

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Выпуск 22 № 2 Год 2012

СОДЕРЖАНИЕ

Методы и средства анализа и моделирования стохастических систем интегрированной логистической поддержки И. Н. Сеницын, А. С. Шаламов, И. В. Сергеев, В. И. Сеницын, Э. Р. Корепанов, В. В. Белоусов, Е. С. Агафонов, В. С. Шоргин	3
Автоматизация отладки программ для рекуррентного операционного устройства Р. А. Зеленов, А. А. Прокофьев, В. Н. Волчек	29
Встроенные средства отладки рекуррентного операционного устройства А. А. Прокофьев, Р. А. Зеленов, В. Н. Волчек	38
Модель потоковой архитектуры на примере распознавателя слов Д. В. Хилько, Ю. А. Степченко	48
Автоматизация обработки информации для обнаружения аномалий формы рельсовых нитей С. К. Дулин, И. Н. Розенберг, В. И. Уманский	58
Методические подходы к выбору сенсорных технологий для ситуационных центров на основе классификации видов деятельности пользователей С. А. Денисов, А. А. Зацаринный, В. А. Кондрашев, К. Г. Чупраков	79
Математические и программные средства моделирования световых приборов И. И. Байнева, В. В. Байнев	95
Цитирование документов в патентах как индикатор взаимосвязи областей науки и технологий О. С. Кожунова	106
Интеллектуальные механизмы семантического поиска в сети Интернет И. П. Кузнецов, М. М. Шарнин, А. Г. Мацкевич	129
Метод классификации информации на основе иерархических тегов и его реализация на примере Семейного архивного фонда И. М. Адамович, О. И. Волков, Н. А. Маркова	146

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Выпуск 22 № 2 Год 2012

СОДЕРЖАНИЕ

Методы обработки сигналов для локализации невосполнимых областей головного мозга <i>Т. В. Захарова, С. Ю. Никифоров, М. Б. Гончаренко, М. А. Драницына, Г. А. Климов, М. Ш. Хазиахметов, Н. В. Чаянов</i>	157
Архитектура гибридной рекомендательной системы GEFEST (Generation–Expansion–Filtering–Sorting–Truncation) <i>Ю. С. Нефедова</i>	176
Непараметрическое оценивание функции плотности смесей вероятностных законов с помощью EM-алгоритма <i>В. Ю. Королев, А. Ю. Корчагин, О. А. Морева</i>	197
О состоятельности оценок параметров масштабных смесей нормальных распределений, получаемых с помощью сеточных методов <i>А. Л. Назаров</i>	227
Оптимизация ресурсов вычислительного комплекса со слабосвязанной архитектурой <i>М. Я. Агаларов</i>	244
Виртуальные машины как средство оптимизации учебного процесса <i>В. С. Петрухин, А. Ю. Шнейдер, С. А. Лучкин</i>	256
Электронные образовательные ресурсы: эффективность применения <i>А. А. Федосеев</i>	263
Анализ международных образовательных стандартов в области информационных технологий <i>В. А. Сухомлин</i>	278
Информационные характеристики органических молекул <i>И. М. Гуревич, М. П. Евстигнеев</i>	308
Abstracts	326
Об авторах	335
About Authors	338
Авторский указатель за 2012 г.	341
2012 Author Index	344

МЕТОДЫ И СРЕДСТВА АНАЛИЗА И МОДЕЛИРОВАНИЯ СТОХАСТИЧЕСКИХ СИСТЕМ ИНТЕГРИРОВАННОЙ ЛОГИСТИЧЕСКОЙ ПОДДЕРЖКИ*

*И. Н. Синицын¹, А. С. Шаламов², И. В. Сергеев³, В. И. Синицын⁴,
Э. Р. Корепанов⁵, В. В. Белоусов⁶, Е. С. Агафонов⁷, В. С. Шоргин⁸*

Аннотация: Рассматриваются компьютерные методы и средства анализа и моделирования для одного из новейших направлений в области экономики послепродажного обслуживания изделий наукоемкой продукции (ИНП) длительного использования: интегрированной логистической поддержки (ИЛП). Особое внимание уделено проблемам аналитического моделирования и минимизации затрат на эксплуатацию ИНП в непрерывно-дискретных и эрдитарных системах.

Ключевые слова: инструментальное программное обеспечение; стохастические системы; эрдитарные системы; стохастический ударный импульс; аналитическое и статистическое моделирование; фильтрация; MATLAB

1 Введение

В начале 2012 г. в издательстве ТОРУС ПРЕСС вышла книга [1]. В книге дано систематическое изложение теоретических основ одного из новейших направлений в области экономики послепродажного обслуживания ИНП длительного использования — ИЛП. Приводятся также результаты работ, выполненных в институте проблем информатики Российской академии наук в рамках Программы «Интеллектуальные информационные технологии, системный анализ и автоматизация» (проект 1.7) Отделения нанотехнологий и информационных технологий РАН.

* Работа выполнена при поддержке программы ОНИТ РАН «Интеллектуальные информационные технологии, системный анализ и автоматизация» (проект № 1.7).

¹ Институт проблем информатики Российской академии наук, sinitsin@dol.ru

² Институт проблем информатики Российской академии наук, a-shal5@yandex.ru

³ Институт проблем информатики Российской академии наук, isergeev@ipiran.ru

⁴ Институт проблем информатики Российской академии наук, vsinitsin@ipiran.ru

⁵ Институт проблем информатики Российской академии наук, ekorepanov@ipiran.ru

⁶ Институт проблем информатики Российской академии наук, vbelousov@ipiran.ru

⁷ Институт проблем информатики Российской академии наук, eagafonov@ipiran.ru

⁸ Институт проблем информатики Российской академии наук, vshorgin@ipiran.ru

Интегрированная логистическая поддержка — это система научно-исследовательских, проектно-конструкторских, организационно-технических, производственных и информационно-управленческих технологий, средств и практических мероприятий, используемых (применяемых) в течение жизненного цикла (ЖЦ) ИИП, направленных на достижение минимальных затрат по обслуживанию и ремонту ИИП при обеспечении требуемых характеристик и показателей функционального качества и технической готовности продукции при ее эксплуатации.

Излагаемые новые научные подходы позволяют кардинально реформировать традиционные системы создания и эксплуатации ИИП путем внедрения методов рационального и, по возможности, оптимального управления процессами расходования временных, материальных, трудовых и др. ресурсов на всех стадиях ЖЦ изделий по критериям экономической целесообразности и эффективности.

Первыми на этот путь вступили западные страны, когда в 90-х г. XX в. были резко сокращены бюджеты военных ведомств, что вынудило их искать возможности экономии финансовых ресурсов, сохраняя при этом основные показатели обороноспособности. Были выработаны новые стандарты по созданию вооружений и военной техники (ВВТ), обеспечивающие снижение стоимости ее ЖЦ. Контракты, заключаемые с поставщиками ВВТ, стали содержать требования по неукоснительному соблюдению правил и процедур, обеспечивающих заданные характеристики и показатели качества и экономичности продукции. В настоящее время на этот путь переводятся и отрасли экономики гражданского назначения. В основу серии этих международных стандартов легла методология CALS — Continuous Acquisition and Life cycle Support — непрерывная информационная поддержка поставок и ЖЦ, или ИПИ — интегрированная поддержка изделий. Эта методология позволяет создавать виртуальные предприятия (ВП) из фирм — участников ЖЦ ИИП, интегрируя их на информационной основе.

Основные принципы ИЛП и интеграции ВП изложены в стандартах MIL STD 1388 (США) и DEF STAN 00-60 (общеевропейский стандарт).

Стандарт DEF STAN 00-60 тесно связан со спецификациями ASD (AeroSpace and Defense Industries Association of Europe — Европейской ассоциации аэрокосмической и оборонной промышленности):

- S1000D “International specification for technical publications utilizing a Common Source Data Base” («Международная спецификация на технические публикации, использующие общую исходную базу данных»);
- S2000M “International specification for Material management” («Международная спецификация на организацию управления материально-техническим обеспечением») и др.

Важно отметить, что в этих стандартах отсутствуют указания на методы и средства, позволяющие достигать реального минимума затрат.

Участники ВП, по сути, самостоятельно пытаются решать свои проблемы или прибегают к услугам фирм — разработчиков программных средств для решения

частных (сервисно-ориентированных) задач по управлению процессами. При этом, как правило, информация об использовании каких-либо методов решения этих задач остается закрытой, что привносит неопределенность при оценке покупателем гарантий достоверности таких программ. Таким образом, напрашивается вывод об определенной неполноте современного содержания СALS, проявляющейся в отсутствии теоретических обоснований путей достижения провозглашаемых ею целей. Для полноценного использования преимуществ данной методологии необходима разработка теории систем ИЛП, включающей в себя единый комплекс современных методов и средств информационного и математического моделирования, оценивания и оптимального управления процессами на всех стадиях ЖЦ комплексной системы «ИНП – системы послепродажного обслуживания (СППО)», вносящими основной вклад в ее стоимость и качество.

В книге, помимо общих понятий, вводящих читателя в курс дела, даются постановки и решения практически по всему комплексу задач ИЛП, обеспечивающих управление стоимостью ЖЦ ИНП и ее минимизацию.

В ч. 1 «Интегрированная логистическая поддержка» (гл. 1–3), излагаются элементы СППО изделий, рассматриваются основы современной ИЛП. Особое внимание уделяется управлению ЖЦ, СППО и их информационному моделированию.

Часть 2 (гл. 4–6) посвящена теории гибридных стохастических систем и компьютерной поддержке исследований и разработок. Приводятся необходимые сведения из теории вероятностей, стохастического анализа и стохастических уравнений. Даются вероятностные распределения процессов в гибридных стохастических системах. Рассматриваются задачи, методы и средства компьютерной поддержки исследований и разработок.

В ч. 3 «Основы математического моделирования, анализа и синтеза систем послепродажного обслуживания» (гл. 7–9) рассматриваются вопросы математического моделирования процессов в СППО, моделирование СППО со смешанными потоками расходования, восстановления и пополнения ресурсов, а также статистический анализ и синтез СППО.

Часть 4 (главы 10–12) посвящена определению и анализу показателей экспортного потенциала ИНП при проектировании. Проводится анализ технико-экономических факторов, определяющих экспортный потенциал ИНП. Рассматриваются вопросы разработки математической модели процессов эксплуатации ИНП для оценки показателей экспортного потенциала, а также оценки влияния эксплуатационно-технических характеристик на экспортный потенциал ИНП.

В заключительной части 5 (гл. 13 и 14) рассматриваются задачи управления поддержкой послепродажного обслуживания на стороне заказчика и на стороне поставщика, а также моделирование инвестиционных процессов ИЛП в условиях современных финансовых рынков.

Таким образом, как следует из краткой аннотации книги, основные решения, позволяющие существенно повысить эффективность методов управления стоимостью ЖЦ ИНП, основываются на современных методах и средствах математического моделирования и оптимизации процессов послепродажного обслуживания ИНП.

Итак, помимо информационного моделирования процессов управления стоимостью ЖЦ ИНП в соответствии со стандартами CALS, весьма важными являются задачи интеллектуальной поддержки управления послепродажным обслуживанием ИНП. Однако методы их решения часто непригодны в условиях стохастической постановки. Фактически нерешенной остается часть проблемы, связанная с автоматизированным планированием потребностей в ресурсах в условиях нестационарного стохастического потребления и выработкой вариантов принимаемых решений по материально-техническому снабжению эксплуатации изделий, обеспечивающая минимальные затраты финансовых средств при заданном уровне исправности парка наукоемкой продукции, что и является главной задачей ИЛП.

Причиной этой нерешенности является невысокий уровень существующих в настоящее время методов математического моделирования процессов послепродажного обслуживания, существенно уступающий уровню моделирования динамики самих изделий в процессе их использования по назначению. Система послепродажного обслуживания, включающая в себя собственно ИНП, а также подсистемы технического обслуживания и ремонта (ТОиР), поставок оборудования и запасных частей (ЗЧ), складирования и транспортировки, обучения персонала и др., оказывается неравнопрочной в смысле уровня математического моделирования, а отсюда и по уровню автоматизации процессов. Обобщая [1–3], приведем ряд основных научно-технических решений для непрерывно-дискретных и эредитарных систем.

2 Описание процессов в непрерывно-дискретных системах послепродажного обслуживания

Представим СППО в виде ориентированного графа, изображенного на рис. 1. Вершины графа соответствуют $1, 2, \dots, n$ состояниям, в которых может находиться некий одноименный материальный ресурс, в том числе финальные ИНП. Количество единиц ресурса в этих состояниях в каждый момент времени $t > 0$ будем обозначать переменными соответственно $Y = [Y_1(t), \dots, Y_n(t)]^T$. Естественно считать эти функции времени составляющими вектора $Y(t)$ размерности n , являющегося по существу фазовым для данной системы.

Дуги графа соответствуют переходам ресурса из состояний k в состояния h ($k, h = 1, 2, \dots, n; k \neq n$). Некоторые состояния ресурса соответствуют его нахождению в системах массового обслуживания (СМО) по восстановлению

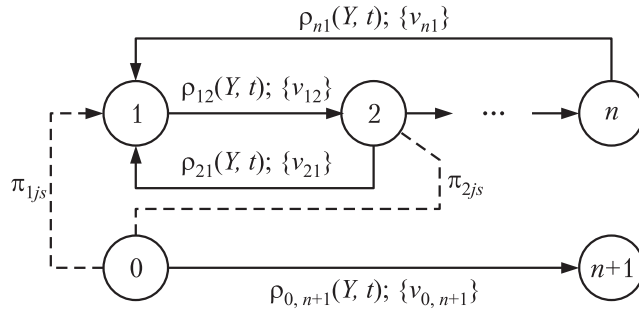


Рис. 1 Общий вид графа системы

до необходимой кондиции после использования, по подготовке к повторному использованию и т. д. Если каналы обслуживания идентичны, то общая производительность СМО определяется числом каналов. На графе это отражено обозначениями вида $\rho_{kh}(Y, t)$, т. е. нелинейная функция в общем случае имеет векторный аргумент. На рис. 2 для примера приведен график зависимости суммарной производительности системы, занимающейся ремонтом ИНП, от количества $Y_1(t)$ ремонтируемых изделий и числа A ремонтных групп. Здесь учитывается, что каждая из них может заниматься одновременно ремонтом только одного изделия с производительностью λ_0 , т. е. поток на выходе каждого канала ремонта является ординарным и имеет интенсивность, равную λ_0 (количеству изделий в единицу времени).

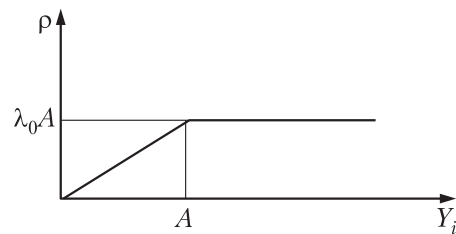


Рис. 2 Нелинейный вид функции производительности системы ремонта

Возможна другая ситуация, когда на выходе каждого канала СМО будет наблюдаться поток «групп» (порций) ИНП или других видов ресурсов случайной численности (объема) ν_{kh} . При этом запись $\rho_{kh}(Y, t); \{\nu_{kh}\}$, сделанная при разметке графа на рис. 1, означает, что из состояния k в состояние h переходят «группы» материальных ресурсов численностью $\{\nu_{kh}\}$ с интенсивностью $\rho_{kh}(Y, t)$. Под интенсивностью в данном случае понимается количество групп ресурсов, появляющихся на выходе СМО в единицу времени.

Применяя теорию гибридных стохастических систем, основанную на эквивалентной статистической линейаризации [1], придем к следующим уравнениям состояния, а также уравнениям для вероятностных моментов первого и второго порядка:

$$\frac{dY_i(t)}{dt} = \Phi_i(Y(t)) + \sum_{k,h=0}^n \sum_{j_{kh}=1}^{\infty} s_{khi}(t_{j_{kh}})\delta(t - t_{j_{kh}}), \quad i = 1, 2, \dots, n; \quad (1)$$

$$\begin{aligned} \dot{m}(t) &= \mathbf{M} \left[\Phi(Y(t)) + \widehat{S}_{kh}(\nu)\rho \right]; \\ \dot{\theta}(t) &= \mathbf{M} \left\{ \left[\Phi(Y(t)) + \widehat{S}_{kh}\rho \right] Y^{\circ T} + Y^0 \left[\Phi^T(Y(t)) + \rho^T \widehat{S}_{kh}^T \right] + \right. \\ &\quad \left. + \widehat{S}_{kh} \text{diag}(\rho) \widehat{S}_{kh}^T \right\}; \quad (2) \end{aligned}$$

где \mathbf{M} — символ математического ожидания; \mathbf{T} — символ транспонирования; $\Phi(Y(t)) = \{\Phi_i(Y(t))\}$ — регулярная составляющая; $\widehat{S}_{kh} = S_{kh}^T$ — матрицы, характеризующие групповые и одинарные переходы; $m(t)$ — $(n \times 1)$ -вектор математического ожидания процесса $Y(t)$; $Y^\circ(t) = Y(t) - m(t)$; $\theta(t)$ — $(n \times n)$ -матрица ковариаций векторного процесса $Y(t)$; $\rho = \rho(Y, t)$ — $(m \times 1)$ -вектор, составленный из интенсивностей $\rho_{kh}(Y, t)$; $\text{diag}(\rho(Y, t))$ — $(m \times m)$ -диагональная матрица вектора ρ , составленного из $\rho_{kh}(Y, t)$.

При этом предполагается, что в рассматриваемой модели СППО существуют неслучайные последовательности вида $\{t_{j1}, t_{j2}, \dots, t_{jr}, j = 1, 2, \dots\}$, образующие неординарные регулярные потоки (потоки групп). Их действие в СППО означает, что процесс $Y(t)$, характеризующий наличие ресурсов в состояниях $1, 2, \dots, n$, претерпевает скачкообразные изменения в моменты времени t_{js} , j — номер момента времени внутри последовательности, s — номер последовательности, $s = 1, 2, \dots, r$. В общем случае скачкообразные изменения могут происходить у каждой компоненты вектора $Y(t)$ в моменты времени t_{js} . Эти изменения могут быть записаны в виде матричного уравнения

$$Y(t_{js}) = DY(t_{js} - 0) + \Pi(t_{js}), \quad (3)$$

где $D = D(t_{js})$ — $(n \times n)$ -матрица, как правило, единичная; $\Pi(t_{js}) \in R^n$ — вектор скачкообразных изменений вектора $Y(t)$. Для дальнейшего его удобно определить в виде $\Pi(t_{js}) = \Pi_{js} = [\pi_{1js}, \pi_{2js}, \dots, \pi_{njs}]^T$, где π_{ijs} — пополнение СППО в состоянии i ($i = 1, 2, \dots, n$) в моменты времени t_{js} .

Необходимо отметить, что часть регулярных потоков может быть связана с внутрисистемными процессами накопления и расходования ресурсов, а часть — с пополнением СППО из внешних источников. При этом один поток может характеризовать пополнение ресурсов, а другой — их потерю, связанную, например, с природными катаклизмами. Для моделирования процессов пополнения, если источник, условно говоря, неисчерпаем, целесообразно присоединить к про-

странству состояний СППО псевдосостояние «0», подразумевая под ним этот источник. Так, у графа, изображенного на рис. 1, при пополнении СППО в моменты времени t_{js} ресурсами в состоянии 1 объемами π_{1js} , а в состоянии 2 объемами π_{2js} появятся дополнительные связи 0–1 и 0–2, что отражено пунктирными линиями.

Для определения обновляющихся начальных условий в уравнениях (1) и (2), описывающих вектор математических ожиданий $m(t)$ составляющих $Y(t)$ и ковариационную матрицу $\theta(t)$, применим к (3) известные операции определения математического ожидания. В результате получим следующие выражения для математических ожиданий

$$m(t_{js}) = Dm(t_{js} - 0) + M \{ \Pi(t_{js}) | Y(t_{js} - 0) \}$$

и ковариаций

$$\begin{aligned} \theta(t_{js}) = & D\theta(t_{js} - 0)D^T + DM \left\{ Y^\circ(t_{js} - 0)\Pi^T(t_{js}) | Y(t_{js} - 0) \right\} + \\ & + M \left\{ \Pi(t_{js})Y^\circ(t_{js} - 0) | Y(t_{js} - 0) \right\} D^T + M \left[\Pi^\circ(t_{js})\Pi^{\circ T}(t_{js}) \right], \end{aligned}$$

где $Y^\circ(t) = Y(t) - m(t)$, $\Pi^\circ(t) = \Pi(t) - m_\Pi(t)$ — центрированные векторы.

3 Анализ и моделирование процессов управления запасами материальных средств при поддержании заданного уровня технической готовности парка изделий наукоемкой продукции

Эту задачу рассмотрим на примере парка однотипных ИНП, состояние которого описывается графом (рис. 3).

Вершины графа отображают возможные состояния составных частей изделия (СЧИ) в системе ППО: 1 — исправные СЧИ на складе; 2 — исправные СЧИ, эксплуатируемые в составе изделия; 3 — неисправные СЧИ, находящиеся на восстановительном ремонте у потребителя; 4 — СЧИ в состоянии списания; 5 — исправные СЧИ в составе ИНП, находящиеся на профилактических работах; 6 — СЧИ в составе ИНП, находящихся на капитальном ремонте; 7 — неисправные СЧИ, накопленные на складе у потребителя для отправки в ремонт; 8 — СЧИ, находящиеся на восстановительном ремонте у поставщика. В обороте находится определенное количество СЧИ каждого типа. В силу различных случайных факторов все они распределяются случайным образом между указанными выше состояниями.

Введем в рассмотрение фазовый вектор системы $Y(t)$, составляющие которого $Y_i(t)$, $i = 1, \dots, n$, суть количество СЧИ данного типа, находящихся в i -м состоянии.

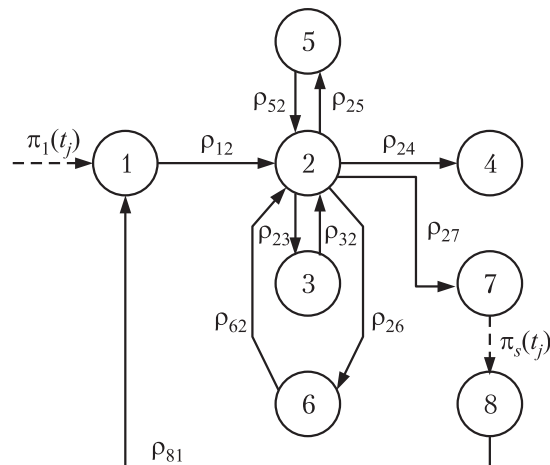


Рис. 3 Система послепродажного обслуживания заказчика

Примем следующие допущения.

1. Составные части изделия в состояниях 3, 4 считаются неисправными, а в состоянии 5 — исправными.
2. При неисправном состоянии в любом случае СЧИ подлежит замене за счет запасов склада, а далее либо отправляется в ремонт, либо списывается.
3. Ремонт СЧИ может быть восстановительным и выполняться в оперативном порядке у потребителя.
4. Сложные виды ремонта СЧИ осуществляются на оборудовании поставщика. Отправка в ремонт производится периодически партиями, которые накапливаются на складе в течение обусловленного периода времени.
5. Плановый капитальный ремонт ИПП выполняется на специализированных ремонтных предприятиях. Отправка изделий на ремонт осуществляется по мере наступления сроков. Запасные части, необходимые для проведения капремонта, поставляются из внешних источников и в данной модели не учитываются. Их стоимость входит в стоимость ремонта.
6. Пополнение объемов ЗЧ осуществляется периодически партиями определенного размера, обеспечивающими оптимальный уровень запаса при заданном среднем коэффициенте исправности изделий, рассчитанном на заданный период функционирования.
7. Критерий оптимальности содержит свободные и несвободные параметры. Свободные параметры учитывают возможность выбора стратегии и тактики использования изделий, а отсюда — оптимизацию доходов и расходов.

Несвободные параметры характеризуют качественные, ценовые и другие факторы, присущие изделию и системе.

Дефицит ЗЧ приводит к простоям, что снижает средний коэффициент исправности. Простой вызывает также нахождение СЧИ в состояниях обслуживания и ремонта.

Стрелками показаны дуги графа, описывающие переходы СЧИ и изделия в целом из одного состояния в другое. Обозначения дуг графа цифрами вида $k-h$ отмечают направления переходов из состояния k в состояние h . Состояния, в которых над СЧИ осуществляются какие-либо действия, соответствуют нахождению их в определенных технологических фазах, реализуемых СМО. В данном случае такими СМО являются система технического обслуживания, система восстановительного ремонта СЧИ у потребителя, система капитального ремонта изделий, система восстановительного ремонта СЧИ у поставщика и склад, который принимает и отпускает СЧИ. Считается, что переходы $k-h$ обусловлены случайными пуассоновскими потоками событий (общепринятая гипотеза применительно к СМО).

Одним из важнейших показателей ИЛП является средний коэффициент технической готовности (исправности) на заданном промежутке времени $[0, T]$, который применительно к рассматриваемой системе определяется по формуле

$$\bar{K}_n(T) = \frac{1}{T} \int_0^T \frac{Y_2(\tau) + Y_5(\tau)}{N} d\tau, \quad (4)$$

где $Y_2(t)$ и $Y_5(t)$ — количество исправных СЧИ; N — находящееся в эксплуатации количество СЧИ данного типа, строго согласованное с количеством эксплуатируемых ИНП (с учетом ИНП, находящихся в капитальном ремонте).

Для приведения к общему виду расширим фазовый вектор $Y(t)$ путем введения переменной $Y_9(t)$:

$$Y_9(t) = \bar{K}_n(t).$$

С целью определения вероятностных характеристик среднего значения коэффициента исправности парка СЧИ на периоде $[0, t)$ заменим в (4) конечный предел T на переменный — t и продифференцируем по этой переменной. Получим дифференциальное уравнение, описывающее динамику изменения среднего коэффициента исправности:

$$\dot{Y}_9(t) = -\frac{1}{t} Y_9(t) + \frac{1}{t} \frac{Y_2(t) + Y_5(t)}{N}.$$

Как видно, правая часть этого уравнения содержит квазипуассоновские ступенчатые процессы $Y_2(t)$, $Y_5(t)$.

В качестве критерия эффективности комплексной системы «поставки запчастей – эксплуатация – техническое обслуживание – ремонт – списание» примем выражение:

$$J = C_k \int_0^T (\bar{K}_и - \bar{K}_{и.тр})^2 d\tau, \quad (5)$$

где $\bar{K}_и(t)$ — среднее на интервале $[0, t)$ значение коэффициента исправности; $\bar{K}_{и.тр}$ — требуемое значение среднего на периоде $[0, T]$ коэффициента исправности; C_k — коэффициент.

Физический смысл принятого критерия означает, что в процессе оптимизации параметров указанной комплексной системы будут определены такие их значения, при которых площадь под кривой модуля разности $\bar{K}_и(t) - \bar{K}_{и.тр}$ будет наименьшей. При этом текущее значение $\bar{K}_и(t)$ будет минимально отличаться от $\bar{K}_{и.тр}$. В дальнейшем будем называть этот критерий критерием оптимальности.

Можно также использовать модернизированную форму критерия оптимальности, учитывающую требуемый уровень гарантии соблюдения заданного коэффициента готовности:

$$J^* = C_k \int_0^T [(\bar{K}_и - \mu\sigma_{ки}^2) - \bar{K}_{и.тр}]^2 d\tau, \quad (6)$$

где $\sigma_{ки}$ — среднеквадратическое отклонение среднего на периоде коэффициента исправности; μ — квантиль нормального распределения случайной переменной $\bar{K}_и(t)$, соответствующий заданной гарантийной вероятности.

Обозначим $Y_{10}(t) = J$ и продифференцируем (5) по времени. При этом, учитывая (6), предварительно в правой части интеграла заменим верхний предел на t .

В результате получим дифференциальное уравнение, описывающее текущее на периоде $[0, t]$ значение критерия J :

$$\dot{Y}_{10}(t) = C_k (Y_9(t) - \bar{K}_{и.тр})^2 = C_k [Y_9^2(t) - 2Y_9(t)\bar{K}_{и.тр} + \bar{K}_{и.тр}^2].$$

Исходя из анализа графа состояний (см. рис. 3) и оценки количества переменных в системе, имеем следующие выражения для S , $\bar{\rho}$ и Φ^T , входящие в (1) и (2):

$$S = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix}; \bar{p} = \begin{bmatrix} \lambda_0 Y_2^* (m_F + K_F Y_1^0) \\ \lambda_{23} Y_2 \\ \lambda_{24} Y \\ \lambda_{25} Y_2 \\ \lambda_{26} Y \\ \lambda_{27} Y_2 \\ K_{03} + K_{13} \overset{\circ}{Y}_3 \\ K_{05} + K_{15} \overset{\circ}{Y}_5 \\ K_{06} + K_{16} \overset{\circ}{Y}_6 \\ K_{08} + K_{18} \overset{\circ}{Y}_8 \end{bmatrix}. \quad (7)$$

Здесь через m_F , K_F , K_{0i} и K_{1i} обозначены коэффициенты статистической линеаризации нелинейных функций, причем

$$\Phi^T(Y(t)) = \left[0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \frac{Y_2 + Y_5}{tN} - \frac{Y_9}{t} Y_9^2 - 2Y_9 \bar{K}_{и.тр} + \bar{K}_{и.тр}^2 \right]. \quad (8)$$

Полученные выражения (7) и (8) положены в основу математического обеспечения инструментального программного комплекса прогнозирования потреб-

The screenshot shows a software window titled "Динамика процессов" (Dynamics of Processes). It contains several input fields for parameters:

- Входные данные (Input Data):**
 - Штатное количество ФИ (шт): 50
 - Кол-во исправн СЧИ (шт): 50
 - Количество СЧИ на складе (шт): 1
 - Ки заданный: 0.8
 - Интервал прогноза (в годах): 10
 - Периодичность пополнения (сут): 365
 - Неработка на отказ (час): 600
 - Неработка СЧИ между кап. ремонтами: 1000
 - Доля СЧИ в заводской ремонт: 0.88
 - Доля СЧИ в безвоз. потери: 0.02
 - Количество СЧИ на зав. ремонте: 0
 - Неработка СЧИ в год (час): 200
 - Неработка ФИ между проф. раб. (час): 100
 - Неработка ФИ между кап. ремон. (час): 1000
 - Объем партии СЧИ для отправки в зав. ремонт (шт): 1
 - Количество ФИ на кап. ремонте: 0
 - Производительность кап. ремонта СЧИ/год: 15
 - Производительность проф. раб. на ФИ/мес: 8
 - Производительность кап. ремонта ФИ/год: 10
 - Производительность зав. ремонта СЧИ/год: 15
- Выбор параметров (Parameter Selection):**
 - Все
 - Количество неисправных ФИ
 - Количество ЭЧ на складе
 - Количество СЧИ в эксплуатации
 - Количество СЧИ на кап. ремонте
 - Списанные СЧИ
 - Количество ФИ на проф. обслуж.
 - Количество ФИ на капит. ремонт.
 - Количество СЧИ на заводск.рем.
 - Ср. коэффициент исправности
- Financial Parameters:**
 - Цена (тыс.\$): 50
 - Коеф-т эскалации цены: 0.05
 - Стоимость НП (тыс.\$):
 - Суммарная стоимость (тыс.\$):
- Inventory Table:**
 - Программная поставка запасных частей
 - Table with columns: год (year) 3, 4, 5, 6, 7, 8, 9, 10 and rows: к-во (quantity), Цена (price), Стоимость (cost).

