$ – ценовая функция (вогнутая).

В процессе решения задачи необходимо минимизировать моделирующую функцию. Чем меньше значение функции $F(x)$, тем более ценным является решение x.

Сортировка популяции производится на основе полученных знаний. После того, как процедура сортировки будет закончена, получаем популяцию

$X = \{X1, X2, X3, \dots XS\}$,

в которой $F(X1) \le F(X2) \le ..... \le F(XS)$, где S – размер популяции.

Отсев неудачных особей после получения потомков является необходимой процедурой. Он формирует из общего множества родителей и потомков популяцию следующего поколения более высокого качества и направляет весь процесс эволюции в нужное русло. Основной классической процедурой отбора является пропорциональный отбор по значению целевой функции индивидов. Основную проблему отсева представляет нахождение компромисса между разнообразием генетического материала и качеством индивидов в популяции.

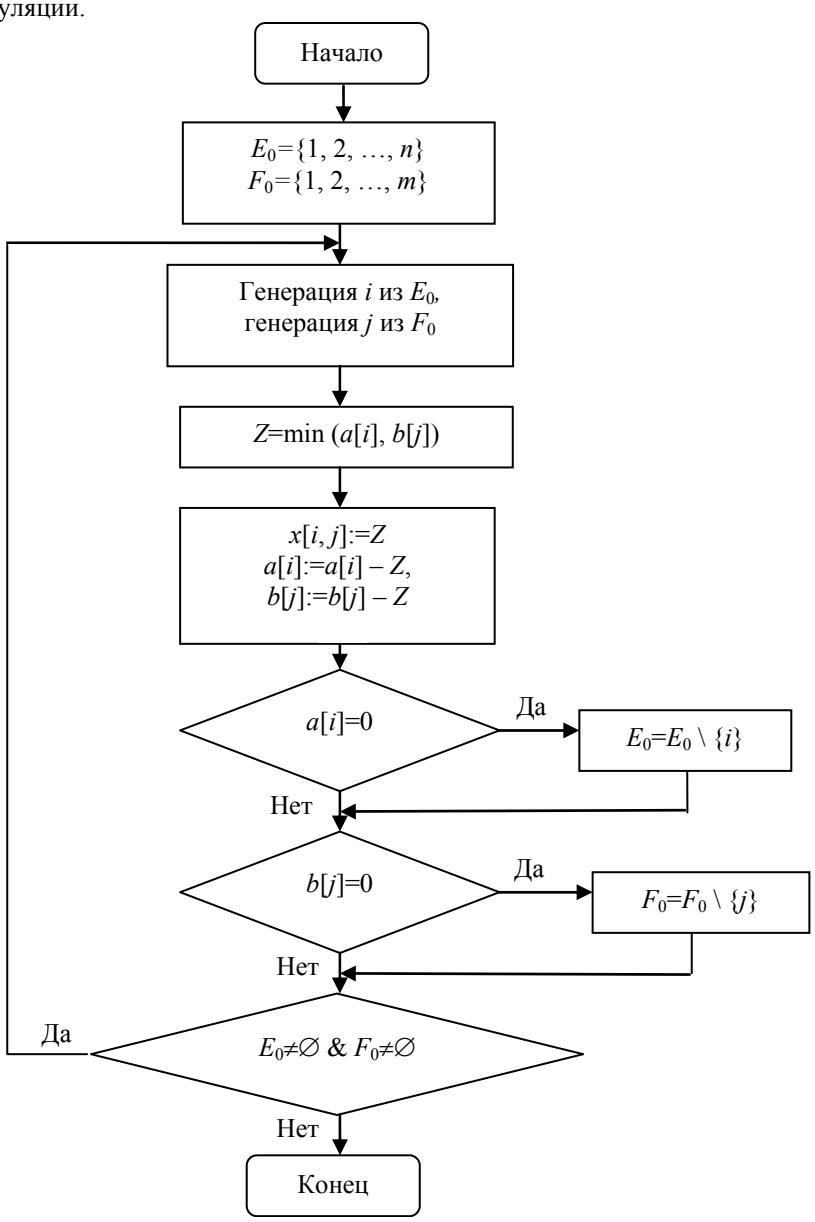


Рисунок. Блок-схема алгоритма формирования начальной популяции

В настоящей работе предлагается введение в генетический алгоритм макроэволюции, т.е. развития нескольких субпопуляций. Сложность этого подхода заключается в том, что после некоторого числа поколений хромосомы в отдельной популяции становятся очень похожими. Работа не с одной, а с несколькими начальными популяциями позволяет получить большее разнообразие генетического материала и возможность вести эволюцию в отдельных субпопуляциях, в каждой из которых можно получить решения, недостижимые в одной популяции.

Если в пределах популяции свободное скрещивание особей нарушается на длительное время, то в процесс эволюции включается третий фактор – изоляция. В результате действия изоляции образуются генотипически отличающиеся друг от друга субпопуляции. Скрещивание представителей этих субпопуляций оказывает положительный эффект на процесс эволюции. Включение в схему генетического алгоритма факторов изоляции и миграции при решении задач оптимизации позволит во многих случаях выходить из локальных оптимумов.

Апробация разработанного алгоритма

Для оценки эффективности разработанного алгоритма сравним полученные данные с результатами работы алгоритмов, созданных авторами [7–13]. Сравнение алгоритмов проводилось на тестовом примере при $n = 10$ источников и $m = 20$ потребителей документов. Результаты сравнения приведены в таблице.

Алгоритм	Возможность работы с вогнутыми функциями	Найденный оптимум	Вычислительная сложность	Источник
Йенсена-Барнеса	да	1635	$o(\exp(n))$	[7]
С.С. Лебедева	нет	–	$o(n^2)$–$o(n^3)$	[8]
Л.П. Падалко	нет	–	$o(n^2)$	[9]
С.Л. Уткина	да	1644	$o(n^3 – n^4)$	[10]
В.Р.Хачатурова, В.М. Монтлевича	да	1601	$o(n^3 – n^4)$	[11]
В.Р.Хачатурова, С.Л. Уткина	да	1538	$o(n^4)$	[12]
Хоанга Туй	да	1548	$o(n^3)$	[13]
Разработанный	да	1537	$o(n^2)$	[14–17]

Таблица. Сравнение различных алгоритмов решения нелинейной ТЗ

Заключение

Разработан генетический алгоритм решения задачи оптимизации документооборота на предприятии, позволяющий находить хорошее приближенное решение и имеющий приемлемую вычислительную сложность. Разработаны модифицированные операторы инициализации, кроссинговера и мутации, позволяющие получать допустимые решения исходной задачи. Предложенный алгоритм можно использовать для оптимизации документооборота при создании информационных систем управления предприятия.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (гранты № 12-01-00474а и № 12-07-00376а).

Литература

1. Зыков А.Г., Безруков А.В., Немолочнов О.Ф., Поляков В.И., Андронов А.В. Графо-аналитические модели вычислительных процессов в САПР // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 4 (74). – С. 116–120.
2. Райзберг Б.А., Лозовский Л.Ш., Стародубцева Е.Б. Современный экономический словарь. – 5-е изд., перераб. и доп. – М.: ИНФРА-М, 2006. – 495 с.
3. Курейчик В.М. Биоинспирированный поиск с использованием сценарного подхода // Известия Южного федерального университета. Технические науки. – 2010. – Т. 108. – № 7. – С. 7–13.
4. Курейчик В.М., Кажаров А.А. Применение пчелиного алгоритма для раскраски графов // Известия Южного федерального университета. Технические науки. – 2010. – Т. 113. – № 12. – С. 30–36.
5. Курейчик В.М., Кажаров А.А. Использование роевого интеллекта в решении NP-трудных задач // Известия Южного федерального университета. Технические науки. – 2010. – Т. 120. – № 7. – С. 30–36.
6. Лунгу К.Н. Линейное программирование. Руководство к решению задач. – М.: ФИЗМАТЛИТ, 2005. – 128 с.
7. Йенсен П., Барнес Д. Потоковое программирование. – М.: Радио и связь, 1984. – 392 с.
8. Лебедев С.С. Конечный метод решения нелинейных задач транспортного типа // Экономика и математические методы. – М.: Наука, 1965. – Т. 1. – Вып. 1. – С. 71–82.
9. Падалко Л.П. Алгоритм решения нелинейной транспортной задачи // Экономика и математические методы. – М.: Наука, 1968. – Т. 4. – Вып. 6. – С. 962–964.
10. Уткин С.Л. Вычислительные алгоритмы и опыт минимизации вогнутой функции на выпуклом многограннике. – М.: ВЦ АН СССР, 1990. – 29 с.
11. Хачатуров В.Р., Монтлевич В.М. Решение нелинейных производственно-транспортных задач с неделимыми потребителями. – М.: ВЦ АН СССР, 1987. – 22 с.
12. Хачатуров В.Р., Уткин С.Л. Решение многоэкстремальных задач вогнутого программирования аппроксимационно-комбинативным методом. – М: ВЦ АН СССР, 1988, – 38 с.
13. Хоанг Туй. Вогнутое программирование при линейных ограничениях // Доклады Академии наук СССР. – М., 1964. – Т. 159. – С. 32–35.
14. Чернышев Ю.О., Басова А.В. Теоретическое обоснование возможности решения вогнутой распределительной задачи с помощью методов эволюционного моделирования // Информационные техноло-

гии в науке, образовании, телекоммуникации и бизнесе. Материалы юбилейной XXX международной конференции. – Украина, Ялта–Гурзуф, 2003. – С. 147–149.
15. Басова А.В. Генетический метод отыскания глобального минимума многоэкстремальных задач // Математические методы в технике и технологиях. Сборник трудов XVI международной научной конференции. – СПб, 2003. – Т. 2. – С. 141–143.
16. Басова А.В. Новые методы отыскания глобального минимума многоэктремальных задач // Фундаментальные и прикладные проблемы приборостроения, информатики, экономики и права. Научные труды VI международной научно-практической конференции. Книга «Информатика». – М., 2003. – С. 31–35.
17. Басова А.В., Белявский П.Г. Генетические методы решения задачи оптимального планирования грузовых морских перевозок // Вестник Донского государственного технического университета. – 2011. – Т. 11. – № 5. (56). – С. 630–633.

Чернышев Юрий Олегович – Филиал военной академии связи (г. Краснодар), доктор технических наук, профессор, myvnn@list.ru

Басова Алина Викторовна – Донской государственный технический университет, кандидат технических наук, доцент, зав. сектором, abasova@donstu.ru

Панасенко Павел Александрович – Филиал военной академии связи (г. Краснодар), офицер службы защиты государственной тайны, We_panasenko_777@mail.ru

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

7 МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ

УДК 007.51

ОЦЕНКА РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В ОБЛАЧНЫХ СЕРВИСАХ НА ОСНОВЕ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

И.А. Зикратов, С.В. Одегов, А.В. Смирных

В работе рассмотрены теоретические аспекты построения защищенных облачных сервисов для обработки информации различной степени конфиденциальности. Предложен новый подход к обоснованию состава средств защиты информации в распределенных вычислительных структурах, заключающийся в представлении задачи оценки рисков как экстремальной задачи принятия решений. Доказывается, что использование метода линейного программирования минимизирует риски информационной безопасности при заданных значениях показателей защищенности с соблюдением экономического баланса на содержание средств защиты и стоимости предоставляемых услуг. Приводится пример, иллюстрирующий полученные теоретические результаты.

Ключевые слова: риски, информационная безопасность, облачные вычисления, целевая функция, угрозы.

Введение

Проблема обеспечения информационной безопасности (ИБ) стоит в ряду первостепенных задач при проектировании информационно-телекоммуникационных систем. Практика показывает, что в настоящее время ядро сетевой архитектуры изменилось с локализованных автономных вычислений на среду распределенных вычислений, что многократно увеличило ее сложность. Широкое распространение получили облачные вычисления – модель обеспечения сетевого доступа к общему пулу ресурсов, которые могут быть оперативно освобождены с минимальными эксплуатационными затратами [1].

Практика показывает, что пользователи облачных технологий передают на аутсорсинг функции хранения и обработки информации различной степени конфиденциальности. В этом случае к поставщику услуг могут предъявляться требования к обеспечению приемлемых уровней рисков ИБ, которые будут зависеть от ценности информационных активов. Следовательно, внедрение облачных технологий формирует актуальную проблему создания методологического подхода для обеспечения ИБ, в частности – снижения рисков ИБ в распределенных системах обработки и хранения данных.

Указанная задача решается путем:

- выявления, анализа и оценки рисков;
- снижения их до приемлемого уровня;
- внедрения адекватных механизмов именно для тех систем и процессов, для которых они необходимы [2].

Выполнение этих этапов позволяет сделать систему безопасности экономически результативной, актуальной и способной реагировать на возникающие угрозы.

Существующие международные стандарты в области менеджмента риска ИБ допускают использование как количественных, так и качественных методов оценки рисков. Наиболее известный вариант решения этой задачи, предлагаемый стандартами, состоит в умножении вероятности реализации угрозы на значение величины ущерба с последующим сопоставлением полученного значения с заданной шкалой [3]. Задача снижения риска в общем виде трактуется как действия, предпринятые для уменьшения вероятности, негативных последствий или того и другого вместе, связанных с риском [4].

Таким образом, поставщикам услуг облачных сервисов необходимо, с одной стороны, обеспечить обслуживание существующего потока заявок на обработку информации различной степени конфиденциальности, экономическую результативность облачного сервиса, рациональное распределение ресурсов в облаке, и, с другой стороны, предпринять действия для снижения, а по возможности – минимизации рисков ИБ.

В настоящей работе деятельность по минимизации рисков ИБ в облачных сервисах рассмотрена как решение экстремальной задачи, и предложен метод ее формализации на основе математического аппарата для этого класса задач – линейного программирования (ЛП).

Как известно, ЛП представляет собой набор переменных $\boldsymbol{x} = (x_1, x_2, \dots x_n)$ и функции этих переменных $f(\boldsymbol{x}) = f(x_1, x_2, \dots x_n)$, которая носит название ***целевой*** функции [5, 6]. Ставится задача: найти экстремум (максимум или минимум) целевой функции $f(\boldsymbol{x})$ при условии, что переменные $\boldsymbol{x}$ принадлежат некоторой области G:

$$\begin{cases} f(\boldsymbol{x}) => extr \\ \boldsymbol{x} \in G \end{cases}.$$

В зависимости от вида функции $f(\boldsymbol{x})$ и области G различают разделы математического программирования: квадратичное программирование, выпуклое программирование, целочисленное программирование и т.д. ЛП характеризуется тем, что

– функция $f(\boldsymbol{x})$ является линейной функцией переменных $x_1, x_2, \dots x_n$;
– область G определяется системой линейных равенств или неравенств.

Учитывая особенности метода ЛП, представим формальную постановку задачи для оценки рисков ИБ в облаке, исходя из следующих рассуждений.

Оценка рисков ИБ в облачном сервисе

Пусть в облаке может обрабатываться информация различной степени конфиденциальности $s = 1, 2, \dots, M$, где M – количество степеней конфиденциальности информации. Следовательно, облачный сервис должен иметь в своем составе ресурсы различного уровня защищенности, $k = 1, 2, \dots, M$. Очевидно, что согласно мандатной модели разграничения доступа при обработке информации s-й степени конфиденциальности на ресурсе k-го уровня защищенности должно выполняться требование

$$s \le k. \tag{1}$$

Для решения задачи квантизации ресурсов облака по уровням защищенности используется подход, представленный в работе [7].

Пусть известны стоимости затрат $C_k^з$ на содержание единицы ресурса различных уровней защищенности $C_1^з, C_2^з, \dots, C_M^з$, и заданы стоимости обработки единицы информации $C_s^о$ различной степени конфиденциальности $C_1^о, C_2^о, \dots, C_M^о$. Известен также поток заявок $\boldsymbol{I} = \{I_1, I_2, \dots, I_s\}$ на обработку информации различной степени конфиденциальности.

Наиболее распространенный способ вычисления риска (R) определяется следующим произведением [4]:

$$R = \sum_i P_i C_i^у ,$$

где P_i – вероятность успешной реализации i-й угрозы; $C_i^у$ – оценка ущерба (стоимости) при успешной реализации i-й угрозы; $i = 1..n$ – количество вероятных угроз.

Введем ограничение. Будем рассматривать систему при наличии одной угрозы. Очевидно, что величину ущерба при реализации угрозы на информацию различной степени конфиденциальности необходимо оценивать дифференцированно. Тогда, зная величины ущерба $C_s^у$ при воздействии угрозы на информацию s-й степени конфиденциальности, и, оценив вероятности реализации угрозы рассматриваемого типа применительно к ресурсу k-го уровня защищенности, построим матрицу рисков $\boldsymbol{r}$:

$$\begin{bmatrix} r_{11} & \cdots & r_{1s} \\ \vdots & \ddots & \vdots \\ r_{k1} & \cdots & r_{ks} \end{bmatrix}, \tag{2}$$

где r_{sk} – риск ИБ при воздействии угрозы на ресурс i-го класса, обрабатывающий информацию j-й степени конфиденциальности.

Требуется определить, сколько единиц ресурсов каждого уровня x_k надо иметь в составе облачного сервиса, чтобы риски ИБ при воздействии заданной угрозы были минимальны. С учетом требования (1) целевая функция принимает следующий вид:

$$x_1 r_{11} + x_2 r_{21} + x_2 r_{22} + x_3 r_{31} + x_3 r_{32} + x_3 r_{33} + .. + x_M r_{MM} \to \min .$$

При этом система ограничений должна обеспечивать выполнение следующих условий:
– обеспечение обслуживания всего потока заявок

$$\sum_i x_i I_i^d \ge \boldsymbol{I},$$

– облачный сервис в искомой конфигурации должен быть экономически рентабелен

$$\sum_i \sum_j x_i C_j^о > \sum_i x_i C_i^з.$$

Последнее ограничение обусловлено тем обстоятельством, что допускается обработка информации низкой степени конфиденциальности на ресурсах высших уровнях защищенности, так, что стоимость обработки для заказчика услуг при этом не должна возрастать. Исходя из этого, для степени конфиденциальности $s = i$, при обработке на ресурсе уровня $k = j$ ($j>i$), стоимость обработки должна быть равной $C_i^о$. Очевидно, система ограничений должна быть дополнена требованиями целостности и не отрицательности величин, исходя из их физического смысла.

Таким образом, в рамках предлагаемого подхода должны быть решены следующие задачи:
– квантизация ресурсов облачного сервиса по уровням защищенности;
– идентификация уязвимостей и угроз;
– классификация обрабатываемой информации по степеням конфиденциальности;
– количественная оценка вероятности реализации угрозы и влияние на состояние облачного сервиса потенциальных угроз;
– определение экономического баланса между затратами на содержание системы защиты информации (СЗИ) ресурса облачного сервиса и предоставляемой стоимостью услуг.

Пример оценки рисков ИБ при использовании ЛП

Проиллюстрируем полученный результат тривиальным примером. На рисунке представлен фрагмент распределенной автоматизированной системы в составе которой имеются ресурсы различной степени защищенности – группы *A*, *B* и *C*, в зависимости от СЗИ, реализованной на данном ресурсе:

– 1-я группа (*A*) – ресурсы с низким уровнем защищенности;
– 2-я группа (*B*) – ресурсы с высоким уровнем защищенности;
– 3-я группа (*C*) – защищенные ресурсы.

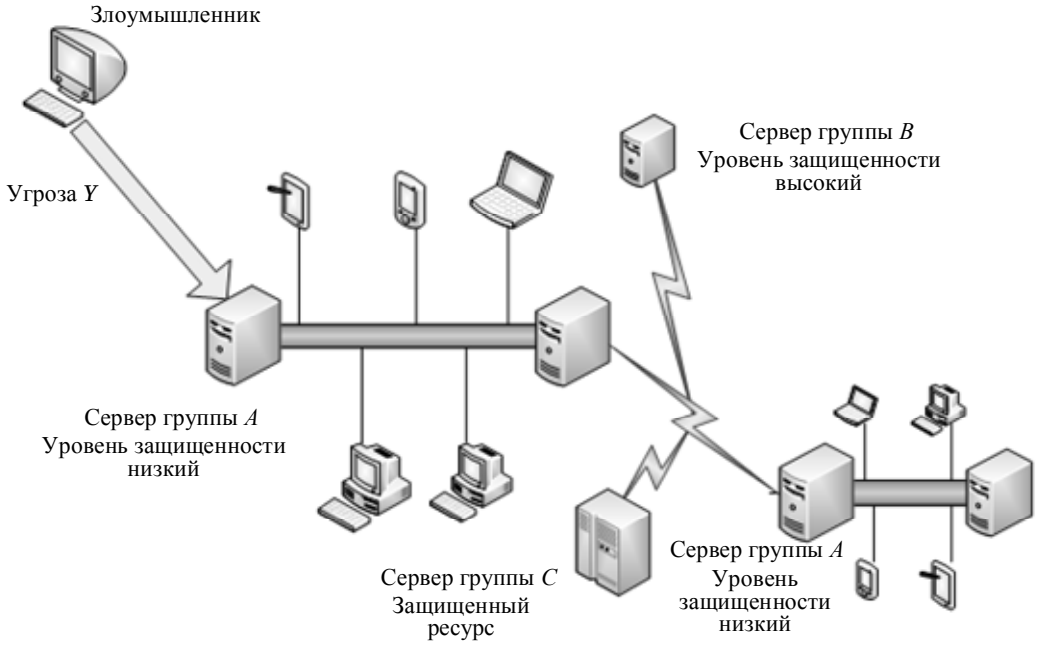


Рисунок. Фрагмент распределенной автоматизированной системы

Стоимость затрат на содержание единицы ресурса *n*-ой группы: $A = 1$, $B = 3$, $C = 8$ условных единиц. Пусть в рассматриваемой системе обрабатывается информация различных категорий конфиденциальности l_s ($s = 1, 2, 3$):

– l_1 – открытая информация;
– l_2 – конфиденциальная информация;
– l_3 – критически важная информация.

Согласно неравенству (1), информация категории l_1 обрабатывается на ресурсах любой группы, информация категории l_2 может обрабатываться на ресурсах групп B и C. Информация категории l_3 обрабатывается только на ресурсах группы C. Стоимость обработки единицы информации *s*-ой степени конфиденциальности составляет $l_1 = 2$, $l_2 = 5$ и $l_3 = 10$.

Масштабирование ресурсов, требуемых для обработки информации *s*-ой степени конфиденциальности осуществляется при помощи заявок $\boldsymbol{I}$. Поток заявок составляет не менее: $I_1 = 200$; $I_2 = 80$; $I_3 = 40$ единиц.

Пусть из общей статистики угроз вида Y известна вероятность реализации угрозы P_i для каждого типа ресурсов P_A, P_B и P_C. Очевидно, что P_i реализации угрозы зависит от уровня защищенности ресурса, но не зависит от степени конфиденциальности информации *s*, обрабатываемой на ресурсе. Составим матрицу рисков (2). Пусть значения вероятности P_i и уровня ущерба $C_s^{\text{у}}$ для различных групп ресурсов и степеней конфиденциальности защищаемой информации: $P_A = 1$, $P_B = 0{,}5$ и $P_C = 0{,}1$ и $C_1^{\text{у}} = 2$, $C_2^{\text{у}} = 5$ и $C_3^{\text{у}} = 10$. Тогда элементы матрицы рисков r_{is} будут иметь значения, представленные в таблице.

Проведем расчет количества емкости заявок I_n. На ресурсах группы *A* могут обрабатываться категории информации $I_1 = 200$; на ресурсах группы *B* обрабатываются заявки I_1 и $I_2 = 280$ и на ресурсах группы *C* могут обрабатываться только заявки категории $I_3 = 80$. Общее количество заявок I_n составляет 320. Решение задачи симплекс-методом дает следующий результат:

– количество единиц ресурса группы *A* $x_1 = 40$ единиц;
– для группы *B* $x_2 = 180$ единиц;
– для группы *C* $x_3 = 100$ единиц.

Анализ результатов показывает, что полученное решение:

– содержит однозначно трактуемые количественные оценки состава ресурсов, различной степени защищенности в облачном сервисе;
– обеспечивает поток заявок на обработку информации различной степени конфиденциальности;
– не противоречит известным моделям разграничения доступа.

	Открытая информация (l_1)	**Конфиденциальная информация (l_2)**	**Критическая информация (l_3)**
Ресурс группы *A*	2	5	10
Ресурс группы *B*	5	2,5	5
Ресурс группы *C*	0,2	0,5	1

Таблица. Матрица рисков ИБ

При этом обеспечивается минимизация рисков ИБ для заданных показателей защищенности и матрицы потерь при соблюдении экономического баланса на содержание СЗИ и стоимости предоставляемых услуг на обработку конфиденциальной информации.

Заключение

Широкое распространение облачных вычислений на рынке поставщиков IT-услуг приводит к необходимости совершенствования научно-методического аппарата для построения систем защиты информации. В работе впервые предложен метод для количественного обоснования состава ресурсов различного уровня защищенности в облачном сервисе. Предлагаемый новый подход позволяет поставщику услуг на основе матрицы рисков, требований заказчиков к обеспечению конфиденциальности, целостности и доступности информации, соблюдения экономических интересов осуществлять оценку и минимизацию рисков информационной безопасности в облачных сервисах. Выполнив подобные расчеты для всех видов угроз безопасности в соответствии с моделью угроз можно принимать обоснованные решения для конфигурации емкости облачных сервисов и ресурсов различного уровня защищенности.

Применимость методики обусловлена использованием апробированного математического аппарата, непротиворечивостью полученных результатов, а также отражает требования международных стандартов в области оценки рисков. Кроме того, разработанный метод оценки рисков соответствует основным постулатам мандатной модели разграничения доступа.

Дальнейшим направлением научной работы являются:
1. оценка чувствительности полученных результатов, так как известным недостатком линейного программирования является высокая чувствительность к изменению исходных данных;
2. возможность адаптации предложенного метода к ролевой модели разграничения доступа, как наиболее распространенной в облачных сервисах.

Литература

1. Jansen W., Grance T. Guidelines on Security and Privacy in Public Cloud Computing. Draft NIST Special Publication 800-144. – Gaithersburg, 2011. – 52 p.
2. Catteddu D., Hogben G. Cloud Computing: Benefits, risks and recommendations for information security. – Heraklion: ENISA, 2009. – 125 p.
3. Марков А., Цирлов В. Управление рисками – нормативный вакуум информационной безопасности [Электронный ресурс]. – Режим доступа: http://www.osp.ru/os/2007/08/4492873/, свободный. Яз. рус. (дата обращения 30.11.2012).
4. ГОСТ Р ИСО/МЭК 27005-2010. Информационная технология. Методы и средства обеспечения безопасности. Менеджмент риска информационной безопасности. – Введ. 30.11.2010. – М.: Стандартинформ, 2011. – 51 с.
5. Хемди А. Таха. Введение в исследование операций. – М.: Вильямс, 2007. – 912 с.
6. Вентцель Е.С. Введение в исследование операций. – М.: Советское радио, 1964 – 391 с.
7. Зикратов И.А., Одегов С.В. Оценка информационной безопасности в облачных вычислениях на основе байесовского подхода // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 4 (80). – С. 121–126.

Зикратов Игорь Алексеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, доцент, зав. кафедрой, zikratov@cit.ifmo.ru

Одегов Степан Викторович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, odegov.sv@gmail.com

Смирных Александр Валентинович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ст. преподаватель, smirnykh_av@spb.power-m.ru

УДК 004.056

АЛГОРИТМ КОНТРОЛЯ ЦЕЛОСТНОСТИ ДЕПЕРСОНАЛИЗИРОВАННЫХ ДАННЫХ В ИНФОРМАЦИОННЫХ СИСТЕМАХ ПЕРСОНАЛЬНЫХ ДАННЫХ

Ю.А. Гатчин, О.А. Теплоухова, А.С. Куракин

Рассматривается проблема обеспечения целостности информации, обрабатываемой в информационных системах персональных данных, при использовании алгоритма деперсонализации данных. Предлагается использовать алгоритм контроля целостности, основанный на использовании хэш-функции, для подтверждения неизменности деперсонализированных данных, хранящихся в информационных системах персональных данных.

Ключевые слова: контроль целостности, хэш-функция, алгоритм деперсонализации, персональные данные, информационные системы персональных данных.

Введение

На сегодняшний день обеспечение безопасности персональных данных (ПДн) является одной из острейших проблем в информационной сфере и взаимоотношениях государства, юридических и физических лиц. Выполнение требований к защищенности ПДн при их обработке в информационных системах, устанавливаемых федеральным законом «О персональных данных» [1], является очень дорогостоящей задачей. Для снижения данных требований [2] к уровню защищенности информационных систем персональных данных (ИСПДн), как правило, применяют способы обезличивания, позволяющие существенно сократить расходы на обеспечение их информационной безопасности и снизить вероятность угроз несанкционированного доступа соответственно. Перспективным способом обезличивания персональных данных является алгоритм деперсонализации [3], применение которого делает невозможным определить принадлежность ПДн их владельцу.

Необходимо отметить, что даже после применения алгоритма деперсонализации ПДн остается актуальной угроза нарушения целостности данных, хранящихся в ИСПДн, направленная на изменение или искажение информации, приводящее к нарушению ее качества. Таким образом, для того чтобы оператор ИСПДн был уверен в том, что загружаемые им данные не были искажены, необходима реализация механизма контроля целостности (КЦ) деперсонализированных данных.

Существующие аппаратно-программные комплексы, включающие в свой функционал КЦ данных, являются достаточно дорогостоящим решением, которое не могут позволить бюджеты небольших организаций. Эффективным способом нейтрализации угрозы нарушения целостности ПДн является предлагаемый в настоящей работе механизм, основанный на использовании хэш-функций в процессе деперсонализации данных, позволяющий оператору отслеживать нарушение целостности деперсонализированных данных. Этот не применявшийся ранее подход не требует больших финансовых затрат и при совместном использовании с алгоритмом деперсонализации позволяет эффективно повысить уровень защищенности ИСПДн от НСД.

Целью данной работы является разработка алгоритма КЦ данных, который при совместном использовании с алгоритмом деперсонализации данных в ИСПДн обеспечивал бы обнаружение нарушения целостности хранящейся в постоянном запоминающем устройстве (ПЗУ) информации при ее загрузке в оперативном запоминающем устройстве (ОЗУ).

Описание алгоритма

Перспективным способом решения данной проблемы является использование алгоритма контроля целостности деперсонализированных данных. Предлагаемый алгоритм основан на вычислении хэш-значения от открытых данных при сохранении информации в ПЗУ до применения алгоритма деперсонализации и после применения обратного алгоритма при загрузке данных в ОЗУ. На основании сравнения полученных хэш-значений делается вывод о целостности загружаемой информации.

Рассмотрим таблицу D, содержащую персональные данные субъектов, обрабатываемых в ИСПДн. Во время работы оператора таблица D хранится в ОЗУ в персонализированном виде, т.е. представляет собой открытую информацию (ОИ). До того, как будет применен алгоритм деперсонализации при сохранении данных в ПЗУ, вычислим хэш-значение от ОИ: $h(D)$. После этого применим алгоритм деперсонализации данных, в результате чего будет получена таблица D', представляющая собой файл с закрытой информацией (ЗИ). Добавим вычисленное ранее хэш-значение $h(D)$ к полученному файлу, применив операцию конкатенации, и получим значение $F(D', h(D))$, которое сохраняется в ПЗУ.

Для того чтобы загрузить персональные данные в ОЗУ необходимо извлечь из хранящегося в ПЗУ файла $F(D', h(D))$ хэш-значение $h(D)$, после чего применяем к таблице D' обратный алгоритм деперсонализации, в результате чего получаем таблицу D''. Вычислим новое хэш-значение от полученной таблицы: $h(D'')$ и сравним со значением, которое было извлечено из хранящегося в ПЗУ файла. В случае если $h(D)$ совпадает с $h(D'')$, принимаем что таблица D'' совпадает с исходной таблицей ПДн D, и загружаем ее в ОЗУ для дальнейшей работы. В противном случае, если полученное хэш-значение $h(D'')$ не совпало

с $h(D)$, алгоритм должен выдать сообщение оператору о том, что была нарушена целостность файла с деперсонализированными данными.

Хэш-значение от таблицы ПДн D должно вычисляться после каждого применения к ней обратного алгоритма деперсонализации. При очередном сохранении данных в ПЗУ должен формироваться файл с ЗИ в результате конкатенации таблицы D и нового хэш-значения $h'(D)$.

При получении сообщения о нарушении КЦ оператор может принять решение о повторной проверке целостности деперсонализированных данных. Если вновь вычисленное хэш-значение от таблицы D'' не будет совпадать с извлеченным значением $h(D)$, оператор должен сообщить об этом администратору информационной безопасности. После чего в ОЗУ загружается резервная копия деперсонализированной таблицы D'.

Практическая реализация алгоритма

Использование в предлагаемом алгоритме КЦ хэш-функции обусловлено следующими ее преимуществами:

1. размер входного файла при вычислении хэш-функции может быть произвольным в отличие от шифрования данных;
2. при вычислении хэш-функции не требуется ключ, который необходимо хранить и вводить, что делает реализацию алгоритма очень простой.

Используемая функция

$$h(M): \{0,1\}^* \to \{0,1\}^n, n \in N,$$

где $M \in \{0,1\}^*$ – произвольное цифровое сообщение, должна быть односторонней хэш-функцией, что означает выполнение следующих условий [4]:

1. значение функции h определено для любого цифрового сообщения $M \in \{0,1\}^*$;
2. для любого цифрового сообщения $M \in \{0,1\}^*$ функция h имеет фиксированный порядок $n \in$ для любого N;
3. для любого $M \in \{0,1\}^*$ значение $h(M)$ вычисляется за полиномиальное время;
4. для любого $M_1 \in \{0,1\}^*$ вычислительно сложно найти сообщение $M_2 \in \{0,1\}^*$, такое, что $M_1 \neq M_2$, $h(M_1) = h(M_2)$;
5. вычислительно невозможно (за разумное время) найти пару (M_1, M_2), $M_1 \neq M_2$, $M_i \in \{0,1\}^*$, $i = 1,2$, такую, что $h(M_1) = h(M_2)$.

Свойства 4, 5 являются важнейшими криптографическими свойствами, обеспечивающими стойкость однонаправленных хэш-функций. В случае если злоумышленнику известен алгоритм построения хэш-функции, первоначальное сообщение $M \in \{0,1\}^*$ и хэш-значение $h(M)$, то вероятность $P(k,n)$ того, что среди сообщений $N_1, \dots, N_k \in \{0,1\}^*$ существует номер $i = 1, \dots, k \in N$, такой что $h(M) = h(N_i)$ равна:

$$P(k,n) = 1 - (1 - 2^{-n})^k \approx k2^{-n}.$$

Таким образом, для осуществления взлома фиксированного значения для 128-битных хэш-значений (например, MD4 (Message Digest 4), MD5(Message Digest 5)) потребуется перебрать около $\mathbf{2^{120}}$ текстов, что за обозримое время не представляется возможным.

Необходимо отметить, что кроме метода прямого перебора существуют более эффективные атаки на хэш-функции. Так, например, взлом за линейное время (туннельный эффект) [5] основывается на том, что большинство хэш-функций используют сдвиговую функцию: текст M разбивается на блоки M_i, затем происходит итерационный процесс подсчета хэш-значения $h_i = f(h_{i-1}, M_i)$, $i \in N$. В силу небольшой длины раунда можно применить дифференциальный анализ и найти такой текст ΔD, что

$$h(M_i + \Delta D) = h(M_i).$$

Ввиду вышесказанного, в предлагаемом алгоритме КЦ рекомендуется использовать одностороннюю хэш-функцию в совокупности с методом повышения ее криптостойкости, например, метод перестановок. В этом случае хэш-функция будет строиться по правилу:

$$h(M) = h(M)||h(\pi_1(M)) \dots ||h(\pi_k(M)).$$

В данном случае $\pi_i, i = 1, \dots, k,$ – произвольные перестановки текста M. Полученное хэш-значение будет хэш-функцией как конкатенация хэш-значений однонаправленной хэш-функции. Данный метод эффективен против туннельного эффекта, поскольку после перестановки полученный текст отличается от исходного не на туннельный эффект ΔD, что приводит к отличию хэш-значений данных текстов. Кроме того, данный метод может быть эффективно распараллелен, что является важным преимуществом для использования его на маломощных компьютерах.

Для получения хэш-значения таблицы D рекомендуется использовать MD5 – 128-битный алгоритм хеширования, предназначенный для вычисления хэш-функций от сообщений произвольной длины. Данный алгоритм является улучшенной в плане безопасности версией MD4.

На вход алгоритма MD5 поступает входной поток данных, хэш которого необходимо найти. Длина сообщения может быть любой (в том числе нулевой). Запишем длину сообщения в L. Это число целое

и неотрицательное. Кратность каким-либо числам необязательна. После поступления данных идет процесс подготовки потока к вычислениям. Сначала дописывают единичный бит в конец потока (байт 0×80), затем необходимое число нулевых бит. Входные данные выравниваются так, чтобы их новый размер L' удовлетворял условию:

$$L' = 512 \times N + 448.$$

Далее выравненные данные разбиваются на блоки (слова) по 32 бита, и каждый блок проходит 4 раунда из 16 операторов. Для 4 раундов используются следующие функции от трех аргументов:

1 раунд: $F(X,Y,Z) = (X \wedge Y) \vee (\overline{X} \wedge Z)$;

2 раунд: $G(X,Y,Z) = (X \wedge Z) \vee (\overline{Z} \wedge Y)$;

3 раунд: $H(X,Y,Z) = X \oplus Y \oplus Z$;

4 раунд: $I(X,Y,Z) = Y \oplus (\overline{Z} \vee X)$.

Так как множество аргументов хэш-функции счетно, а значения имеют определенный порядок, неизбежно возникновение коллизии – получение одного и того же хэша для двух разных входных значений. При использовании в данном случае одного из самых универсальных методов поиска коллизий – атаки «дней рождений», отыскание коллизии для хэш-функции разрядности n потребует в среднем около $2^{\frac{n}{2}}$ операций, что для функции MD5 составляет 2^{64}. Такая вычислительная сложность нахождения коллизий позволяет гарантировать очень низкую вероятность подмены таблиц с ПДн злоумышленником, которая не была бы обнаружена алгоритмом контроля целостности.

При практической реализации алгоритма также целесообразно хранить отдельно резервные копии деперсонализированных данных. Применение таких мер позволит избежать потери ПДн, обрабатываемых в ИСПДн, в случае, если злоумышленник получит доступ к ЗИ, хранящейся в ПЗУ, и целостность данных будет нарушена.

Заключение

Предложенный алгоритм является наиболее перспективным и оптимальным решением задач по контролю целостности персональных данных, хранящихся и обрабатываемых в информационных системах персональных данных.

Данный алгоритм обладает следующими преимуществами:

1. размер файла с персональными данными, поступающего на вход алгоритма, может быть произвольным;
2. для вычисления используемой в алгоритме хэш-функции не требуется ключ, который необходимо хранить и вводить, что значительно упрощает программную реализацию;
3. для применения данного решения не требуется дополнительных аппаратных средств, используемый алгоритм вычисления хэш-функции MD5 входит в состав открытых библиотек криптопримитивов, таких как OpenSSl, Crypt.dll;
4. алгоритм может быть реализован в качестве программной надстройки над алгоритмом деперсонализации персональных данных.

Литература

1. Федеральный закон Российской Федерации от 27 июля 2006 г. № 152-ФЗ. О персональных данных. Принят Государственной Думой Федерального Собрания Российской Федерации 8 июля 2006 г.: одобрен Советом Федерации Федерального Собрания Российской Федерации 14 июля 2006 г. // Российская газета. – 2006. – 29 июля.
2. Об утверждении Положения об обеспечении безопасности персональных данных при их обработке в информационных системах персональных данных: Постановление Правительства Российской Федерации от 17 ноября 2007 г. № 781 г. Москва // Российская газета. – 2007. – 21 ноября.
3. Куракин А.С. Алгоритм деперсонализации персональных данных // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 6 (82). – С. 130–135.
4. Девянин П.Н., Михальский О.О., Правиков Д.И. Теоретические основы компьютерной безопасности: Учебное пособие для вузов. – М.: Радио и связь, 2000. – 192 с.
5. Лёвин В.Ю. О повышении криптостойкости однонаправленных хеш-функций // Фундаментальная и прикладная математика. – 2009. – Т. 15. – № 5. – С. 171–179.

Гатчин Юрий Арменакович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, зав. кафедрой, gatchin@mail.ifmo.ru

Теплоухова Ольга Александровна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, teplouhovaoa@gmail.com

Куракин Александр Сергеевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, nirt@mail.ru

УДК 625.042

СИСТЕМА ПРЕДОТВРАЩЕНИЯ ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЙ НА ЖЕЛЕЗНОЙ ДОРОГЕ НА БАЗЕ ГЕОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

В.В. Алексеев, В.С. Коновалова, А.А. Минина

Предложена система предотвращения чрезвычайных ситуаций на основе геоинформационных технологий, которая позволяет оценивать влияние на безопасность движения дефектов железнодорожного полотна, техногенных объектов (расположенные вблизи исследуемого участка железной дороги фабрики, заводы) и природных факторов (воздействие стихии). Предложены процедуры построения и подготовки карты местности для систематизации и отображения информации. Разработан алгоритм обработки измерительной информации на основе вейвлет-преобразования, позволяющий производить определение параметров дефектов железнодорожного полотна сразу после измерений и не требующий накопления информации для последующей расшифровки. Предложены методики оценки влияния природных факторов и техногенных объектов на безопасность движения с целью оперативного представления информации для выработки управленческих решений.

Ключевые слова: железнодорожное полотно, геоинформационные системы, дефект, чрезвычайные ситуации, вейвлет-преобразование.

Введение

При решении вопросов обеспечения безопасности железнодорожного (ЖД) движения необходимо учитывать особенности железных дорог: их протяженность, распределенность по территории России и сопредельным государствам, наличие возможных разнообразных дефектов полотна, неизбежное влияние на состояние железной дороги природных факторов и техногенных объектов, сезонность, климат, рельеф местности и др. Причем информация о точном местоположении дефектов, возмущающих объектов, их размеры и степень воздействия существенно влияют на качество управленческих решений, принимаемых и реализуемых на различных уровнях. Следует также учитывать, что инфраструктура железной дороги распределена неравномерно географически и достаточно удалена, что обусловливает особенность сбора и обобщения информации.