Рис. 4 Форма ввода-вывода данных

ностей (ПКПП) в ЗЧ и планирования программы поставок на произвольный период эксплуатации.

На рис. 4 показана форма ввода-вывода данных в ПКПП, где часть информации касается финального изделия (ФИ), а другая часть касается одной (выбранной из номенклатурного перечня) компоненты ИНП.

В нижней части рисунка показана начальная цена одной ЗЧ в текущем году, коэффициент эскалации, стоимость начальной поставки, а также таблица, содержащая:

- объемы пополнения ЗЧ на каждый год из заданного периода эксплуатации;
- цена поставки одной ЗЧ по годам с учетом коэффициента эскалации;
- стоимость поставляемых объемов по годам эксплуатации.

В отдельном окне показывается суммарная стоимость поставок ЗЧ по данной компоненте с учетом начальной поставки.

Программный комплекс позволяет решать задачу определения необходимых пополнений ЗЧ по каждой компоненте и в целом для парка ИНП на заданном периоде эксплуатации.

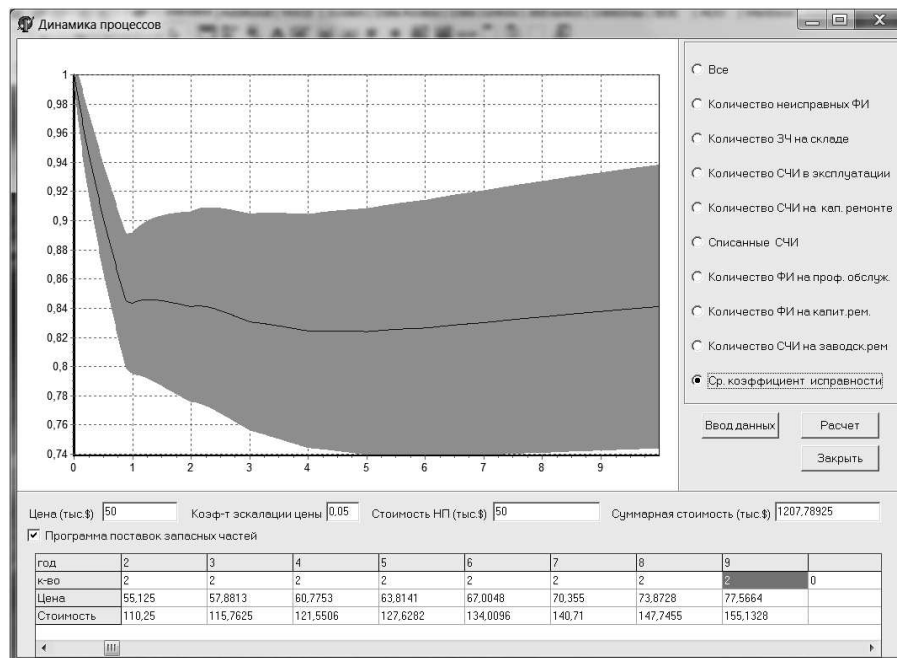


Рис. 5 График коэффициента исправности СЧИ, соответствующий произвольной программе поставок

Объемы пополнения ЗЧ по каждой компоненте считаются оптимальными, если прогнозируемый при этих объемах уровень исправности на этом периоде эксплуатации ИПП минимально отличается от заданного и при этом стоимость поставок, осуществленных на этом периоде, минимальна.

В ПКПП определение оптимальных объемов пополнения (оптимальной программы поставок на заданный период времени) осуществляется автоматически в соответствии с методом градиентного поиска. Критерием оптимизации является выражение (7). Как показали дальнейшие результаты, минимум значений критерия совпадает одновременно с минимальной стоимостью поставок ЗЧ. Это — уникальное свойство выбранного критерия. Излишки ЗЧ при этом на складе являются минимальными.

На рис. 5 показан один из графиков процессов — средний коэффициент исправности на периоде 10 лет в сравнении с заданным, равным 0,8. Показана верхняя и нижняя граница возможных значений реального среднего коэффициента исправности. Этот график соответствует произвольно введенным данным программы поставок ЗЧ заданного типа по годам: 2, 2, 2, 2, 2, 2, 2, 2, 2. При этом график прогнозируемого среднего коэффициента исправности парка изделий проходит выше заданного значения 0,8. Стоимость такой программы составит в среднем \$ 1207,8.

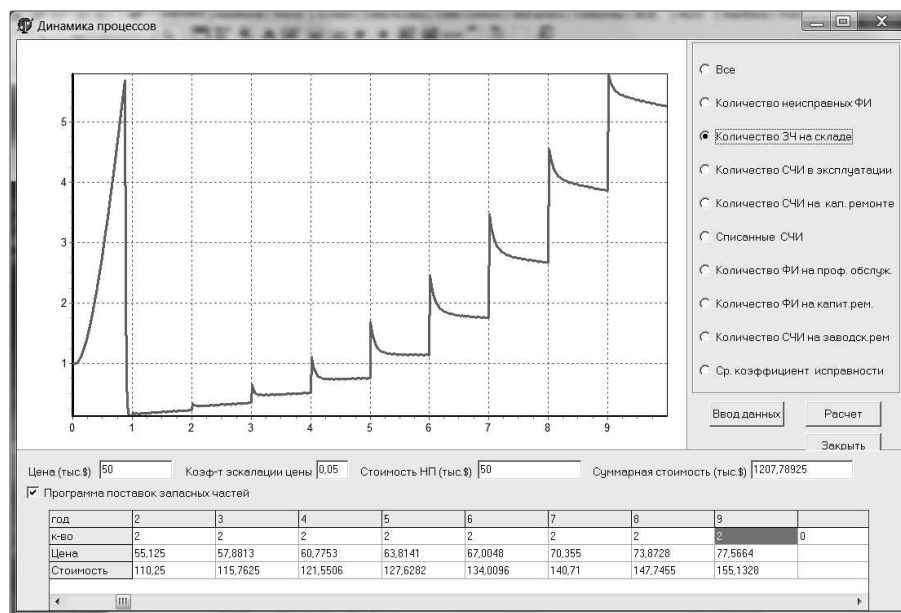


Рис. 6 Прогнозируемое количество ЗЧ на складе

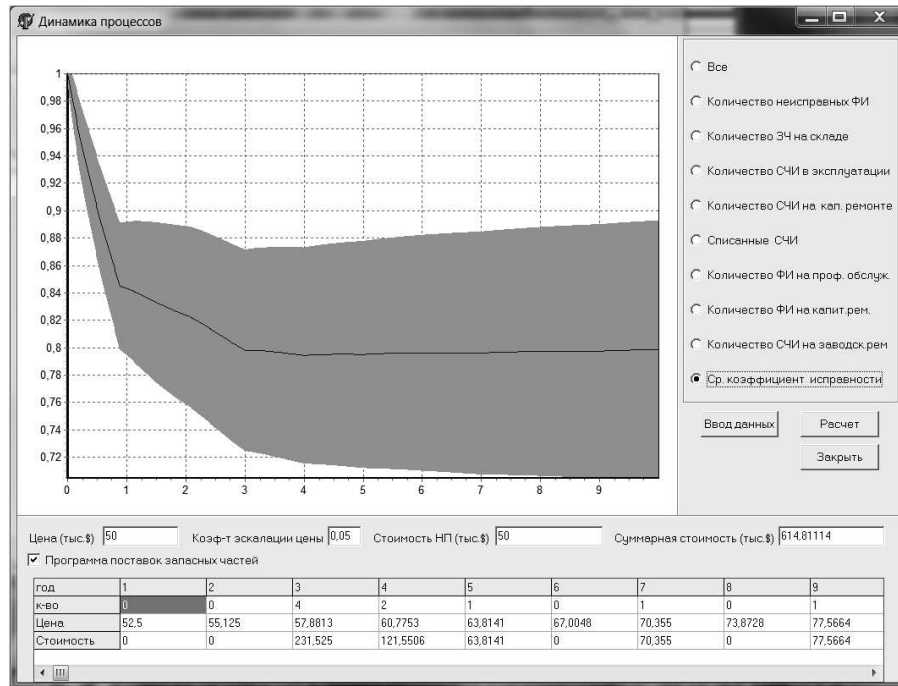


Рис. 7 Средний коэффициент исправности при оптимальной программе поставок

Помимо среднего коэффициента исправности, на графике показаны доверительные границы этого коэффициента, соответствующие доверительной вероятности, равной 0,9.

На рис. 6 показан график ожидаемого изменения среднего количества ЗЧ данного типа на складе при реализации данной (неоптимальной) программы. Очевидна избыточность ЗЧ на данном временном периоде, приводящая к завышенным затратам. На рис. 6 в строке «год» начало отсчета идет от цифры 2. Это несоответствие порождено горизонтальным скроллингом, скрывшим в данном случае часть информации.

В процессе работы алгоритма оптимизации, основанного на применении процедуры градиентного поиска, получаем оптимальную программу поставок ЗЧ (0, 0, 4, 2, 1, 0, 1, 0, 1), при которой средний коэффициент исправности минимально отличается от заданного, равного 0,8 (рис. 7). График прогнозируемого среднего коэффициента исправности показывает практически полное совпадение с заданным на интервале времени после 3-го года эксплуатации. Предложенная программа является оптимальной. Стоимость ее составляет \$ 674,81114, что в 2 раза меньше ранее предложенной произвольной программы поставок ЗЧ.

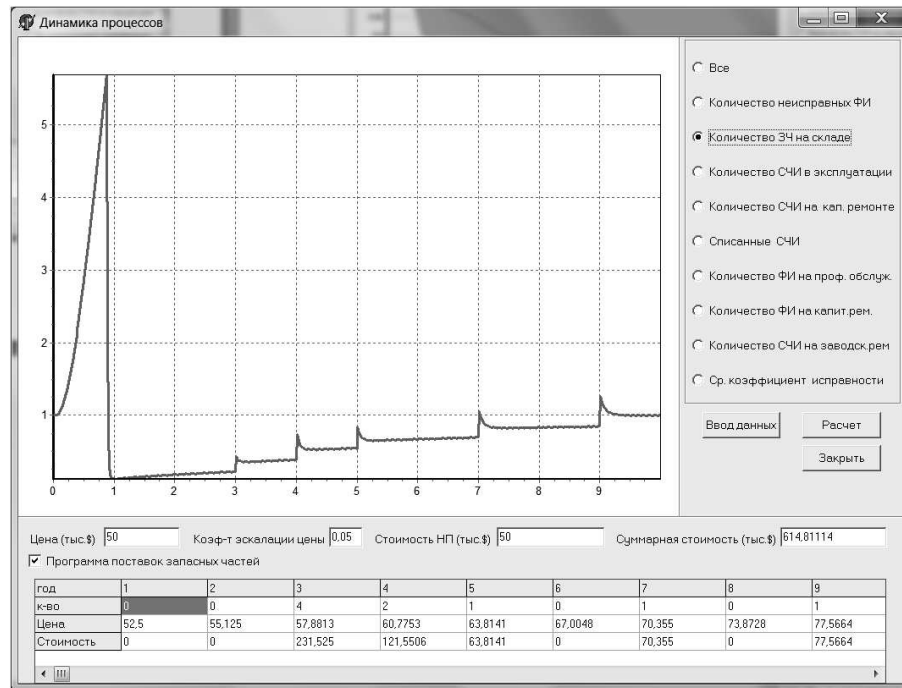


Рис. 8 Избыточное среднее количество ЗЧ на складе при оптимальной программе поставок

Отметим, следуя рис. 7, что на первом году эксплуатации происходит резкое снижение коэффициента исправности. Это объясняется тем, что весь парк ИПП исправен и максимально используется по назначению. Уменьшение коэффициента исправности ниже заданного значения 0.8 предотвращается за счет ЗЧ, накопленных после ремонта в течение года эксплуатации. Это результат использования стратегии материально-технического обеспечения, в соответствии с которой выдача ЗЧ со склада осуществляется при ее наличии на складе и наличии заявки от эксплуатирующей организации, но не ранее, чем коэффициент исправности приблизится (уменьшится) к заданному значению. Избыточное количество ЗЧ на складе при оптимальной программе поставок составит не более одной ЗЧ в течение 10 лет (рис. 8). Необходимо отметить, что алгоритм оптимизации программы поставок работает последовательно на промежутках времени, равных 1 году. При этом начальным условием для очередного года является оптимальное значение поставки в предыдущем году.

Оптимальное планирование на один год осуществляется с помощью этого же алгоритма, но с учетом реальных остатков запасов, хранящихся на складе от

предыдущих поставок, и ЗЧ, поступающих из ремонта. Кроме того, учитывается, что планирование осуществляется как минимум за полгода до начала очередного года поставки. Это вносит некоторые коррективы в алгоритм оптимизации.

Поставка заказчикам подобных комплексов прогнозирования и оптимизации позволит им формировать оптимальную годовую заявку на поставку ЗЧ для проведения как плановых, так и неплановых работ по ТОиР для выполнения плана использования парка ИНП с учетом реальных показателей своей СППО, а также на основе данных, сообщаемых поставщиком в соответствии с контрактом.

4 Особенности моделирования эредитарных стохастических систем интегрированной логистической поддержки

Эредитарные системы, также известные как наследственные системы, системы с последствием или системы с запаздыванием времени, содержат в одном или нескольких из своих звеньев запаздывание во времени начала изменения выходной величины (после начала изменения входной переменной). Эта величина запаздывания остается постоянной во всем последующем ходе процесса и называется временем запаздывания. Если объект характеризуется некоторым параметром, различным по своему значению в разных точках объекта, то говорят, что значения такого параметра распределены (по объекту). Когда таких параметров несколько, объект рассматривается как система с распределенными параметрами. Поведение систем с распределенными параметрами описывается дифференциальными уравнениями в частных производных, дифференциально-разностными, интегральными и иными сложными математическими соотношениями.

Эредитарные стохастические системы служат адекватными моделями в СППО. Для решения задач офлайн- и онлайн-обработки информации в нестационарных эредитарных стохастических системах разработан широкий набор методов и вычислительных средств [4–10].

Особую практическую значимость для оценки воздействий нестационарных, в частности ударных, процессов имеют методы, основывающиеся на сведениях стохастических интегродифференциальных уравнений, описывающих эредитарные стохастические системы, к стохастическим дифференциальным уравнениям с последующим использованием фильтров Калмана и Пугачёва. Интегродифференциальные стохастические уравнения и методы их сведения описываются в работах [5, 6, 8, 10]. Синтез фильтров Пугачёва подробно рассмотрен в работе [8]. Решение задач онлайн-обработки информации в эредитарных стохастических системах в условиях ударных воздействий (УВ) имеет специфические особенности [10]. Рассмотрим их применительно к элементам СППО.

Рассмотрим эредитарную стохастическую систему, описываемую векторным интегродифференциальным уравнением (в смысле Ито) следующего вида [6]:

$$\dot{X} = a(X, t) + \int_{t_0}^t a_1(X(\tau), \tau, t) d\tau + \left[b(X, t) + \int_{t_0}^t b_1(X(\tau), \tau, t) d\tau \right] \dot{W},$$

$$X(t_0) = X_0. \quad (9)$$

Здесь $X \in R^p$ — вектор состояния; $\dot{W} \in R^q$ — векторный белый шум интенсивности $v = v(t)$, не зависящий от начальных условий X_0 ; $W = W(t)$ — случайный процесс второго порядка с независимыми приращениями;

$$a = a(X, t), \quad a : R^p \times R \rightarrow R^p; \quad a_1 = a_1(X(\tau), \tau, t), \quad a_1 : R^p \times R \times R \rightarrow R^p;$$

$$b = b(X, t), \quad b : R^p \times R \rightarrow R^{pq}; \quad b_1 = b_1(X(\tau), \tau, t), \quad b_1 : R^p \times R \times R \rightarrow R^{pq}.$$

Функции a_1 и b_1 представим в виде:

$$a_1 = A(t, \tau)\varphi(X(\tau), \tau); \quad b_1 = \sum_{h=1}^N B_h(t, \tau)\psi_h(X(\tau), \tau),$$

где эрмитарные ядра $A(t, \tau)$ и $B_h(t, \tau)$ — $(p \times p)$ - и $(p \times q)$ -матричные функции; $\psi_h(X(\tau), \tau)$ — $(q \times q)$ -матричная функция; $\varphi(X(\tau), \tau)$ — вектор-функция размерности $(p \times 1)$. Допустим, что эрмитарные ядра $A(t, \tau) = \{A_{ij}(t, \tau)\}$, $i, j = \overline{1, p}$, и $B_h(t, \tau) = \{B_{hij}(t, \tau)\}$, $h = \overline{1, N}$, $i, j = \overline{1, p}$, $j = \overline{1, q}$, удовлетворяют условиям физической реализуемости и условиям затухающей эрмитарности (или асимптотической устойчивости):

$$A_{ij}(t, \tau) = 0, \quad B_{hij}(t, \tau) = 0, \quad \forall \tau > t;$$

$$\int_{-\infty}^{\infty} |A_{ij}(t, \tau)| d\tau < \infty, \quad \int_{-\infty}^{\infty} |B_{hij}(t, \tau)| d\tau < \infty.$$

В том случае, когда эрмитарные ядра удовлетворяют условиям

$$A_{ij}(t, \tau) = \tilde{A}_{ij}(\xi), \quad B_{hij}(t, \tau) = \tilde{B}_{hij}(\xi), \quad \xi = t - \tau,$$

говорят об инвариантной стационарной эрмитарности.

Важный класс эрмитарных ядер — сингулярные ядра

$$A_{ij}(t, \tau) = A_{ij}^+(t)A_{ij}^-(\tau), \quad B_{hij}(t, \tau) = B_{hij}^+(t)B_{hij}^-(\tau).$$

В случае, когда функции $a(X, t)$ и $\varphi(X(\tau), \tau)$ линейны относительно $X(t)$, а $b(X, t)$ и $\psi_h(X(\tau), \tau)$ не зависят от $X(t)$, т. е. когда они представимы в виде

$$a(X, t) = a_{1t}X_t + a_{0t}; \quad \varphi(X(\tau), \tau) = a_{1\tau}^H X_\tau + a_{0\tau}^H;$$

$$b(X, t) + \int_{t_0}^t b_1(X(\tau), \tau, t) d\tau = c_{0t},$$

получим из (9) линейную эредитарную стохастическую систему вида

$$\dot{X} = a_{1t}X_t + a_{0t} + \int_{t_0}^t A(t, \tau) [a_{1\tau}^H X_\tau + a_{0\tau}^H] d\tau + c_{0t}\dot{W}.$$

Если эредитарность стохастическая, функции $b(X, t)$ и $\psi_h(X(\tau), \tau)$ линейны относительно $X(t)$ и представимы в виде

$$b(X, t) = c_{1t}X_t + c_{0t}, \quad \psi(X(\tau), \tau) = c_{1\tau}^H X_\tau + c_{0\tau}^H,$$

то вместо (9) получим следующее уравнение:

$$\dot{X} = a_{1t}X_t + a_{0t} + \int_{t_0}^t A(t, \tau) [a_{1\tau}^H X_\tau + a_{0\tau}^H] d\tau +$$

$$+ \left[c_{1t}X_t + c_{0t} + \sum_{h=1}^N \int_{t_0}^t B_h(t, \tau) [c_{1\tau}^H X_\tau + c_{0\tau}^H] d\tau \right] \dot{W}.$$

В [6] разработаны два способа приведения уравнений эредитарных стохастических систем к дифференциальным стохастическим уравнениям вида

$$\dot{Z} = c(Z, t) + l(Z, t)\dot{W}, \quad Z(t_0) = Z_0,$$

где Z — расширенный вектор состояния дифференциальной стохастической системы, соответствующей исходной эредитарной стохастической системе; $c(Z, t)$ и $l(Z, t)$ — известные функции Z и t . Начальные условия $Z(t_0) = Z_0$ вектора состояния не зависят от белого шума W .

В прикладной механике [10] силовым (или жестким) ударом называют результат воздействия на тело сил бесконечно малой продолжительности, вызывающих

мгновенное изменение его скорости на конечную величину. В технике ударозащищенных систем (УЗС) понятие удара трактуется более широко. Ударом или ударным воздействием называют силовое или кинематическое возмущение относительно малой продолжительности и относительно большой интенсивности. В качестве эталона интенсивности при этом выбирают нормативную величину, снижение или увеличение которой составляет цель защиты от УВ. В качестве эталона длительности принимают период собственных колебаний защищаемого вычислительного оборудования (ВО).

Анализ поведения УЗС при УВ показывает, что имеется ряд параметров УВ, в наибольшей степени влияющих на движение защищаемого оборудования. К их числу относятся: пиковое значение σ_0 , длительность τ и полный импульс S_0 УВ $\sigma(t)$, представляющего при кинематическом возмущении закон изменения ускорения основания (рис. 9). Эти величины связаны соотношением

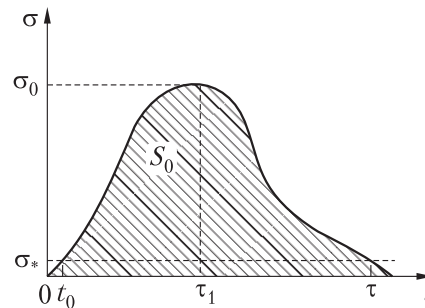


Рис. 9 Параметры ударного импульса

$$S_0 = \int_0^{\tau} \sigma(t) dt.$$

В реальных условиях длительность всякого удара ограничена, поэтому считают $\sigma(t) = 0$ при $t < 0$ и $t > \tau$. Вместе с тем для удобства аналитического описания формы импульса часто используют аппроксимирующие функции, не обращающиеся тождественно в нуль ни при каком конечном t . При этом иногда вводят понятие о длительности на уровне σ_* , считая этой длительностью промежуток времени $t_0 < t < \tau_*$, где $\sigma(t)$ удовлетворяет условию $|\sigma(t)| \geq \sigma_*$.

По величине полного импульса S_0 УВ подразделяют на удары с приращением и без приращения скорости. Импульсы ударов с приращением скорости называют односторонними, если за время удара $\sigma(t)$ сохраняет знак. В описании односторонних импульсов различают фронт нарастания и фронт спада импульса с длительностями соответственно τ_1 и τ_2 .

Важную роль при моделировании играют стохастические УВ, содержащие различные случайные возмущения. В таблице собраны некоторые типовые детерминированные ударные импульсы и стохастические ударные импульсы, уравнения которых содержат аддитивную $\eta(t)$ и мультипликативную $\xi(t)$ случайные функции. В данной работе рассматриваются только нормально распределенные случайные функции $\eta(t)$ и $\xi(t)$.

Типовые УВ с аддитивными и мультипликативными помехами

№	Детерминированный ударный импульс $\sigma(t)$	Стохастический ударный импульс $\sigma(t)$
1	$\sigma_0 \sin\left(\frac{\pi t}{\tau}\right)$	$\sigma_0(1 + \xi(t)) \sin\left(\frac{\pi t}{\tau}\right) + A\eta(t)$
2	$\sigma_0 \left(1 - \cos 2\pi \frac{\pi t}{\tau}\right)$	$\sigma_0(1 + \xi(t)) \left(1 - \cos 2\pi \frac{\pi t}{\tau}\right) + A\eta(t)$
3	$\begin{cases} 2\sigma_0 \frac{t}{\tau}, & 0 \leq t \leq \frac{\tau}{2} \\ 2\sigma_0 \left(1 - \frac{t}{\tau}\right), & \frac{\tau}{2} < t \leq \tau \end{cases}$	$\begin{cases} 2\sigma_0(1 + \xi(t)) \frac{t}{\tau} + A\eta(t), & 0 \leq t \leq \frac{\tau}{2} \\ 2\sigma_0(1 + \xi(t)) \left(1 - \frac{t}{\tau}\right) + A\eta(t), & \frac{\tau}{2} < t \leq \tau \end{cases}$
4	$\sigma_0 \cos\left(\frac{\pi t}{2\tau}\right)$	$\sigma_0(1 + \xi(t)) \cos\left(\frac{\pi t}{2\tau}\right) + A\eta(t)$
5	$\sigma_0 e^{-at}$	$\sigma_0(1 + \xi(t)) e^{-at} + A\eta(t)$
6	$\sigma_0 a e t e^{-at}$	$\sigma_0(1 + \xi(t)) a e t e^{-at} + A\eta(t)$

В Институте проблем информатики Российской академии наук накоплен большой опыт разработки проблемно-ориентированных диалоговых систем и библиотек для анализа и синтеза стохастических систем [11–14]: «СтС-Анализ» (1989–1990, 2004–2007), «СтС-Фильтр» (1991, 2004–2007), «СтС-Модель» (1993, 2007), NALIB (1991–1993), «СтС-СМА» (2005–2007), а также «СтИТ-КР» (2007–2010) на основе канонических разложений случайных функций. В примерах данной работы представлены результаты, полученные с помощью модуля «СтС-УВ», который является развитием «СтС-Фильтр».

Рассмотрим следующую типовую эрдитарную стохастическую систему:

$$a\ddot{X} + b \int_0^t \dot{X}(\tau) e^{-\beta|t-\tau|} d\tau + c \int_0^t X(\tau) e^{-\gamma|t-\tau|} d\tau = Q(t), \quad X_0 = 0, \quad (10)$$

где a, b, c, β и γ — константы. Расширим вектор состояния системы следующим образом: $X_1 = X, X_2 = \dot{X}, \dot{X}_1 = X_2$ и введем следующие обозначения:

$$\frac{b}{a} = 2\varepsilon_0\omega_0; \quad \frac{c}{a} = \omega_0^2; \quad \frac{Q(t)}{a} = \sigma(t).$$

В результате из (10) получим следующую систему дифференциальных уравнений:

$$\left. \begin{aligned} \dot{X}_1 &= X_2; & \dot{X}_2 &= -2\varepsilon_0\omega_0 Y_1 - \omega_0^2 Y_2 + \sigma(t); \\ \dot{Y}_1 &= -\beta Y_1 + X_2; & \dot{Y}_2 &= -\gamma Y_2 + X_1; \\ X_1(0) &= X_2(0) = Y_1(0) = Y_2(0) = 0. \end{aligned} \right\} \quad (11)$$

Пример 1. Анализ эредитарной системы по априорным данным. Проведем вычислительный эксперимент для системы (11) на промежутке времени $t \in [0, 300]$, при этом в качестве входного воздействия возьмем полусинусоидальный ударный импульс:

$$\sigma(t) = \begin{cases} \sigma_0(1 + \xi(t)) \sin\left(\frac{\pi}{\tau}\right) + \eta(t), & t < \tau; \\ \eta(t), & t > \tau, \end{cases} \quad (12)$$

где $\xi(t)$ и $\eta(t)$ — независимые гауссовские белые шумы с нулевыми математическими ожиданиями и одинаковыми дисперсиями, равными 0,1. Зададим следующие значения параметров системы (11): $\omega_0 = 0,1$, $\sigma_0 = 1$, $\tau = 20\pi$. На рис. 10 изображена реализация входного воздействия $\sigma(t)$ и стандартное квадратическое отклонение $\sqrt{D_\sigma}$.

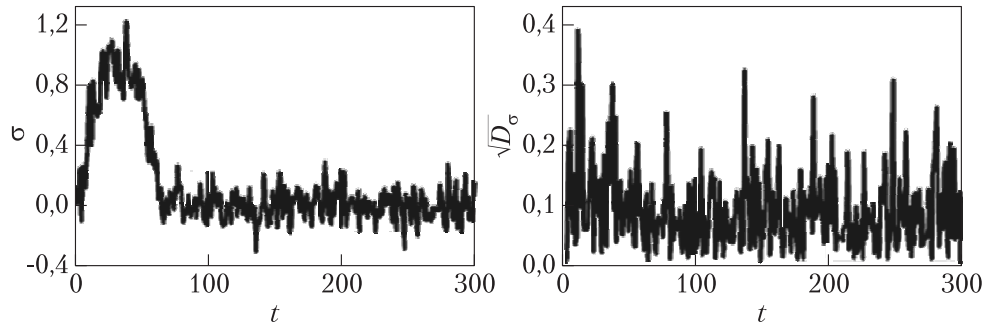


Рис. 10 Входное воздействие $\sigma(t)$ и стандартное квадратическое отклонение $\sqrt{D_\sigma}$

Результаты вычислительного эксперимента для $\varepsilon_0 = 0,1$ представлены на рис. 11, *a* и 12, *a*. На рис. 11, *б* и 12, *б* представлены результаты для $\varepsilon = 0,5$.