В этой связи привлекательным становится применение в системах предотвращения чрезвычайных ситуаций (ЧС) геоинформационных технологий (ГИС). Примером такой системы может служить программный продукт ArcGIS американской компании ESRI [1].

В России на данный момент нет систем предотвращения ЧС на транспорте, охватывающих все аспекты, влияющие на железную дорогу. Существующие системы – путеизмерительные вагоны, исследующие искажения геометрии рельсовых нитей, не способны охватить всю сеть достаточно протяженных железных дорог России. «Ручные» методы контроля полотна железной дороги являются ресурсоемкими и не оперативными. Вместе с этим оценке влияния на железную дорогу природных факторов и техногенных объектов не уделяется сколь-нибудь существенного внимания.

На сегодня ГИС находят применение в таких областях как геологоразведка и полевые изыскательные работы, мониторинг технологических режимов работы нефте- и газопроводов, проектирование магистральных трубопроводов, моделирование и анализ последствий аварийных ситуаций. В земельном кадастре ГИС применяются при ведении кадастрового учета земель и зеленых насаждений, городском планировании и проектировании объектов. В природопользовании ГИС эффективны при управлении лесозаготовками, ведении лесных кадастров, планировании подходов к лесу и проектировании дорог. В вопросах охраны окружающей среды ГИС используют для оценки и мониторинга состояния природной среды, планировании природоохранных мероприятий, моделировании экологических катастроф и анализе их последствий. На транспорте ГИС могут найти применение при управлении транспортной инфраструктурой и ее развитии, управлении парком транспортных средств и для нахождения кратчайшего расстояния, управления движением, оптимизации маршрутов и анализа грузопотоков. Исходя из этого, авторам представляется, что применение ГИС для задачи предотвращения ЧС на железной дороге является логичным и современным решением. Это позволит объединить информацию от всех видов измерителей состояния ЖД полотна, работать с пространственно-распределенной информацией для ее сквозной, интегральной и согласованной обработки. Это обеспечивает полной, достоверной и оперативной пространственной информацией лицо, уполномоченное принимать управленческие решения.

Целью разрабатываемой геоинформационной системы предотвращения ЧС является сбор информации и ее визуальное представление для оценки влияния дефектов ЖД полотна, природных факторов, факторов воздействия на железную дорогу техногенных объектов (фабрики, заводы), расположенных вблизи железной дороги, с целью повышения безопасность движения. В настоящей работе предложен и реализован программно вариант геоинформационного комплекса.

Структурная схема системы предотвращения ЧС на базе ГИС

При разработке ГИС сформулированы основные функциональные задачи системы:

– сохранение измерительной информации для обеспечения возможности сравнения с последующими измерениями, т.е. наличие базы данных (БД);

– наличие в своем составе набора базовых алгоритмов для обработки измерительной информации от датчиков с целью измерения параметров дефектов ЖД полотна;
– наличие методики экспертной оценки степени влияния найденных дефектов, а также природных факторов и техногенных объектов на безопасность движения;
– отображение результатов анализа на карте;
– оперативное предоставление информации лицу, принимающему решения.

Структурная схема системы представлена на рис. 1. Разработанное программно-алгоритмическое обеспечение подобной системы, созданное авторами, в полной мере удовлетворяет сформулированным задачам.

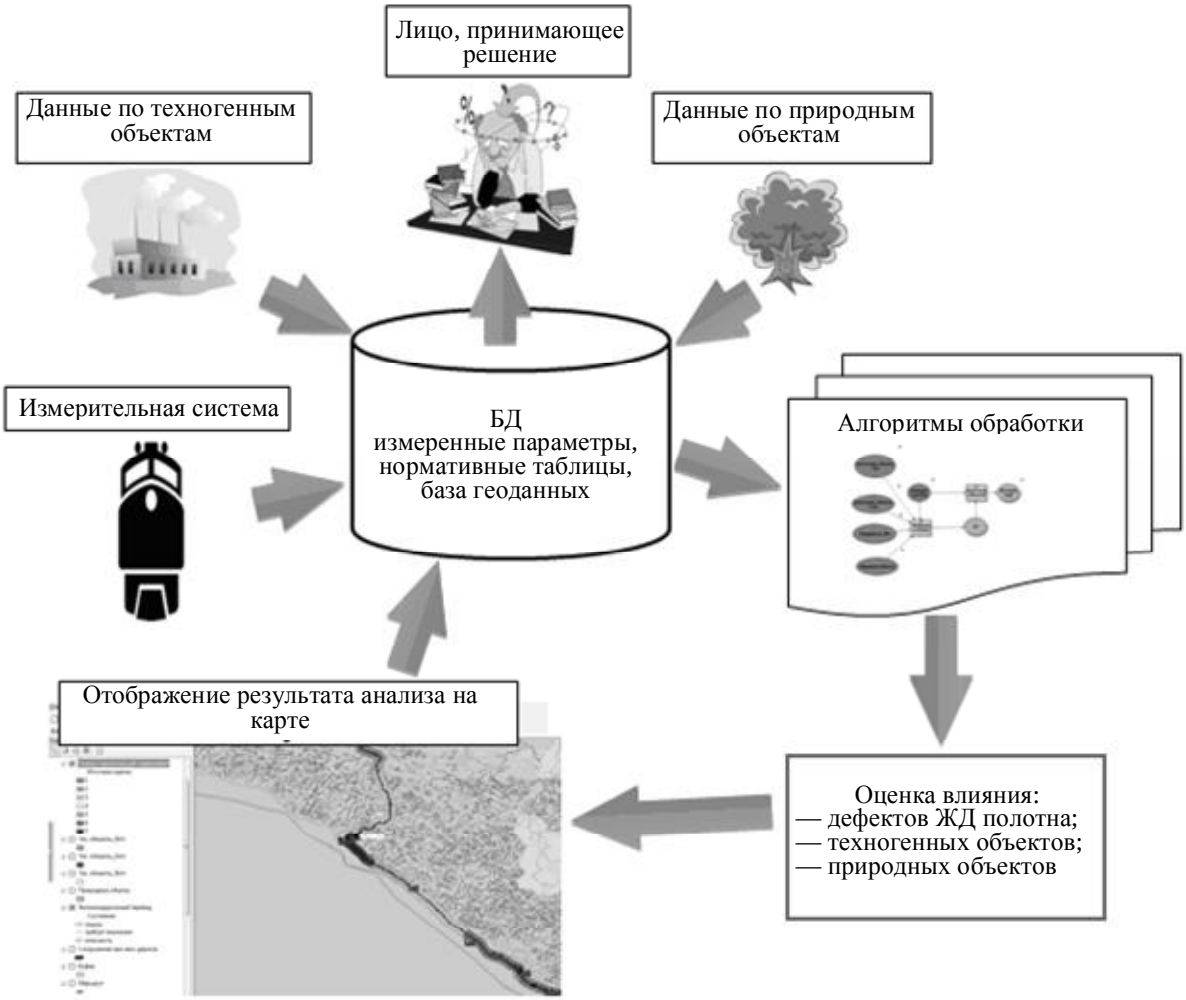


Рис. 1. Структурная схема системы предотвращения ЧС на железной дороге на базе ГИС

Первичным измерительным преобразователем (ПИП), является вибродатчик, установленный на буксе колесной пары [2]. Сигнал с ПИП подвергается обработке с использованием вейвлет-преобразования с целью выделения локальных составляющих, характеризующих параметры дефектов ЖД полотна. Информация о дефектах (местоположение, протяженность, глубина) аккумулируется в БД. Аналогичным образом информация о природных факторах и техногенных объектах, нормативные таблицы и геоданные (топооснова, географическая модель железной дороги, линейная схема маршрута, модели объектов и др.) заносятся в БД.

В дальнейшем анализируется содержимое БД и формируются оценки влияния на безопасность движения дефектов ЖД полотна, техногенных объектов и природных факторов, находящихся вблизи исследуемого участка железной дороги. Результаты анализа отображаются на карте и предоставляются лицу, ответственному за принятие решений.

Оценка влияния дефектов ЖД полотна, природных факторов и техногенных объектов на железную дорогу

Для оценки влияния на железную дорогу дефектов ЖД полотна, природных факторов и техногенных объектов рассматриваемый участок обращения локомотива разбивается на подучастки, на протяжении которых скорость движения локомотива имеет строго фиксированное значение. На данном участке

создается модель железной дороги. Каждый подучасток модели имеет свой идентификационный номер, привязка к которому позволяет оценить влияние дефектов ЖД полотна, природных факторов и техногенных объектов на безопасность движения локомотива.

Необходимо учитывать, что ряд объектов не располагается непосредственно на железной дороге, а находится в отдалении от нее. Разработанные программные модули позволяют строить на модели железной дороги так называемые «буферные зоны». Радиус зон определяется экспериментально. Считается, что все объекты, попавшие в «буферную зону» и характеризуемые показателем воздействия, будут влиять на безопасность подучастка железной дороги (рис. 2).

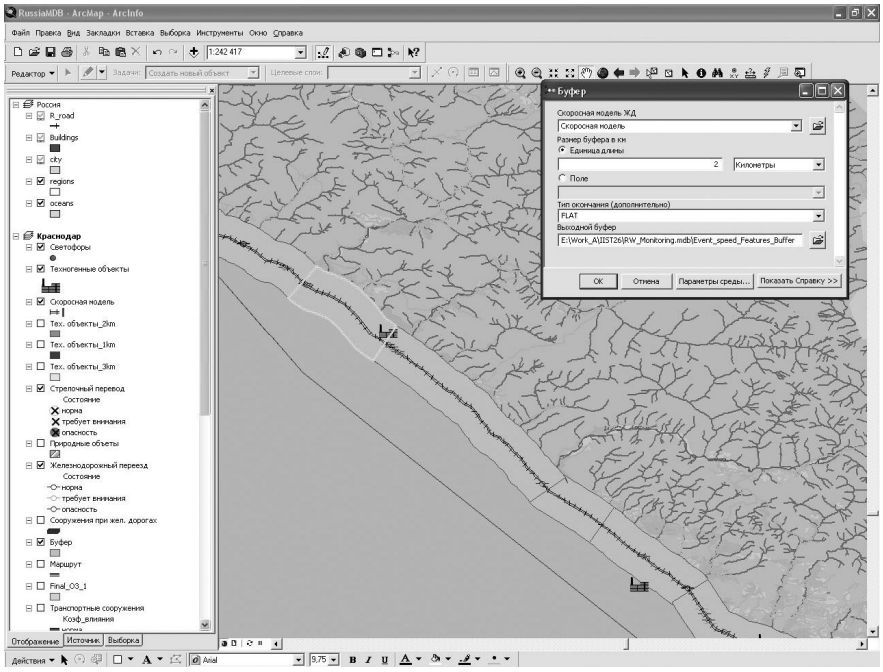


Рис. 2. Пример формирования «буферных зон»

Данные, поступающие от измерительной системы, формируют базу параметров (дефектов), которая позволяет обращаться как к информации, полученной после текущего проезда, так и к ранее полученным результатам контроля состояния пути. В зависимости от степени и количества обнаруженных отступлений от установленных в нормативно-технической документации рассматриваемому участку железной дороги дается балловая оценка его состояния по каждому из показателей.

Обнаруженные дефекты разделяются на сколы и просадки. Также определяется местоположение стыков. Каждый вид дефектов проявляется в измерительном сигнале в виде локальной составляющей в соответствующем частотном диапазоне, для их выделения и разделения, как было отмечено ранее, используется алгоритм дискретного вейвлет-преобразования [2]. Для обеспечения исследования в режиме реального времени применен алгоритм скользящего дискретного вейвлет-преобразования. Этот алгоритм позволяет производить определение состояние ЖД полотна в момент измерения, т.е. отказаться от практикуемого в настоящее время измерения, происходящего в два этапа – измерение и последующая расшифровка результатов.

Размеры отступлений от норм, правила содержания пути и обнаруженные дефекты сопоставляются с допускаемыми размерами и установленными требованиями, предъявляемыми к обеспечению безопасности движения поездов в зависимости от установленных скоростей. Результатом этого является присвоение отрезку пути I, II, III или IV группы отступлений. В зависимости от этого выносятся рекомендации о частичном или полном ремонте ЖД пути, снижении скорости движения ЖД состава или закрытии пути как аварийно-опасного. Кроме того, учитываются срок эксплуатации и материалы.

Для оценки влияния природных факторов и техногенных объектов на железную дорогу по поступившим в базу данным предлагается относить все объекты и факторы, оказывающие влияние на железную дорогу, к одному из трех показателям воздействия:

1. слабое влияние;
2. среднее влияние;
3. сильное влияние.

Для этого прибегают к экспертным оценкам, которые основаны на априорных знаниях о рассматриваемых факторах. В силу того, что объекты могут оказывать воздействие на некотором расстоянии, вокруг них строятся зоны влияния, учитывающие скорость и направление ветра. Далее проводится оценка влияния природных факторов и техногенных объектов на безопасность движения для каждого под-участка.

В качестве примера рассмотрим техногенные объекты (предприятия и заводы), расположенные вблизи железной дороги. Их влияние можно оценить на основе априорных знаний (АЗ), которые являются исходными данными для построения моделей ЖД сооружений и полотна, а также окружающих природных факторов и техногенных объектов [3]:

$$АЗ_{ind} = (A_c, d, t_m, f_b, n_t, tr_f, ...), \quad (1)$$

где A_c – деятельность предприятия (насколько вредное воздействие оказывает деятельность предприятие на окружающую среду); d – удаленность предприятия от железной дороги; t_m – время создания предприятия; f_b – частота возникновения аварий, сбоев в работе предприятия; n_t – применение новых технологий в работе предприятия; tr_f – использование очистных сооружений на предприятии и другие. В выражении (1) представлен минимальный состав данных, который необходим для реализации системы. Количество данных может быть расширено, что может способствовать уточнению получаемых результатов. Полный список формирует эксперт (специалист), занимающийся данной проблемой.

На основе анализа АЗ, представленных в (1), каждому предприятию и заводу присваивается показатель воздействия на железную дорогу. В силу того, что влияние предприятия распространяется на некотором расстоянии, вокруг каждого из них строятся «зоны влияния» (рис. 3). Зона наименьшего радиуса считается самой опасной и имеет максимальный показатель воздействия, установленный экспертами. Следующая по радиусу зона менее опасна, но, в зависимости от вида предприятия, может быть значимой при оценке. Третья зона – самого большого радиуса – считается практически безопасной. Ее необходимо вводить и учитывать при анализе влияния разного рода химических, нефтеперерабатывающих, а также атомных предприятий.

Оценка влияния природных факторов (воздействие стихии) на состояние рельсового пути осуществляется по трехбалльной шкале в результате учета уровня их опасности и удаленности от ЖД полотна. Состав АЗ для природных факторов на примере ЧС, которые могут иметь место в районе рассматриваемого подучастка железной дороги, представлен в следующем виде:

$$АЗ_{em} = (Sp, ex, d, f, d_r ...), \quad (2)$$

где Sp – вид ЧС; ex – протяженность ЧС; d – удаленность ЧС от железной дороги; f – частота появления ЧС в данном районе (по статистическим данным); d_r – степень риска ЧС. Каждый тип ЧС оценивается некоторой степенью риска (чаще всего от 1 до 3, где 1 – наиболее опасная ЧС, а 3 – менее опасная для окружающей среды и человека) [4].

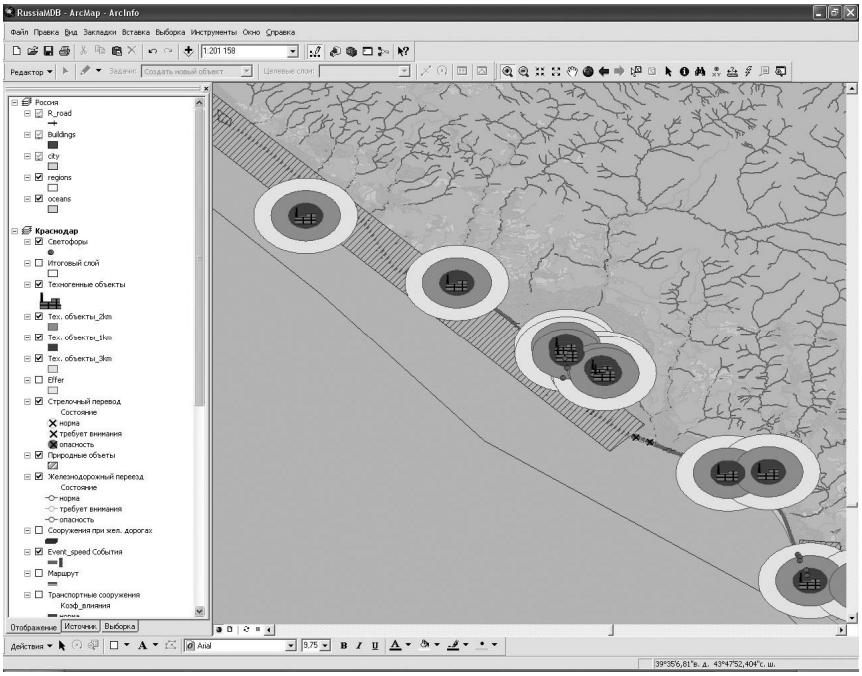


Рис. 3. Зоны влияния предприятий и природных факторов

По АЗ, представленным в (2), каждому природному фактору присваивается показатель воздействия на железную дорогу. Для формирования оценок состояния на базе ГИС созданы модели, реализующие алгоритмы обработки данных, и позволяющие в автоматическом режиме формировать итоговую оценку влияния дефектов ЖД полотна, техногенных объектов и природных факторов в зоне каждого подучастка железной дороги. Информация визуализируется на карте и может быть предоставлена лицу, принимающему решения с целью выработки управляющих рекомендаций.

Заключение

Предложена структурная схема системы предотвращения чрезвычайных ситуаций на железной дороге, реализующая геоинформационные технологии. Для предложенной модели разработан алгоритм обработки измерительной информации с применением вейвлет-преобразования, что позволяет определять параметры дефектов железнодорожного полотна в режиме реального времени и не требует накопления информации для последующей расшифровки.

Предложены методики оценки влияния природных факторов и техногенных объектов на безопасность движения с целью оперативного представления комплексной информации для выработки управленческих решений.

Реализовано программно-алгоритмическое обеспечение системы, позволяющее систематизировать и визуализировать обработанную информацию.

Разработанная система предотвращения чрезвычайных ситуаций на железной дороге на базе геоинформационных технологий обеспечивает:

- сбор и обработку информации по природным факторам, техногенным объектам и от измерительной системы;
- анализ данных с использованием заложенных алгоритмов обработки;
- оценку влияния дефектов железнодорожного полотна на безопасность движения;
- оценку влияния техногенных (фабрики и заводы) объектов и природных (воздействие стихии) факторов на безопасность движения;
- отображение результатов анализа по требуемым запросам к базе данных на карте;
- предоставление оперативной информации лицу, принимающему решения.

Работа выполнена при поддержке Министерства образования и науки Российской Федерации (грант № 14.B37.21.1540).

Литература

1. Андрианов В. ГИС и транспорт // ArcReview. – 2007. – № 3 (42). – С. 1–2.
2. Алексеев В.В., Боронахин А.М., Калякин И.В., Коновалова В.С., Подгорная Л.Н. Измерение характеристик железнодорожного полотна с помощью измерительной системы, построенной на базе миро механических акселерометров // Приборы. – 2011.– № 12 (138). – С. 22–29.
3. Минина А.А. Влияние техногенных и природных объектов на состояние железной дороги // Тезисы доклада 11-ой научно-практической конференции пользователей ГИС «От полюса до полюса». 21–22 мая 2012. – С. 57–63.
4. Безопасность жизнедеятельности в чрезвычайных ситуациях. Методические указания по изучению курса. – Хабаровск, 2000 [Электронный ресурс]. – Режим доступа: http://edu.dvgups.ru/metdoc/enf/bgd/bgd_chs/metod/bgd_chs/akchur1.htm, свободный. Яз. рус. (дата обращения 28.11.2012).

Алексеев Владимир Васильевич – Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ), доктор технических наук, профессор, зав. кафедрой, VVAlekseev@mail.eltech.ru

Коновалова Вера Сергеевна – Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ), кандидат технических наук, ассистент, vskonovalova@inbox.ru

Минина Анастасия Андреевна – Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ), кандидат технических наук, ассистент, aaminina@mail.ru

УДК 003.26.09

ЦИФРОВЫЕ ВОДЯНЫЕ ЗНАКИ В ГРАФИЧЕСКИХ ФАЙЛАХ

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

Рассмотрена задача генерации цифровых водяных знаков для графических файлов. Представлена математическая модель генерации цифровых водяных знаков. Проанализирован алгоритм внедрения сообщений. Предложено применение разработанного стеганоалгоритма для решения задачи проверки авторского права на конкретный файл мультимедиа.

Ключевые слова: ЦВЗ, форматные методы, стеганоалгоритмы пространственной области, стеганоалгоритмы области преобразования, мультимедиа, медиа-пространство, авторское право.

Введение

Использование цифровых форматов мультимедиа в настоящее стало повсеместным [1]. Но наряду с этим в современном информационном обществе, исследования и разработки в области стеганографии

становятся все более популярными. Это связано с тем, что существуют проблемы управления цифровыми ресурсами и контроля использования прав собственности на компьютерные файлы. Отсюда возникает актуальнейшая задача сокрытия информации в условиях развитой инфраструктуры сетевого общения пользователей – интернет-участников открытого и неконтролируемого взаимодействия в медиа-пространстве.