Пример 2. Синтез дискретного фильтра Пугачёва для эредитарной системы. В ходе данного вычислительного эксперимента был выполнен синтез дискретного линейного фильтра Пугачёва для системы и УВ из примера 1. На подготовительном этапе проводилась дискретизация методом Эйлера уравнений (11). Также был взят дискретный аналог ударного импульса (12). На

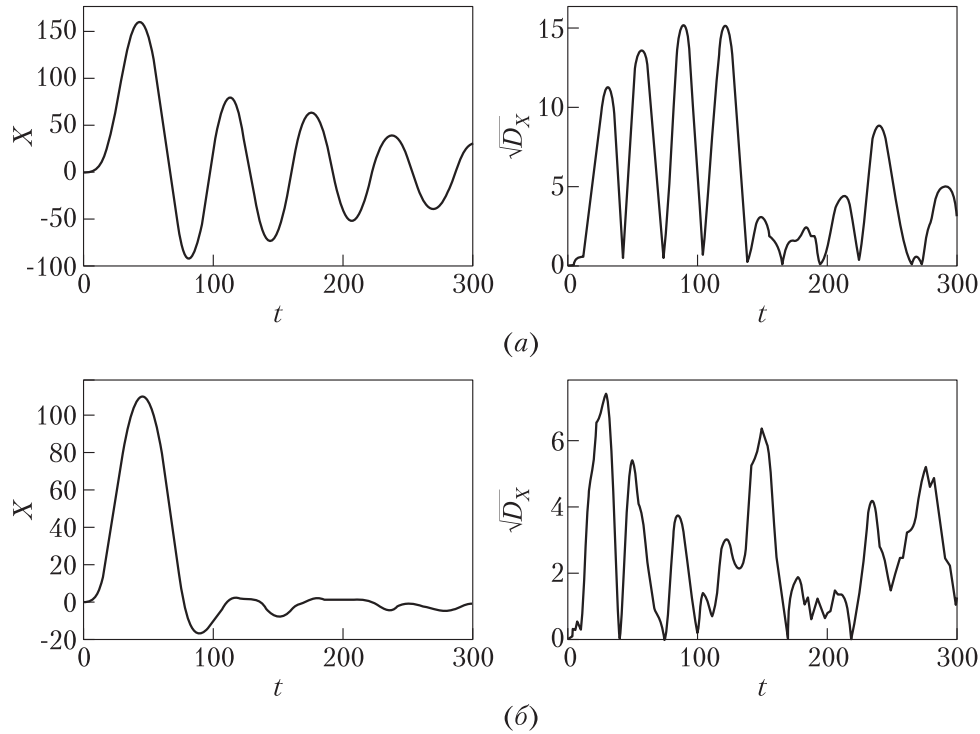


Рис. 11 Координата X и стандартное квадратическое отклонение координаты $\sqrt{D_X}$ системы (11) при $\varepsilon_0 = 0,1$ (а) и $0,5$ (б)

рис. 13–19 представлены результаты моделирования для наблюдаемой величины X_k : $\bar{Y}_k = X_{1k} + 0,1\bar{V}_k$, где \bar{V}_k — дискретный белый шум, и при следующих значениях параметров: $\omega_0 = 0,1$, $\varepsilon_0 = 0,1$, $\sigma_0 = 1$, $\tau = 20\pi$.

Проведенные вычислительные эксперименты для эредитарных систем при различных стохастических УВ показали следующие результаты:

- колебания координаты и ускорения с течением времени затухают;
- значение величины демпфирования ε_0 влияет на скорость затухания колебаний (затухание происходит быстрее при увеличении ε_0);
- предел изменения стандартного отклонения координаты $\sqrt{D_X}$ уменьшается с ростом величины демпфирования;
- для линейной эредитарной системы компоненты ковариационной матрицы ошибки фильтрации позволяют идентифицировать вид ударного импульса (в примере 2 видна структура полусинусоидального импульса).

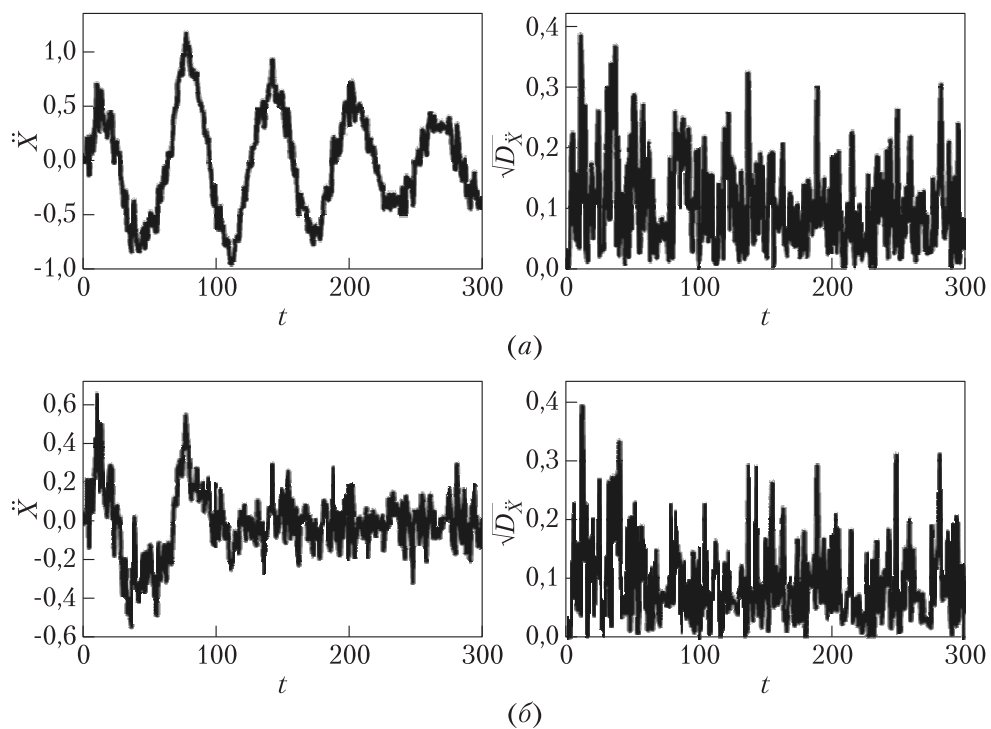


Рис. 12 Ускорение \ddot{X} и стандартное квадратическое отклонение ускорения $\sqrt{D\ddot{X}}$ системы (11) при $\varepsilon_0 = 0,1$ (а) и $0,5$ (б)

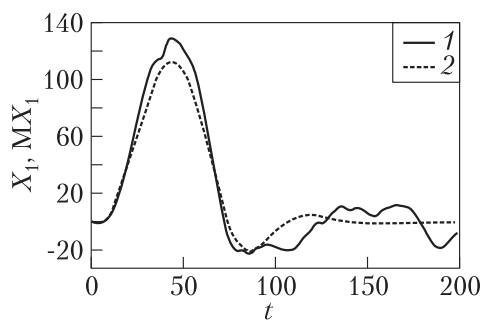


Рис. 13 Величина $X_1(t)$ и ее математическое ожидание $MX_1(2)$

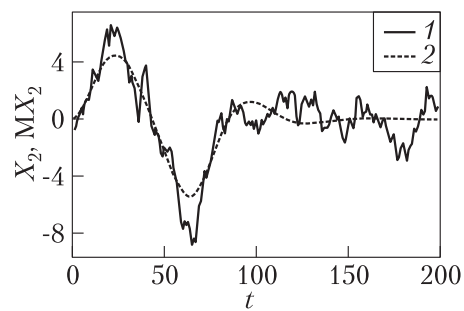


Рис. 14 Величина $X_2(t)$ и ее математическое ожидание $MX_2(2)$

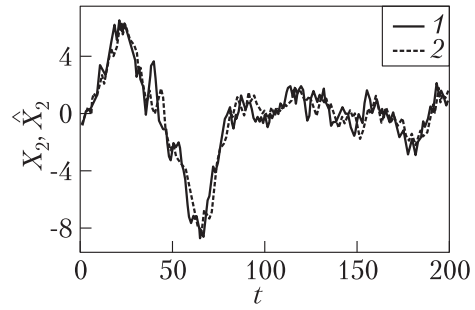
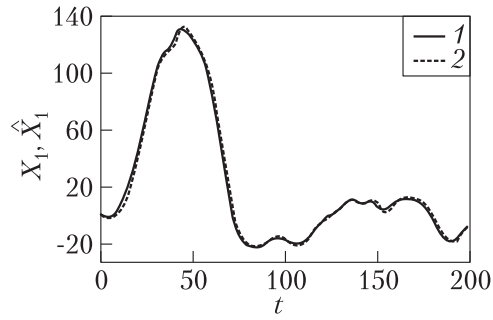


Рис. 15 Величина $X_1(t)$ и оценка $\hat{X}_1(t)$

Рис. 16 Величина $X_2(t)$ и оценка $\hat{X}_2(t)$

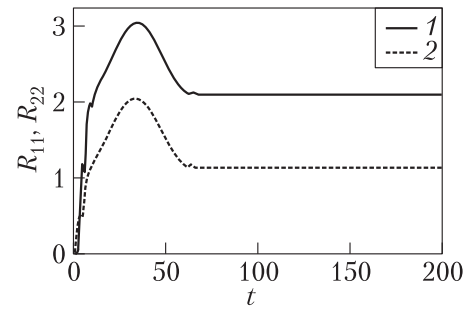
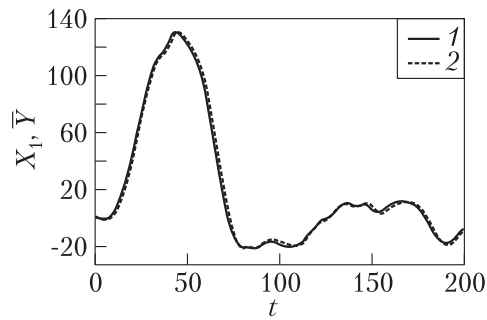


Рис. 17 Измеряемая величина $X_1(t)$ и измерение $\bar{Y}(t)$

Рис. 18 Ошибки оценивания $R_{11}(t)$ и $R_{22}(t)$

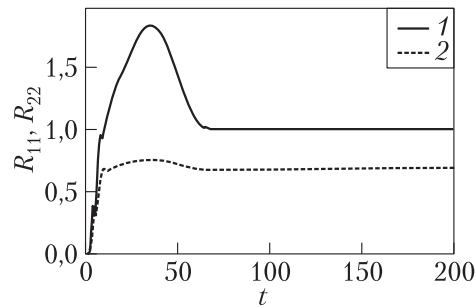


Рис. 19 Ошибки оценивания $R_{33}(t)$ и $R_{44}(t)$

5 Заключение

Таким образом, создание системы управления стоимостью ЖЦ ИНП и минимизация этой стоимости требует решения как минимум следующих основных проблем:

- создание нормативной базы данных паспортизированных СЧИ, содержащей информацию, необходимую для планирования бюджетных показателей СППО (назначенные ресурсы, нормы трудозатрат на ТОиР, ограничения при транспортировке и многие другие);
- создание оперативной базы данных, содержащей информацию по мониторингу реальных данных послепродажного обслуживания (наработку изделий, надежность, остаточные ресурсы, дисциплину поставок (соблюдение сроков и качества) ЗЧ, дисциплину ремонта и др.);
- разработка математической модели СППО парка ИНП данного типа и заданной численности;
- разработка комплекса программных средств.

На базе методов непрерывно-дискретных и эрдитарных стохастических систем разработан комплекс прикладных методов и средств анализа и моделирования процессов в системах ИЛП. Приведены результаты вычислительных экспериментов.

Литература

1. *Синицын И. Н., Шаламов А. С.* Лекции по теории систем интегрированной логистической поддержки. — М.: ТОРУС ПРЕСС, 2012.
2. *Синицын И. Н., Синицын В. И., Корепанов Э. Р., Белоусов В. В., Сергеев И. В., Басилашвили Д. А.* Опыт моделирования эрдитарных стохастических систем // Кибернетика и высокие технологии XXI века (С&Т 2012): Тр. XIII Междунар. научно-технич. конф. — Воронеж: Саквоее, 2012. С. 346–357.
3. *Синицын И. Н., Шаламов А. С., Сергеев И. В.* Проблемы моделирования и минимизации затрат на эксплуатацию изделий наукоемкой продукции на современном этапе // Кибернетика и высокие технологии XXI века (С&Т 2012): Тр. XIII Междунар. научно-технич. конф. — Воронеж: Саквоее, 2012. С. 358–369.
4. *Колмановский В. Б., Носов В. Р.* Устойчивость и периодические режимы регулируемых систем с последействием. — М.: Наука, 1981.
5. *Синицын И. Н.* Stochastic hereditary control systems // Проблемы управления и теории информации, 1986. Т. 15. № 4. С. 287–298.
6. *Пугачев В. С., Синицын И. Н.* Теория стохастических систем. — М.: Логос, 2000.
7. *Chukwi E. N.* Stability and time-optimal control of hereditary systems: With application to the economic dynamics of the US. — Singapore: World Scientific, 2001.
8. *Синицын И. Н.* Фильтры Калмана и Пугачева. — 2-е изд., перераб. и доп. — М.: Логос, 2007.

9. *Chang M. H.* Stochastic control of hereditary systems and applications. — Singapore: World Scientific, 2008.
10. *Сеницын И. Н., Сергеев И. В.* Методическое обеспечение измерения, контроля и испытаний вычислительного оборудования в условиях ударных воздействий // Технические и программные средства систем управления, контроля и измерения (УКИ-2010): Тр. конф. — М.: ИПУ РАН, 2010. Секция 4. С. 1–12.
11. *Сеницын И. Н.* Стохастические системы. Теория и стохастические информационные технологии (I) // История науки и техники. — М.: Научтехлитиздат, 2008. № 7. С. 9–12.
12. *Сеницын И. Н.* Стохастические системы. Теория и стохастические информационные технологии (II) // Системы высокой доступности, 2008. № 2. С. 25–35.
13. *Сеницын И. Н.* Канонические представления случайных функций и их применение в задачах компьютерной поддержки научных исследований. — М.: ТОРУС ПРЕСС, 2009.
14. *Сеницын И. Н., Корепанов Э. Р., Белоусов В. В., Шоргин В. С., Макаренкова И. В., Конашенкова Т. Д., Агафонов Е. С., Семендяев Н. Н.* Развитие компьютерной поддержки статистических исследований систем высокой точности и доступности // Системы и средства информатики, 2011. Вып. 21. № 1. С. 7–37.

АВТОМАТИЗАЦИЯ ОТЛАДКИ ПРОГРАММ ДЛЯ РЕКУРРЕНТНОГО ОПЕРАЦИОННОГО УСТРОЙСТВА*

Р. А. Зеленов¹, А. А. Прокофьев², В. Н. Волчек³

Аннотация: Показаны проблемы, встающие перед разработчиком низкоуровневых программ, и способы их решения на примере среды проектирования программного обеспечения для рекуррентного операционного устройства (РОУ). Описывается реализация средств уменьшения числа ошибок на этапе программирования, а также инструмента автоматической локализации ошибок, приводящих к генерации исключений в аппаратуре.

Ключевые слова: интегрированная среда разработки; автоматизация отладки; потоковая архитектура; рекуррентность

1 Введение

В области разработки низкоуровневых программ, предназначенных для выполнения на микропроцессорах, средства поддержки программирования представлены различными средами разработки либо для ассемблерных языков, либо для языков программирования микроконтроллеров. Данные программные продукты включают в себя оболочку, предоставляющую удобный интерфейс пользователя, компиляторы и отладочные средства. Инструменты отладки, как правило, предоставляют следующую функциональность: пошаговый просмотр, отображение информации о текущих значениях переменных и трассировка всего хода вычислений.

Каких-либо автоматизированных средств локализации ошибок, приводящих к возникновению исключений или прерываний в процессоре, в широком применении не используется. Следовательно, процесс отладки ложится полностью на плечи программиста и эффективность отладки зависит от его опыта и квалификации.

В Институте проблем информатики Российской академии наук ведется разработка многоядерной потоковой рекуррентной вычислительной системы (МПРВС)

* Работа выполнена при частичной финансовой поддержке по Программе фундаментальных исследований ОНИТ РАН на 2011 г., проект 1.5.

¹ Институт проблем информатики Российской академии наук, graf.developer@gmail.com

² Институт проблем информатики Российской академии наук, a.a.prokofyev@mail.ru

³ Институт проблем информатики Российской академии наук, v-volchek@inbox.ru

с нетрадиционной архитектурой [1]. Многоядерная потоковая рекуррентная вычислительная система реализована в виде двухуровневой системы: управляющий уровень, базирующийся на фон-неймановских принципах, и операционный уровень, реализованный в виде РОУ, базирующийся на потоковых (dataflow) принципах. Именно для РОУ требуется разработка низкоуровневых программ-капсул [2], которые представляют собой наборы элементов самодостаточных данных (ЭСД, операндов). Элементы самодостаточных данных — это совокупность данных и инструкций для их обработки.

Для целей программирования капсул была разработана система капсульного программирования и отладки (СКАТ) [3], которую можно отнести к интегрированным средам разработки (IDE). Возможности СКАТ соответствуют функциональности большинства IDE, существующих в мире, т. е. она содержит в себе средства создания и отладки программ, в данном случае капсул.

2 Возможности и недостатки системы капсульного программирования и отладки

В настоящее время СКАТ имеет следующие функциональные возможности:

- (1) создание капсул с помощью визуального конструктора капсул (ВКК);
- (2) отладка капсул:
 - просмотр хода вычислительного процесса:
 - файл-отчет выполнения моделирования;
 - временные диаграммы;
 - пошаговый просмотр всех ключевых регистров и памяти;
 - пошаговая отладка — изменение значений определенных регистров в режиме пошагового просмотра;
- (3) подготовка файлов для выполнения моделирования вычислительного процесса:
 - генерация служебных файлов, необходимых для запуска моделирования в симуляторе ModelSim;
 - создание, редактирование, добавление и удаление файлов VHDL-модели РОУ;
 - формирование содержимого буферной памяти (БП) [4];
- (4) выполнение моделирования в ModelSim [4].

Подробное описание указанных возможностей представлено в [3].

Система капсульного программирования и отладки уникальна тем, что при ее разработке закладывались принципы, предоставляющие программисту возможность реконфигурации данной системы при изменениях в спецификации РОУ.

Это выражается в возможности быстрого изменения формата операндов и настройки процедуры отладки в соответствии с тем, какими аппаратными ресурсами обладает РОУ.

Однако СКАТ не лишена недостатков, которые выражаются в следующем:

- на этапе формирования капсулы отсутствуют средства, снижающие вероятность появления ошибок, связанных с неверной семантикой;
- отсутствует инструментарий, позволяющий автоматизировать процесс локализации ошибок, приводящих к появлению исключений в аппаратуре, и тем самым сокращать время отладки капсул.

Текущий этап разработки МПРВС требует модернизации СКАТ таким образом, чтобы вышеуказанные недостатки были устранены. Основная цель развития функциональных возможностей системы заключается в снижении вероятности появления ошибок в капсулах, но если ошибки все же возникли во время выполнения, то необходимо ускорить процесс их исправления путем внедрения автоматизированного средства локализации ошибок. В данной статье будут рассмотрены проблемы, связанные с реализацией данной модернизации, и способы их решения.

3 Снижение вероятности появления ошибок в капсулах

Ошибки во время выполнения программ могут возникать по нескольким причинам [5]:

- (1) неверный синтаксис или семантика языка;
- (2) неверная реализация алгоритма программистом;
- (3) ошибки в инструментах разработки;
- (4) ошибки в аппаратуре, на которой эти программы выполняются.

В данной статье ошибки, обусловленные причинами 2–4, рассматриваться не будут. Как правило, синтаксический анализ программы проводится в начале этапа трансляции. Однако ВКК, с помощью которого создаются капсулы, исключает появление синтаксически неверных инструкций. Данный компонент использует всегда актуальное XML-представление формата операндов.

Семантические ошибки, т. е. логически неверное использование капсульного языка, более вероятны. Следовательно, в СКАТ должны быть встроены инструменты, помогающие избегать появления такого рода ошибок. Однако построение функционала, позволяющего полностью избежать появления ошибок в программах, невозможно.

Для того чтобы разработать инструментарий, позволяющий снизить вероятность появления ошибок на этапе создания капсулы, необходимо:

- определить множество возможных ошибок;
- оценить целесообразность их обнаружения на этапе создания капсулы;
- оценить требуемые для обнаружения ошибок вычислительные и временные ресурсы;
- в случае принятия решения об обнаружении определенных ошибок на этапе создания капсулы, требуется разработать алгоритм работы данной функциональности.

Во время работы над спецификацией РОУ было проведено исследование, результатом которого стало формирование множества возможных исключений, которые могут генерироваться аппаратурой во время выполнения капсулы. Данное множество формировалось исходя из возможных ошибок, обнаруживаемых в функциональных блоках РОУ. Можно привести перечень данных ошибок:

- Значения функциональных полей в операнде не специфицированы. Этот случай исключается из рассмотрения в рамках данной задачи, так как подобная ситуация возможна по причине либо неправильного функционирования ВКК, который не должен давать пользователю сформировать операнд с не специфицированными значениями функциональных полей, либо из-за ошибок в работе аппаратуры.
- Значения функциональных полей двух операндов, которые будут участвовать в какой-либо операции (сцепившиеся операнды) [6], взаимно противоречивы, т. е. РОУ не может корректно работать с данными значениями.
- Отсутствие определенных операндов в той части капсулы, в которой они должны быть расположены.
- Неверная настройка распределения операндов по секциям.
- Некоторые функциональные поля в операндах имеют такие значения, что не ясен дальнейший путь передачи данных.

Взаимную противоречивость функциональных полей сцепившихся операндов возможно обнаруживать в СКАТ на этапе создания капсулы только лишь для первой пары операндов каждой секции. Это обусловлено тем фактом, что во время выполнения капсулы вторая и последующие пары операндов для каждой секции выбираются исходя из контекста вычислительного процесса, а для его формирования требуется запустить данный вычислительный процесс. Следовательно, вводить функционал обнаружения данного типа ошибок на этапе создания капсулы нецелесообразно по причине низкой эффективности данной логики для капсул, которые содержат больше восьми операндов. Капсулы, как правило, содержат гораздо большее число операндов.

Отсутствие операндов в строго определенных частях капсулы можно считать нарушением структуры капсулы. Данная ошибка возможна в большинстве

случаев только вследствие неправильного программирования, следовательно, целесообразно возложить проверку данных ситуаций на СКАТ до выполнения капсулы в РОУ. Здесь следует ввести некоторое разделение ошибок данного типа.

В первом случае нарушение структуры (принятого формата) капсулы возможно при несоблюдении правил расположения операндов в определенных частях капсулы. Например, некоторые управляющие операнды могут находиться только в начале капсулы.

Во втором случае возможна ситуация, когда в операнде из управляющей части капсулы указаны номера-ссылки на операнды в несуществующей части капсулы — выход за пределы диапазона количества операндов в капсуле либо указание на ту часть капсулы, в которой не может быть данных операндов.

Неверная настройка распределения операндов по секциям, а также значения функциональных полей, которые не позволяют определить направление пересылки данных, могут быть обнаружены только во время выполнения капсулы, так как имеет место зависимость от контекста вычислительного процесса.

Таким образом, принимается решение о том, что на этапе создания капсулы целесообразно обнаруживать только ошибки, связанные с нарушением структуры капсулы. В условиях применения современного аппаратного обеспечения персональных компьютеров вычислительные ресурсы, требуемые для реализации данной проверки, невелики по сравнению со временем выполнения моделирования капсулы в РОУ и затрачиваемых при этом оперативной памяти и процессорных мощностей.

Переходя к реализации обнаружения вышеуказанных ошибок, необходимо учитывать, что СКАТ разрабатывалась с требованием гибкой настройки на изменяющуюся спецификацию РОУ. Следовательно, несмотря на жесткое определение формата капсулы в данный момент, следует закладывать возможность его быстрого изменения. Таким образом, формат капсулы необходимо специфицировать в таком виде, который обеспечит возможность его изменения без перекомпиляции СКАТ.

Проверку на соответствие капсулы специфицированному формату лучше всего проводить перед началом трансляции ее в БП. Это обусловлено тем, что при программировании вначале могут создаваться содержательные операнды, над которыми будут происходить вычисления, а уже после этого — конфигурационные и управляющие, которые разместятся в начале капсулы.

Проанализировав различные варианты реализации спецификации на формат капсулы, разработчики СКАТ пришли к выводу, что лучше всего для этого подходит использование языка регулярных выражений. Суть предлагаемого способа заключается в том, что с помощью регулярного выражения описывается корректная структура капсулы. Затем в строку собираются все сокращенные обозначения операндов капсулы и проверяется соответствие получившейся строки

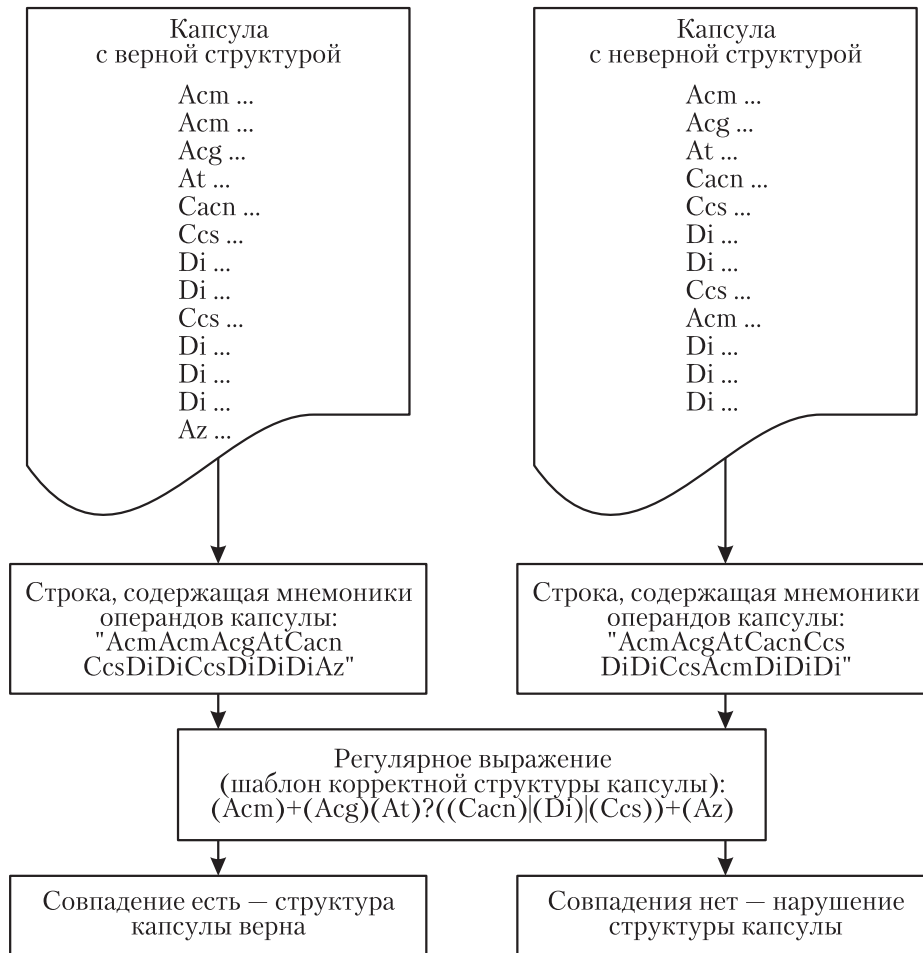


Рис. 1 Пример анализа структуры капсулы на соответствие шаблону

шаблону. Если соответствие найдено, то это означает, что структура капсулы верна.

Пример данной проверки показан на рис. 1. Операнды в капсулах обозначены сокращениями — мнемоникой, многоточие означает функциональные поля. В данном случае важно не назначение операндов, а тот факт, что корректная капсула должна начинаться с Acm-операндов, а заканчиваться операндом Az.

Для того чтобы определить, корректно ли указаны ссылки на операнды в управляющей части капсулы, необходимо проанализировать значения этих ссылок. В случае если они содержат величину, выходящую за границу количества

операндов в капсуле, это считается ошибкой. Также ошибочной будет ситуация, когда ссылки указывают на операнды, тип которых недопустим в данном месте (например, ссылка на операнд из управляющей части капсулы).

В случае обнаружения вышеуказанных ошибок пользователю должно выдаваться сообщение, содержащее тип ошибки и указание на операнд, содержащий ошибочные данные.

4 Автоматизация процесса локализации обнаруженных ошибок

В РОУ реализована логика обнаружения и обработки исключений, возникающих в ходе вычислительного процесса. Эти исключения могут возникнуть во время моделирования выполнения капсулы по причине вышеописанных ошибок. Следовательно, необходимо иметь в СКАТ инструментарий, позволяющий программисту быстрее локализовать обнаруженные ошибки. Описываемый инструмент называется компонентом обработки исключительных ситуаций (КОИС).

Общий алгоритм работы КОИС можно представить в виде трех этапов:

- (1) сбор отладочной информации об исключении;
- (2) анализ информации и локализация ошибки;
- (3) если место возникновения ошибки определено, то выдача пользователю этой, а также всей собранной отладочной информации.

Под отладочной информацией подразумевается тип ошибки, время и место ее обнаружения. Эта информация хранится во внутренних регистрах блока обработки исключений в РОУ, следовательно, может быть получена при анализе файла-отчета, созданного в результате моделирования выполнения капсулы в ModelSim.

Именно на этапе анализа собранной отладочной информации возможно внедрение автоматизированных средств локализации обнаруженных ошибок. После проведения исследования возможности создания подобного инструментария, было принято решение о том, что это реализуемо в разумные сроки, и был разработан алгоритм, согласно которому должен функционировать поиск мест возникновения ошибок.

Суть предлагаемого алгоритма заключается в том, что во время обработки файла-отчета, полученного после моделирования, могут быть обнаружены все зафиксированные аппаратурой ошибки в виде сгенерированных исключений. По типу каждого исключения может быть определен конкретный вид ошибки. Как правило, ошибки могут возникнуть по причине таких значений функциональных полей в операндах, которые приводят к нарушению правильной семантики программы. Далее происходит поиск в капсуле исходных операндов, содержащих неверные значения функциональных полей.

В СКАТ файл-отчет моделирования транслируется во внутреннее представление [3], и оно используется для осуществления пошаговой отладки, во время которой пользователю предоставляются в графическом (удобном для восприятия) виде значения данных в определенных узлах РОУ на каждом шаге выполнения капсулы. Если обнаружено появление исключения, то значения, по которым был определен факт исключения, подсвечиваются.