Сокрытие информации в медиа-пространстве обычно производят при помощи стеганографических алгоритмов. Существует несколько задач, для решения которых используют такие алгоритмы, например:

1. обеспечение тайны переписки (postal privacy);
2. общение удаленных абонентов, обменивающихся цифровыми массивами информации;
3. общение удаленных абонентов в открытых сетевых структурах;
4. достижение скрытности хранимой информации большого объема.

Одним из наиболее эффективных методов защиты мультимедийной информации является встраивание в защищаемый объект невидимых меток – цифровых водяных знаков (ЦВЗ). Название этот метод получил известного способа защиты ценных бумаг, в том числе и денег, от подделки. Термин «digital watermarking» был впервые применен в работе [2]. В отличие от обычных водяных знаков ЦВЗ могут быть не только видимыми, но и (как правило) невидимыми. Невидимые ЦВЗ анализируются специальным декодером, который выносит решение об их корректности.

Стегосистемы ЦВЗ, в частности, должны выполнять задачу защиты авторских и имущественных прав на электронные сообщения при различных попытках активного нарушителя искажения или стирания встроенной в них аутентифицирующей информации. Формально говоря, системы ЦВЗ должны обеспечить аутентификацию отправителей электронных сообщений. Подобная задача может быть возложена на криптографические системы электронной цифровой подписи (ЭЦП) данных, но в отличие от стегосистем ЦВЗ, известные системы ЭЦП не обеспечивают защиту авторства не только цифровых, но и аналоговых сообщений в условиях, когда активный нарушитель вносит искажения в защищаемое сообщение и аутентифицирующую информацию. Иные требования по безопасности предъявляются к стегосистемам, предназначенным для скрытия факта передачи конфиденциальных сообщений от пассивного нарушителя. Также имеет свои особенности обеспечение имитостойкости стегосистем к вводу в скрытый канал передачи ложной информации [3, 4].

Основные требования, предъявляемые к ЦВЗ для графических файлов, представлены в [5, 6].

Математическая модель генерации ЦВЗ

При формальном представлении генерации ЦВЗ в виде математической модели воспользуемся общепринятой записью:

$$\varphi : X \to Y,$$

где φ – отображение (функция); X – область определения; Y – область значений. Введем следующие обозначения: $Y_{\text{ЦВЗ}}$ – множество ЦВЗ; $X_{\text{КЛЮЧ}}$ – множество ключей, $X_{\text{КОНТЕЙНЕР}}$ – множество контейнеров; $X_{\text{СООБЩЕНИЙ}}$ – множество скрываемых сообщений. Тогда генерация ЦВЗ может быть представлена в виде

$$F : X_{\text{КОНТЕЙНЕР}} \times X_{\text{КЛЮЧ}} \times X_{\text{СООБЩЕНИЙ}} \to Y_{\text{ЦВЗ}},$$

или

$$y_{\text{ЦВЗ}} = F\left(x_{\text{КОНТЕЙНЕР}}, x_{\text{КЛЮЧ}}, x_{\text{СООБЩЕНИЙ}}\right),$$

где $y_{\text{ЦВЗ}} \in Y_{\text{ЦВЗ}}$, $x_{\text{СООБЩЕНИЙ}} \in X_{\text{СООБЩЕНИЙ}}$, $x_{\text{КЛЮЧ}} \in X_{\text{КЛЮЧ}}$, $x_{\text{КОНТЕЙНЕР}} \in X_{\text{КОНТЕЙНЕР}}$. Функция F (отображение) может быть произвольной, но на практике требования робастности ЦВЗ накладывают на нее определенные ограничения. Формально это можно записать так:

$$y_{\text{ЦВЗ}} = F\left(x_{\text{КОНТЕЙНЕР}}, x_{\text{КЛЮЧ}}, x_{\text{СООБЩЕНИЙ}},\right) \approx F\left(x_{\text{КОНТЕЙНЕР}} + \varepsilon, x_{\text{КЛЮЧ}}, x_{\text{СООБЩЕНИЙ}}\right),$$

т.е. незначительно измененный контейнер не приводит к изменению ЦВЗ. Кроме того, функция F часто является составной:

$$F = T \circ G,$$

где $G : X_{\text{КЛЮЧ}} \times X_{\text{СООБЩЕНИЙ}} \to X_{\text{КОД}}$ и $T : X_{\text{КОНТЕЙНЕР}} \times X_{\text{КОД}} \to Y_{\text{ЦВЗ}}$; $\circ$ – суперпозиция.

Функция G может быть реализована при помощи криптографически безопасного генератора псевдослучайных последовательностей с $x_{\text{КЛЮЧ}}$, в качестве начального значения.

Отсчеты ЦВЗ принимают обычно значения из множества {–1, 1}, при этом для отображения {0,1} → {–1, 1} можно применить двоичную относительную фазовую модуляцию ФМн-2 (Binary Phase Shift Keying – BPSK) [7]. Данный вид модуляции нашел очень широкое применение ввиду высокой помехоустойчивости и простоты модулятора и демодулятора. Оператор T модифицирует кодовые слова $X_{\text{КОД}}$, в результате чего получается ЦВЗ – $Y_{\text{ЦВЗ}}$. На этот оператор не накладывают условие существования у

него обратного, так как соответствующий выбор G уже гарантирует необратимость F. Функция T должна быть выбрана так, чтобы незаполненный контейнер $x_{\text{КОНТЕЙНЕР}_0} \in X_{\text{КОНТЕЙНЕР}}$, заполненный контейнер $x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}} \in X_{\text{КОНТЕЙНЕР}}$ и незначительно модифицированный заполненный контейнер $x'_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}} \in X_{\text{КОНТЕЙНЕР}}$ порождали бы один и тот же ЦВЗ:

$$T\left(x_{\text{КОНТЕЙНЕР}_0}, x_{\text{КОД}}\right) = T\left(x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}}, x_{\text{КОД}}\right) = T\left(x'_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}}, x_{\text{КОД}}\right),$$

т.е. она должна быть устойчивой к малым изменениям контейнера.

Процесс встраивания ЦВЗ $y_{\text{ЦВЗ}}(i,j)$ в исходное изображение $x_{\text{КОНТЕЙНЕР}_0}(i,j)$ можно описать как суперпозицию двух сигналов:

$$\Psi : X_{\text{КОНТЕЙНЕР}} \times Y_{\text{ЦВЗ}} \times X_{\text{МАСКА}} \to X_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}},$$

или

$$x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}}(i,j) = x_{\text{КОНТЕЙНЕР}_0}(i,j) \oplus x_{\text{МАСКА}}\, y_{\text{ЦВЗ}}(i,j)\, p(i,j),$$

где $X_{\text{МАСКА}}$ – маска встраивания ЦВЗ, учитывающая характеристики зрительной системы человека, служит для уменьшения заметности ЦВЗ; $p(i,j)$ – проектирующая функция, зависящая от ключа; знаком $\oplus$ – обозначен оператор суперпозиции, включающий в себя, помимо сложения, усечение и квантование.

Проектирующая функция осуществляет «распределение» ЦВЗ по области изображения. Ее использование может рассматриваться, как реализация разнесения информации по параллельным каналам. Кроме того, эта функция имеет определенную пространственную структуру и корреляционные свойства, использующиеся для противодействия геометрическим атакам.

Одним из наиболее важных устройств в стегосистеме является стегодетектор. В зависимости от типа он может выдавать двоичные либо М-ичные решения о наличии/отсутствии ЦВЗ (в случае детектора с мягкими решениями). Рассмотрим вначале более простой случай «жесткого» детектора стего. Обозначим операцию детектирования через D. Тогда

$$D : X_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}} \times Y_{\text{ЦВЗ}} \to \{0,1\},$$

$$D\left(x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}}, y_{\text{ЦВЗ}}\right) = D\left(x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}}, F\left(x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}}, x_{\text{КЛЮЧ}}, x_{\text{СООБЩЕНИЙ}}\right)\right) =$$

$$= \begin{cases} 1, \text{ если } y_{\text{ЦВЗ}} \text{ есть} \\ 0, \text{ если } y_{\text{ЦВЗ}} \text{ нет} \end{cases}.$$

В качестве детектора ЦВЗ обычно используют корреляционный приемник.

Предположим, что у половины пикселей изображения значение яркости увеличено на 1, а у остальных – осталось неизменным, или уменьшено на 1. Тогда:

$$x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}} = x_{\text{КОНТЕЙНЕР}_0} + y_{\text{ЦВЗ}},$$

где $y_{\text{ЦВЗ}} = F\left(x_{\text{КОНТЕЙНЕР}_0}, x_{\text{КЛЮЧ}}, x_{\text{СООБЩЕНИЙ}}\right)$. Коррелятор детектора ЦВЗ вычисляет величину

$$x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}} \cdot y_{\text{ЦВЗ}} = \left(x_{\text{КОНТЕЙНЕР}_0} + y_{\text{ЦВЗ}}\right) \cdot y_{\text{ЦВЗ}} = x_{\text{КОНТЕЙНЕР}_0} \cdot y_{\text{ЦВЗ}} + y_{\text{ЦВЗ}} \cdot y_{\text{ЦВЗ}}.$$

Так как $y_{\text{ЦВЗ}}$ может принимать значения ±1, то $x_{\text{КОНТЕЙНЕР}_0} \cdot y_{\text{ЦВЗ}}$ будет мало, а $y_{\text{ЦВЗ}} \cdot y_{\text{ЦВЗ}}$ будет всегда положительно. По этой причине величина $x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}} \cdot y_{\text{ЦВЗ}}$ будет очень близка к $y_{\text{ЦВЗ}} \cdot y_{\text{ЦВЗ}}$. Следовательно, можно определить вероятность неверного обнаружения стего, как дополнительную (комплементарную) функцию ошибок от корня квадратного из отношения $y_{\text{ЦВЗ}} \cdot y_{\text{ЦВЗ}}$ («энергии сигнала») к дисперсии значений пикселей яркости («энергия шума»).

Для случая мягкого детектора и закрытой стегосистемы имеем две основные меры похожести:

$$\delta = \frac{x_{\text{КОНТЕЙНЕР}_0} \cdot x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}}}{\left\| X_{\text{КОНТЕЙНЕР}_0} \right\| \left\| X_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}} \right\|}$$ – нормированный коэффициент взаимной корреляции и

$$\delta = N - \sum_{i}^{N} x_{\text{КОНТЕЙНЕР}_0}(i) \cdot x_{\text{КОНТЕЙНЕР}_{\text{ЗАПОЛНЕННЫЙ}}}(i)$$ – расстояние по Хэммингу.

В детекторе возможно возникновение двух типов ошибок. А именно, существует вероятность того, что детектор не обнаружит имеющийся ЦВЗ и есть вероятность ложного нахождения ЦВЗ в пустом контейнере (вероятность ложной тревоги). Снижение одной вероятности приводит к увеличению другой. Надежность работы детектора характеризуют вероятностью ложного обнаружения. Система ЦВЗ должна быть построена таким образом, чтобы минимизировать вероятности возникновения обеих ошибок, так как каждая из них может привести к отказу от обслуживания.

Алгоритм внедрения сообщений

При разработке системы скрытой передачи и стеганоалгоритма, представленной авторами в [8], обнаружились трудности, касающиеся межформатных преобразований. Поскольку JPEG – формат сжатия с потерями, то они (потери) в общем случае не позволяют восстановить встроенное сообщение, поскольку восстановление происходит после процедур межформатных преобразований JPEG – RGB BMP – JPEG. В разработанной системе указанная проблема была решена. Процесс внедрения сообщения включает ряд превентивных мер. Непосредственно внедрение происходит по алгоритму, представленному на рис. 1.

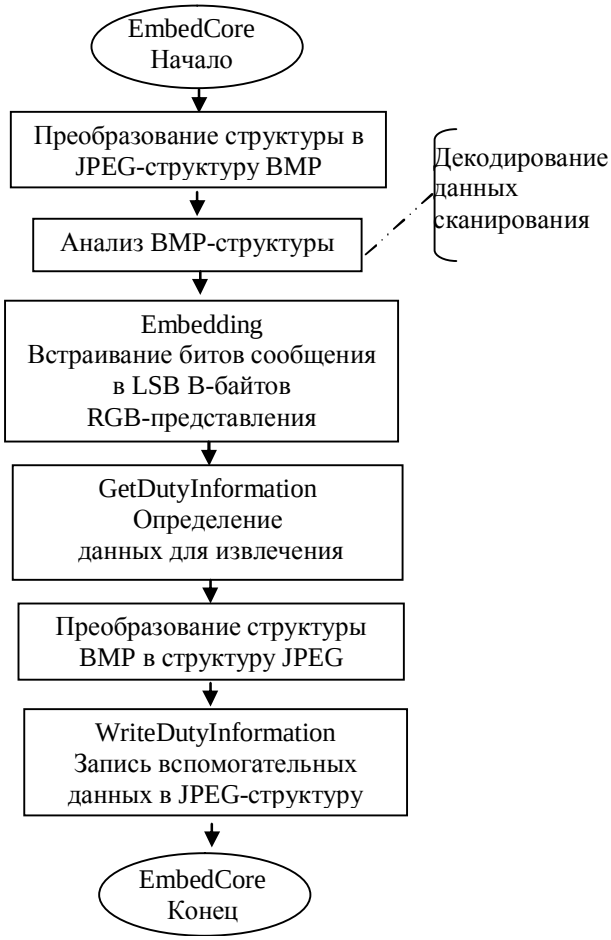


Рис. 1. Блок-схема алгоритма внедрения сообщения

На первом этапе производится преобразование потока данных JPEG в поток данных BMP. При этом увеличивается размер потока за счет изменения принципа кодирования информации об 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 в себе.

Известно, что система человеческого зрения обладает особенностью слабой чувствительности к изменениям в оттенках синего цвета, поэтому для встраивания используются B-составляющие RGB-структур [9]. На самом деле, человеческий глаз также редко может отследить изменения в наименьшем значащем бите красной и зеленой компоненты RGB-структуры. Для минимизации объема изменения пространственной области в режиме работы по умолчанию используется только младший бит такого байта, что до минимума снижает вероятность обнаружения изменения даже на изображениях с большой площадью заливки синего цвета. Простейший способ замены битов – последовательная замена в каждом *b*-байте – представлен на рис. 2.

Поскольку JPEG – формат сжатия с потерями, необходимо учесть этот факт для возможности извлечения сообщения на принимающей стороне. Разработанный механизм компенсирования потерь при межформатных преобразованиях был представлен авторами в [8].

Для определения факта наличия скрытого сообщения пять выходных файлов JPEG с внедренными сообщениями различной длины были проверены хорошо известной программой Stegdetect. Эта програм-

ма детектирования факта встраивания ориентирована на поиск байтовых сигнатур, выдающих стеганографическое вмешательство.

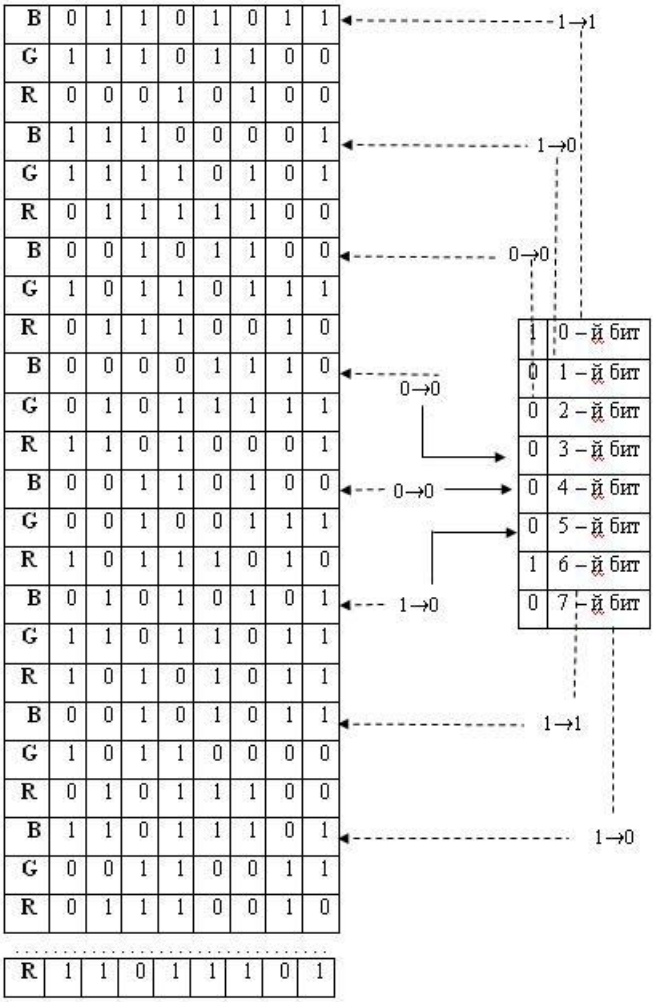


Рис. 2. Последовательная замена битов

В результате проведенного эксперимента программа Stegdetect не смогла корректно указать на факт внедрения данных с помощью разработанного алгоритма. Визуально определить этот факт также не представляется возможным. Визуальный анализ проводился группой людей при наличии оригинального файла JPEG без внедренного сообщения. Определить, в какой из двух файлов встроены данные, им не удалось.

Заключение

Для графических изображений с точки зрения защиты авторского права на их файлы принципиально необходимо реализовывать автоматическое подписывание файлов с целью опубликования информации об авторе. Это может быть текст или иная графическая информация, размещенная в какой-либо (например, нижней) части изображения, однозначно ассоциирующаяся с личностью автора-правообладателя. Такие «метки» служат неопровержимой ссылкой на источник, предоставивший конкретный графический файл. Внедрение в изображения цифровых водяных знаков, позволяющих подтвердить и проверить права разработчика на данный файл мультимедиа, является также эффективной защитной мерой для соблюдения прав интеллектуальной собственности. Такие метки могут быть различным образом расположены в мультимедийном файле и служить противодействием для таких неправомерных деяний, например, как подмена авторства и отказ от авторства. Представленный алгоритм позволяет решить эту задачу. Но необходимо отметить тот факт, что данный алгоритм не проверялся на устойчивость к различного рода искажениям изображения. Эта задача входит в план ближайших научных исследований авторов данной работы.

Литература

1. Сидоркина И.Г., Коробейников А.Г., Кудрин П.А. Алгоритм распознавания трехмерных изображений с высокой детализацией // Вестник Марийского государственного технического университета. – 2010. – № 2 (9). – С. 91–99.

2. Osborne C., van Schyndel R., Tirkel A. A Digital Watermark // IEEE In-tern. Conf. on Image Processing. – 1994. – P. 86–90.
3. Ramkumar M. Data Hiding in Multimedia. – PhD Thesis. – New Jersey Institute of Technology, 1999. – 72 p.
4. Simmons G. The prisoner`s problem and the subliminal channel // Proc. Workshop on Communications Security (Crypto′83). – 1984. – P. 51–67.
5. Барсуков В.С. Романцов А.П. Компьютерная стеганография: вчера, сегодня, завтра. Технологии информационной безопасности XXI века // Специальная техника. – 1998. – № 4 [Электронный ресурс]. – Режим доступа: http://st.ess.ru, свободный. Яз. рус. (дата обращения – 10.10.2012).
6. Bender W., Gruhl D., Morimoto N., Lu A. Tehniques for data Hiding // IBM Systems Journal. – 1996. – V. 35. – № 3&4. – P. 313–336.
7. Скляр Б. Цифровая связь. Теоретические основы и практическое применение: Пер. с англ. – М.: Вильямс, 2003. – 1104 с.
8. Коробейников А.Г., Кувшинов С.С., Блинов С.Ю., Лейман А.В., Нестеров С.И. Разработка стеганоалгоритма на базе форматных и пространственных принципов сокрытия данных // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – № 1 (77). – С. 116–119.
9. Марр Д. Зрение: информационный подход к изучению представления и обработки зрительных образов: Пер. с англ. – М.: Радио и связь, 1987. – 400 с.

Коробейников Анатолий Григорьевич – Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н.В.Пушкова РАН, доктор технических наук, профессор, зам. директора, Korobeynikov_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, stasblino@yandex.ru

Лейман Альберт Владимирович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, аспирант, allxxl@yandex.ru

Кутузов Илья Михайлович – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, formalizator@gmail.com

8 ЭКОНОМИКА И ФИНАНСЫ. МЕНЕДЖМЕНТ

УДК 330

МЕТОДИЧЕСКИЕ ПОЛОЖЕНИЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ ДЕЯТЕЛЬНОСТИ ТЕЛЕКОММУНИКАЦИОННОГО ПРЕДПРИЯТИЯ

О.А. Цуканова, Т.С. Лаврентьева

Уточнены методические положения оценки эффективности деятельности для телекоммуникационных предприятий, предложены подходы, предусматривающие определенную последовательность оценки эффективности и использование системы идентификаторов, отражающих уровень достигнутой эффективности.

Ключевые слова: телекоммуникационные услуги, эффективность, эффект, внутренняя эффективность, внешняя эффективность, EBITDA, рентабельность, экономическая эффективность, социальная эффективность.

Введение

Активное развитие телекоммуникаций на сегодняшний день является условием роста экономики и предопределяет вектор на становление постиндустриального общества. Развитие телекоммуникаций непосредственно связано с развитием экономики: чем выше уровень развития экономики, тем выше уровень развития телекоммуникаций. При этом наблюдается также связь в обратном направлении – рост телекоммуникационной отрасли, помимо увеличения числа рабочих мест, за счет технологичности и инновационности характера деятельности увеличивает эффективность других отраслей экономики.

Таким образом, телекоммуникационный рынок – один из самых динамично развивающихся сегментов высокотехнологической отрасли в России, что связано, в первую очередь, с ростом популярности сети Интернет. По прогнозам экспертов, к концу 2014 г. проникновение Интернета в среду 25–44-летних пользователей достигнет 97–99%. При этом в Санкт-Петербурге и в целом на Северо-Западном направлении диагностируется самый высокий уровень информатизации в России [1].

На сегодняшний день можно утверждать, что от уровня развития телекоммуникаций зависит темп экономического роста рынка услуг, скорость инновационного развития общества. Следовательно, проблема анализа эффективности деятельности предприятий, функционирующих в сфере телекоммуникаций, является актуальной.

Непрерывная и комплексная оценка эффективности деятельности хозяйствующих субъектов необходима для того, чтобы оценить динамику развития предприятия, разработать комплекс мер по минимизации влияния негативных факторов. При этом требуется разрабатывать методические положения для оценки с учетом специфики деятельности предприятия, так как общие подходы и методики, которые используются для всех предприятий, обладают минимальной практической ценностью.

Целью настоящей работы является формирование методических положений для системной оценки эффективности деятельности телекоммуникационных предприятий в России.

Сущность понятия «эффективность хозяйственной деятельности»

Эффективность – это комплексное понятие, включающее в себя различные типы и аспекты его оценки. Эффективность тесно связана с понятием «результативность» и должна рассматриваться как система индикаторов и характеристик, в совокупности отражающих это понятие, включая:

- результативность как степень достижения системой поставленных перед ней целей;
- экономичность как степень оптимального использования системой нужных ресурсов;
- качество как степень соответствия сервиса требованиям, запросам и ожиданиям;
- прибыльность как соотношение между валовыми доходами и суммарными издержками;
- производительность как соотношение предоставляемых услуг и уровень затрат;
- инновационное сопровождение деятельности, учитывающее современные тенденции отрасли.

Таким образом, в соответствии с интегральным подходом к общей оценке эффективности деятельности коммерческих организаций, эффективной может быть признана деятельность организации, имеющая инновационный характер, обеспечивающая заданный уровень прибыльности и поддерживающая соответствующий уровень качества.

Общее понятие эффективности интегрирует в себе следующие виды частных ее показателей:

- социальная эффективность – степень достижения поставленных целей, определенного социального эффекта в процентном или абсолютном выражении (как конкретный результат в отношении всего населения или группы людей, получивших выгоду), а также степень удовлетворения общественным запросам (уровень информатизации, степень охвата абонентов предоставляемых услуг);
- экономическая эффективность, выражающаяся в отношении полезных конечных результатов функционирования объекта предпринимательской деятельности к затраченным ресурсам;

– социально-экономическая эффективность – соотношение полученного социального эффекта и произведенных затрат;
– экономия, учитывающая мероприятия по снижению и более экономичному потреблению трудовых, финансовых и материальных ресурсов.

Для анализа эффективности деятельности телекоммуникационного предприятия следует учитывать ряд особенностей. Во-первых, это специфика создаваемого продукта, который, в отличие от продукта промышленного производства, не имеет вещественной формы. Во-вторых, это неотделимость во времени процесса потребления услуг связи от процесса их производства. Особенно ярко это проявляется в телефонной связи, где сам процесс передачи телефонного сообщения происходит с участием абонентов, т.е. совпадает с процессом потребления. Третьей особенностью телекоммуникаций является то, что, в отличие от промышленности, где предмет труда подвергается вещественному изменению, поступает в сферу обращения, становится товаром и только затем потребляется, в производственном процессе связи информация как предмет труда должна подвергаться только пространственному перемещению, т.е. изменению ее местоположения. Всякое другое изменение означает ее искажение, потерю потребительских свойств и потребительной ценности и наносит ущерб пользователю. В-четвертых, существует прямая зависимость успешности проектов сервисов и деятельности по оказанию телекоммуникационных услуг от уровня инфраструктурной мощности организации – развитости сети и телекоммуникационной инфраструктуры.

Особенности комплексной оценки эффективности деятельности предприятия

По мнению авторов, при оценке эффективности деятельности телекоммуникационного предприятия целесообразно использовать как количественные, так и качественные показатели.

Использование комплекса показателей для анализа эффективности деятельности продуцента услуг с целью ее повышения предполагает нескольких уровней анализа, результаты которого дадут объективное представление об общей эффективности деятельности предприятия на рынке.

Первый уровень анализа содержит исследование количественных показателей, отражающих эффект и эффективность деятельности предприятия в целом. По итогам исследования хозяйственной деятельности типичного представителя рынка телекоммуникационных услуг авторами была предложена система качественных и количественных показателей оценки эффективности деятельности предприятия.

В качестве количественных характеристик целесообразно использовать показатели ликвидности и платежеспособности, рентабельности и доходности, деловой активности, финансовой устойчивости.

Рентабельность – показатель экономической эффективности деятельности организации, характеризующий уровень отдачи от затрат и степень использования средств [2]. На телекоммуникационных предприятиях для оценки итогов хозяйственной деятельности используют следующие виды рентабельности: рентабельность активов, рентабельность собственного капитала, валовая рентабельность реализованной продукции, операционная рентабельность реализованной продукции, чистая рентабельность реализованной продукции.

Специфичным и важным показателем, недостаточно описанным в отечественной литературе, следует считать показатель EBITDA (Earnings Before Interest, Tax, Depreciation and Amortization), характеризующий финансовый результат хозяйственной деятельности организации до вычета обслуживания долга, амортизационных отчислений и налога на прибыль.

Показатель EBITDA используется для долгосрочной оценки эффективности операций компании. Поскольку EBITDA представляет прибыль компании, освобожденную от влияния налогового окружения и способов финансирования, а также от влияния организации учета (в части амортизации), он позволяет успешно сравнивать эффективность работы компании в разные периоды. Отражая общее состояние бизнеса в компаниях, косвенно характеризуя уровень менеджмента и его способность эффективно управлять производственно-коммерческой деятельностью в рыночной среде, EBITDA представляет особый интерес для инвесторов, в том числе иностранных.

Одной из важных составляющих эффективности экономической системы является эффективность капитальных вложений. Она выражается отношением полученного эффекта к капитальным вложениям, вызвавшим этот эффект, т.е. экономический эффект, приходящийся на один рубль инвестиций, обеспечивших этот эффект.

Второй уровень должен содержать анализ качественных показателей эффективности деятельности предприятия, в целом отражающих общую атмосферу внутри предприятия. К таким показателям, которые могут быть получены посредством проведения опроса, можно отнести степень удовлетворенности сотрудников членством в трудовом коллективе и справедливостью оплаты труда; мнение о конкурентоспособности продукции и услуг, их цене и качестве; точка зрения работников на условия труда и социальные программы внутри предприятия.

Третий уровень исследования должен содержать качественные результаты анализа информационной эффективности деятельности продуцента услуг – данные об использовании хозяйствующим субъ-

ектом современных информационных продуктов для оптимизации управления ресурсами, о работе с потребителями, поставщиками, степени интеграции в сетевую экономику. Недостаточный учет уровня информационной эффективности хозяйственной деятельности предприятия в перспективе приведет к негативным последствиям – отставанию от передовых отраслевых и мировых тенденций, отсутствию возможности снижения себестоимости продукции в долгосрочном периоде. При этом достаточно сложно провести количественный анализ эффективности внедрения информационных систем на предприятии в краткосрочном периоде времени.

Четвертый уровень анализа включает данные о степени удовлетворенности хозяйственной деятельностью предприятия групп первичных стейкхолдеров – покупателей (физических и юридических лиц), поставщиков услуг, оборудования и пр., а также посредников (рекламных агентств, контрагентов по оказанию маркетинговых и консалтинговых услуг, финансовых посредников, предприятий по оказанию различного рода инфраструктурных услуг и пр.). Уровень удовлетворенности покупателей товаров и услуг телекоммуникационной отрасли зависит от конкурентоспособности продукции, т.е. соотношения ее цены и качества. В ходе анализа конкурентоспособности выпускаемой продукции и услуг целесообразно сравнить продукцию конкурентов с большими объемами сбыта на рынке с продукцией конкурентов, имеющих меньший объем сбыта, на основе результатов которого можно определить направления улучшения конкурентной позиции выпускаемых товаров и услуг. Степень удовлетворенности поставщиков и посредников зависит, в первую очередь, от оперативности и своевременности оплаты за предоставленные товары и услуги. Степень удовлетворенности первичных групп стейкхолдеров находится в тесной зависимости от непосредственных количественных показателей эффективности деятельности продуцента сферы услуг, полученных на первом этапе исследования [3, 4].

Эффективность сервисной деятельности – это способность предприятия оказывать оптимальное количество услуг определенного качества по сравнению с ожиданиями клиентов и имеющимся спросом на данные услуги при оптимальном использовании затрат.

Система анализа эффективности предоставления услуг в сфере телекоммуникации и сервисной деятельности представляет собой группу идентификаторов, состоящую из следующих основных элементов: затраты, результаты предоставления услуг, процесс обслуживания, спрос на услуги, внутренняя эффективность, внешняя эффективность, потенциальная эффективность, интегральная эффективность сервисной деятельности. Данные комплексные показатели являются также интегральными и рассчитываются на основе введенных специальных метрик, учитывающих особенности инфраструктуры предоставляемых услуг, например, степень удовлетворенности клиентов, эффективность управления инцидентами и пр.

Внутренняя эффективность сервисной деятельности – это оптимальное использование затрат в данный период времени по сравнению с возможными затратами при постоянной величине объема оказываемых услуг. Затраты включают в себя две составляющие: затраты оператора и затраты клиента на реализацию базового соглашения о предоставляемом уровне услуг.

Внешняя эффективность сервисной деятельности – это оптимальная величина выпуска услуг определенного качества в данный период времени по сравнению с ожиданиями клиентов организации при постоянной величине затрат на производство данного объема услуг. Показатель может рассчитываться на основе таких метрик, как соотношение мощности каналов и затрат на содержание инфраструктуры [5].

Результаты процесса обслуживания двойственны, так как они включают в себя количественный результат (объем услуг) и качественный результат (качество услуги и процесса обслуживания). Механизм управления услугами должен базироваться на оценке внешней эффективности (чем выше качество предоставляемых услуг и сервиса, тем выше финансовые показатели эффективности).

Потенциальная эффективность сервисной деятельности – это способность предприятия производить максимально возможный объем услуг определенного качества по сравнению с имеющимся спросом на данные услуги. Количественный результат сервисной деятельности обусловлен спросом. Если спрос совпадает с производственной мощностью оператора, то потенциальная эффективность оптимальна. Если спрос чрезмерный, то потенциальная эффективность выходит за пределы оптимума, т.е. организация выполняет больший объем услуг, чем потенциально возможно, однако здесь может возникнуть негативный эффект в виде ухудшения качества услуг. Если спрос ниже потенциальной мощности, то потенциальная эффективность низка по сравнению с оптимальной.

Таким образом, внутренняя (затратная) эффективность или эффективность использования ресурсов – это одна сторона сервисной эффективности, внешняя эффективность или способность генерировать выручку из этих ресурсов – другая сторона, при этом воздействие на эффективность сервиса оказывает и потенциальная эффективность – величина спроса на услуги телекоммуникационного предприятия.

Показатели-метрики, которые целесообразно использовать при таком подходе: число случаев нарушений предоставления сервиса, процент требуемых технологических изменений в инфраструктуре сервиса, затраты на предоставление услуг, степень удовлетворенности клиентов, рассчитанная как по количеству качеству обращений в компанию, объему продаж и т.д.

Пятый уровень анализа должен содержать данные о социальной эффективности деятельности продуцента телекоммуникационных услуг в обществе, его полезности для разных физических и юридических лиц, степени удовлетворенности групп вторичных стейкхолдеров.

Данный подход к оценке эффективности может в полной мере, по мнению авторов, использоваться только с точки зрения применения в компаниях политики клиентоориентированности, постоянного анализа уровня сервиса, систематической политики повышения эффективности использования технологий и ресурсов.

Заключение

Таким образом, в работе авторами перечислена совокупность показателей для оценки эффективности хозяйственной деятельности телекоммуникационного предприятия, предложена последовательность оценки эффективности. При этом анализ эффективности проводится как с точки зрения экономической целесообразности, так и с точки зрения социальной ориентации предприятия. Данный подход можно применять на практике для комплексного анализа эффективности деятельности хозяйствующего субъекта, задействованного в сфере услуг телекоммуникаций.

Литература

1. Аналитический бюллетень «Интернет в России. Методика и основные результаты исследования ФОМ». – 2011. – В. 32. [Электронный ресурс]. – Режим доступа: http://bd.fom.ru/report/cat/smi/smi_int/pressr_130611/, свободный. Яз. рус. (дата обращения 09.01.2012).
2. Райзберг Б.А., Лозовский Л.Ш., Стародубцева Е.Б. Современный экономический словарь. – 5-е изд., перераб. и доп. – М.: ИНФРА-М, 2006. – 495 с.
3. Цуканова О.А. Особенности применения различных инструментов продвижения услуг на рынке // Научно-технический вестник СПбГУ ИТМО. – 2011. – № 4 (74). – С. 142–145.
4. Цуканова О.А. Стратегическое управление предприятиями сферы рекламно-издательских услуг: особенности функционирования, направления рационализации хозяйственной деятельности. – СПб: Изд-во ООО «Студия «НП-Принт», 2012. – 138 с.
5. Бовыкин В.Н. Новый менеджмент. – М.: Экономика, 2003. – 321 с.

Цуканова Ольга Анатольевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат экономических наук, доцент, zoa1999@mail.ru

Лаврентьева Татьяна Сергеевна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, студент, t.nelarina@mail.ru

УДК 336.011, 336.71

ИННОВАЦИОННАЯ ДЕЯТЕЛЬНОСТЬ СУБЪЕКТОВ БАНКОВСКОЙ СИСТЕМЫ КАК ФОРМА РЕАЛИЗАЦИИ ЭКОНОМИЧЕСКИХ ИНТЕРЕСОВ

Т.Я. Красноперова

Рассматривается актуальная проблема инновационного развития банковской системы Российской Федерации как формы реализации экономических интересов. Предлагаются пути активизации инновационной деятельности посредством координации экономических интересов субъектов банковской деятельности.

Ключевые слова: инновационное развитие, банковская система, взаимодействие экономических интересов, инновации, рентабельность активов.

Введение

Характер обусловленности инновационной деятельности российской банковской системы экономическими интересами представляет особенный интерес в связи с государственной и экономической значимостью банковской деятельности для становления рыночной экономики в России. Ключевым понятием в анализе специфики категории экономических интересов и организации инновационной деятельности в банковской сфере является финансовое посредничество (financial intermediary). Согласно теории финансового посредничества, банки рассматриваются российскими и зарубежными исследователями как экономические институты, осуществляющие трансформацию финансовых контрактов, при этом проис-

ходит взаимодействие экономических интересов различных субъектов, главным координатором которых выступает непосредственно банк, а опосредованно – государство. Данный подход получил развитие в работах таких российских экономистов, как Ю.И. Коробов, А.М. Тавасиев, Л.В. Кох, К.Р. Тагирбеков [1–4] и других. Деятельность банков как финансовых посредников является сферой концентрации экономических интересов всех экономических субъектов, причем как на национальном, так и на общемировом уровне, в связи с усилением влияния мировой финансовой глобализации и конкуренции с иностранными банками. Трансформация экономических интересов хозяйствующих субъектов в условиях переходной рыночной экономики, современный механизм реализации, их противоречия и способы разрешения анализируется в работах М.А. Мариновой, А.В. Тарасова, И.Н. Урусовой и других [5–7].

Дальнейшее развитие банковской системы и ее отдельных элементов (банков) требует анализа специфики их экономических интересов и разработки методов нейтрализации противоречий экономических интересов в целях обеспечения инновационного развития. Цель данного исследования – определить место инновационной деятельности банков в удовлетворении их экономических интересов и пути их координации.

Специфика экономических интересов банков

Экономические интересы банков возникают в объективной реальности, но форма их реализации осложнена внешними и внутренними факторами, что вносит элемент субъективности и неопределенности. Форм реализации интересов, кроме непосредственно инновационной деятельности, может быть множество, что порождает неопределенность, а это, в свою очередь, – риски. Наличие неопределенности и различного рода рисков обусловлено субъективной стороной экономических интересов банков и их несовпадением различной степени, вплоть до противоречий. Система экономических интересов банков и банковской системы действует эффективно до тех пор, пока объективный характер соответствует производственным отношениям и учитывается всеми субъектами банковского процесса. Степень расхождений, вплоть до противоречий экономических интересов банка и взаимодействующих с ним субъектов, обусловлена преобладанием субъективной стороны экономических интересов, результатом чего является повышение уровня неопределенности, влияющей на состояние инновационной среды и уровень инновационного потенциала.

Рис. 1. Взаимосвязь инновационной деятельности и экономических интересов

Наличие неопределенности в рамках реализации экономических интересов банка предопределяет неизбежное наличие рисков в банковской деятельности. Риски банка детерминированы значительно бóльшим количеством переменных, чем у других хозяйствующих субъектов, в связи с финансовым по-

средничеством банка. Удовлетворяя свои интересы, банк одновременно должен обеспечить экономические интересы клиентов. Сложность банковской деятельности состоит в том, что она должна учитывать и обеспечивать одновременно экономические интересы нескольких субъектов – государства и общества (в виде налогов и содействия экономическому развитию хозяйствующих субъектов), собственников денежных средств в распоряжении банка, заемщиков и собственно банка. Любой из хозяйствующих субъектов, реализуя свои экономические интересы, вступает в систему экономических отношений через затраты различных видов ресурсов, сопоставляя их с полученными результатами. Основной мерой реализации экономических интересов является экономическая эффективность. Инновационная деятельность, как и экономические интересы, проявляется в столкновении объективно существующих рисков с их субъективной оценкой. В результате каждый субъект лишен однозначно заданных параметров, обеспечивающих успешную реализацию его экономических интересов, так как действует в условиях неопределенности и альтернативности. Банковская система и конкретные банки, действуя в сфере услуг, а не в сфере товарного производства, подвержены рискам в большей степени, так как противоречия экономических интересов взаимодействующих субъектов усугубляются неосязаемостью и сложностью прогнозирования результатов банковских услуг во времени, что особенно остро проявляется в процессе инновационной деятельности. Следовательно, инновационные процессы в банковской деятельности позволяют реализовать экономические интересы не только банка-инноватора, но и целого комплекса других взаимодействующих субъектов – клиентов: заемщиков, вкладчиков; государства и т.д. (рис. 1).

Противоречия экономических интересов взаимодействующих субъектов в сфере банковских услуг обостряются в связи с тем, что объектом экономических интересов является не товар, удовлетворяющий какую-либо конкретную потребность, а деньги как всеобщий эквивалент, причем основная их масса является собственностью клиентов, а не банков. Деньги представляют совокупный триединый экономический интерес для коммерческих банков: это одновременно и средство удовлетворения экономических интересов, и предмет труда, и средство труда. Это обстоятельство усиливает значение инновационного развития банков и в целом банковской системы в процессе удовлетворения их экономических интересов.

Инновационная деятельность как форма удовлетворения экономических интересов субъектов банковской деятельности: проблемы и пути решения

Движущей силой инновационного процесса является удовлетворение экономических интересов банков и их клиентов, вызванное рыночной конкуренцией. Преодоление расхождений экономических интересов взаимодействующих субъектов посредством инновационной деятельности является условием сохранения их целостности и действенности. Эти расхождения не должны доходить до уровня противоречий, при котором эффективная инновационная деятельность невозможна.

Наряду с инновационной деятельностью, важным фактором, влияющим на удовлетворение экономических интересов, является государственная поддержка некоторых банков. Банки с государственным участием в уставном капитале или финансовой поддержкой со стороны государства могут позволить себе превалирование собственных экономических интересов в ущерб интересам партнеров, что нарушает рыночную конкуренцию и снижает их заинтересованность в инновационном развитии.

Внедрение инноваций, с одной стороны, увеличивает риски в связи с неопределенностью их результатов, с другой стороны, снижает их в связи с получением конкурентного преимущества инноватора. Уровень проявления банковских рисков при внедрении инноваций зависит от степени расхождения экономических интересов банка и взаимодействующего с ним субъекта, но с корректировкой на соотношение уровня их экономической безопасности или государственного участия в капитале банка. Степень удовлетворения экономических интересов субъектов прямо пропорционально зависит от уровня его экономической безопасности.

На взгляд автора, взаимодействие экономических интересов банка и других субъектов в условиях рыночной экономики в основном поддается саморегулированию под действием экономических законов, но степень государственного вмешательства необходимо регулировать. Реализация экономических интересов банков подчиняется следующим основным принципам: адекватность экономическим закономерностям, системность, комплексность, приоритетность, взаимозависимость. Эти же принципы важны и для осуществления полноценной эффективной инновационной деятельности. В отсутствии единой индустриальной политики банки не имеют стимулов активно участвовать в модернизации предприятий и отраслей и финансировании их инновационного развития. Посыл же к государству очевиден: безотлагательно определиться с формированием индустриальной политики и не бояться поддерживать рост экономики в целом, а не только отдельных монопольных бизнесов [8].