Пользователь может запустить автоматический поиск места возникновения ошибки. Он работает по принципу обратной трассировки, т. е. на шаге, на котором обнаружено исключение, собирается информация о текущем операнде и начинается обратный проход по шагам и отслеживание преобразований, которые породили данный операнд. Упрощенно процесс можно описать следующим образом:

1. Шаг № 3. Обнаружено исключение. Функциональное поле ФП операнда ОЗ содержит неверное, с точки зрения семантики, значение. Операнд ОЗ находится в функциональном блоке ФБЗ.
2. Анализ шага № 2. Операнд ОЗ является результатом действий над операндами О2 и О1 в функциональном блоке ФБ2. Функциональное поле ФП попало в ОЗ из О2.
3. Анализ шага № 1. Операнд О2 находится в функциональном блоке ФБ1. Сюда он попал из БП, которая содержит в себе капсулу.
4. Поиск в исходной капсуле операнда О2. Именно его функциональное поле ФП стало причиной исключения в ФБЗ на шаге № 3.
5. Осуществляется выдача пользователю сообщения о том, что ошибка находится в ФП в операнде О2 в исходной капсуле и необходимо его изменить.

Приведенная последовательность вычисления места возникновения ошибки сильно упрощена, так как зачастую ошибочные значения функциональных полей находятся сразу в двух операндах и необходимо проводить обратную трассировку для каждого из них. Чем больше номер шага, тем, соответственно, больше итераций требуется для прохода через все шаги к искомому операнду в исходной капсуле.

5 Заключение

В ходе проведенной модернизации в СКАТ были внедрены следующие функции:

1. Проверка соответствия структуры капсулы шаблону, заданному в виде регулярного выражения.
2. Проверка корректности ссылок на операнды, которые содержатся в управляющей части капсулы.

3. Автоматический поиск в капсуле исходного операнда, содержащего такие значения функциональных полей, которые приводят к генерации исключения в РОУ в процессе выполнения капсулы.

В результате вышеуказанных изменений была исключена возможность запуска на выполнение в РОУ капсулы, имеющей нарушение в своей структуре, а также значительно сократилось время отладки в случаях, когда необходимо найти и исправить ошибку, приводящую к генерации исключений.

Литература

1. *Степченков Ю. А., Петрухин В. С.* Особенности гибридного варианта реализации на ПЛИС рекуррентного обработчика сигналов // Системы и средства информатики. Доп. вып. — М.: ИПИ РАН, 2008. С. 118–129.
2. *Зеленов Р. А., Степченков Ю. А., Волчек В. Н., Хилько Д. В., Шнейдер А. Ю., Прокофьев А. А.* Система капсульного программирования и отладки // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1. С. 24–30.
3. *Степченков Ю. А., Волчек В. Н., Петрухин В. С., Прокофьев А. А., Зеленов Р. А.* Механизмы обеспечения поддержки алгоритмов цифровой обработки речевых сигналов в рекуррентном обработчике сигналов // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1. С. 30–46.
4. ModelSim Tutorial. http://www.actel.com/documents/modelsim_tutorial_ug.pdf.
5. *Тэллес М., Хсих Ю.* Наука отладки. — М.: Кудиц-образ, 2003. 560 с.
6. *Степченков Ю. А., Волчек В. Н., Петрухин В. С., Прокофьев А. А., Зеленов Р. А.* Цифровой сигнальный процессор с нетрадиционной рекуррентной потоковой архитектурой // Проблемы разработки перспективных микро- и наноэлектронных систем — 2010: Сб. трудов. — М.: ИПИМ РАН, 2010. 694 с.

ВСТРОЕННЫЕ СРЕДСТВА ОТЛАДКИ РЕКУРРЕНТНОГО ОПЕРАЦИОННОГО УСТРОЙСТВА*

А. А. Прокофьев¹, Р. А. Зеленов², В. Н. Волчек³

Аннотация: Рассмотрены проблемы обеспечения наблюдаемости и управляемости внутренних состояний рекуррентного операционного устройства (РОУ). Рассмотрены вопросы внедрения средств, позволяющих разработчику максимально быстро обнаружить, локализовать и исправить ошибки в РОУ, предложены наиболее эффективные, учитывающие архитектурную новизну и специфику его организации решения.

Ключевые слова: отладка; аппаратура; рекуррентность; потоковая архитектура

1 Введение

В Институте проблем информатики Российской академии наук ведется разработка многоядерной потоковой рекуррентной вычислительной системы (МПРВС) с нетрадиционной архитектурой. Данная архитектура является развитием потокового подхода. В текущей реализации осуществляется взаимодействие двух уровней: функции управляющего уровня возложены на фон-неймановский процессор, а в качестве операционного уровня выступает РОУ, имеющее в своем составе четыре процессорных ядра. Управляющий уровень осуществляет взаимодействие с периферийными устройствами и организует подготовку данных для дальнейшей их обработки в РОУ [1, 2].

В качестве исполняемых программ для РОУ выступают капсулы [3]. Это программы, представляющие собой набор элементов самодостаточных данных (операндов), содержащих данные и инструкции для их обработки, а также инструкции настройки режимов работы РОУ. Капсула хранит рекуррентно сжатый алгоритм решения конкретной задачи, который разворачивается в ходе выполнения в РОУ.

Реализуется МПРВС в базе ПЛИС (программируемой логической интегральной схемы), в качестве языка описания аппаратуры используется VHDL.

*Работа выполнена при частичной финансовой поддержке по Программе фундаментальных исследований ОНИТ РАН на 2011 г., проект 1.5.

¹Институт проблем информатики Российской академии наук, a.a.prokofyev@mail.ru

²Институт проблем информатики Российской академии наук, graf.developer@gmail.com

³Институт проблем информатики Российской академии наук, v.volchek@inbox.ru

На текущем этапе проведена верификация как отдельных функциональных блоков, так и всего РОУ в целом. Подготовлен набор капсул, реализующих алгоритмы, требующиеся для решения задачи распознавания слов диктора [4]. На следующем этапе предстоит комплексная отладка разработанного программного и аппаратного обеспечения на инструментальной плате фирмы Altera. Так как МПРВС является системой реального времени и корректность ее функционирования зависит от времени выполнения отдельных программ и скорости работы аппаратуры, невозможно ограничиться лишь отладкой VHDL-модели, необходимо также иметь возможность проверки и отладки устройства непосредственно в аппаратуре (в реальных условиях).

Процесс отладки в аппаратуре является итерационным. Сначала требуется проверить работоспособность каждого из блоков в отдельности, заменяя сопряженные с ним блоки заглушками (имитирующими входные воздействия, например, путем подачи на входы отлаживаемого блока определенной последовательности данных). Затем проверяется взаимодействие отдельных пар блоков, после чего можно приступить к комплексной отладке.

В данной статье рассматриваются вопросы обеспечения наблюдаемости и управляемости внутренних состояний РОУ, а также проблемы реализации функций отладки, связанные с особенностями архитектуры МПРВС, а именно:

- графодинамическим представлением алгоритмов — отсутствуют понятия адреса команды и данных [5];
- параллельным выполнением ветвей графа алгоритма.

2 Функции отладки в рекуррентном операционном устройстве

Функции отладки в РОУ направлены на поиск и устранение ошибок в программной и аппаратной частях системы. Источниками ошибок могут служить как неверные логические предпосылки при создании спецификаций, сложность учета всех возможных факторов, влияющих на работу системы, так ошибки инженеров, занимающихся проектированием и реализацией аппаратуры и программного обеспечения. Чтобы упростить процесс выявления и локализации ошибок, необходимо реализовать определенный набор функций, которые позволят провести отладку.

Существует пять основных приемов комплексной отладки микропроцессорной системы [6]:

- (1) останов функционирования системы при возникновении определенного события;
- (2) чтение / изменение содержимого памяти или регистров системы;
- (3) пошаговое отслеживание поведения системы;

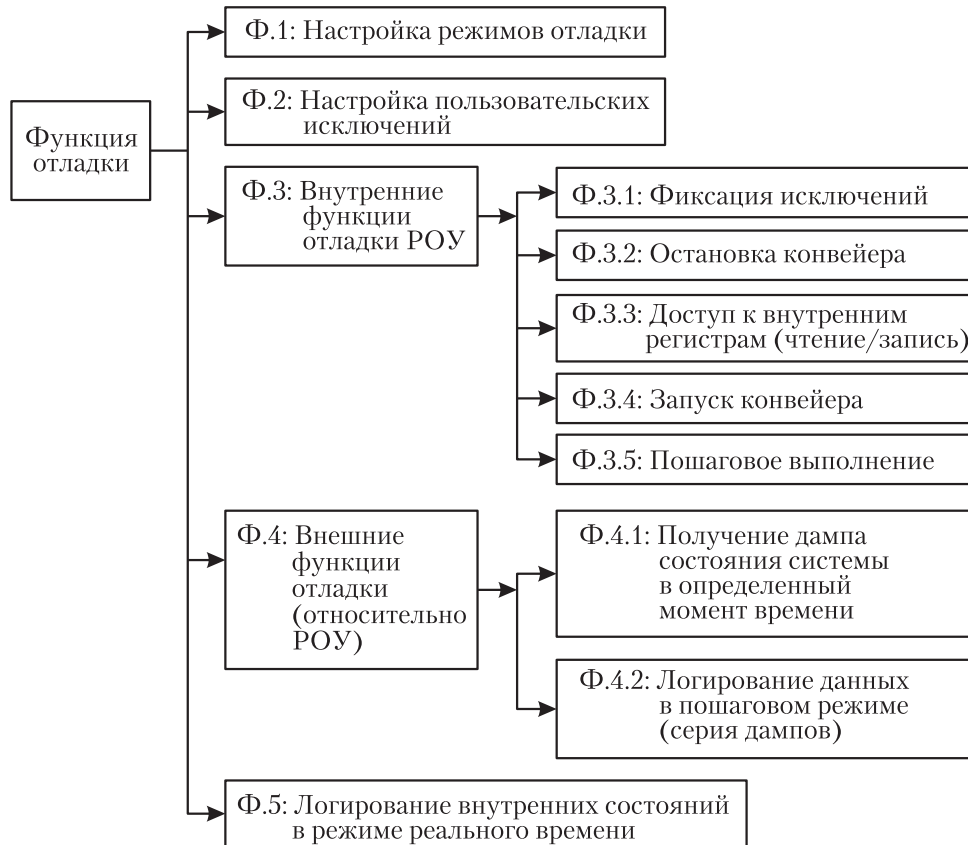


Рис. 1 Структура функции отладки

- (4) отслеживание поведения системы в реальном времени;
- (5) временное согласование программ.

На рис. 1 представлена структура функций отладки, которые обеспечивают возможность использования приемов, описанных выше. Реализация этих функций должна удовлетворять критериям:

- полноты отладочной информации;
- легкости получения и анализа информации;
- минимизации аппаратных затрат на реализацию с целью снизить влияние на характеристики устройства.

3 Реализация встроенных средств отладки рекуррентного операционного устройства

Функции отладки и обнаружения исключительных ситуаций объединены в РОУ в виде единого функционального блока отладки и фиксации исключений (БОФИ). Такая реализация позволяет оперативно управлять состоянием конвейера, избегать потери данных и нарушения хода вычислений при возникновении исключения. Для обеспечения доступа к внутренним регистрам и памяти внутри РОУ организована единая отладочная шина. По ней передается следующая информация:

1. Флаг работы в режиме отладки (1 бит). При единичном значении работа конвейера приостанавливается, давая возможность последовательно обратиться ко всем необходимым внутренним регистрам и памяти.
2. Флаг разрешения записи (1 бит). При единичном значении разрешает изменение значения регистра или ячейки памяти, к которой осуществляется обращение.
3. Адрес (32 бита). Определяет, с каким конкретно регистром или ячейкой памяти осуществляется взаимодействие.
4. Данные для записи (32 бита).
5. Считанные данные (32 бита).

Введение режимов отладки требует обязательной корректировки VHDL-кода, а именно частей, описывающих синхронные процессы. На рис. 2 дополнительная логика выделена графически. Условно описание процесса можно разбить на три блока:

- (1) аппаратный сброс (строки 2–4, рис. 2);
- (2) режим отладки (строки 5–24);
- (3) нормальный режим работы (строки 25–27).

Рассмотрим часть, описывающую режим отладки, подробнее. Из условия на строке 5 видно, что режим отладки активируется путем установки единичного значения сигнала `debug_mode`. Данный сигнал должен быть глобальным и одновременно останавливать все синхронные процессы в устройстве, чтобы не затронуть ход вычислений. В режиме отладки обеспечивается возможность получить доступ к внутренним регистрам: в зависимости от адреса на шину данных (`debug_out_data`) подается либо значение регистра, либо высокий импеданс. Обязательным условием является обеспечение уникальности адресов с целью предотвращения коллизий.

```

1 PROCESS (clk,rst,debug_mode) BEGIN
2   IF rst='1'THEN
3     tag_data_reg  <= (others => '0');
4     debug_out_data <= (others => 'Z');
5   ELSIF clk'event AND clk='1' AND debug_mode='1' THEN
6     IF debug_address(31 DOWNT0 30)=core_num THEN
7       CASE debug_address(29 DOWNT0 0) IS
8         WHEN "000000000000000000000000000010000" =>
9           debug_out_data <= tag_data_reg(31 DOWNT0 0);
10          IF debug_wren='1' THEN
11            tag_data_reg <= tag_data_reg(34 DOWNT0 32)
12              & debug_in_data;
13          END IF;
14          WHEN "000000000000000000000000000010001" =>
15            debug_out_data <= "00000000000000000000000000000000"
16              & tag_data_reg(34 DOWNT0 32);
17            IF debug_wren='1' THEN
18              tag_data_reg <= debug_in_data(2 DOWNT0 0)
19                & tag_data_reg(31 DOWNT0 0);
20            END IF;
21          WHEN others =>
22            debug_out_data <= (others => 'Z');
23          END CASE;
24        END IF;
25     ELSIF clk'event AND clk='1' THEN
26       -- Тело процесса
27     END IF;
28 END PROCESS;

```

Рис. 2 Пример синхронного процесса

Из рис. 2 видно, что адресное пространство логически разделено:

- старшие биты указывают номер вычислительного ядра (строка 6);
- средняя часть определяет регистр (строки 8, 15);
- младшие биты определяют часть регистра, если его длина превышает 32 разряда.

Таким образом, БОФИ получает доступ к остановке и запуску конвейера (реализуются функции Ф.3.2 и Ф.3.4 с рис. 1), а также возможность чтения/записи внутренних регистров (Ф.3.3). Пошаговое выполнение (Ф.3.5) является ничем иным, как автоматическим переходом в режим отладки после запуска устройства в нормальном режиме в течение определенного количества тактов. Очевидно, что в режиме отладки работа БОФИ должна контролироваться извне, так как процедуры считывания данных выполняются специальным ПО или программистом. Кроме этого, необходимо обеспечить возможность настройки режимов отладки (Ф.1).

Для доступа к БОФИ с хост-компьютера используется интерфейс JTAG [7], а для доступа со стороны управляющего уровня — 32-разрядные шины данных и адреса. Взаимодействие управляющего уровня с БОФИ необходимо для реализации служебных процедур, когда требуется доступ к внутренним регистрам РОУ.

В нормальном режиме работы РОУ БОФИ воспринимает поступающие данные как инструкции (установка настроечных регистров либо требование перейти в режим отладки). В режиме же отладки интерфейсные шины напрямую соединяются с внутренними шинами отладки, возврат в нормальный режим инициализируется подачей нулевого значения адреса.

Что касается фиксации исключений (Ф.3.1), то в результате анализа функциональных блоков РОУ выявлены и специфицированы ситуации, которые могут негативно повлиять на ход вычислительного процесса. К ним относятся ошибки, возникающие в результате обнаружения недопустимых значений функциональных полей; попыток передачи данных на одну шину из различных источников; из-за противоречивости значений функциональных полей в парах операндов; при обращении к несуществующей ячейке памяти; при переполнении в результате вычислений и т. п. Все подобные ситуации внесены в перечень исключений с классификацией по типам, приоритетности и катастрофичности последствий. Наличие ошибки, входящей в перечень выявленных исключений, фиксируется непосредственно в каждом из функциональных блоков и передается в БОФИ, где, в свою очередь, генерируется информационный код, содержащий в себе данные о месте, типе и времени ошибки. При необходимости может быть произведена остановка конвейера и сгенерирован сигнал прерывания.

Так как в архитектуре РОУ (потока данных) отсутствует понятие адреса команд, определение места на потоковом графе алгоритма, в котором находится вычислительный процесс, составляет определенные трудности. Требуется введение дополнительных аппаратных счетчиков, стартующих по определенным системным событиям:

- счетчик шагов выполнения капсулы — стартует по началу выполнения капсулы, инкрементируется после выполнения очередного вычислительного шага;
- индексные регистры — служат для считывания операндов из капсулы, по их значениям можно определить, какая капсула выполняется в данный момент;
- счетчик циклов выполнения алгоритма — запускается при старте вычислений, инкрементируется после прохождения очередного цикла вычислений.

Сочетание значений этих счетчиков позволяет определить временную составляющую в информации о зарегистрированном в ходе вычислений исключении.

С целью упрощения отладки логика фиксации исключительных ситуаций расширена возможностью настраивать условия пользовательских исключений (Ф.2):

- достижение определенного узла в графе вычислительного процесса;
- возникновение какого-либо отладочного события, например:
 - запись операнда в память совпадений;
 - перенаправление операндов;
 - формирование перехода;
 - исполнение кода операции.

Отладочные условия настраиваются и активируются через описанный выше интерфейс БОФИ.

4 Сбор отладочной информации

Сбор информации о ходе вычислительного процесса и предоставление ее пользователю в удобном для анализа виде является необходимой частью процесса отладки. Объем доступной на кристалле ПЛИС памяти недостаточно для хранения огромных массивов данных, представляющих собой дампы состояний внутренних регистров РОУ, полученных в течение нескольких вычислительных шагов. Единственный возможный выход — использование внешней памяти. Это может быть, например, оперативная память на отладочной плате либо память жесткого диска хост-компьютера. В любом случае без передачи управления на внешние относительно РОУ устройства обойтись невозможно (Ф.4), поэтому после фиксации исключения осуществляется останов конвейера и путем генерации прерывания контроль над РОУ передается управляющему уровню.

Управляющий уровень принимает решение о дальнейших действиях:

- сохранить состояния определенного набора регистров в оперативную память и продолжить вычисления (Ф.4.1);
- продолжить вычисления в пошаговом режиме, выполняя логирование состояний системы после каждого шага (Ф.4.2);
- передать управление на хост-компьютер. Программа-отладчик в автоматическом режиме либо при участии пользователя начнет сбор информации. При необходимости также может быть выполнена корректировка внутренних состояний регистров и памяти.

В случае необходимости сбора отладочной информации в режиме реального времени (без остановки конвейера) наиболее эффективным решением является применение встроенного логического анализатора, предоставляемого разработчиками ПЛИС [8]. Применение данного инструмента базируется на использовании незадействованной в проекте памяти ПЛИС. Такая память — ограниченный ресурс, поэтому на время отладки целесообразно по максимуму сократить используемую проектом память (например, уменьшить объем буферной памяти

и памяти совпадений). К плюсам логического анализатора относятся простота использования и отсутствие необходимости вносить в проект какие-либо изменения. Недостатком можно считать то, что каждое переопределение логируемых данных требует перепрограммирования ПЛИС.

5 Визуализация процесса отладки

Лишь обладать информацией о ходе вычислительного процесса недостаточно, необходимо предоставить ее пользователю в удобном и наглядном виде. Оптимальным вариантом будет, если от пользователя не потребуются дополнительных усилий для освоения инструментов встроенной отладки. Поэтому в качестве визуального интерфейса взаимодействия со встроенными средствами отладки РОУ целесообразно использовать интерфейс отладки капсул системы капсульного программирования и отладки (СКАТ) [3].

Взаимодействие СКАТ с аппаратурой осуществляется при помощи интерфейса JTAG [7]. С точки зрения пользователя, отладка в аппаратуре выглядит точно так же, как и отладка капсул на VHDL-модели РОУ. Единственное отличие заключается в том, что изменения внутренних состояний могут быть выполнены только на текущем шаге (при пошаговой отладке) либо не могут быть применены совсем (если СКАТ работает в режиме анализа собранной информации).

Так как работа носит научно-исследовательский характер и имеется большая вероятность внесения изменений в аппаратную часть РОУ, нужно обеспечить возможность легкой подстройки СКАТ под новую структуру аппаратуры. Поэтому информация об адресах внутренних регистров и памяти РОУ предоставляется для СКАТ в виде XML-файла, что позволяет легко модифицировать перечень доступных для отладки регистров без перекомпиляции.

6 Автоматизация диагностирования ошибок

Процесс отладки, а именно установка причины возникновения ошибки, имеет достаточно большую аналитическую интеллектуальную составляющую. Чтобы снизить трудоемкость и ускорить процесс определения, на каком уровне возникает ошибка (управляющий уровень, РОУ, интерфейс взаимодействия уровней), диагностирование можно автоматизировать с помощью СКАТ следующим образом:

1. СКАТ считывает из аппаратуры текущее состояние РОУ, включая часть буферной памяти, соответствующую капсуле, в ходе выполнения которой была зафиксирована ошибка.
2. Структура капсулы (типы операндов и значения функциональных полей) сравнивается с эталоном (капсулой, подготовленной пользователем). Если

отклонений не найдено, то переход к шагу 3. В противном случае можно сделать вывод, что ошибка возникла в процессе записи шаблона капсулы в буферную память, следовательно, необходимо проверить работу управляющего уровня и его синхронизацию с РОУ.

3. Капсула, абсолютно идентичная той, что была выполнена в аппаратуре, запускается на программной модели РОУ.
4. Результат выполнения в модели сравнивается с данными, считанными из аппаратуры. Если они совпадают, то ошибка, скорее всего, не связана с физической реализацией и имеет систематический характер, а дальнейшую отладку можно проводить на VHDL-модели РОУ.

Таким образом, в автоматическом режиме можно выяснить, влияет ли реализация в аппаратуре на появление ошибки, что помогает резко сократить область поиска.

7 Заключение

Разработанные средства отладки РОУ позволяют ускорить процесс обнаружения и локализации ошибок. Это достигается за счет введения в аппаратуру дополнительного блока отладки, взаимодействие с которым осуществляется через интерфейс JTAG, таким образом обеспечивается доступ к внутренним регистрам и памяти устройства. С помощью визуального интерфейса программы СКАТ пользователь легко может в пошаговом режиме анализировать внутреннее состояние РОУ и при необходимости модифицировать его. Средства автоматизации диагностирования ошибок, встроенные в СКАТ, позволяют ускорить локализацию ошибки за счет определения уровня на котором она возникла: управляющий уровень либо интерфейс его взаимодействия с РОУ; ошибка в логике работы РОУ или в капсуле; ошибка, связанная с особенностями реализации в аппаратуре (например, ошибки синхронизации).

Литература

1. *Степченко Ю. А., Петрухин В. С.* Особенности гибридного варианта реализации на ПЛИС рекуррентного обработчика сигналов // Системы и средства информатики. Доп. вып. — М.: ИПИ РАН, 2008. С. 118–129.
2. *Степченко Ю. А., Волчек В. Н., Петрухин В. С., Прокофьев А. А., Зеленов Р. А.* Цифровой сигнальный процессор с нетрадиционной рекуррентной потоковой архитектурой // Проблемы разработки перспективных микро- и наноэлектронных систем — 2010: Сб. трудов. — М.: ИПИМ РАН, 2010. 694 с.
3. *Зеленов Р. А., Степченко Ю. А., Волчек В. Н., Хилько Д. В., Шнейдер А. Ю., Прокофьев А. А.* Система капсульного программирования и отладки // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1. С. 24–30.

4. *Степченков Ю. А., Волчек В. Н., Петрухин В. С., Прокофьев А. А., Зеленов Р. А.* Механизмы обеспечения поддержки алгоритмов цифровой обработки речевых сигналов в рекуррентном обработчике сигналов // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1. С. 30–46.
5. *Филин А. В.* Динамический подход к выбору архитектуры вычислительных устройств обработки сигналов // Системы и средства информатики. — М.: Наука, 2001. Вып. 11. С. 247–261.
6. *Воробьев Н. В., Горбунов В. Л., Горячев А. В. и др.* Микропроцессоры: В 3-х кн. Кн. 3. Средства отладки, лабораторный практикум и задачник: Уч-к для вузов / Под ред. Л. Н. Преснухина. — М.: Высшая школа, 1986. 351 с.
7. Altera, Virtual JTAG (sld_virtual_jtag). Megafunction User Guide. http://www.altera.com/literature/ug/ug_virtualjtag.pdf.
8. Altera, Quartus II Handbook v11.1.0. Vol. 3: Verification. http://www.altera.com/literature/hb/qts/qts_qii5v3.pdf.

МОДЕЛЬ ПОТОКОВОЙ АРХИТЕКТУРЫ НА ПРИМЕРЕ РАСПОЗНАТЕЛЯ СЛОВ*

Д. В. Хилько¹, Ю. А. Степченков²

Аннотация: Рассмотрена имитационная модель потоковой рекуррентной архитектуры (МПРА), реализуемой в виде сверхбольших интегральных схем (СБИС) на кристалле ПЛИС (программируемой логической интегральной схемы) фирмы Altera. Предлагаемая модель позволяет решать задачи отладки функциональных блоков архитектуры до их внедрения в аппаратуру и служит инструментом разработки и отладки программ. Описаны ключевые аспекты архитектуры, специализированного языка программирования и их отражение в модели.

Ключевые слова: имитационная модель; потоковая архитектура; рекуррентность; программирование

1 Введение

В ИПИ РАН ведутся работы по созданию нетрадиционной рекуррентной архитектуры, предназначенной для реализации параллельных вычислений ограниченной размерности в области цифровой обработки сигналов. Для экспериментальной апробации предлагаемой архитектуры разрабатывается VHDL-модель и СБИС на ее основе — рекуррентный обработчик сигналов (РОС). Он исполняется в гибридном, двухуровневом варианте с ведущим фон-неймановским процессором на управляющем (верхнем) уровне (УУ) и рядом потоковых процессоров на нижнем уровне — рекуррентном операционном устройстве (РОУ) [1]. Апробация РОС осуществляется на примере программы распознавателя слов.

Архитектура РОУ радикально отличается по основным моментам не только от классической архитектуры фон Неймана, но и от других нетрадиционных параллельных архитектур. Для существующих традиционных и нетрадиционных компьютерных архитектур характерно наличие двух потоков: активного потока инструкций и пассивного потока данных. В РОУ оба потока сливаются в один общий поток самодостаточных данных, в котором, помимо собственно

*Работа выполнена при частичной финансовой поддержке по Программе фундаментальных исследований ОНИТ РАН на 2012 г. (проект 1.5) и Программе фундаментальных исследований Президиума РАН (проект 16).

¹Институт проблем информатики Российской академии наук, dhilko@yandex.ru

²Институт проблем информатики Российской академии наук, YStepchenkov@ipiran.ru

обрабатываемых данных, содержится управляющая и служебная информация, необходимая для их обработки (в существующих архитектурах кодируемая в форме инструкций) [2, 3].

Для достижения максимальной производительности предлагаемой архитектуры разрабатывается соответствующее программное обеспечение (ПО), оптимизированное для выполнения в ее среде. В качестве основного инструмента программирования РОС выступает специализированная среда СКАТ [4], использующая VHDL-модель как основное средство исполнения и отладки программ. Однако эта версия СКАТ обладает ограниченным набором утилит автоматизации процесса разработки ПО. Специфические особенности, не свойственные другим архитектурам, делают невозможным применение в полном объеме известных методов и технологий программирования для разработки программ на существующем низкоуровневом языке. Необходим специализированный язык программирования высокого уровня и компилятор для него.

Дополнительным препятствием для успешного программирования в среде МПРА является отсутствие завершенной, полнофункциональной реализации РОС (VHDL-модели, СБИС) в данной работе исследовательского характера.

В результате анализа указанных проблем было решено создать дополнительную, обособленную модель РОС, которая позволила бы отлаживать его программы, анализировать его поведение и локализовывать блоки и функции VHDL-модели, требующие доработки.

2 Ключевые особенности рекуррентного обработчика сигналов и языка его программирования

Рекуррентный обработчик сигналов относится к классу потоковых архитектур, но, в отличие от них, тегируемые данные являются самодостаточными (рекуррентно разворачивающимися). Теги содержат некоторую начальную сжатую информацию, обеспечивающую выполнение требуемой процедуры. Каждый следующий шаг процедуры рекуррентно самоопределяется, в том числе с учетом результата предыдущего шага.

В разрабатываемой архитектуре в состав центрального процессорного устройства (ЦПУ) включено автономное устройство преобразования тегов (ПТ), которое обладает возможностью саморазвертки рекуррентного вычислительного процесса (ВП). Устройство ПТ представляет собой относительно простую комбинационную схему, содержащую средства настройки на предметную область. В памяти находится только набор операндов (элементов самодостаточных данных) и начальных значений их функциональных полей, устройством ПТ динамически подвергаемых рекуррентной развертке. Такое представление программы называется капсулой. Более подробно с предлагаемой архитектурой можно ознакомиться в работе [2].

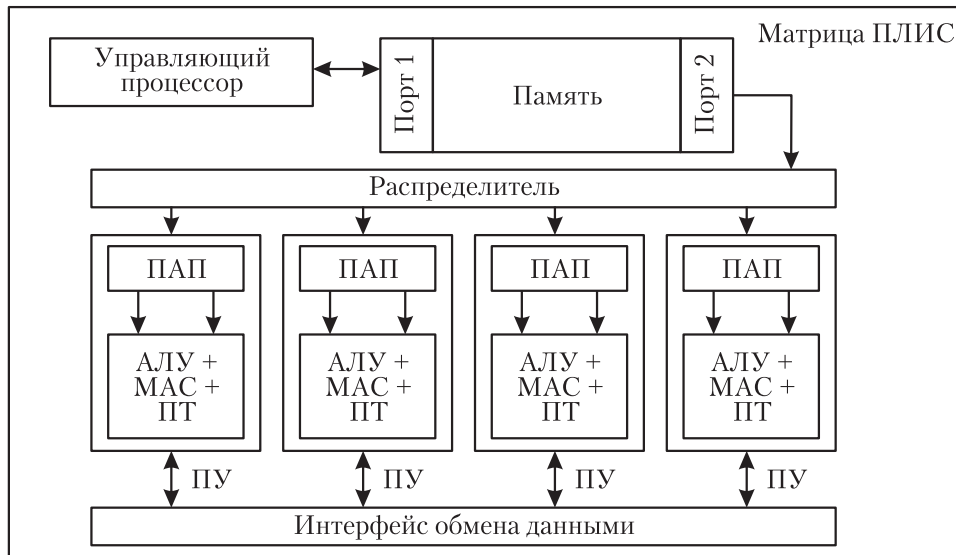


Рис. 1 Структура РОС

В настоящее время РОС реализуется в однокристальном исполнении и имеет четыре процессорных ядра (ПЯ), способных функционировать параллельно. Таким образом, можно говорить о параллелизме на уровне ПЯ. Однако это не единственный уровень параллелизма, задействованный в РОС. Специфика исполнения архитектуры позволяет выделить достаточно глубокий конвейер, каждая из ступеней которого может функционировать независимо. Также можно выделить параллелизм на самом низком уровне — в вычислительных блоках и регистрах. В настоящем исполнении РОС поддерживает несколько разновидностей суперскалярных режимов вычислений [5].