В последние годы наблюдается, с одной стороны, централизованная выдача льготных кредитов подконтрольным государству банкам в целях стимулирования ими инновационных процессов в экономике, а с другой стороны, завышение банками реальных процентных ставок по кредитам и присвоение рисковой надбавки, которая должна доставаться заемщику. В результате тормозится процесс воспроизводства инвестиций, снижается эффективность инновационной деятельности предприятий или она ста-

новится экономически нецелесообразной, если стоимость кредита превышает или дает незначительную маржу в сравнении с уровнем рентабельности активов предприятий (рис. 2). По данным Росстата и Центрального банка России, уровень рентабельности активов превышал средние процентные ставки по кредитам сроком до 1 года только в период 2006–2007 г.г., т.е. кредиты банков были экономически выгодными для предприятий и могли способствовать их модернизации.

Рис. 2. Соотношение уровня рентабельности активов предприятий и средних процентных ставок по кредитам сроком до 1 года нефинансовым предприятиям, %

Таким образом, условия кредитования не соответствуют экономическим интересам заемщиков и не могут стимулировать их инновационное развитие, что, в свою очередь, снижает инновационные возможности банков. Вышеуказанные действия свидетельствуют об отсутствии согласованных действий элементов экономической системы и деформации координации экономических интересов.

Отечественная банковская система вслед за государством в лице Центрального банка Российской Федерации, размещающего активы в ценные бумаги иностранных эмитентов, не дает возможности реальному сектору экономики производить модернизацию производства, разрабатывать и внедрять инновации, назначая банковский процент, превышающий уровень рентабельности предприятий реального производства. Финансирование эффективных инвестиционных проектов технического перевооружения предприятий способно обеспечить их оздоровление, а вместе с тем улучшить кредитную базу банков. Расхождение вплоть до противоречий экономических интересов банков и субъектов реального сектора экономики, выражающееся в кредитном «голоде» предприятий, недостаточном финансировании модернизации и внедрения инноваций является социально-экономической проблемой, так как следствием его является сокращение рабочих мест и понижение уровня заработной платы, сокращение платежеспособности населения. Недостаточность «длинных» денег у банков в связи со «стерилизацией» государством денежной массы вопреки экономическим интересам субъектов национальной экономики приводит к тому, что банки вынуждены в целях выживания отдавать преимущество кредитованию физических лиц, так как процентные ставки почти в два раза выше. В результате объемы кредитования физических лиц растут на 20–30% быстрее, чем предприятий. Средства, которые могли бы содействовать инновационному развитию предприятий, создавать высокооплачиваемые места для рабочих, используются банками для удовлетворения субъективных экономических интересов.

Заключение

Таким образом, необходима единая стратегия гармонизации экономических интересов всех субъектов банковской деятельности, которая будет направлена на обеспечение финансирования инновационной деятельности как исходной базы экономического сотрудничества. Экономические интересы – движущая сила инновационной деятельности, которая является одной из форм ее реализации. Детерминантом инновационного процесса выступает инновационная среда, т.е. условия внешней и внутренней среды бытия субъекта, носящие объективно-субъективный характер. Система экономических интересов и инноваций действует эффективно до тех пор, пока преобладает их объективный характер, воплощающий всеобщий интерес. Необходимо формирование инновационной среды, способствующей внедрению инноваций как одного из факторов удовлетворения экономических интересов, а также координация экономических интересов субъектов банковской системы и национальной экономики.

Литература

1. Банковские операции: Учебное пособие / Под ред. Ю.И. Коробова. – М.: Магистр, 2007. – 446 с.
2. Кох Л.В. Эффективность инновационной деятельности банка: теория и методология. – СПб: Изд-во Политех. ун-та, 2008. – 194 с.

3. Банковское дело: управление и технологии / Под ред. А.М. Тавасиева. – М.: ЮНИТИ-ДАНА, 2001. – 863 с.
4. Организация деятельности коммерческого банка: Пособие / Под ред. К.Р. Тагирбекова. – М.: ИНФРА-М: «Весь мир», 2003. – 848 с.
5. Маринова М.А. Экономические интересы в процессе их трансформации в условиях переходной российской экономики: Дис. … канд. экон. наук: 22.00.03. – М., 2001. – 155 с.
6. Тарасов А.В. Экономические интересы субъектов хозяйственной деятельности: противоречия и способы их разрешения: Дис. … канд. экон. наук: 08.00.01. – Воронеж, 2006. – 183 с.
7. Урусова И.Н. Современный механизм реализации экономических интересов хозяйствующих субъектов: монография. – Чебоксары: Изд-во Чувашского государственного университета, 2006. – 123 с.
8. Доклад Президента Ассоциации российских банков Г.А. Тосуняна // Деньги и кредит. – 2012. – № 5. – С. 7–8.

Красноперова Танзиля Яуфатовна – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Международный институт бизнеса и права, кандидат экономических наук, доцент, krasntat@mail.ru

9 НОВЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

УДК 519.86

АРХИТЕКТУРА ЛИНГВИСТИЧЕСКОГО ТРЕНАЖЕРА ДЛЯ ЭКСПРЕСС-ОСВОЕНИЯ НАВЫКОВ ОБЩЕНИЯ НА ИНОСТРАННОМ ЯЗЫКЕ

В.В. Киселев, О.Е. Елисеева, Ю.И. Ковалёнок, Ю.Н. Хитрова

Рассматривается архитектура лингвистического тренажера, реализуемого в виде сетевого программного комплекса для обеспечения возможности экспресс-освоения навыков общения на иностранном языке. Основное внимание уделено определению формальных моделей, лежащих в основе центрального компонента архитектуры, обеспечивающего реализацию гибких стратегий обучения вне зависимости от применяемой методики обучения.

Ключевые слова: лингвистический тренажер, модель обучения, модель учебных воздействий, модель обучаемого.

Введение

Известно, что самым сложным в процессе изучения иностранного языка является овладение речевыми навыками – говорением и аудированием. Единственно возможным решением указанной проблемы является постоянная тренировка произношения звуков, слов, фраз иностранного языка. Однако это требует дополнительных временных затрат, а времени в условиях современного, интенсифицированного, динамичного общества всегда не хватает. Применение компьютерных технологий позволяет повысить эффективность обучения путем использования тренажеров, которые, с одной стороны, дают возможность учащимся тренировать навыки говорения в любое удобное для них время, а с другой – в значительной степени облегчают труд преподавателей языка. Для ускорения процесса «выхода в речь» разрабатываются всевозможные интенсивные методики, часть из которых базируется на том, что основу изучения иностранного языка составляет вводный фонетико-корректировочный или фонетико-грамматический курс, в рамках которого учащиеся, наряду с изучением звукового состава изучаемого иностранного языка, одновременно овладевают базовым лексическим и грамматическим набором знаний. Использование указанных методик позволяет учащимся практически после первых вводных занятий не просто произносить отдельные звуки, но и строить простейшие фразы, рассказывать о себе и пр.

В работе не ставится задача анализа тех или иных методик обучения языку. Целью работы является построение архитектуры лингвистического тренажера (ЛТ) и рассмотрение ее основных компонентов. ЛТ реализуется в виде сетевого программного комплекса с использованием технологий искусственного интеллекта и обеспечивает освоение навыков понимания на слух и беглого говорения на иностранных языках. Одной из основных особенностей ЛТ является реализация гибких стратегий обучения для различных учащихся в соответствии с их возможностями, успехами и пожеланиями. Кроме того, как и в некоторых современных системах обучения иностранному языку, обеспечивается возможность записи и анализа речи учащегося с целью повышения эффективности обучения.

Архитектура лингвистического тренажера

На рис. 1 представлена архитектура разрабатываемого ЛТ, которая разбита на 4 основных блока.

1. Блок реализации обучения, в рамках которого осуществляется моделирование процесса обучения иностранному языку, а именно: изучение фонетики, лексики и грамматики иностранного языка на теоретическом и практическом уровнях. Основное внимание при этом уделяется именно практической составляющей, которая заключается в выполнении учащимся ряда упражнений для тренировки навыков говорения и аудирования.
2. Блок служебных функций, в рамках которого реализуется ряд модулей для обеспечения надежного функционирования системы в целом.
3. Блок формирования контента, в задачи которого входит обеспечение возможностей гибкого пополнения необходимой учебной информацией модели учебных воздействий в рамках блока реализации обучения.
4. Блок интерфейса с пользователем, в рамках которого учащимся и преподавателям предоставляются разные возможности: учащимся – для взаимодействия с блоком реализации обучения ЛТ; преподавателям – для наполнения ЛТ учебными материалами посредством блока формирования контента.

Как указывалось выше, особенностью рассматриваемого ЛТ является обеспечение гибких стратегий обучения для различных учащихся. Исходя из этого, в настоящей работе рассмотрим более подробно формальные модели, которые являются основой блока реализации обучения и обеспечивают указанную функциональность.

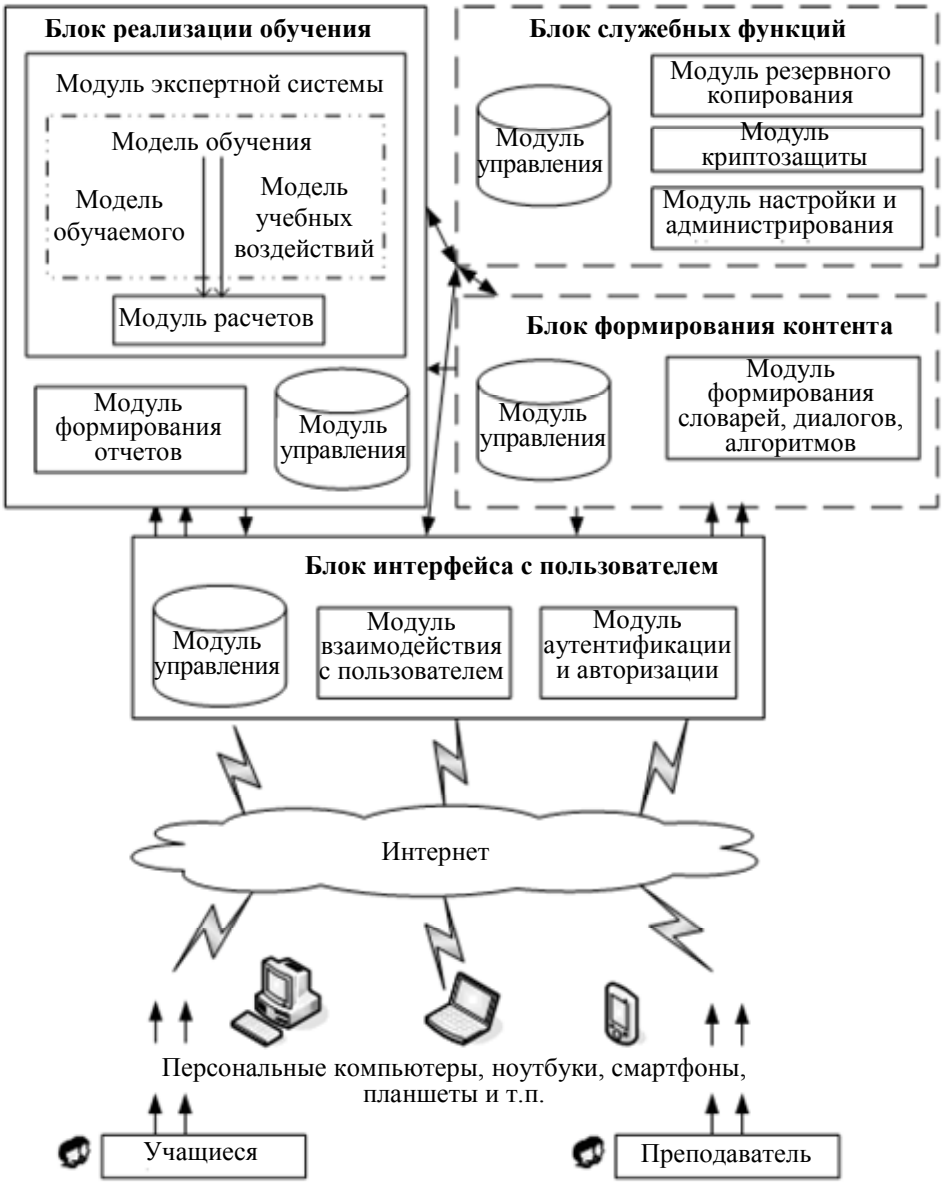


Рис. 1. Архитектура лингвистического тренажера

Формальные модели и технологии

Прежде чем перейти к формальному рассмотрению ЛТ, заметим, что в процессе разработки авторами используются уже ставшие традиционными представления и подходы к созданию так называемых интеллектуальных (экспертных) обучающих систем [1–4]. Однако в связи с тем, что ЛТ предназначен для использования в глобальной сети Интернет и реализуется на базе существующих сетевых технологий (в частности, серверная часть функционирует на облачном сервере Windows Azure), которые имеют определенные объективные технические ограничения, для обеспечения работы системы в режиме реального времени мы несколько упрощаем упомянутые подходы. В связи с этим в задачи модуля экспертной системы входит лишь коррекция курса обучения и формирование оптимальной стратегии для максимально быстрого прохождения курса. Так, например, в рамках избранного подхода успевающим учащимся предлагается меньшее количество упражнений, в отличие от менее успевающих.

В основу проектирования ЛТ положена формальная модель обучения LM (learning model), в рамках которой решаются задачи построения оптимальной стратегии обучения для каждого учащегося. В рамках модели обучения строятся модель учебных воздействий TAM (training actions model) и модель обучаемого SM (student model).

Модель учебных воздействий TAM представляется в виде ориентированного графа, включающего узлы трех типов: T – теоретический компонент; P – практическое (тренировочное) упражнение; C – контролирующее упражнение для оценки уровня достижений учащегося, на основе которого осуществляется поиск пути на графе, соответствующего оптимальному пути прохождения учебного курса конкретным

обучаемым. Каждый узел трактуется в данной модели как отдельный урок курса, который является минимальным учебным воздействием. Урокам (узлам графа), в свою очередь, ставятся в соответствие векторы составляющих процесса изучения иностранного языка:

– **a** (v, g, ph) – один или несколько аспектов языка: лексика (словарь) (v), грамматика (g), фонетика (ph);
– **n** (A, S, R, W) – вид приобретаемых знаний и навыков: аудирование (A), говорение (S), чтение (R), письмо (W).

Координаты обоих указанных векторов могут принимать одно из двух значений: 0 (нет) или 1 (есть). Так, например, если координата v вектора **a** для некоторого узла равна 0, это означает, что в соответствующем уроке нет новой словарной составляющей. Если же указанная координата равна 1, то в данном уроке изучаются новые слова.

На рис. 2 приведен пример графов, описывающих последовательность уроков. На графе, изображенном на рис. 2, а, совокупность нескольких дуг, выходящих из одного и того же узла в различные узлы уроков одного типа (например, из урока теоретического типа $T1$ в узлы тренировочных упражнений $P1$-1, $P1$-2, $P1$-3) означает, что после указанного урока ($T1$) может следовать любой из равнозначных альтернативных уроков ($P1$-1, $P1$-2, $P1$-3). При этом выбор одного из альтернативных уроков может происходить либо случайным образом (например, на начальных стадиях «знакомства» системы с учащимся), либо на основе анализа истории взаимодействия учащегося с ЛТ, которая сохраняется в рамках модели обучаемого. Кроме того, для упрощения рассматриваемой модели и (или) по усмотрению методиста (преподавателя языка, участвующего в создании контента ЛТ) из всех равнозначных уроков может быть выделен один обязательный, который предлагается пройти всем учащимся (рис. 2, б). Все равнозначные уроки связаны друг с другом двунаправленными дугами по схеме «каждый с каждым». Заметим, что для улучшения восприятия на рис. 2 эти дуги изображены в виде ребер и не показаны все бинарные связи.

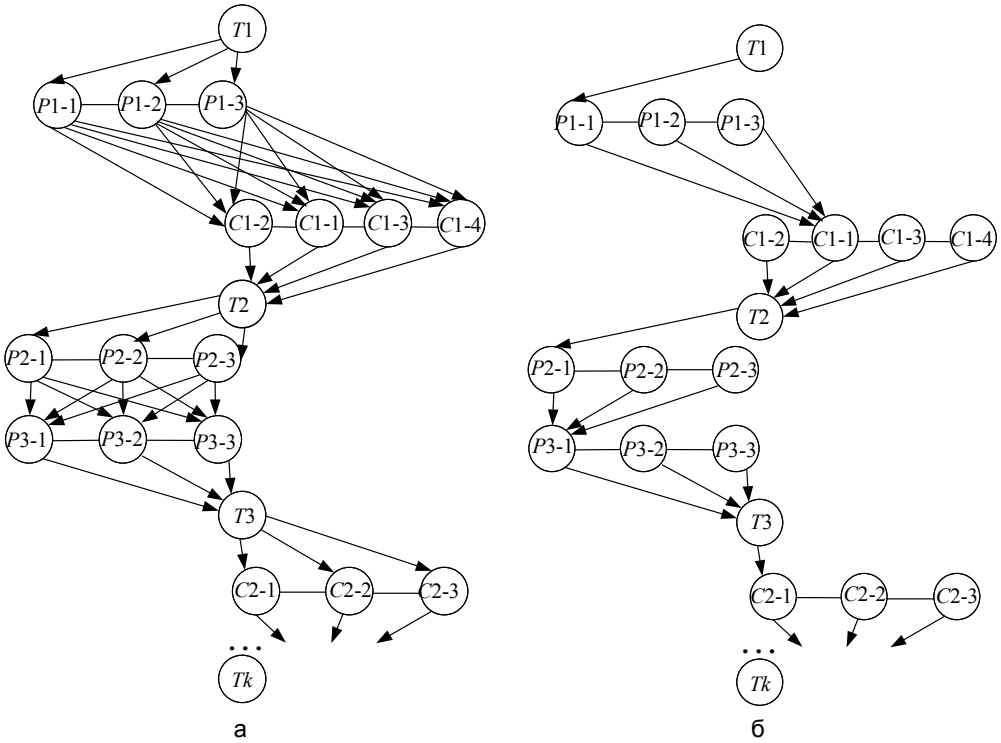


Рис. 2. Графы, описывающие совокупность учебных воздействий: выбор одного из альтернативных уроков (а); выделение обязательных уроков для прохождения учащимися (б)

Как видно из рис. 2, среди уроков теоретического типа Ti альтернативные равнозначные не задаются. Это связано с тем, что теоретический материал рассматривается в системе как строгая, методически обоснованная последовательность обязательных к изучению элементов учебного курса.

Если учащийся успешно выполняет упражнение из совокупности альтернативных равнозначных, то осуществляется переход к следующему уроку (не равнозначному), т.е. к следующему уровню иерархии графа учебных воздействий. Если же упражнение выполнено неуспешно, то учащемуся предоставляется альтернативное равнозначное упражнение (он остается на том же уровне иерархии графа).

Для обеспечения гибкости и индивидуальности обучения, учащимся предоставляется дополнительная возможность пропускать те или иные уроки по своему усмотрению (в этом случае в модели обучаемого делается соответствующая помета, и в случае необходимости учащемуся еще раз предоставляется возможность пройти пропущенный урок).

Описанным выше образом, в зависимости от успехов и пожеланий учащегося, происходит построение оптимального пути прохождения учебного курса индивидуально для каждого пользователя ЛТ. Указанный путь сохраняется в виде истории взаимодействия с ЛТ в модели обучаемого.

Следуя принятой в сфере искусственного интеллекта и создания экспертных систем терминологии, можно в некотором приближении говорить о том, что сформированные на основе рассмотренной модели учебных воздействий ТАМ-структура и содержание учебного материала являются частью общей базы знаний (БЗ) ЛТ.

С учетом того, что целью ЛТ является экспресс-обучение навыкам говорения и аудирования, подавляющее большинство упражнений предполагает тренировку именно этих навыков. Для повышения эффективности данного процесса в системе используются речевые технологии, а именно, средства воспроизведения, записи, анализа и распознавания речи [5]. При этом компоненты воспроизведения речи включаются в состав как теоретической, так и практической части учебного курса. В процессе создания ЛТ для обеспечения качества обучения иностранная речь записывается носителями соответствующего языка. Очевидно, что такое речевое наполнение учебного материала создает дополнительные условия для самостоятельной работы учащихся в индивидуальном ритме. Средства записи и анализа речи используются для расширения разнообразия и повышения эффективности тренировочных и контрольных упражнений. Так, например, благодаря возможностям распознавания речи, для тренировки произношения в ЛТ реализуются упражнения, предполагающие речевой ввод ответов (однозначных) на заданные вопросы. Указанная функциональная возможность существенно отличает предлагаемый ЛТ от большинства существующих на сегодняшний день электронных ресурсов обучения иностранному языку.

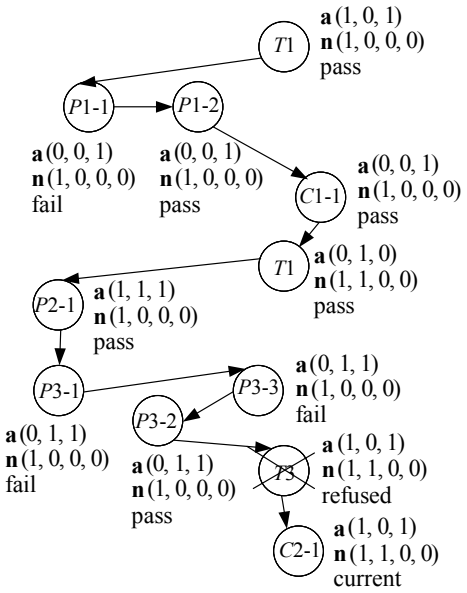


Рис. 3. Пример графа истории взаимодействия обучаемого с ЛТ

В модели обучаемого SM фиксируется следующая информация: личные данные (имя, возраст, пол, начальный уровень владения языком, родной язык, изучаемый язык/языки, логин, пароль и т.п.); история взаимодействия с ЛТ в виде пройденного в процессе обучения пути на графе учебных воздействий с необходимыми пометами о достигнутых результатах, отказах и пр.

На рис. 3 приведен пример графа истории взаимодействия учащегося с ЛТ. Из рисунка видно, какие знания и навыки и в каких аспектах языка приобрел учащийся. Эта информация копируется из БЗ модели учебных воздействий и становится фрагментом БЗ модели обучаемого. На рассматриваемом примере графа также отмечено, с каким успехом были пройдены те или иные уроки: pass – успешно (для теоретических уроков это значение по умолчанию); fail – неуспешно (и вслед за этим осуществлен переход к равнозначному альтернативному упражнению); refuse – учащийся отказался от прохождения урока; current – текущее состояние прохождения курса (либо учащийся в текущий момент времени проходит данный урок, либо остановился на нем и завершил сеанс работы с ЛТ).

Как показано на рис. 1, в состав блока реализации обучения входит модуль расчетов, задачей которого является расчет оптимального пути на графе учебных воздействий. Фактически, следуя описанным выше упрощениям, алгоритмы данного модуля заключаются в анализе успешности выполнения упражнений и выборе следующего урока в соответствии с этим.

Целью модуля формирования отчетов является анализ БЗ модели обучаемого (истории взаимодействия с ЛТ) и вывод пользователю карты достижений, на которой в наглядном виде показано, какие уро-

ки и с каким успехом он прошел, какие знания и навыки при этом приобрел. По данной карте учащийся, в частности, сможет увидеть, где у него есть пробелы в знаниях.

Заключение

Ключевым отличием представленной в работе архитектуры лингвистического тренажера является ее ориентация на гибкие, индивидуальные стратегии обучения, благодаря использованию современных технологий искусственного интеллекта и речевых технологий. В отличие от большинства существующих на сегодняшний день электронных образовательных ресурсов, разработка лингвистического тренажера не зависит от изучаемого языка и используемой методики преподавания. Благодаря этому на основе предложенной архитектуры возможно построение совокупности лингвистических тренажеров для приобретения навыков общения на различных языках.

Настоящая работа выполнена в рамках опытно-конструкторских работ по теме «Создание компьютерного лингвистического тренажера для экспресс-освоения навыков общения на иностранном языке» при финансовой поддержке Министерства образования и науки Российской Федерации, государственный контракт № 07.524.11.4017.

Литература

1. Brusilovsky P. Student Model Centered Architecture for Intelligent Learning Environments // Proc. of Fourth International Conference on User Modeling. – Hyannis, MA, USA: User Modeling Inc., 1994. – P. 31–36.
2. Lane C.H. Intelligent Tutoring Systems: Prospects for Guided Practice and Efficient Learning // Whitepaper for the Army's Science of Learning Workshop. – Hampton, VA, 2006. – P. 1–11.
3. Рыбина Г.В. Обучающие интегрированные экспертные системы: некоторые итоги и перспективы // Искусственный интеллект и принятие решений. – 2008. – № 1. – С. 22–46.
4. Голенков В.В., Гулякина Н.А., Тарасов В.Б., Елисеева О.Е. и др. Интеллектуальные обучающие системы и виртуальные учебные организации. – Минск: БГУИР, 2001. – 488 с.
5. Киселев В.В., Вишнякова М.А. Речевая аналитика в интеллектуальных диалоговых системах // Материалы международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» (Open Semantic Technologies for Intelligent Systems, OSTIS-2011). – Минск: БГУИР, 2011. – С. 231–237.

Киселев Виталий Владимирович – ООО «Речевые технологии», г. Минск, Беларусь, руководитель, kiselev-v@speechpro.com

Елисеева Ольга Евгеньевна – Белорусский государственный университет, кандидат технических наук, доцент, volga.eliseeva@gmail.com

Ковалёнок Юрий Иванович – ООО «Речевые технологии», г. Минск, Беларусь, специалист, kavalionak-y@speechpro.com

Хитрова Юлия Наримановна – ООО «ЦРТ», коммерческий директор, julia@speechpro.com

SUMMARY

REVIEW ARTICLE

ADAPTIVE OUTPUT CONTROL: SUBJECT MATTER, APPLICATION TASKS AND SOLUTIONS

A. Bobtsov, V. Nikiforov

The problem of adaptive output control for parametric and functionally uncertain plants is considered. Application examples illustrating the practical use of the discussed theory are given along with the mathematical formulation of the problem. A brief review of adaptive output control methods, by both linear and non-linear systems, is presented and an extensive bibliography, in which the reader will find a detailed description of the specific algorithms and their properties, is represented. A new approach to the output control problem - a method of consecutive compensator - is considered in detail.

Keywords: adaptive control, robust control, output control, complex dynamic systems, nonlinear systems.

Alexei Bobtsov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Dean, Department Head, D.Sc., Professor, bobtsov@mail.ifmo.ru

Vladimir Nikiforov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Vice-chancellor, D.Sc., nikiforov@mail.ifmo.ru

UP-TO-DATE ISSUES OF CONTROL SYSTEMS THEORY

FROM THE EDITORIAL BOARD

NRU IFMO Vice-chancellor, the Chief Editor of "Scientific and Technical Journal of Information Technologies, Mechanics and Optics", Doctor of Technical Sciences, Professor, Vladimir Nikiforov is going to celebrate his 50th anniversary in March 2013. Mr. V. Nikiforov's colleagues and the editorial board send their jubilee greetings and wish him further creative success! Scientific works, kindly put at editorial disposal by colleagues and pupils of Mr. V. Nikiforov, are collected in this section.

CONTROL APPROACH FOR NONLINEAR PLANT WITH PARAMETRIC UNCERTAINTIES AND INPUT DELAY

A. Bobtsov, A. Pyrkin

Stabilization problem for unstable nonlinear plant with parametric uncertainties and input delay is considered. A new approach is proposed that makes it possible to identify plant parameters and design the stabilization algorithm with state-feedback predictor.

Keywords: control under delay conditions, nonlinear systems, identification, predictor.

Alexei 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, Associate professor, PhD, a.pyrkin@gmail.com

SUBOPTIMAL CONTROL OF NONLINEAR MULTI-AGENT SYSTEMS

I. Furtat

The article deals with the issue of robust suboptimal control for multi-agent systems which model is represented by differential equations with Lipschitz nonlinearity. It is assumed that only scalar subsystem outputs are available to measurement. An algorithm of decentralized control is obtained providing uncertainties compensation and subminimization of performance index with given accuracy.

Keywords: multi-agent system, robust control, decentralized control, optimal control, disturbances compensation.

Igor Furtat – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, cainenash@mail.ru

DYNAMICAL SYSTEMS STABILITY STUDY BY THE PROPERTIES OF THE VECTOR PHASE SPEED FLOW THROUGH A CLOSED CONVEX SURFACE

I. Furtat

The paper deals with stability analysis of autonomous dynamical systems by vector phase speed flow through a closed convex surface. Conditions on the sign of the vector phase speed flow for stability of dynamical systems are proposed.

Keywords: vector field flow, stability, the second Lyapunov's theorem.

Igor Furtat – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, cainenash@mail.ru

HYBRID ALGORITHM OF FREQUENCY IDENTIFICATION FOR MULTI-SINUSOIDAL SIGNAL

A. Bobtsov, A. Vedyakov, S. Kolyubin, A. Pyrkin

Identification problem of unknown frequencies for multi-sinusoidal signal on two harmonics case example is considered. Based on cascade reduction method, identification algorithm is proposed to accelerate the process of parametric convergence of the unknown frequencies for multi-sinusoidal signal.

Keywords: sinusoidal signal, reduction, identification.

Alexei Bobtsov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Dean, Department Head, D.Sc., Professor, bobtsov@mail.ifmo.ru

Alexei Vedyakov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, vedyakov@gmail.com

Sergei Kolyubin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, s.kolyubin@gmail.com

Anton Pyrkin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, a.pyrkin@gmail.com

INTEGRAL DIFFERENTIATING SPEED-GRADIENT ALGORITHMS IN MULTIPLE SYNCHRONIZATION ISSUES FOR VIBRATION UNITS

A. Fradkov, O. Tomchina, V. Galitskaya, D. Gorlatov

Integral differentiating speed-gradient algorithms and their special cases (proportional, integral and proportional-integral speed-gradient algorithms) are described. The definition of an approximate multiple frequency coordinate synchronization is given and integral differentiating algorithms of multiple synchronization control for vibration units are designed. The results of comparative computer simulation for proportional and proportional-integral algorithms are shown, proving that the proposed algorithms have a number of advantages over the known ones.

Keywords: integral differentiating algorithms, speed gradient, multiple synchronization, vibration installation, vibration transport.

Alexander Fradkov – Institute of Problems in Mechanical Engineering, the Russian Academy of Sciences, Head of Laboratory, D.Sc., Professor, fradkov@mail.ru

Olga Tomchina – Saint Petersburg State Polytechnic University, Associate professor, PhD, Associate professor, otomchina@mail.ru

Viktoria Galitskaya – Saint Petersburg State Polytechnic University, postgraduate, Kvika86@mail.ru

Dmitry Gorlatov – Saint Petersburg State Polytechnic University, postgraduate, dprog@mail.ru

DISCRETE-TIME ROBUST CONTROL FOR ROBOT-MANIPULATOR

S. Gusev

The problem of robot-manipulator control is considered under uncertainty in robot dynamic parameters and input noise. The linear discrete-time regulator is proposed for desired motion stabilization of robot-manipulator. It is shown that decreasing the sampling time it is possible to make the tracking error as small as necessary.

Keywords: robot-manipulator, robust control, linear regulator, discrete time.

Sergei Gusev – Saint Petersburg State University, Associate professor, PhD, senior scientific researcher, gusev@ieee.org

ANALYSIS AND PROCESSING OF ELECTROMYOGRAM SIGNALS

K. Zimenko, A. Borgul', A. Margun

A method of electromyogram signals processing and identification for implementation in rehabilitation devices control is given. The method is based on the high-frequency components filtration which improves the signal/noise ratio; also it is based on the wavelet analysis for signal preprocessing and motion type classification by taught artificial neural network. Obtained accuracy of motion type classification is 94%.

Keywords: electromyogram, neural network, signal processing, wavelet transform.

Konstantin Zimenko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, kostyazimenko@gmail.com

Alexander Borgul' – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, borgulalexandr@gmail.com

Alexei Margun – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, alexeimargun@gmail.com

IMPLEMENTATION OF INDUSTRIAL ROBOT-MANIPULATOR SETUPS IN EDUCATIONAL PROCESS

A. Bobtsov, S. Kolyubin, A. Pyrkin

The article deals with the most important aspects of industrial robot-manipulator setups implementation as an educational tool for the workshop part of advanced technologies in robotics and control engineering courses.

Keywords: robot-manipulator, control system, educational applications.

Alexei Bobtsov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Dean, Department Head, D.Sc., Professor, bobtsov@mail.ifmo.ru
Sergei Kolyubin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, s.kolyubin@gmail.com
Anton Pyrkin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, a.pyrkin@gmail.com

ON OPTIMIZATION OF UNMANNED AERIAL VEHICLE APPLICATION FOR SEARCH AND TRACKING OF OBJECTS

R. Abdulov, N. Abdullayev, H. Asadov

Critical analysis of the known method of signal arrival time difference on the basis of UAV is carried out. Research results show the possibility of this method optimization and further synthesis of the new one. Optimization of the known search and tracking method based on the group of UAV makes it possible to give practical recommendations on the synthesized new method application.
Keywords: optimization, UAV, search, location, measurements, correlation.
Rauf Nusrat ogly Abdulov – S&R Institute, Ministry of Defense Industry, Azerbaijan, Deputy Chief Engineer, Rabdulov2009@rambler.ru
Novruz Almamed ogly Abdullayev – S&R Institute, Ministry of Defense Industry, Azerbaijan, Section head, PhD, senior scientific researcher, Naabdullayev@mail.ru
Hikmat Hamid ogly Asadov – S&R Institute of Aerospace Informatics, Azerbaijan, Section head, D.Sc., Associate professor, asadzade@rambler.ru

OPTICAL AND OPTICAL ELECTRONIC SYSTEMS. OPTICAL TECHNOLOGIES

HOMOGENIZERS OF LASER RADIATION FABRICATED BY LASER GRAVING OF POLYMER FILMS

V. Balya, I. Denisyuk, E. Barausova

The article deals with homogenizers of laser radiation based on crossing micro-prism structures and random diffusers by laser ablation of thermo-sensitive film and subsequent nano-imprint copying.
Keywords: homogenizer, laser ablation, micro-prism arrays, random diffusers.
Vera Balya – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, Vera_Balya@mail.ru
Igor Denisyuk – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, denisiuk@mail.ifmo.ru
Ekaterina Barausova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, barausik@yandex.ru

PHOTONICS AND OPTICAL INFORMATICS

METAFILMS CHARACTERISTICS FOR THE TERAHERTZ FREQUENCY RANGE AT THE GEOMETRIC PARAMETERS SCALING

Yu. Terekhov, M. Khodzitskiy, G. Belokopytov

The reflection and transmission coefficients of metafilms composed of U-shaped resonators were obtained in the frequency range of 0.1 - 1 THz by numerical finite-element simulation method. The side length, the gap depth and the period of resonators were proportionally scaled. The experimental spectra of the transmission coefficients for metafilms with U-shaped resonators were received. The re-tuning of the resonant modes was shown from 0.12 till 0.51 THz at the geometric parameters scaling.
Keywords: metamaterials, U-shaped resonator, electrodynamic similarity, terahertz frequency range, plasmonic resonance, metafilm.
Yury Terekhov – Lomonosov Moscow State University, postgraduate, yury.terekhov@phys.msu.ru
Mikhail Khodzitskiy – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, assistant, PhD, khodzitskiy@yandex.ru
Gennady Belokopytov – Lomonosov Moscow State University, Associate professor, PhD, Associate professor, gvb@phys.msu.ru

TERAHERTZ RADIATION INFLUENCE ON THE GROWING OF NEURITES

M. Tsurkan, O. Smolyanskaya, N. Bryantseva

Our research was devoted to the impact of broadband pulsed THz radiation in the frequency range of 0.05 to 2 THz on the neurite growth in the sensory ganglia of 10-12-day chicken embryos. Dependence of changes in functional responses of cells on the average output power has been found.
Keywords: terahertz (THz) radiation, therapy, medicine, cell growth, stimulation, organotypic tissue culture, neurite.
Maria Tsurkan – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, tsurkan.maria@yandex.ru

Olga Smolyanskaya – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, o_smolyanskaya@mail.ru
Natalia Bryantseva – Tomsk National Research State University, junior scientific researcher, PhD, brjantseva@rambler.ru

FEMTOSECOND PHOTON ECHO ON SIZE-INHOMOGENEOUS QUANTUM DOTS OF ZINC OXIDE AT ROOM TEMPERATURE

N. Vashurin, I. Popov, S. Putilin, V. Sidorova, S. Stepanov, N. Sushentsov

The article deals with the discovery of femtosecond signals of self-diffraction and photon echo in a zinc oxide film of 100 nm thickness, consisting of different-sized quantum dots, at room temperature. Experimental results, proving the Autler-Townes effect demonstration on exciton transitions of quantum dots, are obtained.

Keywords: femtosecond photon echo, self-diffraction, zinc oxide film, magnetron sputtering, quantum dots, Autler-Townes effect.

Nikita Vashurin – Volga State University of Technology, postgraduate, Nickita_Vashurin@mail.ru
Ivan Popov – Volga State University of Technology, Professor, D.Sc., Professor, popov@volgatech.net
Sergey Putilin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Senior Researcher, PhD, Senior Researcher, seputilin@yandex.ru
Vera Sidorova – Mari State University, Associate professor, PhD, Associate professor, veranig@yandex.ru
Sergei Stepanov – Mari State University, postgraduate, stepan_mail@mial.ru
Nikolai Sushentsov – Volga State University of Technology, Department head, PhD, Associate professor, sni@mari-el.ru

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES

SIMULATION ERRORS OF HIGH-LOADED SYSTEMS IN GPSS WORLD

T. Aliev

The article presents research results of random number generators (RNG) influence on simulation accuracy of high-loaded queuing systems in the GPSS World simulation system. It is shown that for certain combinations of RNGs, called problem generators, the inaccuracy of the simulation results of highly loaded systems may reach tens of percent. The analysis of problem combinations revealed factors causing high level of inaccuracy in simulation results.

Keywords: simulation, high-loaded system, random number generators (RNG), simulation error, GPSS World.

Taufik Aliev – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, aliev@d1.ifmo.ru

HIERARCHY OF ALGEBRAIC BAYESIAN NETWORK GLOBAL STRUCTURES AS A SYSTEM OF GRAPHS AND HYPERGRAPHS

A. Filchenkov

Algebraic Bayesian networks are logical and probabilistic graphical models which make it possible to perform logical and probabilistic inference for knowledge uncertainty formalized via scalar and interval truth values of propositional formulae. Global structures of algebraic Bayesian network are overviewed; their systematization based on hypergraph representation is proposed and functional hierarchy of global structures is revealed. The proposed systematization gives the possibility to apply graph and hypergraph theory methods for solving a number of problems in analysis of algebraic Bayesian network global structures, in particular — to propose and justify a criterion for identification of its primary structure acyclicity, — and also gives the theoretical basis for algorithms of specified networks automated learning.

Keywords: algebraic Bayesian network, adjacency graphs, machine learning, global structure, probabilistic graphical models.

Andrei Filchenkov – SPIIRAS, junior scientific researcher; Saint Petersburg State University, postgraduate, aaafil@mail.ru

TAG LINES EXTRACTION AND RANKING IN TEXT ANNOTATION

S. Popova, I. Khodyrev

The article deals with comparative analysis of two approaches for tag lines ranking in text annotation task. The first approach is based on a weight evaluation of extracted phrases using TextRank algorithm, the second one is based on tf-idf estimation. Experiments were carried out on the base of INSPEC dataset collection. Experiment descriptions and comparative results are given. It was shown experimentally that tf-idf approach gives better results than the one based on TextRank.

Keywords: annotation, tag lines extraction and ranking, annotation quality evaluation.

Svetlana Popova – Saint Petersburg State University, Saint-Petersburg State Polytechnic University, senior lecturer, svp@list.ru
Ivan Khodyrev – «Vismart», Ltd., Deputy Director for Science, kivan.mih@gmail.com

DATA VERIFICATION IN ISSUE SUPERVISING SYSTEMS BY PRODUCTION RULES

R. Katerinenko, I. Bessmertny

The paper proposes a method of data verification in issues tracking systems by means of production rules. This model makes it possible to formulate declaratively conditions that the information containment should comply with and apply reasoning procedures. Practical application of proposed verification system in a real software development project is described.

Keywords: production rules, logical reason, issue tracking systems, verification.

Roman Katerinenko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, innuWorks@gmail.com

Igor Bessmertny – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate Professor, PhD, Associate Professor, igor_bessmertny@hotmail.com

AGGLOMERATIVE CLUSTERING OF SOUND RECORD SPEECH SEGMENTS BASED ON BAYESIAN INFORMATION CRITERION

O. Kudashev

This paper presents the detailed description of agglomerative clustering system implementation for speech segments based on Bayesian information criterion. Numerical experiment results with different acoustic features, as well as the full and diagonal covariance matrices application are given. The error rate DER equal to 6.4% for audio records of radio «Svoboda» was achieved by means of designed system.

Keywords: speech segments clustering, Bayesian variation analysis, speech technologies.

Oleg Kudashev – «Speech Technology Center», Ltd., programmer; Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, kudashev@speechpro.com

PROGRAM SYSTEM AND INFORMATION METADATA BANK OF TERTIARY PROTEIN STRUCTURES

T. Nikitin, Yu. Porozov

The article deals with the architecture of metadata storage model for check results of three-dimensional protein structures. Concept database model was built. The service and procedure of database update as well as data transformation algorithms for protein structures and their quality were presented. Most important information about entries and their submission forms to store, access, and delivery to users were highlighted. Software suite was developed for the implementation of functional tasks using Java programming language in the NetBeans v.7.0 environment and JQL to query and interact with the database JavaDB. The service was tested and results have shown system effectiveness while protein structures filtration.

Keywords: program system, information bank, protein structure, PDB, structure quality, checking server, metadata.