На рис. 1 приведена структура двухуровневой архитектуры РОС, с подробным описанием которой можно ознакомиться в работе [1].

Как было представлено выше, РОС имеет свой уникальный язык программирования, называемый капсульным. В последующих разделах будет приведен пример программы на капсульном языке. Предлагаемому языку в чистом виде (без вспомогательных утилит и подходов) свойственны те же недостатки, что и другим языкам ассемблерного типа:

1. Поскольку алгоритм является параллельным и рекуррентно свернутым, проследить ход ВП практически невозможно в отсутствие специальных навыков или инструментальных средств.
2. При создании архитектуры и капсульного языка преследовалась цель минимизировать размер потока самодостаточных данных. С одной стороны,

это достигается благодаря рекуррентной свертке, а с другой — при помощи специальных конструкций языка. Поэтому по структуре и размеру капсулы невозможно заранее оценить время ее исполнения (в условных вычислительных шагах). При условии, что капсула написана корректно, минимальная длина — ее преимущество, оборачивающееся потерей наглядности.

Поиск возможных путей устранения этих недостатков привел к необходимости построения моделей программ. Развитие точных методов в программировании вызвало возникновение различных формальных моделей программ, в том числе и моделей параллельных программ. Среди них можно выделить: операторные схемы программ, сети Петри, UCLA-графы и др. [6]. Эти модели позволяют представить алгоритм в виде графовых структур, демонстрирующих последовательные и параллельные участки как вычислительного, так и управляющего процесса.

Развитием моделей параллельных программ для РОС стала концепция графодинамики [7]. Согласно этой концепции программа представляется в виде генерируемой последовательности графов ВП. Содержание графов зависит от настройки преобразователя тегов на предметную область. В РОС задано жесткое ограничение на функциональность преобразователя тегов, благодаря чему возможна обратная трассировка от конечного графа (развернутого алгоритма) к свернутому представлению (капсуле). Такая концепция определяет последовательность разработки капсул: построение динамических графов с последующей сверткой конечного графа в капсулу.

3 Имитационная модель рекуррентного обработчика сигналов

Разработка модели преследует несколько целей, наиболее значимые из которых: отладка и доработка VHDL-модели РОС и ее реализации в виде СБИС; разработка и отладка ПО в условиях постоянно совершенствующейся спецификации архитектуры; исследование реакций модели на эволюционное развитие МПРА.

Для построения модели РОС был выбран метод имитационного моделирования, позволяющий охватить все интересующие разработчиков аспекты функционирования РОС как системы в целом, а также функционирование ПО в его среде.

Предлагаемый вариант архитектуры РОС требует представления множества алгоритмов разрабатываемого ПО в виде двухуровневой структуры. К первому уровню можно отнести те алгоритмы, реализация которых нецелесообразна в среде РОУ (например, алгоритмы управления данными, алгоритмы управления ВП и др.). Ко второму уровню относятся алгоритмы, реализующие вычислительные задачи высокой степени параллелизма и сложности. Это является еще одной причиной разработки имитационной модели.

В настоящее время ведутся работы над двумя версиями модели РОС, различными по своей функциональности и назначению. Первая, условно названная «черный ящик» — упрощенная модель, фактически преобразованная в автомат. Вторая, названная «имитатор РОС» — полнофункциональная модель, описывающая поведение основных блоков РОС и позволяющая отслеживать протекание ВП.

А. Модель «черный ящик»

Эта версия модели разрабатывается с целью автоматизации процессов отладки «имитатора РОС», VHDL-модели и аппаратной однокристалльной реализации РОС, а также алгоритмов программы распознавателя слов, исполняемых в среде РОС.

Модель является одним из компонентов полнофункциональной версии и позволяет определить корректность или некорректность входного набора данных для заданной входной программы (капсулы) и в случае его корректности — вычислить соответствующий выходной набор данных. Полученный выходной набор отправляется на сравнение с результатами работы «имитатора РОС» или аппаратной реализации РОС (в зависимости от объекта отладки).

Концептуальная структура модели представлена на рис. 2. Она содержит два основных компонента: анализатор входной капсулы и вычислитель.

Анализатор входной капсулы осуществляет ее синтаксический разбор с целью поиска различных операндов, несущих входные данные от УУ. При обнаружении всех необходимых операндов формируется комплект текущих входных данных.

Вычислитель является простым автоматом, осуществляющим поиск комплекта выходных данных в матрице для заданной капсулы (алгоритма) и текущего комплекта входных данных. В случае если соответствия не найдено, фиксируется

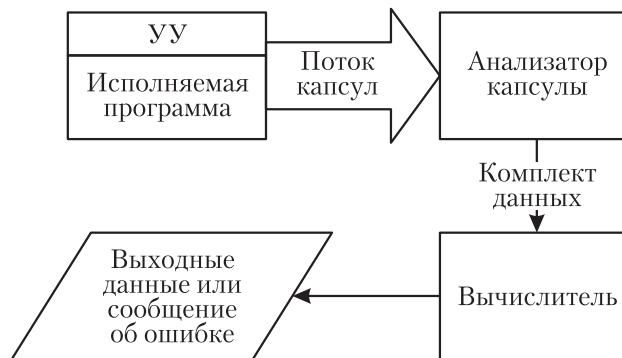


Рис. 2 Структура модели «черный ящик»

факт несоответствия структуры капсулы формату потока данных, поступающих с УУ.

Важно отметить, что указанная матрица формируется заранее программистом, в текущей версии модели — вручную. При дальнейшем развитии в состав модели будет введена утилита формирования матриц.

Б. Модель «имитатор РОС»

Данная версия модели на текущем этапе разработки ориентирована на отладку ПО, разрабатываемого для РОС. Разработчики стремятся к тому, чтобы структура модели максимально соответствовала спецификации РОС на логическом уровне. Данный подход позволяет использовать модель для исследования поведения функциональных блоков и локализации объектов архитектуры, если появится необходимость их доработки. Другими словами осуществляется глубокий анализ капсул и элементов архитектуры и проверка:

- (1) соответствия структуры капсулы алгоритма формату потока данных, поступающих с УУ;
- (2) соответствия поведения «имитатора РОС» и/или аппаратной реализации РОС требованиям, заложенным в спецификации;
- (3) корректности и эффективности реализации исполняемой капсулы;
- (4) целесообразности и эффективности функционирования конкретных механизмов архитектуры.

Расширенная версия данной модели предоставляет следующие возможности:

- (1) мультипроектность — возможность работы сразу с несколькими проектами отлаживаемых программных систем (например, с несколькими различными проектами распознавателя слов);
- (2) создание, редактирование и исполнение программ (капсул) на языке капсульного программирования;
- (3) подключение и использование специализированного компилятора для языка капсульного программирования (в настоящее время находится в разработке);
- (4) автоматическое и пошаговое моделирование процесса исполнения капсулы или последовательности капсул; эта функция позволит также осуществлять полную трассировку ВП, а при необходимости и корректировать его с сохранением всех изменений в лог;
- (5) использование модели «черный ящик» для автоматизированного самотестирования и контроля результатов вычисления капсул.

Концептуальная структура модели, содержащая шесть основных компонентов, представлена на рис. 3.

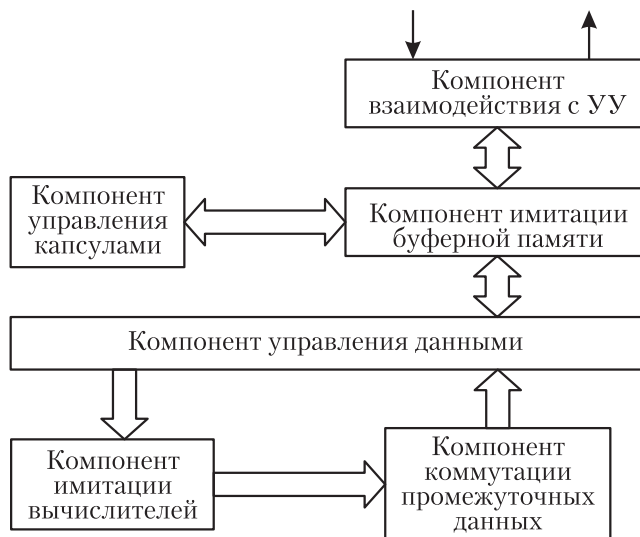


Рис. 3 Структура модели «имитатор РОС»

Компонент взаимодействия с УУ обеспечивает интерфейс, совместимый с ПО УУ. Введение этого компонента позволяет подключать имитационную модель к реальному управляющему процессору.

Компонент имитации буферной памяти совместно с компонентом управления капсулами реализует блок «Память» (см. рис. 1): обеспечивает хранение внутреннего представления капсулы в памяти модели и организует доступ к нему.

Компонент управления капсулами предоставляет сервисные функции для работы с ними (создание, редактирование), а также преобразует исходную капсулу во внутреннее представление модели.

Компонент управления данными — ключевое логическое звено модели — реализуется в виде блока «Распределитель» и четырех блоков «ПАП» (см. рис. 1). Он обеспечивает рассылку операндов в четыре параллельных потока вычислений и образование пар операндов.

Компонент имитации вычислителей, содержащий четыре блока «АЛУ + МАС + ПТ» (см. рис. 1), организует дешифрацию и вычисления в каждом блоке «Вычислитель» в псевдопараллельном режиме (за один такт синхронизации модели последовательно обрабатывают четыре вычислителя).

Компонент коммутации промежуточных данных (блок «Интерфейс обмена данными» на рис. 1) осуществляет сбор и перераспределение промежуточных данных между выходными буферами модели и компонентом управления данными.

Реализация всех указанных функций и компонентов позволит создать универсальное программное средство, пригодное как для отладки ПО, так и для отладки и усовершенствования архитектуры РОС.

4 Пример применения модели

В качестве примера приводится фрагмент капсулы вычисления четырех натуральных логарифмов. Здесь капсула составлена корректно и заполнена корректным набором данных. Ожидаемый результат модели — выходной набор данных. Фрагмент капсулы приводится без комментариев, так как эти данные являются ноу-хау и не раскрываются разработчиками.

```
@d0=-1 Acg: Cx=3 Cp=4 Cc=d Cr= Ce=ecs Cs=ei;
@d0=-1 At: Tc= Tm= Tu= Tr= Tn=0 To=0 Ts= Te= Ta= [Sm=2 Dr=1111 ];
@d0=-1 Casn: [Oc=*sa Ou=l Sm=0 Dr=1111 Ds=o De= ];
@d0=-1 Ccs: [Oc= Ou=k Sm=0 Dr=1111 Ds= De=] Cj=ri Cl=> Cd=Cb=Cs=0 Cm=;
@d0 = 0 Di: V = 6454 [Oc = *sa Ou = l Sm = 1 Dr = 0001 Ds = n De =];
@d0 = 1 Di: V = 16002 [Oc = *sa Ou = l Sm = 1 Dr = 0010 Ds = n De =];
@d0 = 2 Di: V = 11170 [Oc = *sa Ou = l Sm = 1 Dr = 0100 Ds = n De =];
@d0 = 3 Di: V = 8917 [Oc = *sa Ou = l Sm = 1 Dr = 1000 Ds = n De =];
- - -
@d0=-1 Di: V=ra [Oc=e Ou=k Sm=0 Dr=1111 Ds=o De= ];
- - -
@d0=-1 Di: V=rc [Oc=e Ou=e Sm=0 Dr=1111 Ds=n De= ];
@d0=-1 Az:
```

В данном примере входные данные выделены курсивом. Если для наглядности обозначить

$$g1 = 6454, \quad g2 = 16\,002, \quad g3 = 11\,170, \quad g4 = 8917,$$

то отладочные запуски распознавателя слов с целью извлечения промежуточных значений натуральных логарифмов указанных величин приводят к следующим результатам:

$$\ln(g1) = 7187, \quad \ln(g2) = 21\,947, \quad \ln(g3) = 13\,659, \quad \ln(g4) = 10\,409.$$

Результаты запуска модели «черный ящик» отражены на рис. 4 и совпадают с ожидаемыми.

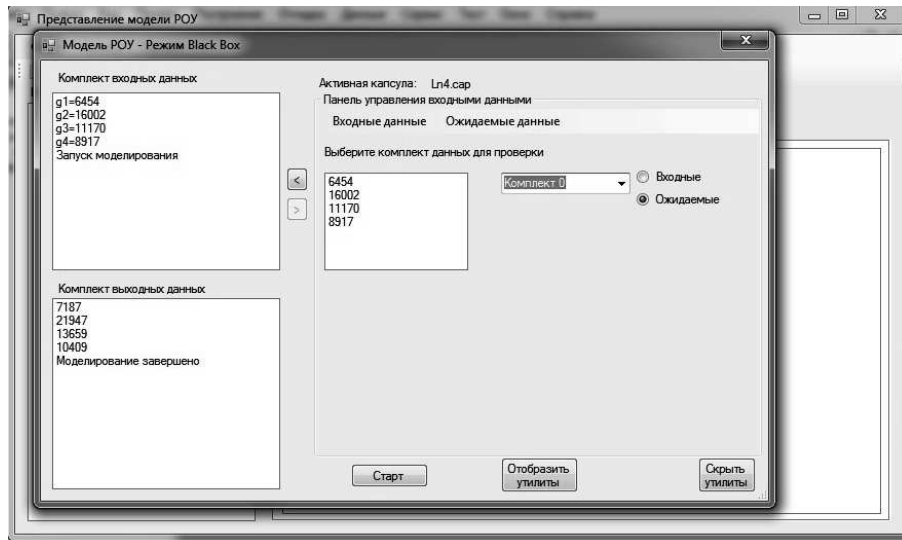


Рис. 4 Результат запуска модели «черный ящик»

5 Заключение

Для повышения эффективности и качества процесса разработки полнофункциональной VHDL-модели РОС и СБИС на ее основе создана имитационная модель РОС, отражающая требования спецификации.

Модель наглядно демонстрирует процессы, протекающие внутри РОС, служит хорошим средством обнаружения ошибок в ПО и поддерживает два режима работы, каждый из которых имеет уникальное назначение и область применения. Оба режима созданы, чтобы предоставлять удобный инструментарий программистам и разработчикам аппаратной реализации РОС.

Разработчики модели надеются, что она станет мощным инструментом отладки и проектирования сложных внутренних механизмов РОС.

Литература

1. Степченко Ю. А., Петрухин В. С. Особенности гибридного варианта реализации на ПЛИС рекуррентного обработчика сигналов // Системы и средства информатики. Доп. вып. — М.: ИПИ РАН, 2008. С. 118–129.
2. Степченко Ю. А., Петрухин В. С., Филин А. В. Рекуррентное операционное устройство для процессоров обработки сигналов // Системы и средства информатики. — М.: Наука, 2001. Вып. 11. С. 283–315.

3. Хилько Д. В., Степченков Ю. А. Вопросы программируемости многоядерной вычислительной архитектуры с единым потоком для эффективной реализации рекуррентных вычислений // Многоядерные процессоры и параллельное программирование; Системы обработки сигналов на базе ПЛИС и цифровых сигнальных процессоров: Сб. ст. Регион. науч.-практ. конф. — Барнаул: Алтайский ун-т, 2011. С. 86–92.
4. Зеленов Р. А., Степченков Ю. А., Волчек В. Н., Хилько Д. В., Шнейдер Ю. А., Прокофьев А. А. Система капсульного программирования и отладки // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1. С. 24–30.
5. Степченков Ю. А., Волчек В. Н., Петрухин В. С., Прокофьев А. А., Зеленов Р. А. Механизмы обеспечения поддержки алгоритмов цифровой обработки речевых сигналов в рекуррентном обработчике сигналов // Системы и средства информатики. — М.: ТОРУС ПРЕСС, 2010. Вып. 20. № 1. С. 31–47.
6. Питерсон Дж. Теория сетей Петри и моделирование систем / Пер. с англ. — М.: Мир, 1984. 264 с.
7. Филин А. В. Динамический подход к выбору архитектуры вычислительных устройств обработки сигналов // Системы и средства информатики. — М.: Наука, 2001. Вып. 11. С. 247–261.

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ИНФОРМАЦИИ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ ФОРМЫ РЕЛЬСОВЫХ НИТЕЙ

С. К. Дулин¹, И. Н. Розенберг², В. И. Уманский³

Аннотация: На основе анализа экспериментальных данных предложены аналитические модели обработки информации, поступающей от измерителей инерциально-спутниковой системы, фиксирующих аномалии формы рельсовой нити (АФРН). Модели описываются нестационарными дифференциальными уравнениями с неизвестными параметрами. Полученные уравнения оценки вектора состояния интегрированной навигационной системы (НС) позволили разработать метод идентификации параметров АФРН, свободный от необходимости расширения размерности фильтра и допущения о стационарности идентифицируемых параметров и обеспечивающий повышение точности идентификации за счет использования вероятностных критериев, нелинейно зависящих от апостериорной плотности вероятности (АПВ) вектора состояния.

Ключевые слова: инерциально-спутниковая система; аномалия формы рельсовой нити; апостериорная плотность вероятности вектора состояния

1 Введение

Ключевой задачей интеллектуализации средств обработки информации о состоянии технических объектов является создание информационно-измерительных систем, позволяющих осуществить интеграцию пространственно-распределенной информации, с которой имеет дело географическая информационная система (ГИС), с данными высокоточного спутникового позиционирования, объединив их в единой геоинформационной базе данных отрасли. Это позволит рассматривать все виды геоинформационных ресурсов во взаимосвязи и оперативно получать информацию необходимого вида и содержания.

Оценка параметров АФРН по измерительной информации, поступающей от приборов — гироскопов и акселерометров, — не может быть решена без учета соответствующих априорных данных с требуемой достоверностью в реальном

¹Институт проблем информатики Российской академии наук, s.dulin@ccas.ru

²ОАО Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте (НИИАС), I.Rozenberg@gismps.ru

³ЗАО Интехгеотранс, umanvi@yandex.ru

масштабе времени в силу того, что гироскопы и акселерометры выдают измерения векторов скорости и ускорения со случайными ошибками и моменты времени появления на рельсовой нити АФРН априори не известны — они случайны, случайность имеет место также и в значениях параметров АФРН.

Поэтому данная задача может быть поставлена и разрешена только как динамическая статистическая задача одновременного обнаружения и оценивания параметров АФРН при исходных данных в виде систем стохастических дифференциальных уравнений вектора состояния локомотива, стохастических дифференциальных уравнений наблюдения, описывающих связь измерений с параметрами АФРН, и критерия оптимальности идентификации в виде апостериорного функционала при условии полученных измерений в текущие моменты времени.

В представленной работе на основе экспериментальных данных впервые разработаны аналитические модели реакции измерителей инерциально-спутниковой системы на АФРН, описываемые нестационарными дифференциальными уравнениями с неизвестными параметрами, определяющими характер АФРН. Полученные модели послужили основой для создания оригинального метода идентификации параметров АФРН в реальном времени, свободного от недостатков традиционных методов адаптивной фильтрации [1, 2] — необходимости расширения размерности фильтра и допущения о стационарности идентифицируемых параметров. Метод обеспечивает повышение точности идентификации за счет использования вероятностных критериев, нелинейно зависящих от АПВ состояния.

Проведенный анализ причин и условий возникновения АФРН показал, что метод обнаружения АФРН должен работать по критерию совместного обнаружения альтернативных сложных гипотез, представляющих возможные типы АФРН на различных участках движения локомотива при аддитивном воздействии помех. Достоверность обнаружения АФРН при этом определяется вероятностью превышения корреляционным интегралом порогового уровня, устанавливаемого по допустимой вероятности ложного обнаружения АФРН из-за помех при движении локомотива.

Отличительной особенностью разработанных методов и алгоритмов оперативного мониторинга АФРН по данным бортовой инерциально-спутниковой системы от результатов предыдущих исследований является использование критерия совместного обнаружения и оценивания параметров альтернативных сложных гипотез и метода субоптимальной идентификации параметров АФРН в реальном времени. Повышение точности идентификации обеспечивается за счет использования вероятностных критериев, нелинейно зависящих от АПВ вектора состояния.

Несмотря на то что конечной целью исследований ставилась лишь разработка методов и алгоритмов оперативного мониторинга АФРН, с целью проверки до-

стоверности обнаружения и оценки параметров АФРН были проведены реальные эксперименты с использованием четырех датчиков ADIS 16364 [3]. Все четыре датчика составляли единую систему регистрации, показания которой совместно обрабатывались в общем вычислительном комплексе для двух значений скорости движения «Сапсана»: 120 км/ч на криволинейных участках пути (поворотах) и 250 км/ч на прямолинейных участках. Величина отклонения от нормы конфигурации рельсовых линий на аномальных участках принималась равной 20, 8 и 4 мм. Учитывались две крайние категории просадок и искривлений пути в плане, а именно I (наименее опасная) и IV (наиболее опасная), т. е. при скорости движения в 120 км/ч просадки и искривления не должны превышать амплитуды в 8 мм для I категории, а для IV категории просадки и искривления не должны превышать амплитуды в 20 мм. Тем самым разработанные методы и алгоритмы обнаружения и оценивания параметров АФРН были апробированы в реальном масштабе времени и в реальных условиях движения локомотива.

Технология мониторинга АФРН внедрена в Комплексную систему пространственных данных инфраструктуры железнодорожного транспорта (КСПД ИЖТ) ОАО «Росжелдорпроекта», что подтверждено соответствующим актом о внедрении.

2 Математические модели аномалии формы рельсовой нити

Исходные данные о реальном состоянии рельсовых нитей могут быть получены от интегрированной бортовой НС, фиксирующей линейную и угловую скорость локомотива, а также его ускорения. Значения этих характеристик зависят от типа участка рельсовых нитей: прямолинейный участок, односекционный поворот, стрелочный перевод, уклон. На каждом из этих участков могут иметь место различные АФРН, в том числе односторонние и двухсторонние просадки, искривления пути в плане, а также шумовые воздействия на систему регистрации.

Как видно из рис. 1 и 2, на которых изображены формы эффекта от АФРН, возникающих в процессе эксплуатации железнодорожных путей, в одномерном случае (по оси чувствительности датчика) реакцию измерителя на АФРН достаточно адекватно можно аппроксимировать конечной суммой соответствующим образом взвешенных гауссоид:

$$Z^*(t) = \sum_{i=1}^n \gamma_i \exp [-(t - m_i)^2 \sigma_i] , \quad (1)$$

где γ_i , m_i , σ_i — параметры АФРН, в общем случае нестационарные (кусочно-постоянные во времени) и определяемые в процессе идентификации, или гауссоидной огибающей, модулирующей знакопеременный сигнал, — например,

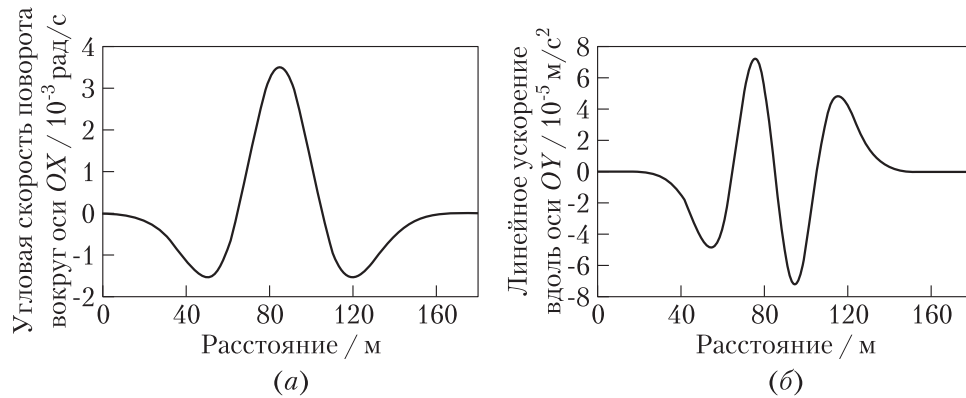


Рис. 1 Выходные сигналы датчиков угловых скоростей (а) и линейного ускорения (б), возникающие при одновременной просадке обеих рельсовых нитей

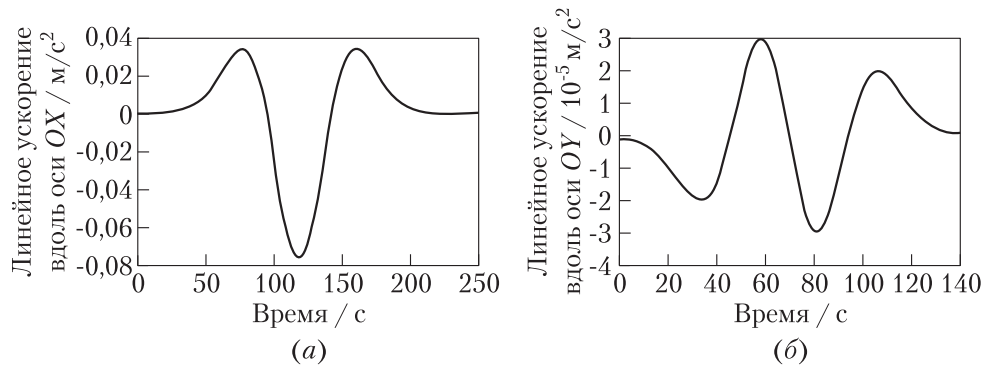


Рис. 2 Изменение линейного ускорения по оси OX (а) и по оси OY (б), которое возникает под воздействием рыскания вправо передней колесной пары

как показано на рис. 1 и 2, низкочастотное ($\omega \approx 0,06-0,1 \text{ с}^{-1}$) гармоническое колебание

$$Z_1^*(t) = \gamma(t) \exp[-(t - m(t))^2 \sigma(t)] \sin \omega t. \quad (2)$$

Выбор гауссоиды как огибающей АФРН объясняется тем обстоятельством, что, несмотря на сложную и изменяющуюся форму, которую принимает рельс в разные моменты времени в процессе движения локомотива и под воздействием нескольких точечных нагрузок (в местах касания колес), координаты точек касания колес практически всегда плавно смещаются вниз в аномальных зонах (например, под «подвешенными» шпалами) и возвращаются в «нулевое» положение на безаномальных участках рельсовой колеи.

Для обеспечения возможности дальнейшего использования существующего аппарата теории нелинейной фильтрации для оценки АФРН представим подобные реакции чувствительных элементов (ЧЭ) на АФРН в виде системы обыкновенных дифференциальных уравнений. Учтем, что при описании параметров сигналов (1) и (2) в силу их кусочно-постоянного характера можно использовать традиционный метод «замороженных коэффициентов» и что функции $z_i = \exp[-(t - m_i)^2 \sigma_i]$, $y = \sin \omega t$ являются решением следующих дифференциальных уравнений:

$$\dot{z}_i = -2z_i(t - m_i)\sigma_i; \quad (3)$$

$$\ddot{y} + \omega^2 y = 0. \quad (4)$$

Система (4) линейно зависит от идентифицируемого параметра $\bar{\omega} = \omega^2$, что существенно упрощает процедуру идентификации, а систему (3) для обеспечения подобной (линейной) зависимости от параметров можно представить в виде

$$\dot{z}_i = -2z_i t \sigma_i + 2z_i m_i \sigma_i = -2z_i t \sigma_i + 2z_i \mu_i,$$

где μ_i — новый параметр, идентификация которого, осуществляемая одновременно с идентификацией параметра σ_i , легко позволяет получить требуемый параметр $m_i = \mu_i / \sigma_i$.

Более того, принимая во внимание тот факт, что гауссоиды, формирующие реакцию ЧЭ на АФРН, распределены во времени последовательно (не перекрываются), обобщенную временную реакцию ЧЭ на АФРН можно представить в виде только одной гауссоиды с кусочно-постоянными во времени априори неизвестными параметрами γ , μ и σ :

$$Z^*(t) = \gamma z, \quad \dot{z} = -2zt\sigma + 2z\mu \quad (5)$$

или как

$$Z_1^*(t) = \gamma(t)z(t)y(t), \quad \dot{z} = -2zt\sigma + 2z\mu, \quad \ddot{y} + \omega^2 y = 0. \quad (6)$$

Так как представление (5) по вычислительным затратам оказывается почти в 3 раза эффективнее (6), то далее будем использовать описание реакции ЧЭ на АФРН в виде (5). При этом также будем учитывать, что, несмотря на общность формы описания реакции ЧЭ, вытекающую из подобия реакции на АФРН всех ЧЭ — акселерометров, датчиков угловых скоростей (ДУСов), измерителей линейной скорости, — параметры этой реакции для разных типов ЧЭ различны. Поэтому в представлении (5) при описании переменных и параметров далее будем использовать следующие индексы: a_i — для акселерометров, V_i — при измерении линейной скорости, ω_i — для ДУСов, $i = X, Y, Z$. В этом случае трехмерный

вектор реакции акселерометров (по всем осям приборного трехгранника) можно представить как

$$\begin{vmatrix} \gamma_{a_x} & z_{a_x} \\ \gamma_{a_y} & z_{a_y} \\ \gamma_{a_z} & z_{a_z} \end{vmatrix} = \begin{vmatrix} z_{a_x} & 0 & 0 \\ 0 & z_{a_y} & 0 \\ 0 & 0 & z_{a_z} \end{vmatrix} \begin{vmatrix} \gamma_{a_x} \\ \gamma_{a_y} \\ \gamma_{a_z} \end{vmatrix} = S_a(z_\alpha) \begin{vmatrix} \gamma_{a_x} \\ \gamma_{a_y} \\ \gamma_{a_z} \end{vmatrix} = S_a(z_\alpha)\gamma_a,$$

где

$$\begin{aligned} \begin{vmatrix} \dot{z}_{a_x} \\ \dot{z}_{a_y} \\ \dot{z}_{a_z} \end{vmatrix} &= \begin{vmatrix} -2z_{a_x}t\sigma_{a_x} + 2z_{a_x}\mu_{a_x} \\ -2z_{a_y}t\sigma_{a_y} + 2z_{a_y}\mu_{a_y} \\ -2z_{a_z}t\sigma_{a_z} + 2z_{a_z}\mu_{a_z} \end{vmatrix} = \\ &= \begin{vmatrix} -2z_{a_x}t & 2z_{a_x} & 0 & 0 & 0 & 0 \\ 0 & 0 & -2z_{a_y}t & 2z_{a_y} & 0 & 0 \\ 0 & 0 & 0 & 0 & -2z_{a_z}t & 2z_{a_z} \end{vmatrix} \begin{vmatrix} \sigma_{a_x} \\ \mu_{a_x} \\ \sigma_{a_y} \\ \mu_{a_y} \\ \sigma_{a_z} \\ \mu_{a_z} \end{vmatrix} = S_{a*} \begin{vmatrix} \sigma_{a_x} \\ \mu_{a_x} \\ \sigma_{a_y} \\ \mu_{a_y} \\ \sigma_{a_z} \\ \mu_{a_z} \end{vmatrix} = S_{a*}\alpha_a; \end{aligned}$$

γ_a и α_a — векторы неизвестных параметров реакции акселерометров на АФРН, или

$$\dot{z}_a = S_{a*}(z_\alpha, t)\alpha_a.$$

Аналогично, с точностью до обозначений, может быть представлен трехмерный вектор реакции ДУСов: $S_d(z_d)\gamma_d$ и $\dot{z}_d = S_{d*}(z_d, t)\alpha_d$.

При измерении вектора реакции в виде линейной скорости z_v учтем, что он представляет собой интеграл от приведенного выше вектора ускорений и, следовательно, в дифференциальной форме может быть описан как $\dot{z}_v = S_a(z_\alpha)\gamma_a$.

3 Математическая модель интегрированной навигационной системы с учетом реакции чувствительных элементов на аномалии форм рельсовой нити

В качестве исходной модели вектора состояния НС будем использовать модель [2], описывающую пространственное движение объекта в параметрах Родрига–Гамильтона. Тогда, учитывая аддитивность реакции акселерометров и ДУСов на АФРН по отношению к основным измеряемым сигналам, модель вектора состояния НС с учетом эффектов от АФРН можно представить следующим образом:

$$\left. \begin{aligned}
 \dot{\lambda} &= \frac{1}{2} \Phi_0(\lambda) \left\{ (r+h)^{-1} \begin{vmatrix} -V_Y \\ V_X \\ 0 \end{vmatrix} + \begin{vmatrix} 2\Omega(\lambda_2\lambda_3 + \lambda_1\lambda_4) \\ \Omega(2\lambda_1^2 + 2\lambda_3^2 - 1) \\ 2\Omega(\lambda_3\lambda_4 - \lambda_1\lambda_2) \end{vmatrix} \right\}; \\
 \dot{\mu} &= \frac{1}{2} \Phi_0(\mu) (Z_d - m_d - S_d\gamma_d - W_d); \\
 \dot{z}_d &= S_{d*}\alpha_d; \quad \dot{a}_d = S_{a*}\alpha_a; \quad \dot{z}_v = S_a(z_\alpha)\gamma_a; \\
 \dot{V}_X &= C_{(1)}(\mu, \lambda)(Z_a - S_a\gamma_a - W_a) + 4\Omega(\lambda_3\lambda_4 - \lambda_1\lambda_2)V_Y - \\
 &\quad - (2\Omega(2\lambda_1^2 + 2\lambda_3^2 - 1) + V_X(r+h)^{-1})V_Z - \\
 &\quad - 4\Omega^2(r+h)(\lambda_2\lambda_3 + \lambda_1\lambda_4)(\lambda_3\lambda_4 - \lambda_1\lambda_2); \\
 \dot{V}_Y &= C_{(2)}(\mu, \lambda)(Z_a - S_a\gamma_a - W_a) - 4\Omega(\lambda_3\lambda_4 - \lambda_1\lambda_2)V_X + \\
 &\quad + (4\Omega(\lambda_2\lambda_3 + \lambda_1\lambda_4) - V_Y(r+h)^{-1})V_Z - \\
 &\quad - 2\Omega^2(r+h)(2\lambda_1^2 + 2\lambda_3^2 - 1)(\lambda_3\lambda_4 - \lambda_1\lambda_2); \\
 \dot{V}_Z &= C_{(3)}(\mu, \lambda)(Z_a - S_a\gamma_a - W_a) + (2\Omega(2\lambda_1^2 + 2\lambda_3^2 - 1) + \\
 &\quad + V_X(r+h)^{-1})V_X - (4\Omega(\lambda_2\lambda_3 + \lambda_1\lambda_4) - V_Y(r+h)^{-1}) + \\
 &\quad + \Omega^2(r+h) (2(\lambda_2\lambda_3 + \lambda_1\lambda_4)^2 - (2\lambda_1^2 + 2\lambda_3^2 - 1)^2) + \\
 &\quad + g_0(r, h, \varphi(\lambda_1, \dots, \lambda_4)); \\
 \dot{h} &= V_Z,
 \end{aligned} \right\} \quad (7)$$

или в векторной форме Ланжевена соответственно:

$$\dot{Y} = F(Y, t) + S(Y, t)\alpha + F_0(Y, t)\xi,$$

где

$$Y = |\lambda^T \mu^T z_d^T z_a^T z_v^T V_X V_Y V_Z h|^T, \quad \xi = |W_d^T W_a^T|^T;$$

векторная и матричная функции $F(Y, t)$ и $F_0(Y, t)$ приведены в [2]:

$$S(Y, t) = \begin{vmatrix} -\frac{1}{2} \Phi_0(\mu)S_d & 0 & 0 & 0 \\ 0 & S_{d*} & 0 & 0 \\ 0 & 0 & S_{a*} & 0 \\ 0 & 0 & 0 & -C(\mu, \lambda)S_a \end{vmatrix}; \quad \alpha = |\gamma_d^T \alpha_d^T \alpha_a^T \gamma_a^T|^T.$$

Для высокоточного решения задачи идентификации параметров АФРН полагаем далее, что наблюдение навигационных параметров осуществляется только с использованием средств измерения спутниковых НС (СНС). Так как время появления АФРН в процессе движения локомотива не известно, то для обеспечения возможности их наблюдения требуется высокая частота позиционирования — существенно выше, чем в ГЛОНАСС (0,5 Гц).

Поэтому в качестве СНС, используемой для решения данной задачи, целесообразно рассматривать GPS (global positioning system), обеспечивающую высокую частоту позиционирования — 20 Гц (в ближайшей перспективе — до 100 Гц), позволяющую практически без потери точности считать процесс наблюдения непрерывным.

В качестве моделей наблюдателей используем информационные модели сигналов кодовых и доплеровских измерений, имеющие в параметрах Родрига–Гамильтона без учета реакции на АФРН следующий вид:

$$\begin{aligned}
 Z_R &= H_R(\lambda, h) + W_{Z_R} = W_{Z_R} + ((\xi_c - 2(r+h)(\lambda_2\lambda_4 + \lambda_1\lambda_3))^2 + \\
 &\quad + (\eta_c - 2(r+h)(\lambda_3\lambda_4 - \lambda_1\lambda_2))^2 + (\zeta_c - (r+h)(2\lambda_1^2 + 2\lambda_4^2 - 1))^2)^{1/2}; \\
 Z_V &= \left[(\xi_c - 2(r+h)(\lambda_2\lambda_4 + \lambda_1\lambda_3))(V_{\xi_c} - [GB^T(\lambda)]_{(1)}V_S) + (\eta_c - \right. \\
 &\quad \left. - 2(r+h)(\lambda_3\lambda_4 - \lambda_1\lambda_2))(V_{\eta_c} - [GB^T(\lambda)]_{(2)}V_S) + \right. \\
 &\quad \left. + (\xi_c - (r+h)(2\lambda_1^2 + 2\lambda_4^2 - 1))(V_{\zeta_c} - [GB^T(\lambda)]_{(3)}V_S) \right] \times \\
 &\quad \times ((\xi_c - 2(r+h)(\lambda_2\lambda_4 + \lambda_1\lambda_3))^2 + \\
 &\quad + (\eta_c - 2(r+h)(\lambda_3\lambda_4 - \lambda_1\lambda_2))^2 + (\zeta_c - (r+h)(2\lambda_1^2 + 2\lambda_4^2 - 1))^2)^{-1/2} + \\
 &\quad + W_{Z_V} = H_V(\lambda, h, V_S) + W_{Z_V}, \quad (8)
 \end{aligned}$$

где $[GB^T(\lambda)]_{(i)}$ — i -я строка матрицы $GB^T(\lambda)$.

Так как АФРН на изменение пространственных координат ЦМ локомотива практически не влияют, то в уравнениях спутниковых измерений учтем далее только влияние АФРН на изменение вектора линейной скорости. Связь векторов линейной скорости в гринвичской и приборной системах координат определяется матрицей конечного поворота $G(\Omega t)D^T(\mu)$, компоненты которой описаны в [2], следовательно, модель сигнала доплеровских измерений с учетом реакции на АФРН имеет вид:

$$\begin{aligned}
 Z_V &= \left[(\xi_c - 2(r+h)(\lambda_2\lambda_4 + \lambda_1\lambda_3))(V_{\xi_c} - [GB^T]_{(1)}V_S - [GD^T]_{(1)}z_v) + \right. \\
 &\quad \left. (\eta_c - 2(r+h)(\lambda_3\lambda_4 - \lambda_1\lambda_2))(V_{\eta_c} - [GB^T]_{(2)}V_S - [GD^T]_{(2)}z_v) \right] +
 \end{aligned}$$

$$\begin{aligned}
 & + (\zeta_c - (r+h)(2\lambda_1^2 + 2\lambda_4^2 - 1)) (V_{\zeta_c} - [GB^T]_{(3)}V_S - [GD^T]_{(3)}z_v) \Big] \times \\
 & \quad \times \left((\xi_c - 2(r+h)(\lambda_2\lambda_4 + \lambda_1\lambda_3))^2 + (\eta_c - 2(r+h)(\lambda_3\lambda_4 - \lambda_1\lambda_2))^2 + \right. \\
 & \left. + (\xi_c - (r+h)(2\lambda_1^2 + 2\lambda_4^2 - 1))^2 \right)^{-1/2} + W_{Z_V} = H_{V^*}(\lambda, \mu, z_v, h, V_S) + W_{Z_V}, \quad (9)
 \end{aligned}$$

где $[GD^T]_{(i)}$ — i -я строка матрицы $GD^T(\mu)$.

Для упрощения последующей записи будем использовать аналогично вышеизложенному векторное представление уравнений спутниковых измерений (8) и (9) в виде:

$$Z = \begin{vmatrix} Z_R \\ Z_V \end{vmatrix} = \begin{vmatrix} H_R(\lambda, h) \\ H_{V^*}(\lambda, \mu, z_v, h, V_S) \end{vmatrix} + \begin{vmatrix} W_{Z_R} \\ W_{Z_V} \end{vmatrix} = H_{c^*}(Y, t) + \zeta,$$

где ζ — белый гауссовский шум с нулевым средним и матрицей интенсивностей

$$D_{CHC} = \begin{vmatrix} D_{Z_R} & 0 \\ 0 & D_{Z_V} \end{vmatrix}.$$

4 Математическая постановка задачи идентификации аномалии формы рельсовой нити

Цель идентификации АФРН — восстановление истинной реализации вектора АФРН в реальном времени контроля рельсовых путей с требуемой достоверностью при минимальных вычислительных затратах с последующей оценкой ее параметров и классификацией типа, т. е. с установлением принадлежности к конкретному участку рельсового пути: линейному, односекционному повороту, стрелочному переводу, линейному уклону или подъему, выпуклому или вогнутому профилю.

Так как классификация типа АФРН может быть осуществлена только после оценки реализации вектора АФРН и идентификации его параметров, то рассмотрим далее в качестве первоочередной задачу одновременного оценивания вектора параметров движения локомотива с учетом влияния АФРН. Опираясь на результаты разд. 2 и 3, данную задачу сформулируем как задачу текущего стохастического оценивания вектора навигационных параметров объекта

$$\dot{Y} = F(Y, t) + S(Y, t)\alpha + F_0(Y, t)\xi, \quad (10)$$

наблюдаемого СНС,

$$Z = H_{c^*}(Y, t) + \zeta$$

с целью идентификации α — вектора параметров АФРН.

Для решения задачи идентификации необходимо, чтобы искомым вектор $\alpha(t)$ доставлял оптимум некоторому заданному обобщенному вероятностному функционалу J , зависящему от АПВ $\rho(Y, t)$ процесса Y , причем в общем случае нелинейно.

Анализируя физическое существо решаемой задачи, можно сделать вывод о том, что в качестве наиболее адекватной формы минимизируемого функционала J наиболее целесообразно использовать аддитивную совокупность двух функционалов: оптимизация первого из которых $J_1 = \int_Y \Phi_1[p(Y, t)] dY$ должна обеспечить минимум неопределенности (максимальную информативность) идентифицированного вектора α , и второго $J_2 = \int_T \Phi_2[\alpha(t)] dt$ — регуляризацию процедуры идентификации, т. е. минимум его «энергетики» (в соответствии с основными подходами теории регуляризации, а также с принципом Ферма) на заданном (конечном или текущем) интервале времени T , т. е.

$$J = \int_Y \Phi_1[\rho(Y, t)] dt + \int_T \Phi_2[\alpha(t)] dt.$$

При этом в соответствии с постановкой задачи функцию Φ_1 целесообразно выбирать как ядро функционала энтропии Шеннона ($\Phi_1 = -\rho \ln \rho$) или Фишера ($\Phi_1 = -\rho[\partial \ln \rho / d\xi][\partial \ln \rho / d\xi]^T$), а Φ_2 — в виде классической квадратичной, регуляризующей формы, заданной на текущем интервале времени:

$$J_2 = \int_{t_0}^t \alpha^T \alpha dt = \int_{t_0}^t \Phi_2[\alpha] dt.$$

Таким образом, исследуемую задачу можно сформулировать как задачу поиска вектора α , доставляющего минимум функционалу

$$J = \int_Y \Phi_1[\rho] dY + \int_{t_0}^t \Phi_2[\alpha] dt, \quad (11)$$

определенному на множестве функций АПВ ρ , удовлетворяющих решению известного уравнения Стратоновича [2], построенного в соответствии с видом уравнения объекта (10) и его наблюдателя.

Как было отмечено ранее, формирование АПВ ρ на основе решения данного уравнения в многомерном случае представляет собой непростую вычислительную задачу, поэтому используем далее гауссовскую аппроксимацию $\tilde{\rho}$, позволяющую свести задачу построения АПВ к интегрированию системы обыкновенных дифференциальных уравнений ее параметров:

$$\left. \begin{aligned}
 \hat{Y} &= F(\hat{Y}, t) + S(\hat{Y})\alpha + K(\hat{Y}, t) \left[Z - H_{c^*}(\hat{Y}, t) \right] = \\
 &= G_1(\hat{Y}, R, Z, t) + S(\hat{Y})\alpha; \\
 K(\hat{Y}, t) &= R \frac{\partial H_{c^*}^T(\hat{Y}, t)}{\partial \hat{Y}} D_{CHC}^{-1}; \\
 \dot{R}(\hat{Y}, t) &= \frac{\partial F(\hat{Y}, t)}{\partial \hat{Y}} R(\hat{Y}, t) + R(\hat{Y}, t) \frac{\partial F^T(\hat{Y}, t)}{\partial \hat{Y}} + \\
 &+ F_0(\hat{Y}, t) D_\xi F_0^T(\hat{Y}, t) + \frac{\partial S(\hat{Y})}{\partial \hat{Y}} \alpha R + R \alpha^T \frac{\partial S^T(\hat{Y})}{\partial \hat{Y}} - \\
 &- K(\hat{Y}, t) D_{CHC} K^T(\hat{Y}, t) = G_2(\hat{Y}, R, t) + \frac{\partial S}{\partial \hat{Y}} \alpha R + R \alpha^T \frac{\partial S^T}{\partial \hat{Y}}; \\
 \hat{Y}_0 &= M(Y_0), \quad R_0 = M \left\{ (Y_0 - \hat{Y}_0)(Y_0 - \hat{Y}_0)^T \right\}.
 \end{aligned} \right\} \quad (12)$$

Ввиду того, что дальнейшее решение задачи предполагает операции с вектором параметров АПВ $\tilde{\rho}$, а уравнение для R матричное, то, используя приведенные ниже соотношения матричной алгебры, преобразуем уравнения параметров $\tilde{\rho}$ (12) к единой векторной форме:

$$\begin{aligned}
 \hat{Y} &= G_1(\hat{Y}, R, Z, t) + S(\hat{Y})\alpha; \\
 (\dot{R})^{(v)} &= G_2^{(v)}(\hat{Y}, R, t) + (R \otimes E) \left[\frac{\partial \widehat{S}(\hat{Y})}{\partial Y} \right] \alpha + (E \otimes R) \left[\frac{\partial S(\hat{Y})}{\partial Y} \right]^* \alpha = \\
 &= G_2^{(v)}(\hat{Y}, R, t) + G_3(\hat{Y}, R)\alpha; \\
 G_3(\hat{Y}, R) &= (R \otimes E) \left[\frac{\partial \widehat{S}(\hat{Y})}{\partial \hat{Y}} \right] + (E \otimes R) \left[\frac{\partial S(\hat{Y})}{\partial \hat{Y}} \right]^*,
 \end{aligned}$$

где $A^{(V)}$ — операция преобразования матрицы A размерности $m \times n$ в вектор: $A^{(V)} = |a_{11} \ a_{21} \ \dots \ a_{m1} \ a_{12} \ a_{22} \ \dots \ a_{m2} \ \dots \ a_{1n} \ a_{2n} \ \dots \ a_{mn}|^T$; \otimes — кронекеровское произведение; E — единичная матрица соответствующей размерности; $[\widehat{\partial S \hat{Y}} / \partial \hat{Y}]$ — операция вертикальной последовательной переориентации блоков блочной матрицы $B = |B_1 \ B_2 \ \dots \ B_n|$ в матрицу

$$\widehat{B} = \begin{vmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{vmatrix};$$

$[\partial S(\hat{Y})/\partial \hat{Y}]^*$ — операция формирования блочной матрицы B^* путем образования ее блоков из соответствующих строк исходной матрицы B :

$$B^* = \begin{bmatrix} [(B_{(1)}^T)^{\circ}]^T \\ [(B_{(2)}^T)^{\circ}]^T \\ \vdots \\ [(B_{(n)}^T)^{\circ}]^T \end{bmatrix},$$

$B_{(i)}^T$ — i -й столбец B^T , (\circ) — операция, обратная операции преобразования $A^{(V)}$ матрицы A , $(ASC)^{(V)} = (C^T \oslash A)S^{(V)}$, A , S , C — обыкновенные матрицы, $(S\alpha)^{(V)} = \hat{S}\alpha$, $(\alpha^T S^T)^{(V)} = S^* \alpha$.

Полученные уравнения позволяют записать для расширенного вектора параметров

$$Y_* = \begin{bmatrix} \hat{Y} \\ R^{(v)} \end{bmatrix}$$

плотности $\tilde{\rho}$ его зависимость от искомого вектора α в следующей простой форме:

$$\dot{Y}_* = \begin{bmatrix} G_1(Y_*, Z, t) \\ G_2^{(v)}(Y_*, t) \end{bmatrix} + \begin{bmatrix} S(Y_*) \\ G_3(Y_*) \end{bmatrix} \alpha = G_0 + G\alpha. \quad (13)$$

Исходя из изложенного, задачу параметрической идентификации АФРН окончательно сформулируем как задачу поиска вектора α , обеспечивающего оптимум функционала J (11) при условии, что вектор параметров Y_* плотности $\tilde{\rho}(Y, Y_*)$, аппроксимирующей функцию АПВ вектора состояния Y , описывается уравнением (13).

5 Решение задачи идентификации аномалии формы рельсовой нити

Учитывая, что при неотрицательно определенной критериальной функции для обеспечения ее минимального значения в каждый момент времени достаточно, чтобы производная ее по времени, взятая с обратным знаком, имела максимум [1], получаем условие для определения искомого вектора параметров:

$$\max_{\alpha} \{-j\} = \max_{\alpha} \left\{ - \left(\int_{\hat{Y}} \frac{\partial \Phi_1[\tilde{\rho}]}{\partial \rho} \frac{\partial \tilde{\rho}(Y, Y_*)}{\partial Y_*} \dot{Y}_* dY + \Phi_2[\alpha] \right) \right\}. \quad (14)$$

Подставляя в (14) выражение для правой части уравнения (13), имеем следующее уравнение относительно α :

$$\frac{\partial}{\partial \alpha} \left\{ \int_Y \frac{\partial \Phi_1}{\partial \rho} \frac{\partial \tilde{\rho}}{\partial Y_*} dY (G_0 + G\alpha) + \Phi_2[\alpha] \right\} = 0.$$

Из последнего вытекает окончательное уравнение для определения искомого вектора α

$$- \int_Y \frac{\partial \Phi_1}{\partial \rho} \frac{\partial \tilde{\rho}}{\partial Y_*} dY G = \frac{\partial \Phi_2}{\partial \alpha} [\alpha], \quad (15)$$

решение которого осуществляется исходя из конкретного вида функции Φ_2 . Так, для предложенной выше квадратичной формы функции $\Phi_2(\alpha)$ уравнение (15) принимает вид:

$$- \int_Y \frac{\partial \Phi_1}{\partial \rho} \frac{\partial \tilde{\rho}}{\partial Y_*} dY G = 2\alpha^T,$$

откуда

$$\alpha = -\frac{1}{2} G^T \int_Y \frac{\partial \Phi_1}{\partial \rho} \left(\frac{\partial \tilde{\rho}}{\partial Y_*} \right)^T dY. \quad (16)$$

Выражение (16) позволяет легко учесть возможные в общем случае ограничения на вектор α (например, $|\alpha| = \alpha_{\max}$) при известных текущих значениях гауссовской функции $\tilde{\rho}$, формируемых на основе решения уравнений (13) после подстановки в них выражения (16):

$$\dot{Y}_* = G_0 - \frac{1}{2} G G^T \int_Y \frac{\partial \Phi_1}{\partial \rho} \left(\frac{\partial \tilde{\rho}}{\partial Y_*} \right)^T dY. \quad (17)$$

Анализ данного уравнения позволяет сделать следующие выводы, свидетельствующие о преимуществах предложенного метода перед традиционными. Во-первых, рассмотренная процедура идентификации оказывается инвариантной к размерности вектора неизвестных параметров — размерность уравнения (17) сохранилась равной размерности исходной системы оценки вектора состояния (13). По сравнению с системой уравнений традиционной адаптивной фильтрации, использующей предположение о стационарности идентифицируемых параметров и расширяющей вектор состояния за счет уравнений этих параметров, это сокращает размерность интегрируемой системы на $N^2/2 + (3/2 + M)N$, где N — число неизвестных параметров (в рассматриваемом случае $N = 18$), M — размерность

вектора состояния системы (в рассматриваемом случае $M = 12$). Во-вторых, отсутствие допущений о неизменности неизвестных параметров на интервале оценивания позволяет их идентифицировать с большей точностью.

Более того, в силу достаточно простой структуры уравнения (17), аналогичной, по существу, (13), и возможности быстродействующего вычисления определенного интеграла в его правой части (ядро которого известно) с помощью существующих эффективных численных методов, очевидно, что описанный метод идентификации может быть легко реализован в реальном масштабе времени поступления измерительной информации на бортовом вычислителе локомотива на основе текущей информации о навигационных параметрах.

6 Идентификация аномалии формы рельсовой нити при использовании сети дифференциальных поправок

Использование сети дифференциальных поправок позволяет практически точно решать задачу позиционирования локомотива и вычисления его линейной и угловой скорости. Это обстоятельство можно использовать для существенного упрощения решения задачи идентификации АФРН при единственном условии — движении локомотива с заранее известной скоростью, контролируемой, например, той же СНС. В этом случае уравнения вектора состояния НС с учетом АФРН (7) существенно упрощаются:

$$\left. \begin{aligned} \dot{\mu} &= \frac{1}{2} \Phi_0(\mu)(Z_d - m_d - S_d \gamma_d - W_d); \\ \dot{z}_d &= S_{d*} \alpha_d; \quad \dot{z}_a = S_{a*} \alpha_a; \quad \dot{z}_v = S_a(z_\alpha) \gamma_a, \end{aligned} \right\} \quad (18)$$

а для наблюдения за вновь полученным вектором $Y = |\mu^T \ z_d^T \ z_a^T \ z_v^T|^T$ достаточно использовать только сигнал доплеровских измерений, который после учета (вычитания) известных составляющих принимает следующий вид:

$$\begin{aligned} Z_{V*} &= \left[(\xi_c - 2(r+h)(\lambda_2 \lambda_4 + \lambda_1 \lambda_3)) \left(-[GD^T]_{(1)} z_v \right) + \right. \\ &\quad \left. + (\eta_c - 2(r+h)(\lambda_3 \lambda_4 - \lambda_1 \lambda_2)) \left(-[GD^T]_{(2)} z_v \right) + (\zeta_c - \right. \\ &\quad \left. - (r+h)(2\lambda_1^2 + 2\lambda_4^2 - 1)) \left(-[GD^T]_{(3)} z_v \right) \right] \left[(\xi_c - 2(r+h)(\lambda_2 \lambda_4 + \lambda_1 \lambda_3))^2 + \right. \\ &\quad \left. + (\eta_c - 2(r+h)(\lambda_3 \lambda_4 - \lambda_1 \lambda_2))^2 + (\zeta_c - (r+h)(2\lambda_1^2 + 2\lambda_4^2 - 1))^2 \right]^{-1/2} + \\ &\quad + W_{Z_V} = \psi_1(t)[GD^T]_{(1)} z_v + \psi_2(t)[GD^T]_{(2)} z_v + \psi_3(t)[GD^T]_{(3)} z_v + W_{Z_V} = \\ &= H_{V0}(\mu, z_v) + W_{Z_V}, \end{aligned}$$

где $\psi_i(t)$, $i = 1, 2, 3$ — известные функции времени, определяемые по показаниям СНС и информации о скорости локомотива.

Структура фильтра при этом остается аналогичной (12), но существенно сокращается его размерность (соответственно, и для приведенного выше алгоритма идентификации). Рассмотренный случай интересен еще тем, что, во-первых, демонстрирует нетривиальную возможность наблюдения угловых колебаний локомотива (вектора μ) только с помощью СНС, а во-вторых, позволяет решить задачу идентификации АФРН без привлечения показаний акселерометров — только по измерениям ДУСов и СНС.

В завершение анализа решения задачи идентификации АФРН следует заметить, что, несмотря на свою вычислительную простоту, предложенный алгоритм (и метод) идентификации все же усложняет алгоритм текущей оценки вектора навигационных параметров. В связи с этим его целесообразно использовать не в течение всего времени движения, а только при обнаружении АФРН. В свою очередь, обнаружение АФРН легко осуществляется путем контроля характера изменения правой части системы (12) или (18) — при появлении АФРН будут наблюдаться гауссоидальные скачки, легко обнаруживаемые традиционным использованием спектрального анализа, вейвлет-алгоритмов, вычислением корреляционных функций, а также использованием пороговых функций (устройств) или сравнением (в виде отношения правдоподобия) статистики, вычисляемой для получаемых фильтром Калмана оценок аномалии, с пороговым уровнем, установленным по допустимой вероятности ложного обнаружения. При этом особенно эффективно обнаружение АФРН может быть осуществлено при постоянной скорости локомотива, когда правые части уравнений систем (12), (18) или близки к нулю, или имеют весьма простой характер изменения во времени.

7 Имитационное моделирование системы регистрации и идентификации аномалии формы рельсовой нити

В результате проведения предварительных исследований в созданной системе регистрации центральным элементом был определен миниатюрный прибор ADIS 16364 фирмы Analog Device, включающий в себя трехосный датчик линейных ускорений (акселерометр) и трехосный ДУС (гироскоп). Параметры прибора ADIS 16364 были взяты из технического описания [3]. При имитации работы прибора осуществлена калибровка его шумов по результатам натуральных записей с выхода прибора, полученных на движущемся локомотиве.

Для обоснования требований, предъявляемых к системе регистрации, моделирующий комплекс был оснащен алгоритмами обнаружения АФРН и фильтрации помеховых и шумовых воздействий при оценке параметров аномалий. Моделирующий комплекс позволяет оценить требования по имитируемым и реальным потокам информации контрольно-измерительных приборов. При этом

решается задача оценки адекватности моделируемых значений реальным показаниям приборов. Предусмотрен также режим полунатурного моделирования. В этом режиме имитируемые данные используются совместно с реальной информацией: например, шумы в показаниях датчиков берутся реальные, а сигнал, формируемый в момент прохождения локомотивом участка пути с АФРН — имитируемый. Необходимость полунатурного моделирования вызвана тем, что в паспортных данных на прибор указываются лишь основные параметры датчиков, а кажущиеся разработчикам второстепенные параметры не приводятся (например, вид спектральной плотности помех). В результате сигналы датчиков, имитируемые в соответствии с паспортными параметрами, могут в определенных условиях существенно отличаться от реальных.

В основу моделирования геометрии рельсовых нитей взяты аналитические соотношения, приведенные в [4], реально используемые для описания железнодорожного полотна на местности. Наиболее важными участками пути считаются стрелочные переводы и односекционные повороты. Это связано с тем, что сигналы, возникающие в датчиках линейных ускорений и угловых скоростей при прохождении локомотивом указанных участков пути, могут быть в определенной мере похожи на сигналы при наличии АФРН на относительно прямолинейных участках пути.

Сформулируем теперь задачи, которые могут быть (и были) решены с помощью имитационного моделирования в процессе разработки системы обнаружения АФРН.

Обоснование требований к системе регистрации аномалии формы рельсовой нити. Обоснование требований проводится по составу, диапазону возможных значений, по точностным характеристикам, темпу выдачи-обновления информации, формируемой на выходе того или иного регистрирующего датчика о текущем состоянии рельсовых нитей на различных участках.