Timur Nikitin – Saint Petersburg State Polytechnic University, student, tim04k@gmail.com

Yuri Porozov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Laboratory head, Associate professor, PhD, porozov@ifc.cnr.it

FREQUENCY ANALYSIS OF RLE-BLOCKS REPETITIONS IN THE SERIES OF BINARY CODES WITH OPTIMAL MINIMAX CRITERION OF AUTOCORRELATION FUNCTION

A. Kovylin, D. Zlobin, A. Rodionov

The article describes the problem of searching for binary pseudo-random sequences with quasi-ideal autocorrelation function, which are to be used in contemporary communication systems, including mobile and wireless data transfer interfaces. In the synthesis of binary sequences sets, the target set is manning them based on the minimax criterion by which a sequence is considered to be optimal according to the intended application. In the course of the research the optimal sequences with order of up to 52 were obtained; the analysis of Run Length Encoding was carried out. The analysis showed regularities in the distribution of series number of different lengths in the codes that are optimal on the chosen criteria, which would make it possible to optimize the searching process for such codes in the future.

Keywords: pseudo-random sequences, chaotic signals, Barker codes, minimax criterion, autocorrelation function, Run Length Encoding.

Aleksander Kovylin – Far-Eastern Federal University (FEFU), engineer, kalexer@hotmail.com

Dmitriy Zlobin – Far-Eastern Federal University (FEFU), engineer, memrbomel@mail.ru

Aleksander Rodionov – Far-Eastern Federal University (FEFU), Associate Professor, PhD, deodar1618@mail.ru

MINIMAL MODIFICATION OF AUTOMATA-BASED PROGRAMS INDUCED BY CHANGES IN THEIR TEST SCENARIOS

A. Shestakov

The article proposes modification method for automata-based programs induced by changes in their test scenarios. The method can also be used for constructing automata-based programs without using any prior knowledge of a program structure and without using evolutional algorithms (as was proposed by many authors before). The main feature of the method is a minimal requirement for program changes. Its essence lies in the

reduction of the original problem to the problem of finding the shortest path in the acyclic weighted graph constructed from the original automaton and new test scenario.
Keywords: automata-based programming, test scenario, Moore automaton.
Alexei Shestakov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, shestakov@rain.ifmo.ru

PHASE GRADIENT METHOD OF MAGNETIC FIELD MEASUREMENTS IN ELECTRIC VEHICLES

N. Ptitsyna, V. Ismagilov, Yu. Kopytenko, A. Korobeynikov

Operation of electric and hybrid vehicles demands real time magnetic field control, for instance, for fire and electromagnetic safety. The article deals with a method of magnetic field measurements onboard electric cars taking into account peculiar features of these fields. The method is based on differential methods of measurements, and minimizes the quantity of magnetic sensors.
Keywords: electric car, magnetic field, electromagnetic safety, magnetic field measurement, phase gradient method.
Natalia Ptitsyna – Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences St. Petersburg Branch (SPbF IZMIRAN), Senior researcher, PhD, nataliaptitsyna@ya.ru
Valery Ismagilov – Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences St. Petersburg Branch (SPbF IZMIRAN), scientific secretary, PhD, senior scientific researcher, IVS@izmiran.spb.ru
Yuri Kopytenko – Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences St. Petersburg Branch (SPbF IZMIRAN), Director, D.Sc., Professor, office@izmiran.spb.ru
Anatoly Korobeynikov – Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences St. Petersburg Branch (SPbF IZMIRAN), Deputy director for science, D.Sc., Professor, Korobeynikov_A_G@mail.ru

MATERIALS TECHNOLOGY AND NANOTECHNOLOGIES

GALVANIC MAGNETIC PROPERTIES OF BISMUTH THIN FILMS DOPED WITH TELLURIUM MADE BY THERMAL VACUUM EVAPORATION

V. Komarov, D. Matveev, E. Demidov, A. Krushelnitskiy

The influence of n-type impurity of tellurium (concentration range from 0.005 atomic % Te to 0.15 atomic % Te) on galvanic magnetic properties (resistivity, magnetic resistance and Hall constant) of Bi thin films with various thicknesses was studied. The properties were measured in temperature range from 77 to 300 K. It was established that the classical size effect in the films is significant and decreases with higher concentration of Te impurity. The analysis of experimental results was carried out in approximation of the law of Jones-Schoenberg dispersion for Bi films doped with tellurium. Calculation of concentration and mobility of charge carriers in the studied films was made.
Keywords: bismuth, n-type doping, thin films, resistivity, magnetic resistance, Hall constant, the size effect.
Vladimir Komarov – Herzen State Pedagogical University, Associate professor, PhD, Associate professor, va_komar@mail.ru
Daniil Matveev – Herzen State Pedagogical University, postgraduate, danila200586@mail.ru
Evgeniy Demidov – Herzen State Pedagogical University, senior scientific researcher, PhD, demidov_evg@mail.ru
Artemiy Krushelnitskiy – Herzen State Pedagogical University, student, ak.spb.ru@gmail.com

EFFECT OF LASER MICRO STRUCTURING PROCESS ON ELECTROPHYSICAL PROPERTIES OF SIO2/SI SYSTEM

A. Skvortsov, R. Khaletskiy, Huynh Cong Tu

Electrophysical parameters of SiO_2/Si system after irradiation by pulsed ytterbium fiber laser of ILI-1-50 type have been investigated. It is shown that in the silicon-silicon dioxide system under laser radiation a formation of defects occurs responsible for the appearance of different kinds of charges. It is found that laser-induced defects appear at sites remote from the irradiation zone as well. The defect generation rate at the silicon-silicon dioxide interface depends on their position in the energy band gap of silicon and reaches a maximum near the Fermi level.
Keywords: laser irradiations, dislocation, slip lines, micro structuring, capacitance-voltage characteristics.
Albert Skvortsov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, a-skvortsov@yandex.ru
Roman Khaletskiy – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, halecky@yandex.ru
Huynh Cong Tu – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, Picochip912@yahoo.com

COMPUTER-AIDED DESIGN SYSTEMS

SIMPLEX- METHOD AND ARTIFICIAL BASIS METHOD APPLICATION FOR ONBOARD EQUIPMENT DESIGN

M. Deyko, I. Zharinov

The article deals with optimization methods application for design decisions (simplex- method and artificial basis method) in instrument making. Their application for onboard equipment designing is exemplified.

Keywords: optimization, instrument making, simplex-method, artificial basis method.

Mikhail Deyko – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, MadCat_MKII@mail.ru

Igor Zharinov – Saint Petersburg Scientific Design Bureau "Electroavtomatica" n.a. P. A. Efimov, Head of learning scientific center, D.Sc., Associate professor, igor_rabota@pisem.net

ANALYSIS OF WIRELESS SENSOR AND ACTUATOR NETWORKS DESIGN METHODS

D. Kirov, A. Ozhiganov

Existing design methods of information and control systems are considered. Aspect, actor-oriented and platform-based approaches are selected and analyzed in relation to wireless sensor and actuator networks. Generic design processes of investigated systems are described based on these methods. Some improvement directions for them are suggested, their perspectives are discussed as well.

Keywords: wireless sensor and actuator networks, physically embedded intelligent systems, aspect design, actor-oriented design, platform-based design, WSN, WSAN.

Dmitry Kirov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, kdedge@mail.ru

Alexander Ozhiganov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Professor, D.Sc., Professor, ojiganov@mail.ifmo.ru

EVOLUTION MODELING METHODS APPLICATION FOR CIRCULATION OF DOCUMENTS OPTIMIZATION AT THE ENTERPRISE

Yu. Tchernyshev, A. Basova, P. Panasenko, V. Polyakov

A new approach to creation of information control systems at the enterprise is offered. The genetic decision algorithm for the solution of actual problems of nonlinear optimization is developed making it possible to increase the decision quality and speed. The developed procedure of initial population initialization is described. The modified operators of crossing-over, mutation are offered, the new approach to rejection of impractical decisions is justified. Introduction in genetic algorithm of macro evolution, that is the development of several subpopulations, is offered. It is noted that work not with one but with several initial populations gives the possibility to receive a bigger variety of genetic material and conduct evolution in separate subpopulations.

Keywords: optimization genetic algorithm, information systems, transport task.

Yuri Tchernyshev – Branch of the Military Academy of Communications, Professor, D.Sc., Professor, myvnn@list.ru

Alina Basova – Don State Technical University, Section head of post-graduate course, PhD, Associate professor, abasova@donstu.ru

Pavel Panasenko – Branch of the Military Academy of Communications (Krasnodar), Officer of the State Secrets Protection, We_panasenko_777@mail.ru

Vladimir Polyakov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, v_i_polyakov@mail.ru

INFORMATION SECURITY METHODS AND SYSTEMS

INFORMATION SECURITY RISKS OPTIMIZATION IN CLOUDY SERVICES ON THE BASIS OF LINEAR PROGRAMMING

I. Zikratov, S.Odegov, A. Smirnykh

The paper discusses theoretical aspects of secure cloud services creation for information processing of various confidentiality degrees. A new approach to the reasoning of information security composition in distributed computing structures is suggested, presenting the problem of risk assessment as an extreme problem of decision-making. Linear programming method application is proved to minimize the risk of information security for given performance security in compliance with the economic balance for the maintenance of security facilities and cost of services. An example is given to illustrate the obtained theoretical results.

Keywords: risks, information security, cloud computing, objective function, threats.

Igor Zikratov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department Head, D.Sc., Professor, zikratov@cit.ifmo.ru

Stepan Odegov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, odegov.sv@gmail.com

Alexander Smirnykh – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, senior lecturer, smirnykh_av@spb.power-m.ru

INTEGRITY CONTROL ALGORITHM FOR DEPERSONALIZED INFORMATION IN INFORMATION SYSTEMS OF PERSONAL DATA

Yu. Gatchin, O. Teploukhova, A. Kurakin

The article deals with the problem of integrity ensuring for the information processed in information systems of personal data by depersonalization data algorithm. In order to solve this problem, we propose to use integrity algorithm, based on hash functions to confirm the immutability of depersonalized data stored in the information systems of personal data.

Keywords: integrity control, hash function, depersonalization algorithm, personal data, information systems of personal data.

Yuri Gatchin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Department head, D.Sc., Professor, gatchin@mail.ifmo.ru

Olga Teploukhova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, teplouhovaoa@gmail.com

Alexander Kurakin – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, postgraduate, nirt@mail.ru

RAILWAY EMERGENCY SITUATIONS PREVENTION SYSTEM BASED ON GEOGRAPHIC INFORMATION TECHNOLOGIES

V. Alekseev, V. Konovalova, A. Minina

The article deals with the system of emergency situation prevention on railway based on geographic information technologies (GIS), which makes it possible to estimate the influence of railroad defects, industrial objects (factories, plants located near investigated section of the railway) and natural factors (influence of the elements) on traffic safety. The procedures of area maps design for information systematization and display are offered. The algorithm of measurement information processing is developed, based on wavelet transform, that gives the possibility for the definition of railway bed defect parameters immediately after the measurements and does not require accumulation of information for further decryption. Techniques for influence estimation of natural and industrial factors on traffic safety are suggested to provide immediate representation of information for decision-making.

Keywords: railway bed, geographic information systems, defect, emergency situations, wavelet transform.

Vladimir Alekseev – Saint Petersburg State Electrotechnical University "LETI" n.a. V.I. Ul'yanov (Lenin), Department head, D.Sc., Professor, VVAlekseev@mail.eltech.ru

Vera Konovalova – Saint Petersburg State Electrotechnical University "LETI" n.a. V.I. Ul'yanov (Lenin), assistant, PhD, vskonovalova@inbox.ru

Anastasia Minina – Saint Petersburg State Electrotechnical University "LETI" n.a. V.I. Ul'yanov (Lenin), assistant, PhD, aaminina@mail.ru

DIGITAL WATERMARKS IN IMAGE FILES

A. Korobeynikov, S. Kuvshinov, S. Blinov, A. Leyman, I. Kutuzov

The article is about generation problem of digital watermarks for graphic files. Mathematical model of digital watermarks generation is presented. The algorithm of messages introduction is analyzed. Developed steganography algorithm application for copyright check to a concrete file of multimedia is offered.

Keywords: digital watermarks, format methods, steganography algorithms of the spatial domain, steganography algorithms of the conversion domain, multimedia, media environment, copyright.

Anatoly Korobeynikov - Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences St. Petersburg Branch (SPbF IZMIRAN), 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

Il'ya Kutuzov – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, formalizator@gmail.com

ECONOMICS AND FINANCES. MANAGEMENT

METHODOLOGICAL ISSUES OF TELECOMMUNICATION COMPANY PERFORMANCE ESTIMATION

O. Tsukanova, T. Lavrent'yeva

Methodological issues for telecommunication companies performance estimation are specified, approaches are suggested that provide a certain sequence of effectiveness evaluation and the use of identifiers reflecting the level of achieved performance.

Keywords: telecommunications services, efficiency, effect, internal efficiency, external efficiency, EBITDA, profitability, economic efficiency, social efficiency.
Olga Tsukanova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, Associate professor, PhD, Associate professor, zoa1999@mail.ru
Tat'yana Lavrent'yeva – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, student, t.nelarina@mail.ru

INNOVATIVE ACTIVITY OF BANK SYSTEM SUBJECTS AS A FORM OF REALIZATION OF ECONOMIC INTERESTS

T. Krasnoperova

The article deals with the actual problem for innovative development of the Russian Federation banking system as form of economic interests realization.Ways of innovative activity stirring up by coordination of economic interests for bank activity subjects are offered.
Keywords: innovative development, bank system, interaction of economic interests, innovations, profitability of assets.
Tanzilya Krasnoperova – Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, International Institute of Business and Law, Associate professor, PhD, krasntat@mail.ru

NEW TECHNOLOGIES IN EDUCATION

LINGUISTIC SIMULATOR ARCHITECTURE FOR QUICK COMMUNICATION SKILLS MASTERING IN A FOREIGN LANGUAGE

V. Kiselev, O. Eliseeva, Yu. Kovalyonok, J. Khitrova

The article deals with linguistic simulator architecture as a network software solution designed to develop quickly foreign language communication skills. Key attention is paid to formal models which are the basic architecture component providing flexible educational strategies with no dependency on applicable teaching methods.
Keywords: linguistic simulator, learning model, training actions model, student model.
Vitaliy Kiselev – «Speech Technologies» Ltd., Minsk, Head, kiselev-v@speechpro.com
Olga Eliseeva – Belarusian State University, Associate professor, PhD, Associate professor, volga.eliseeva@gmail.com
Yury Kovalyonok – «Speech Technologies» Ltd., Minsk, software testing specialist, kavalionak-y@speechpro.com
Julia Khitrova – «Speech Technologies Center», Ltd., business manager, julia@speechpro.com

Уважаемые читатели!

«Научно-технический вестник информационных технологий, механики и оптики» издается с 2001 года как периодическое научное и научно-образовательное издание. Журнал входит в утвержденный Высшей Аттестационной комиссией «Перечень ведущих рецензируемых научных журналов и изданий, выпускаемых в Российской Федерации, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук». Все публикации в журнале проходят рецензирование. Периодичность издания составляет 6 выпусков в год, объем каждого выпуска - 140 страниц.

Журнал печатает материалы, отражающие достижения науки, техники и технологии в области современного приборостроения и новых информационных технологий в России и за рубежом по следующим направлениям: информационные технологии, информационно-телекоммуникационные системы, элементы и устройства вычислительной техники и систем управления, системный анализ и математическое моделирование, системы автоматизированного проектирования, оптика, оптотехника и приборостроение, физика твердого тела, микроэлектроника, нанотехнологии, мехатроника, методы и системы защиты информации, биомедицинские технологии, экономика и финансы.

Издание зарегистрировано как сериальное в Международном центре ISSN, номер ISSN 2226-1494, а также в Федеральной службе по надзору в сфере связи и массовых коммуникаций, свидетельство ПИ ФС77-47243 от 10 ноября 2011.

На журнал «Научно-технический вестник информационных технологий, механики и оптики» можно оформить подписку в любом отделении связи, а также в редакции журнала (Санкт-Петербург, Кронверкский пр., 49, комн. 330). Подписной индекс 47 197 по каталогу агентства Роспечать (Газеты. Журналы).

Сайт журнала http:// ntv.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.

Во 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: //ntv.ifmo.ru

Подписано к печати 15.12.2012 Тираж 350 экз. Заказ № 31(83)

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 (83) ISSN 2226-1494
NAUČNO-TEHNIČESKIJ VESTNIK INFORMACIONNYH TEHNOLOGIJ, MEHANIKI I OPTIKI

REVIEW ARTICLE
Adaptive output control: subject matter, application tasks and solutions — *A. Bobtsov, V. Nikiforov* — *1*

UP-TO-DATE ISSUES OF CONTROL SYSTEMS THEORY
Control approach for nonlinear plant with parametric uncertainties and input delay — *A. Bobtsov, A. Pyrkin* — *15*
Suboptimal control of nonlinear multi-agent systems — *I. Furtat* — *19*
Dynamical systems stability study by the properties of the vector phase speed flow through a closed convex surface — *I. Furtat* — *23*
Hybrid algorithm of frequency identification for multi-sinusoidal signal — *A. Bobtsov, A. Vedyakov, S. Kolyubin, A. Pyrkin* — *28*
Integral differentiating speed-gradient algorithms in multiple synchronization issues for vibration units — *A. Fradkov, O. Tomchina, V. Galitskaya, D. Gorlatov* — *30*
Discrete-time robust control for robot-manipulator — *S. Gusev* — *38*
Analysis and processing of electromyogram signals — *K. Zimenko, A. Borgul', A. Margun* — *41*
Implementation of industrial robot-manipulator setups in educational process — *A. Bobtsov, S. Kolyubin, A. Pyrkin* — *43*
On optimization of unmanned aerial vehicle application for search and tracking of objects — *R. Abdulov, N. Abdullayev, H. Asadov* — *45*

OPTICAL AND OPTICAL ELECTRONIC SYSTEMS. OPTICAL TECHNOLOGIES
Homogenizers of laser radiation fabricated by laser graving of polymer films — *V. Balya, I. Denisyuk, E. Barausova* — *50*

PHOTONICS AND OPTICAL INFORMATICS
Metafilms characteristics for the terahertz frequency range at the geometric parameters scaling — *Yu. Terekhov, M. Khodzitskiy, G. Belokopytov,* — *55*
Terahertz radiation influence on the growing of neurites — *M. Tsurkan, O. Smolyanskaya, N. Bryantseva* — *60*
Femtosecond photon echo on size-inhomogeneous quantum dots of zinc oxide at room temperature — *N. Vashurin, I. Popov, S. Putilin, V. Sidorova, S. Stepanov, N. Sushentsov* — *65*

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES
Simulation errors of high-loaded systems in GPSS World — *T. Aliev* — *70*
Hierarchy of algebraic Bayesian network global structures as a system of graphs and hypergraphs — *A. Filchenkov* — *75*
Tag lines extraction and ranking in text annotation — *S. Popova, I. Khodyrev* — *81*
Data verification in issue supervising systems by production rules — *R. Katerinenko, I. Bessmertny* — *86*
Agglomerative clustering of sound record speech segments based on Bayesian information criterion — *O. Kudashev* — *90*
Program system and information metadata bank of tertiary protein structures — *T. Nikitin, Yu. Porozov* — *94*
Frequency analysis of RLE-blocks repetitions in the series of binary codes with optimal minimax criterion of autocorrelation function — *A. Kovylin, D. Zlobin, A. Rodionov* — *99*
Minimal modification of automata-based programs induced by changes in their test scenarios — *A. Shestakov* — *104*
Phase gradient method of magnetic field measurements in electric vehicles — *N. Ptitsyna, V. Ismagilov, Yu. Kopytenko, A. Korobeynikov* — *108*

MATERIALS TECHNOLOGY AND NANOTECHNOLOGIES
Galvanic magnetic properties of bismuth thin films doped with tellurium made by thermal vacuum evaporation — *V. Komarov, D. Matveev, E. Demidov, A. Krushelnitskiy* — *113*
Effect of laser micro structuring process on electrophysical properties of SiO_2/Si system — *A. Skvortsov, R. Khaletskiy, Huynh Cong Tu* — *119*

COMPUTER-AIDED DESIGN SYSTEMS
Simplex- method and artificial basis method application for onboard equipment designs — *M. Deyko, I. Zharinov* — *124*
Analysis of wireless sensor and actuator networks design methods — *D. Kirov, A. Ozhiganov* — *129*
Evolution modeling methods application for circulation of documents optimization at the enterprise — *Yu. Tchernyshev, A. Basova, P. Panasenko, V. Polyakov* — *135*

INFORMATION SECURITY METHODS AND SYSTEMS
Information security risks optimization in cloudy services on the basis of linear programming — *I. Zikratov, S.Odegov, A. Smirnykh* — *141*
Integrity control algorithm for depersonalized information in information systems of personal data — *Yu. Gatchin, O. Teploukhova, A. Kurakin* — *145*
Railway emergency situations prevention system based on geographic information technologies — *V. Alekseev, V. Konovalova, A. Minina* — *148*
Digital watermarks in image files — *A. Korobeynikov, S. Kuvshinov, S. Blinov, A. Leyman, I. Kutuzov* — *152*

ECONOMICS AND FINANCES. MANAGEMENT
Methodological issues of telecommunication company performance estimation — *O. Tsukanova, T. Lavrent'yeva* — *158*
Innovative activity of bank system subjects as a form of realisation of economic interests — *T. Krasnoperova* — *161*

NEW TECHNOLOGIES IN EDUCATION
Linguistic simulator architecture for quick communication skills mastering in a foreign language — *V. Kiselev, O. Eliseeva, Yu. Kovalyonok, J. Khitrova* — *166*

SUMMARY — *171*

© Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2013