Проведение сравнительной оценки информативности датчиков параметров движения, предлагаемых различными фирмами-изготовителями. Сравнение осуществляется путем моделирования показаний различных датчиков для одних и тех же условий движения локомотива и типовых АФРН при воздействиях различного вида шумов и помех, связанных с кинематикой движения.

Сравнительный анализ алгоритмов обнаружения аномалии формы рельсовой нити. Путем имитационного моделирования можно сформировать различные условия движения локомотива, в том числе специальные условия, которые могут не встретиться в ограниченных по времени и условиям натурных испытаниях. Это позволяет всесторонне изучить сильные и слабые стороны различных подходов к выявлению АФРН.

Периодическая проверка работоспособности программно-алгоритмического обеспечения штатной системы обнаружения аномалии формы рельсовой нити. Имитационное моделирование целесообразно использовать и для создания систе-

мы тестов, с помощью которых должна производиться периодическая проверка функционирования эксплуатируемой системы регистрации текущего состояния рельсовых нитей.

Конечной задачей моделирования является оценка характеристик качества обнаружения АФРН, так как неточное вычисление параметров АФРН и их координат на рельсовых нитях можно компенсировать дальнейшим исследованием выявленного аномального участка более точными измерительными средствами и системами.

При моделировании рассчитываются показания датчиков для полностью исправного железнодорожного полотна, т. е. без аномалий, а затем в показания датчиков на участках пути, на которых имеют место АФРН, добавляются соответствующие поправки, шумы и помехи.

Анализ возможности обнаружения плавных по форме АФРН проводился путем имитационного моделирования. Вместе с тем уже имеющиеся результаты обработки реальных записей прибора обеспечили возможность задать в модели реальные параметры шумовой составляющей приборов. Вследствие этого удалось более достоверно оценить и возможность обнаружения различных АФРН, которые будут фиксироваться конкретным устройством, сконструированным на основе прибора ADIS 16364.

Дисперсия шума в процессе моделирования получена на основе реальных измерений для локомотива, движущегося со скоростью 40 км/ч; при этом для моделирования наблюдений на скорости, близкой к 60–70 км/ч, величина дисперсии была несколько увеличена, а именно была принята для датчиков линейных ускорений равной $0,05 \text{ м/с}^2$ (в состоянии покоя $0,01 \text{ м/с}^2$) и $0,015 \text{ рад/с}$ для ДУСов (в состоянии покоя $0,0144 \text{ рад/с}$, т. е. $0,83 \text{ град/с}$).

Чтобы сложность проблемы обнаружения АФРН была отчетливо видна, в модель были заложены шумовые параметры реального прибора, а также добавлен шум небольшой мощности, возникающий в процессе движения.

В качестве АФРН рассмотрено только искривление пути в плане (в горизонтальной плоскости OXY), вызывающее горизонтальное «рыскание» локомотива относительно прямолинейного движения вдоль оси OY .

Чтобы оценить принципиальную возможность обнаружения АФРН, воспользуемся критерием обнаружения аномалий на основе сравнения — в виде отношения правдоподобия, статистики, вычисляемой для получаемых фильтром Калмана оценок параметров аномалии, с пороговым уровнем, установленным по допустимой вероятности ложного обнаружения.

8 Структура алгоритма обнаружения аномалии

Так как получаемая функция правдоподобия представима нормальной плотностью распределения вероятностей, то при выборке достаточно большого объема

априорное распределение на множестве значений компонент вектора состояния будет постоянным в области значений функции правдоподобия, достаточно отличных от нуля. Это означает, что для обнаружения аномалии с требуемой информативностью — достоверностью, необходимо иметь выборку измерений, непосредственно связанную через функцию правдоподобия с фактом наличия или отсутствия аномалии в условиях текущего контроля состояния рельсовых нитей. Отсюда непосредственно следует, что факт обнаружения аномалии можно фиксировать на основе сравнения статистики в виде отношения правдоподобия, вычисляемой для получаемых фильтром Калмана оценок параметров аномалии, с пороговым уровнем, установленным по допустимой вероятности ложного обнаружения и условному распределению вероятностей статистики (при условии отсутствия аномалии).

Структура обнаружителя аномалии в реальном времени при нормальных функциях правдоподобия гипотезы $\theta = 0$ (нет АФРН) и альтернативы $\theta = 1$ (есть АФРН) в этом случае имеет классический вид:

$$\begin{aligned} \Lambda_i &= \Lambda_{i-1} \frac{\rho((z_i - Z^*)/\theta = 1)}{\rho(z_i/\theta = 0)} = \\ &= \Lambda_{i-1} \frac{1/(\sqrt{2\pi}\sigma) \exp\{-(z_i - Z^*)^2/(2\sigma^2)\}}{1/(\sqrt{2\pi}\sigma) \exp\{-z_i^2/(2\sigma^2)\}} \geq \pi(\alpha), \end{aligned}$$

где $i = 0, 1, \dots, n$ — временные отсчеты; z_i — показания ЧЭ в i -е моменты времени; Z^* — амплитуда реакции ЧЭ на АФРН; $\pi(\alpha)$ — пороговый уровень, однозначно соответствующий пороговому уровню правила обнаружения АФРН.

С учетом всей выборки в целом эта структура сводится к виду:

$$\sum_{i=0}^n z_i Z^* \geq (Z^*)^2 0,5 + \ln(\pi), \quad (19)$$

т. е. к виду корреляционно-оценочного обнаружителя детерминированного сигнала с априори не известными параметрами.

В (19) левая часть — результат раскрытия выражения для Λ_i , представляющий свертку выборки и реакции ЧЭ на АФРН, или, иначе, согласованный фильтр, а правая — пороговый уровень, однозначно соответствующий пороговому уровню правила обнаружения АФРН.

Моделирование методов обнаружения АФРН осуществлялось по следующей методике. Пусть имеет место некая локальная аномалия рельсовой колеи и на выходе датчика формируется последовательность измерений соответствующего признака. Предварительно эту последовательность подадим на вход порогового устройства. Пороговый уровень устройства установим по мощности шума прибора

системы регистрации. На выходе устройства будет иметь место случайная последовательность из нулей и единиц, например

00011111111010100000000011111110000001101111111011...

По каждой «единичной» серии принимается решение о наличии АФРН, координата АФРН определяется по медиане серии. Заметим, что серии, расположенные на расстоянии, меньшем ширины типовой АФРН, имеет смысл объединять в единую АФРН. Ширина типовой АФРН вычисляется на уровне, установленном в пороговом устройстве.

В качестве опорного полезного сигнала Z^* при моделировании была взята гауссоидальная форма аномалии при минимальной скорости движения, при которой возможно обнаруживать аномалии (40 км/ч). Выше отмечалось, что сигнал от какого-либо датчика при прохождении аномалии рельсовой колеи при разных скоростях локомотива будет одинаковым по форме, но разным по амплитуде и временной протяженности. (Это обстоятельство реализуется правилом (19).) Адаптивные фильтры (12) или (17) удаляют помехи и шум, выделяя сигнал реакции от АФРН.

В результате моделирования процесса обнаружения односторонней и двухсторонней просадки, а также горизонтального рыскания установлено, что наличие трехосного акселерометра и трехосного гироскопа позволяет надежно распознать типовую АФРН, которая обязательно проявится в виде сильного сигнала хотя бы в одном из шести каналов. Результаты фильтрации, рассмотренные выше, можно суммировать в виде табл. 1.

Очень слабый уровень реакции на АФРН в канале угловой скорости по оси OZ объясним, поскольку движение моделировалось вдоль оси OY и только рыскание (отклонение пути в плане) создало сигнал, величина которого тем меньше, чем длиннее вагонная тележка локомотива. По этой же причине имеет место слабый сигнал и в канале угловой скорости по оси OX . Интересно, что из-за относительно малых поперечных размеров вагонной тележки угол поворота локомотива на аномалии по продольной оси OY (крен) отчетливо регистрируется.

Как видно из табл. 1, от каждого типа АФРН очень сильный сигнал формируется только в одном из шести измерительных каналов, а учет информации по остальным измерительным каналам, в которых сигнал от АФРН слабый, лишь незначительно увеличивает вероятность обнаружения АФРН. Таким образом, использование шестикомпонентного измерительного устройства (трехосный акселерометр совместно с трехосным гироскопом) соответствует поставленной задаче обнаружения аномалий рельсовых нитей предложенным трехканальным обнаружителем. Использование только акселерометров или гироскопов резко снижает качество решения задачи гарантированного обнаружения АФРН.

Заметим, что моделировались АФРН, приводящие в основном к отклонениям геометрии пути преимущественно в какой-либо одной плоскости. В реальности

Таблица 1 Сравнение интенсивности обнаруженных сигналов от типовых АФРН в измерительных каналах прибора

Тип измерительного канала	Тип АФРН		
	Односторонняя просадка	Двухсторонняя просадка	Рыскание (отклонение пути в плане)
Линейное ускорение по оси OX	Очень слабый сигнал	Очень слабый сигнал	Сильный сигнал
Линейное ускорение по оси OY	От слабого до среднего сигнала	От слабого до среднего сигнала	От слабого до среднего сигнала
Линейное ускорение по оси OZ	От слабого до среднего сигнала	Сильный сигнал	Очень слабый сигнал
Угловая скорость поворота вокруг оси OX	От слабого до среднего сигнала	От слабого до среднего сигнала	Очень слабый сигнал
Угловая скорость поворота вокруг оси OY	Сильный сигнал	Очень слабый сигнал	Очень слабый сигнал
Угловая скорость поворота вокруг оси OZ	Очень слабый сигнал	Очень слабый сигнал	От слабого до среднего сигнала

АФРН могут быть сложнее по форме и их ориентация в пространстве может не совпадать с плоскостями системы координат. В этом случае сигнал от АФРН будет появляться в нескольких измерительных каналах сразу, разработанный обнаружитель можно подключить к каждому из них и осуществлять обработку информации независимо с последующим принятием решения об обнаружении АФРН по логике «хотя бы по одному каналу обнаружена АФРН». Подходы к анализу сложных аномалий и помех рассмотрены в [5, 6].

Измерительный прибор, кроме трехосных акселерометра и гироскопа, включает в себя устройство навигации на основе GPS-приемника. Это позволяет осуществить как координатную привязку обнаруженных АФРН, так и идентификацию их параметров, используя разработанные выше алгоритмы. Точность топографической привязки порядка 2–3 метров, реализуемая GPS-приемником, обеспечивает надежную локализацию АФРН и для выезда на место ремонтной бригады, и для последующего мониторинга АФРН.

9 Заключение

Предложенный в работе подход предполагает разработку специальных методов интеграции геоинформационных ресурсов (включая данные СНС), обеспечивающих поддержку технологий управления железнодорожным транспортом. Возрастающие требования к совместной обработке пространственных данных

для различных приложений выявили острую потребность в создании геоинформационного пространства. Актуализация геоинформационного пространства, поддерживаемого инструментальной средой ГИС, в современных условиях предполагает разработку и сопровождение интегрированных моделей данных. Эти модели данных должны описывать характеристики железнодорожных объектов и обеспечивать безопасность функционирования их подсистем.

Использование дифференциального описания реакции измерителей на АФРН позволило построить вектор состояния интегрированной НС и вектор его наблюдения с учетом влияния АФРН в форме, применимой для существующих методов нелинейной фильтрации, и получить соответствующие уравнения оценки параметров движения, в т. ч. реакции на АФРН.

Проведенный анализ причин и условий возникновения АФРН показал, что метод обнаружения АФРН должен работать по критерию совместного обнаружения альтернативных сложных гипотез, представляющих возможные типы АФРН на различных участках движения локомотива при аддитивном воздействии помех. Достоверность обнаружения АФРН при этом определяется вероятностью превышения корреляционным интегралом порогового уровня, устанавливаемого по допустимой вероятности ложного обнаружения АФРН из-за помех при движении локомотива. При наличии сети дифференциальных поправок и заданных параметров движения локомотива процедура оценки и идентификации реакции на АФРН может быть существенно упрощена.

Разработанные методы и алгоритмы обнаружения и оценивания параметров АФРН реализуемы программно в реальном масштабе времени и в реальных условиях движения локомотива.

Литература

1. Бабич О. А. Обработка информации в навигационных комплексах. — М.: Машиностроение, 1991. 325 с.
2. Уманский В. И. Модели синтеза высокоточных систем позиционирования локомотива для решения задач тесной и глубокой интеграции бесплатформенных инерциальных и спутниковых навигационных систем // Системы высокой доступности, 2011. Т. 7. № 3. С. 61–76.
3. Six Degrees of Freedom Inertial Sensor ADIS 16364. Analog Devices, Inc. Norwood, MA 02062-9106. USA, 2009. www.analog.com.
4. Уманский В. И. Метод идентификации аномалии рельсового пути // Вестник Тверского ун-та. Сер. Прикладная математика, 2011. № 2(21). С. 55–62.
5. Тихонов В. И., Харисов В. Н. Статистический анализ и синтез радиотехнических устройств и систем. — М.: Радио и связь, 1991. 608 с.
6. Бранец В. Н., Шмыглевский И. П. Введение в теорию бесплатформенных инерциальных навигационных систем. — М.: Наука, 1992. 280 с.

МЕТОДИЧЕСКИЕ ПОДХОДЫ К ВЫБОРУ СЕНСОРНЫХ ТЕХНОЛОГИЙ ДЛЯ СИТУАЦИОННЫХ ЦЕНТРОВ НА ОСНОВЕ КЛАССИФИКАЦИИ ВИДОВ ДЕЯТЕЛЬНОСТИ ПОЛЬЗОВАТЕЛЕЙ

С. А. Денисов¹, А. А. Зацаринный², В. А. Кондрашев³, К. Г. Чупраков⁴

Аннотация: Рассматриваются вопросы применения современных сенсорных технологий для систем ситуационных центров (СЦ) с целью повышения эффективности их функционирования на основе совершенствования интерфейса взаимодействия лица, принимающего решения (ЛПР), с информационно-аналитической системой поддержки принятия решений (ИАС ППР). Предлагаются критерии выбора сенсорных технологий для СЦ на основе проведенной классификации видов деятельности их пользователей.

Ключевые слова: ситуационный центр; сенсорная панель; интерфейс взаимодействия; информационно-аналитическая система

1 Введение

Разработка интерфейса взаимодействия человека с информационными системами — многогранная и сложная научно-техническая задача применения компьютерной техники. Прикосновение как один из естественных путей взаимодействия человека с объектами окружающего мира занимает ведущее место среди способов управления компьютерными системами. Идея взаимодействия с информационными системами посредством прикосновения к экрану появилась в 70-х гг. прошлого века, тогда же были разработаны первые сенсорные средства отображения информации. К настоящему моменту пройден почти полувековой путь развития, создано множество различных технологий фиксации прикосновения к экрану, значительно превосходящих по своим характеристикам исходные технические решения. Активному развитию сенсорных технологий способствует интеграция поддержки сенсорного управления в современные операционные системы (MS Windows, Linux, Apple iOS и Mac OS, Google Android) и разработка приложений, которые ориентированы на работу с помощью прикосновений к экрану, а

¹Институт проблем информатики Российской академии наук, sdn@ipi.ac.ru

²Институт проблем информатики Российской академии наук, azatsarinny@ipiran.ru

³Институт проблем информатики Российской академии наук, vd@ipi.ac.ru

⁴Институт проблем информатики Российской академии наук, chkos@rambler.ru

также непрерывно растущая потребность человека использовать сенсорные технологии, в том числе в своей профессиональной деятельности, особенно на этапе выработки и принятия решения.

Понятие «принятие решения» все чаще связывают с организационно-техническими структурами, называемыми ситуационными центрами.

Эффективность современных ситуационных центров значительно выше, чем у аналогичных структур, разработка которых проводилась 5–10 лет назад. Эта тенденция связана с тем, что, во-первых, ведущие научные институты страны уделяют большое внимание вопросам, связанным с функционированием СЦ, и успешно внедряют разработанные технические и организационные методики в реальные проекты [1–3]. Во-вторых, сами организации сегодня все больше заинтересованы в том, чтобы развернуть мощные средства поддержки принятия решений [4–6]. В-третьих, основу современного СЦ составляет мощная ИАС ЛПР.

Взаимодействие между ИАС ЛПР и ЛПР часто осложняется отсутствием у последнего достаточных навыков взаимодействия с предлагаемыми аппаратно-программными средствами поддержки принятия решений [6, 7]. Возникает задача создания интерфейсов, с помощью которых можно обеспечить достаточный уровень естественности взаимодействия с информационной системой.

По мнению авторов статьи, применение сенсорных технологий в СЦ для организации взаимодействия между ЛПР и ИАС ЛПР позволяет создать гибкий, наглядный, удобный интерфейс взаимодействия, который позволяет ЛПР в максимальной степени сконцентрироваться на осмыслении результатов анализа ситуации, содержания возможных вариантов ее развития и, наконец, непосредственно на принятии решения, а не на технических аспектах организации соответствующего процесса.

В данной статье сформулированы критерии применимости сенсорных технологий в СЦ на основе классификации видов деятельности их пользователей, даны рекомендации по выбору той или иной сенсорной технологии.

2 Виды деятельности современных ситуационных центров

Ситуационный центр — это сложная организационно-техническая структура, вопросам создания и функционирования которой посвящены работы отечественных авторов [2, 5, 6]. В общем виде процесс функционирования СЦ может быть описан совокупностью следующих основных видов деятельности:

- получение/сбор информации;
- анализ и структурирование полученной информации;
- моделирование и выработка решений;
- обсуждение и принятие решения.

Для выполнения вышеупомянутых видов деятельности в СЦ формируются группы пользователей (под пользователями будем понимать должностных лиц, принимающих участие в процессе подготовки и принятия решения). Пользователям выделяются ресурсы (в том числе технические) и создаются структурные единицы.

Обычно в состав СЦ входят следующие структурные единицы [8]:

- диспетчерский пункт (дежурная служба);
- служба экспертов-аналитиков (обычно дополнительно структурируется по решаемым функциональным задачам);
- руководство (структура часто совпадает со структурой службы экспертов-аналитиков);
- орган коллективного принятия решений, функционирующий в ситуационном зале (зале заседаний, совещаний и т. п.).

В диспетчерском пункте происходит сбор, первичная обработка и анализ информации, принятие оперативных решений. Диспетчерский пункт обычно функционирует в круглосуточном режиме работы, пользователями являются дежурный смены и его помощники — профессионалы высокой квалификации в предметной области СЦ, имеющие узконаправленную подготовку в области информационных технологий. Функциональные обязанности дежурной смены в части информационных задач СЦ сводятся к наблюдению за обстановкой и протоколированию значимых событий. Как правило, при выполнении своих служебных обязанностей пользователи диспетчерских пунктов руководствуются принятым алгоритмом действий, выполняя стандартный набор операций, даже в случае возникновения чрезвычайных и нештатных ситуаций.

В условиях диспетчерского пункта сенсорная технология может быть применена в первую очередь для контроля обстановки и управления отображаемым контентом, быстрой навигации между информационными окнами и выполнения стандартных операций. Учитывая методику [9], можно рассчитать, что для выполнения этих задач достаточно использование несколькими пользователями сенсорной панели среднего размера (26"–55") с совмещением сенсорной матрицы с экраном дисплея, отображающего текущую обстановку (комплекс событий) в виде иконок на топологической (картографической) основе. Прикосновение к иконкам обеспечивает управление отображаемым контентом.

На рабочих местах экспертов-аналитиков осуществляется углубленный анализ полученной информации, формируются аналитические отчеты, моделируются возможные варианты развития ситуации и, наконец, вырабатывается ряд обоснованных решений. Функции экспертов-аналитиков, как правило, выполняют высококвалифицированные сотрудники, имеющие углубленную подготовку как в предметной области, так и в области информационных технологий. Основой

деятельности экспертов-аналитиков является сложная работа с информацией, в том числе операции ввода и редактирования на топологической (картографической) основе высокого разрешения, выполнение расчетных задач. Данные операции являются фактором, который препятствует применению сенсорных технологий в работе экспертов-аналитиков, поскольку на данном этапе своего развития сенсорные технологии по скорости ввода и точности позиционирования уступают общепризнанным и повсеместно используемым устройствам ввода и редактирования информации — клавиатуре и манипулятору «мышь».

Руководство в большинстве случаев оперирует материалами, подготовленными группой экспертов-аналитиков. Руководитель принимает решение на основе анализа альтернатив либо самостоятельно, либо на основе обсуждения с группой экспертов. В условиях кабинета руководителя вариантом применения сенсорной технологии может являться решение «сенсорный стол» (панель 46"–70", учитывая методику [9]). Интерактивный сенсорный стол сочетает в себе функции обычного стола и мультимедийной презентационной системы на основе средств отображения информации, совмещенных с сенсорной матрицей. Лицам, участвующим в обсуждении, предоставляется возможность непосредственно во время дискуссий управлять диспозицией, моделировать варианты развития ситуации, вносить изменения в подготовленные варианты решений и протоколировать их, используя интуитивно понятный тактильный интерфейс взаимодействия с системой.

В ситуационном зале происходит концентрация информации, поступающей от приложений СЦ, и ее отображение на средствах отображения информации коллективного пользования (СОИКП) [10]. Аудиторию составляют лица (пользователи), принимающие участие в процессе принятия решений. В формате совещания (заседания) докладчиком (группой докладчиков) представляются донесения, выступления и прочие информационные сообщения, подготовленные группой (группами) экспертов-аналитиков. Далее пользователи участвуют в коллективном обсуждении представленной информации и принимают решения.

Применение современных сенсорных технологий в условиях ситуационного зала должно позволить проводить совещания в интерактивном режиме. Сенсорная технология при этом должна интегрироваться с коллективным экраном (компонентом СОИКП). Участники совещания должны иметь возможность управления отображением информации на СОИКП с помощью индивидуальных интерактивных терминалов участников и выносных сенсорных панелей, расположенных на трибунах докладчиков. Использование интерактивных терминалов повышает общую комфортность и результативность проведения совещания, позволяет автономно проводить совещание без участия обслуживающего персонала и самостоятельно вносить пометки в отображаемые на экране материалы.

Сенсорные средства интерактивного проведения совещания должны позволить председателю:

- наблюдать на персональном терминале список участников совещания с указанием докладчиков и их иллюстративных материалов;
- определять порядок предоставления слова выбранным участникам;
- осуществлять контроль за длительностью выступления участника и длительностью совещания;
- запрашивать у режиссера дополнительную информацию о режиме проведения совещания.

Участникам совещания должна предоставляться возможность:

- управлять показом иллюстративных материалов к докладу с расположенных перед ними интерактивных терминалов;
- выводить иллюстративные материалы на СОИКП, интерактивные терминалы остальных участников совещания и на выносные интерактивные панели;
- наносить пометки поверх изображения иллюстративных материалов.

По ряду оценок (см., например, [7]), применение сенсорной технологии в условиях ситуационного зала может увеличивать скорость усвоения данных слушателями на 60%. Такой прирост обеспечивается за счет концентрации внимания на средство отображения информации, используемое докладчиком, где непосредственно формируется аудиовизуальная динамика доклада.

3 Критерии использования и обзор сенсорных технологий

Решение об использовании сенсорных технологий для СЦ должно быть обоснованным с точки зрения функциональных преимуществ, получаемых за счет той или иной сенсорной технологии.

Сенсорная технология представляется совокупностью трех основных логических компонентов: экрана, сенсорной матрицы и вычислительного ресурса (рис. 1).

Первый и третий компоненты обычно входят в состав средства отображения информации, или дисплея, а второй может быть реализован как на базе технологии самого экрана, так и с использованием внешних для дисплея подсистем. Экран дисплея с точки зрения сенсорной технологии представляет собой плоский источник визуальной информации, подлежащей управлению, с ограниченной разрешающей способностью экрана. Сенсорная матрица — физическая среда, реагирующая на прикосновение некоторым предметом к поверхности экрана. Прикосновение преобразуется в электронный сигнал. Затем набор информации о положении прикосновения, его размерах и, возможно, его силе передается в вычислительный блок. Задачей вычислительного блока в самом общем виде является сопоставление информации, отображаемой на дисплее, и сигналов, полученных от прикосновений к сенсорной матрице, в условиях реального времени.

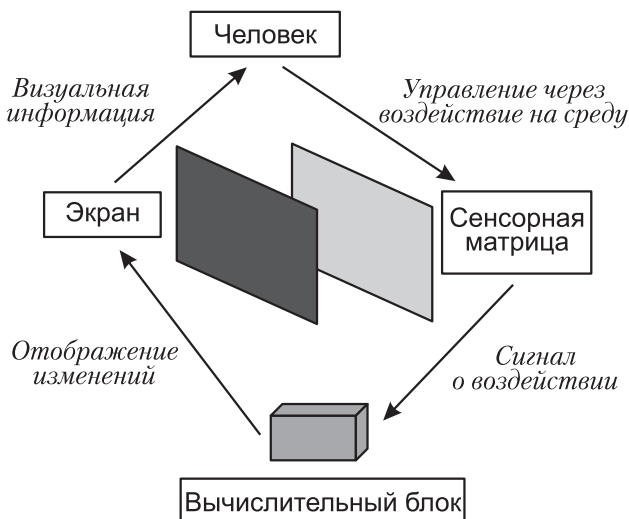


Рис. 1 Логическая схема сенсорной технологии

Учитывая специфику видов деятельности различных групп пользователей в СЦ, можно выделить следующие критерии выбора сенсорных технологий:

- размер экрана (диагональ экрана);
- разрешающая способность экрана;
- разрешающая способность сенсорной матрицы (количество одновременных прикосновений);
- совмещенность экрана и сенсорной матрицы;
- возможность распознавания мультитжестов для управления системой;
- надежность сенсорной матрицы;
- стоимость сенсорной технологии.

Размер сенсорной матрицы определяется ее геометрическими характеристиками — чаще всего длиной и шириной. В случае использования сенсорной технологии, совмещенной с экраном, размеры сенсорной матрицы совпадают с размерами дисплея и имеют прямоугольную форму. В этом случае для определения размера экрана можно использовать методику [9], предусматривающую возможность комфортного наблюдения за информацией на средстве отображения.

Разрешение сенсорной матрицы — количество различных одинарных прикосновений, распознаваемых сенсорной матрицей. Необходимое разрешение экрана зависит от размеров сенсорной матрицы и площади устройства, которым производится прикосновение. Очевидно, что разрешение сенсорной матрицы определяет

минимально осязаемое смещение прикосновения, т. е. чем выше разрешение, тем с большей динамической чувствительностью может работать сенсорный дисплей. С другой стороны, разумно ограничивать разрешение сенсорной матрицы разрешением самого экрана. Для диспетчерского пункта возможно использование сенсорной матрицы с невысоким разрешением, поскольку пользователи в основном визуально контролируют поступающую информацию. Для кабинета руководителя и ситуационного зала рекомендуется использовать сенсорные матрицы с высоким разрешением, поскольку во время обсуждения решений ведется активная работа с совмещенными сенсорными средствами отображения информации.

Совмещенность экрана и сенсорной матрицы. Современные решения позволяют не совмещать сенсорную матрицу и экран. Вынесенная сенсорная матрица может быть реализована с помощью дополнительного экрана, например меньшего размера, который показывает изображение с основного экрана либо его части. Такие решения оказываются удобными в случае, когда основной экран имеет достаточно большие размеры, недоступен для работы непосредственно с ним либо ввиду организационных ограничений (например, при ведении заседаний).

Количество одновременных прикосновений к экрану определяет функциональные возможности работы с контентом, а также способность сенсорного дисплея быть использованным в качестве коллективного тактильного поля, где часто требуется одновременная работа нескольких человек. Хотя такая возможность выглядит достаточно естественно, но ее техническая реализация часто затруднена и в некоторых сенсорных технологиях вообще невозможна.

Стоимость сенсорных решений определяется их рыночным позиционированием и себестоимостью. Стоимость может служить некоторой интегральной оценкой для всех вышеперечисленных параметров, так как эти параметры в серьезной степени влияют на доступность решений.

3.1 Инфракрасная технология

Первой и одной из наиболее простых является инфракрасная технология, которая использует прямоугольные размеры экрана. Ее суть заключается в том, что по горизонтали и вертикали экрана расположены два ряда инфракрасных диодов, а на противоположных им сторонах находятся улавливающие инфракрасный свет сенсоры.

Прикосновение к поверхности экрана сопряжено с тем, что некоторые сенсоры по горизонтали и вертикали перестают получать излучение, что определяется вычислительным блоком как пара координат, где произошло прикосновение к поверхности экрана (рис. 2).

Данная технология была реализована одной из первых в рамках сенсорных экранов еще в 70-х гг. XX в., но ее разрешающая способность, т. е. количество

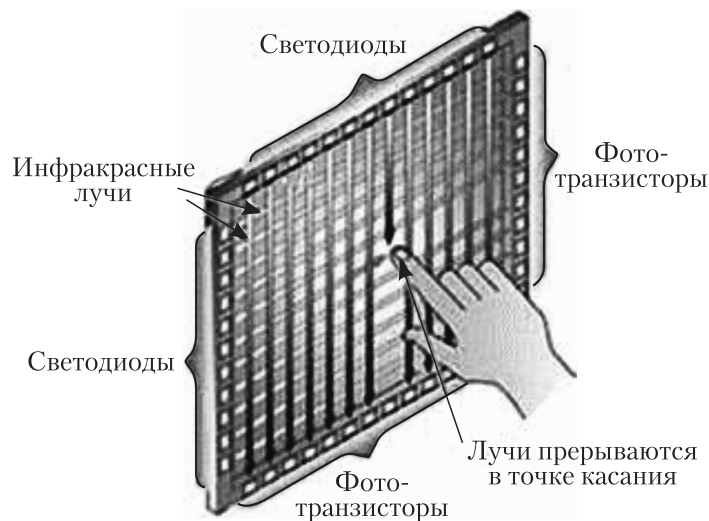


Рис. 2 Инфракрасная сенсорная технология

различных прикосновений, с точки зрения сенсора была очень мала и исчислялась всего парой десятков. Современные инфракрасные экраны могут определить уже более миллиона различных прикосновений к экрану.

Важной особенностью этих сенсорных матриц является отделимость инфракрасных сенсоров от экрана и, как следствие, их высокая надежность и ремонтпригодность. Инфракрасные сенсоры не влияют на качество изображения экрана, их можно чинить в «полевых» условиях. Более того, сенсорный экран может быть создан на базе отдельной сенсорной рамки.

Основным недостатком инфракрасной технологии является ее невысокая разрешающая способность из-за ограничений по расстоянию между излучающими диодами, что сдерживает ее применение на компактных устройствах. Также из-за волновых свойств света существуют ограничения, связанные с размером предмета, которым прикасаются к экрану, — его толщина должна быть не меньше определенного значения в зависимости от размеров экрана и разрешающей способности сенсорной матрицы.

Основной областью применения инфракрасной технологии сегодня являются СОИКП, такие как рекламные киоски и информационные панели.

Сенсорные панели с данной технологией можно использовать в диспетчерской службе СЦ. Использование ее в деятельности других групп пользователей СЦ связано с существенным ограничением функциональных задач, решаемых этими пользователями.

3.2 Матричная технология

Матричная технология по своей сути напоминает инфракрасную, но вместо оптических сенсоров в ней используется сетка из горизонтальных и вертикальных проводников с двух сторон прослойки, которые при прикосновении в некоторой точке замыкаются, тем самым фиксируя координаты воздействия.

Технология хороша тем, что предельно проста и использует доступные компоненты. Но, с другой стороны, технология может обеспечить еще меньшую, чем инфракрасная, разрешающую способность — параллельные проводники нельзя размещать слишком близко друг к другу из-за возможной электрической разрядки. Ввиду постоянного снижения цены на более совершенные электронные компоненты матричная технология уже практически не используется, а вместо нее применяется резистивная.

3.3 Резистивная технология

Намного более высокой разрешающей способностью, в отличие от матричной и инфракрасной технологии, обладает резистивная. В ее основе содержатся два проводящих слоя, нанесенных на стекло и на гибкую внешнюю мембрану. Пространство между проводящими слоями заполнено микроизоляторами, которые при нажатии на гибкую мембрану позволяют проводящим слоям замыкаться. Электроды подключены к мембране и стеклу по перпендикулярным парам сторон (рис. 3).

При нажатии на экран последовательно определяются координаты точки нажатия — X на мембране и Y на стекле — с помощью подачи напряжения и оценки сопротивления, возникающего при нажатии. Обычный резистивный экран может при этом определить несколько десятков миллионов различных прикосновений. Износоустойчивость резистивной сенсорной технологии измеряется 30 млн нажатий в одну точку.

Главные недостатки резистивной сенсорной технологии определяются наличием гибкой мембраны — значительное изменение качества изображения экрана (потеря в яркости и в цветопередаче), а также потеря сенсорных свойств при повреждении мембраны, которое можно нанести с помощью любого достаточно острого предмета. Проблему низкой устойчивости к механическим повреждениям удалось частично решить с помощью совершенствования технологии до пятипроводной схемы, где электроды подключены к углам экрана, резистивный слой мембраны заменен на проводящий, а на стекло нанесен резистивный слой. Порез мембраны в таком случае позволяет дальше пользоваться сенсорными способностями экрана.

Ввиду своей невысокой стоимости, высокой универсальности и точности резистивная сенсорная технология была одной из самых массовых еще 2–3 года назад. Основной областью применения резистивной технологии были сотовые телефо-

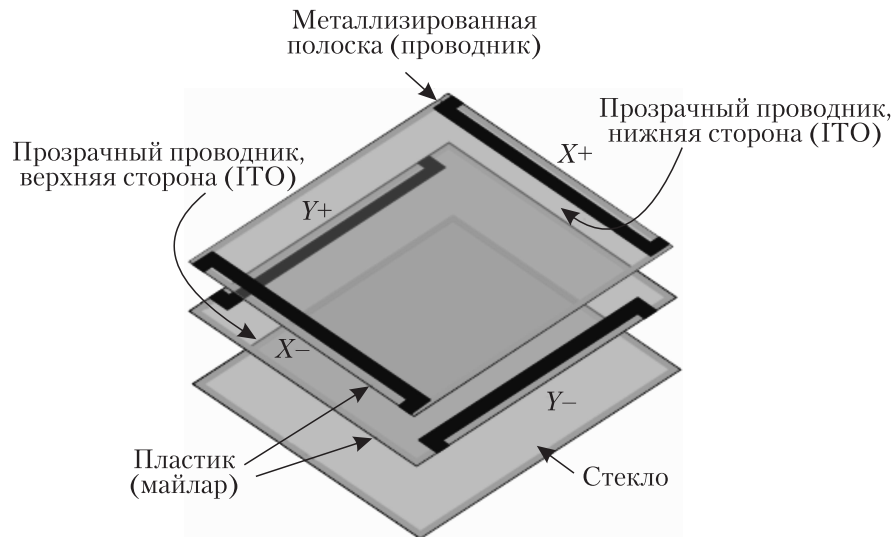


Рис. 3 Принцип работы резистивных сенсорных экранов (ITO — indium tin oxide, оксид индия и олова)

ны, коммуникаторы, POS (point-of-sale) терминалы. Сегодня она замещается более совершенной емкостной технологией.

Тем не менее в СЦ резистивная технология с несущественными ограничениями в решаемых функциональных задачах (по сравнению с емкостной технологией) может применяться в диспетчерском пункте и в зале совещаний (на индивидуальных интерактивных терминалах участников совещания).

3.4 Емкостная и проекционно-емкостная технологии

Принцип работы емкостной сенсорной технологии основан на использовании физиологических свойств человеческого тела — наличия емкостного сопротивления. Экран в емкостной технологии представляет собой стекло с нанесенным на него проводящим слоем, по углам которого находятся электроды. При прикосновении к экрану пальцем происходит утечка тока, которая регистрируется вычислительным блоком, и на основании этих данных осуществляется определение позиции прикосновения к экрану.

Такая конструкция обеспечивает низкую оптическую коррекцию изображения экрана (менее 10%), имеет высокую износоустойчивость, допуская сотни прикосновений в одной точке, а также нечувствительна к изолирующим загрязнениям.

Технология определяет ряд ограничений. Прикасаться к экрану можно только проводящим предметом с емкостным сопротивлением (экран не определит

прикосновение руки, например, в обычной перчатке). Кроме того, проводящий слой, несмотря на свою достаточно высокую прочность, тем не менее может быть поврежден, потому технология предполагает некоторую степень аккуратности со стороны пользователя.

Увеличить степень механической прочности помогло создание проекционно-емкостной технологии, которая, как и емкостная, использует в своей основе стеклянную плоскость, на которую нанесен проводящий слой. Однако, в отличие от обычной емкостной технологии, проводящий слой, состоящий из электродов, дополнительно спрятан за еще одной стеклянной прослойкой, толщина которой может достигать 18 мм.

Прикосновение к стеклу создает набор сопротивлений между точкой прикосновения и ближайшими электродами. Вычислительный блок при подаче напряжения на электроды определяет наибольшее сопротивление и тем самым фиксирует точку прикосновения (рис. 4).

Ключевым преимуществом проекционно-емкостной технологии является возможность реализации функции мультитач, которая в значительной степени увеличивает возможности управления контентом. Существуют также разновидности экранов на проекционно-емкостной технологии, которые могут работать при прикосновении в перчатке и даже определять жесты. Возможности определяются толщиной защитного стекла, числом электродов на проводящем покрытии и ресурсами вычислительного блока.

Основной областью применения емкостной технологии являются платежные и информационные терминалы, банкоматы, устройство «тачпад» ноутбуков, телефонов и планшетов.

В СЦ в ближайшем будущем (с учетом массовости и удешевления) проекционно-емкостная технология сможет применяться для всех групп пользователей.

3.5 Акустические сенсорные технологии — SAW и DST

Применение звуковых сенсоров для реализации сенсорных экранов было впервые произведено в технологии поверхностно-акустических волн (SAW, Surface Acoustic Waves), идея которой является симбиозом инфракрасных и емкостных экранов. Основу такого экрана составляет стеклянная пластина, на углах которой находятся пьезоэлектрические преобразователи, которые создают

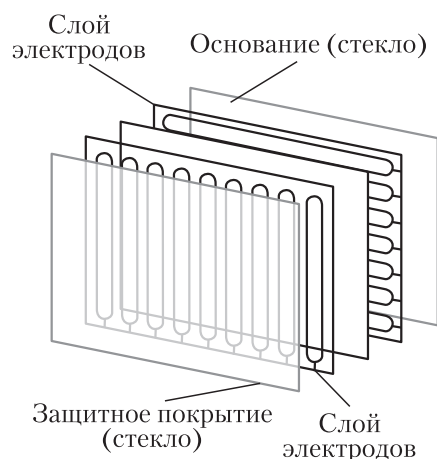


Рис. 4 Принцип работы проекционно-емкостного сенсорного экрана

поток акустических волн на поверхности. По сторонам экрана находятся акустические отражатели. Акустические волны, в случае если нет прикосновения, возвращаются обратно в пьезоэлектрические преобразователи без изменения, что фиксируется вычислительным блоком. При прикосновении к экрану возникает локальное поглощение акустических волн, обеспечивающее информационный сигнал для вычислительного блока.

Важной особенностью таких экранов является возможность фиксировать не только положение прикосновения, но и его силу за счет изменения интенсивности поглощения волн в точке.

Несмотря на относительную сложность, технология поверхностно-акустических волн показала высокую надежность и прочность, обладая при этом невысокой стоимостью. Недостатком технологии является высокая чувствительность к различным вибрациям и электроакустическим шумам, что резко ограничивает ее практическое применение. Так, в частности, технология не может использоваться в портативных устройствах и вычислительных центрах. В настоящее время она применяется в информационных киосках и рекламных дисплеях с диагональю экрана не менее 15".

Более совершенной, по сравнению с предыдущей технологией, оказалась технология DST (Dispersive Signal Technology), разработанная компанией ЗМ. Она определяет прикосновение благодаря непосредственно механическому воздействию на рабочую среду — стеклянную пластину — и последствиям такого воздействия.

Стеклянная пластина в этом случае абсолютно прозрачна в рабочей области экрана, а по ее углам расположены пьезоэлектрические преобразователи, принимающие звуковые колебания, возникающие после прикосновения и затухающие в течение некоторого времени. Для определения прикосновения используется модель распространения акустических волн в ограниченной однородной среде (стекле), которая характеризуется нелинейностью зависимости сигналов, получаемых датчиками, от координат прикосновения. Данная технология промышленным образом реализована одной из последних и требует высокой вычислительной мощности.

Технология DST менее чувствительна к акустическим шумам и вибрациям, а также позволяет определять положение прикосновения с ощутимо более высокой точностью.

В СЦ акустические сенсорные технологии могут применяться для всех групп пользователей. Однако с учетом стоимости акустических сенсорных панелей первоочередным применением этой технологии представляется использование ее для «сенсорного стола» руководства. Возможность фиксирования силы нажатия позволяет более четко акцентировать внимание на вносимые правки лицами, принимающими участие в обсуждении (например, увеличивая толщину проводимых линий).

3.6 «Перьевые» технологии

Ряд технологий основан на прикосновении предметами с определенными свойствами. К их числу принадлежат ультразвуковая и индукционная технологии. Принцип действия первой заключен в создании приемников акустических ультразвуковых волн по углам экрана и использовании специального пера, излучающего ультразвук. По характеру улавливаемого сигнала определяется местоположение пера. Особенностью такой технологии является энергетическая активность пера, а значит, и необходимость использования батареек.

Индукционная технология основана на явлении возникновения тока в проводнике при движении в магнитном поле экрана. Такая технология позволяет создавать сенсорные экраны очень высокой точности, но взамен требует периодической калибровки и обладает высокой себестоимостью. Поэтому данная технология применяется только в дорогих решениях.

Плюс перьевых технологий заключается в том, что, в отличие от пальца, перо может содержать в себе несколько датчиков, фиксирующих, например, ускорения и силу нажатия, что в некоторой степени может упрощать интерфейсы управления.

В СЦ перьевая технология может применяться в ситуационном зале (на индивидуальных интерактивных терминалах участников совещания и председателя).

3.7 Интегрированные сенсоры

В число новых разработок сенсорных жидкокристаллических (ЖК) экранов входят такие, где сенсорные способности «защиты» непосредственно в ЖК-матрицу. Такие решения представлены компаниями Philips и Planar Systems.

Идея, использованная компанией Philips, достаточно проста — в каждый ЖК-пиксел встроен специальный механический ключ-замыкатель. При прикосновении происходит деформация верхней стеклянной пластины, после чего в пикселах образуются подвижные микроконтакты, которые в зоне прикосновения замыкаются. Соответственно, вычислительному блоку остается лишь определить пикселы, где это прикосновение произошло. Важно, что в данном случае нет необходимости встраивать какие-либо дополнительные сенсоры и внешние блоки электроники.

Структура ЖК-матрицы в таком решении остается неизменной, но можно выделить два основных способа создания микроконтактов в пикселах, замыкаемых механическим путем. В первом случае микроконтакты создаются с помощью миниатюрных спейсеров — шариков с проводящим покрытием, диаметр которых меньше, чем аналогичных шариков-спейсеров, которые создают промежуток между покрытиями. Во втором случае подвижный контакт образуется методами фотолитографии и выборочного травления.

Идея, реализованная компанией Planar System, заключается в интегрировании в конструкцию TFT-дисплея специальной матрицы фотодатчиков, которая действует согласованно с TFT-матрицей. Разрешающая способность такой сенсорной матрицы может быть меньше, чем число пикселей, так как фоточувствительные сенсоры могут быть расположены не в каждом пикселе, что обеспечивает снижение стоимости такого сенсорного экрана. Данная технология, несмотря на оптическую природу, способна не реагировать на шумы типа тени, градиента освещенности, а также и на разные условия основного уровня яркости.

Распознавание осуществляется в два этапа — на первом происходит формирование образа прикосновения за счет получения информации с фотодатчиков, а на втором этапе происходит оцифровка полученных данных, определение центра масс, который и становится точкой прикосновения. При таком подходе учитывается и изменение размера и положения теней, что помогает в частности определять наличие «клика». Такая обработка данных осуществляется отдельно по горизонтальной и вертикальной оси, что позволяет уменьшить объем вычислений.

Практические достоинства и недостатки интегрированных технологий пока трудно оценить, поскольку их еще нет в массовом производстве.

3.8 Технологии трехмерных сенсоров

Все вышеперечисленные технологии за редким и весьма условным исключением имеют двумерный характер сенсорной матрицы, т. е. способны идентифицировать положение прикосновения только в рамках плоскости или поверхности второго порядка. Ясно, что такое ограничение для сенсорных технологий естественно лишь в том случае, если идет управление двумерным изображением. Однако технологии отображения информации уже предлагают массовые решения для показа трехмерного контента. Для сенсорных технологий, соответственно, формируется задача этим контентом управлять. Для решения этой задачи, вообще говоря, необходимо формировать тройку, определяющую координаты в трехмерном пространстве, а также характеризовать изменение этой тройки через первую и вторую производную (скорость и ускорение движения). Решение такой задачи может быть произведено с помощью любой из представленных выше идей — оптической, акустической или электрической — с добавлением еще одной размерности сенсоров. Но стоит понимать, что основное ограничение возникает со стороны самих средств отображения — они на самом деле лишь псевдотрехмерные, так как создают воспринимаемое человеком трехмерное изображение на двумерной поверхности.

Тем не менее уже сегодня распространены инструменты управления изображением, имеющие встроенные гироскопы, которые позволяют измерять изменения пространственных координат.

Применение технологии трехмерных сенсоров целесообразно предусматривать в концептуальных описаниях СЦ следующего поколения.

4 Заключение

К настоящему времени сформированы две предпосылки для широкого применения сенсорных технологий в СЦ.

Первая состоит в массовом использовании сенсорных технологий в различных сферах человеческой деятельности, что существенно удешевляет применение этой технологии в системах специального назначения, в том числе в СЦ.

С другой стороны, успешное решение важных задач предметной области в рамках деятельности СЦ способствует увеличению финансовых вложений, направленных на применение современных технологий (в том числе сенсорных панелей) при создании или модернизации СЦ с целью повышения эффективности их функционирования.

Рассмотренные в статье возможности современных сенсорных технологий и критерии их выбора для видов деятельности СЦ согласно проведенной классификации являются основой для широкого применения этих технологий с целью повышения эффективности функционирования СЦ путем совершенствования интерфейса взаимодействия ЛПР с ИАС ППР.

Литература

1. *Зацаринный А. А., Сучков А. В., Босов А. В.* Ситуационные центры в современных информационно-телекоммуникационных системах специального назначения // ВКСС Connect! (Ведомственные корпоративные сети и системы), 2007. № 5(44). С. 64–76.
2. *Зацаринный А. А., Сучков А. В., Босов А. В.* Некоторые общие подходы к формированию функциональных требований к ситуационным центрам и их реализации // Системы и средства информатики. — М.: Наука, 2010. Вып. 20.
3. *Зацаринный А. А., Ионенков Ю. С., Козлов С. В.* Некоторые вопросы проектирования информационно-телекоммуникационных систем. — М.: ИПИ РАН, 2010. 218 с.
4. *Зацаринный А. А.* Особенности системного подхода к проектированию, внедрению и развитию современных корпоративных телекоммуникационных сетей // ВКСС Connect! (Ведомственные корпоративные сети и системы), 2002. № 3. С. 110–116.
5. *Ильин Н. И., Демидов Н. Н.* Основные направления развития ситуационных центров органов государственной власти // ВКСС Connect! (Ведомственные корпоративные сети и системы), 2007. № 6(45). С. 2–9.
6. *Зацаринный А. А., Козлов С. В., Сучков А. П.* Некоторые организационные вопросы создания ситуационных центров специального назначения // Информационные технологии управления информационными ресурсами двойного применения (IV). — М.: ИПИ РАН, 2008. С. 49–58.
7. *Крапивенко А. В.* Технологии мультимедиа и восприятие ощущений. — М.: Бином. Лаборатория знаний, 2009.

8. *Зацаринный А. А., Козлов С. В., Татаринцев Р. В.* Методические подходы к информационному обследованию объектов автоматизации федеральных органов исполнительной власти // Системы и средства информатики, 2011. Вып. 21. № 2. С. 34–50.
9. *Чупраков К. Г.* Исследование и разработка методов построения систем отображения информации для ситуационного центра: Дис. . . . канд. техн. наук. — М.: ИПИ РАН, 2010. 214 с.
10. *Зацаринный А. А., Чупраков К. Г.* Некоторые аспекты выбора технологии для построения систем отображения информации ситуационного центра // Информатика и её применения, 2010. Т. 4. Вып. 3. С. 62–71.

МАТЕМАТИЧЕСКИЕ И ПРОГРАММНЫЕ СРЕДСТВА МОДЕЛИРОВАНИЯ СВЕТОВЫХ ПРИБОРОВ

И. И. Байнева¹, В. В. Байнев²

Аннотация: Рассмотрены особенности световых приборов (СП), светоперераспределяющих устройств, проанализированы задачи и возможные решения их расчета и проектирования. Описаны математические и программные модели, используемые в светотехническом проектировании и расчетах светильников с зеркальными отражателями. Описаны принципы работы с разработанным программным обеспечением для автоматизированной системы моделирования и проектирования световых приборов.

Ключевые слова: моделирование; программа; система; световой прибор; отражатель; проектирование

1 Введение

Световой прибор состоит из множества элементов, таких, как источник(и) света (ИС), оптическая система, пускорегулирующая аппаратура, электроустановочные устройства, конструктивные узлы, параметры которых существенно зависят от особенностей конструкции СП и заметно влияют на его характеристики в целом.

При проектировании и конструировании СП необходимо принимать во внимание особенности этих светотехнических изделий: массовый характер производства и применения СП, материальные и трудовые затраты на их изготовление, монтаж и эксплуатацию, влияние СП на технико-экономические и эстетические характеристики строящихся и реконструируемых объектов. Процесс конструирования и проектирования оказывает решающее влияние на потребительские свойства СП, их надежность, а также экономичность производства и эксплуатации.

С другой стороны, разработчики СП в первую очередь стремятся создавать светильники с максимальной световой отдачей, заданной кривой светораспределения, увеличивая при этом модельный ряд СП под различные задачи клиентов.

Широкое использование ЭВМ и автоматизированных систем проектирования на всех стадиях разработки СП необходимо, чтобы избавить инженеров от

¹Мордовский государственный университет имени Н. П. Огарева, baynevaii@rambler.ru

²Мордовский государственный университет имени Н. П. Огарева, bw14@mail.ru

выполнения трудоемких расчетов, многовариантного анализа и большого объема графических работ.

Несмотря на многолетнюю историю развития и существования СП, а также массовый характер их производства, до сих пор не хватает широко используемых общедоступных программных средств для расчета и моделирования оптических систем. В настоящее время в светотехническом проектировании и производстве используются следующие наиболее известные программные продукты.

Программа LightTools [1] по проектированию, расчету и анализу систем освещения позволяет моделировать различные по сложности системы, визуализировать результаты, производить автоматическую оптимизацию; в ней также предусмотрена интеграция с SolidWorks [2].

SolidWorks — система автоматизированного проектирования, инженерного анализа и подготовки производства изделий любой сложности и назначения. Она представляет собой инструментальную среду, предназначенную для автоматизации проектирования сложных изделий в машиностроении и других областях промышленности. SolidWorks является системой гибридного (твердотельного и поверхностного) параметрического моделирования и предназначена для проектирования деталей и сборок в трехмерном пространстве, а также для оформления конструкторской документации.

Программа оптико-механического моделирования TracePro [3] позволяет создавать трехмерные модели оптико-механических систем и проводить анализ распределения освещенности в системе, учитывая рассеяние, отражение, поглощение и дифракцию света.

AutoCAD — система автоматизированного проектирования и черчения [4], используемая для оформления графической документации в электронном виде.

КОМПАС-3D — система трехмерного моделирования, которая позволяет создавать трехмерные модели деталей и сборок, чертежи и спецификации различной степени сложности [5, 6].

DIALux — профессиональный пакет для работы с освещением [7], производит расчет таких световых характеристик, как яркость, освещенность, показатели блескости, коэффициент естественного освещения и пр., позволяет рассчитывать дневной свет и тени при планировании освещения, принимая во внимание географическое расположение здания, погодные условия и тени от окружающих строений и прочих объектов. DIALux также осуществляет моделирование различных световых сцен в помещениях.

Программный продукт Light-in-Night Road предназначен для расчета уличного освещения, проектирования наружного освещения внегородских автомагистралей, улично-дорожной сети городов и сельских населенных пунктов и связанной с ней инфраструктуры [8]. Программа позволяет проводить расчеты освещения для трех классов объектов: прямолинейных участков дорог и улиц, типовых участков дорог и участков произвольной конфигурации. Она рассчитывает

распределение освещенности и яркости дорожного покрытия, нормируемые значения среднего уровня и равномерности этих характеристик, регламентируемые показатели ослепленности водителя световыми приборами и пр.

Как показывает анализ вышеупомянутых программных средств, их задача заключается в основном в разработке конструкторской документации и проектировании освещения. Конкретным расчетом и моделированием светоперераспределяющих устройств они не занимаются.

Кроме того, стоимость таких программ может составлять тысячи долларов, что не всегда доступно светотехническим фирмам.

2 Особенности оптических систем световых приборов

Основную роль в перераспределении и преобразовании света ИС играет оптическая система СП, состоящая из элементов, участвующих в указанной трансформации света. К оптическим элементам относятся: отражатели, преломлятели, рассеиватели, защитные стекла, экранирующие решетки и кольца, светофильтры, экраны, различные световоды [9].

В основных видах зеркальных отражателей используются оптические свойства параболических, эллиптических, гиперболических, цилиндрических и плоских зеркальных поверхностей, а также комбинированных поверхностей. Все указанные зеркальные поверхности могут быть определены как частные случаи поверхностей второго порядка.

Светильники отличаются большим многообразием форм отражателей, которое обусловлено разными требуемыми кривыми силы света (КСС), характеризующими светораспределение светильника, т. е. зависимость значения силы света от направления, и другими (в частности, эстетическими) соображениями. При

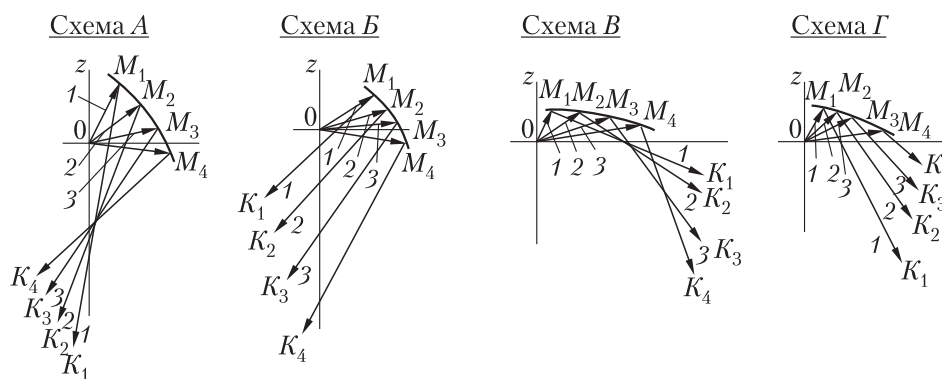


Рис. 1 Схемы хода падающих и отраженных осевых лучей

этом ход падающих и отраженных лучей можно свести к четырем принципиально отличным схемам [10], определяемым характером пересечения отраженных лучей между собой и с оптической осью светильника (рис. 1).

В схеме А отраженные лучи пересекаются и между собой, и с оптической осью; в схеме Б пересекаются с оптической осью, но не пересекаются между собой; в схеме В пересекаются между собой, но не пересекаются с оптической осью; в схеме Г не пересекаются ни между собой, ни с оптической осью.

В реальных светильниках применяются и комбинированные системы. Кроме того, довольно широкое применение (особенно с линейными ИС) нашли параболические, гиперболические, цилиндрические, гиперболические и другие типы отражателей.

3 Математические модели в светотехническом проектировании и расчетах светильников с зеркальными отражателями

Светильники с зеркальными отражателями по своим светотехническим возможностям превосходят другие приборы этого класса, являются более универсальными. Они обладают наибольшим коэффициентом полезного действия, достаточно технологичны. С помощью зеркальных светильников можно создавать разные КСС.

Задачей светотехнического расчета СП, в том числе и светильников, является определение геометрической формы элементов оптической системы СП, обеспечивающей (при совместном действии с выбранным источником света) требуемое светораспределение [10]. При расчете должны быть учтены особенности как светотехнических материалов, используемых для изготовления элементов оптической системы, так и технологических процессов их производства, а также условия работы этих элементов с точки зрения воздействия на них теплового режима СП и окружающей среды.

С помощью отражателя зеркальный светильник должен обеспечивать заданную форму КСС, исходя из определенной формы КСС источника света.

Для расчета формы зеркального отражателя необходимо иметь такое уравнение, которое давало бы зависимость величины радиус-вектора r некоторой точки M от углов φ и ψ , определяющих падающий луч, и углов α и β , определяющих отраженный луч (рис. 2).

Дифференциальное уравнение зеркального отражателя связывает воедино все его основные параметры ($r, \varphi, \psi, \alpha, \beta, p$) [10].

Дифференциальное уравнение профиля отражателя в общем виде:

$$\frac{dr}{r} = \operatorname{tg} i d\varphi,$$

где $i = (\varphi - \alpha)/2$ — угол падения луча в точку M отражателя.

В случае зеркальной поверхности с тороидными зонами профильная кривая составляется из зон постоянной кривизны ($1/R$), имеющих общие нормали в граничных точках, т.е. состоит из отрезков сопряженных дуг, имеющих радиус кривизны R и координаты центра кривизны ($X_{\text{ц}}, Z_{\text{ц}}$) как функции ($r_i, r_{i-1}, \varphi_i, \varphi_{i-1}, \delta_i, \delta_{i-1}$). Эти дуги определяются следующими уравнениями:

$$R = \frac{r_{i-1} \cos \varphi_{i-1} - r_i \cos \varphi_i}{\cos \delta_{i-1} - \cos \delta_i};$$

$$X_{\text{ц}} = X_{i-1} - R \sin \delta_{i-1};$$

$$Z_{\text{ц}} = Z_i - R \cos \delta_i.$$

Иногда лучше образовывать зеркальные поверхности вращением профильных кривых с плавно изменяющейся кривизной, подбирая при этом их параметры таким образом, чтобы они соответствовали необходимой функции $\alpha(\varphi)$. В качестве профильных кривых могут быть применены кривые 2-го порядка — конические сечения, т.е. эллипс, гипербола, парабола. Такие кривые характеризуются фокусом, в который помещают световой центр, эксцентриситетом, углом между осями отражателя.

Если поверхность зеркального круглосимметричного отражателя набрана из параболоидных поясов, то такая кривая не является гладкой, имеет точки излома, при переходе через которые угол α меняется скачкообразно. Чтобы избежать этого, нужно соединять точки границ зон лекальной кривой, что приводит к введению дополнительных уравнений. Для такой профильной кривой

$$r_i = r_{i-1} \frac{1 + \cos(\varphi_{i-1} - \alpha)}{1 + \cos(\varphi_i - \alpha)}.$$

Расчет зеркальных светильников с источниками света с большими светящими телами (люминесцентные лампы, дуговые ртутные лампы и др.) имеет свои сложности и особенности.

Цилиндрические зеркальные отражатели для люминесцентных и других трубчатых ИС характеризуются кривыми силы света в двух плоскостях: в продольной плоскости (параллельной оси лампы) и в профильной (перпендикулярной продольной и проходящей через центр светильника). Обычно светораспределение зеркального цилиндрического светильника задается в виде КСС в профильной

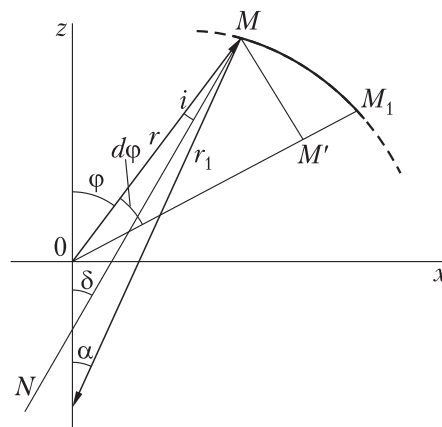


Рис. 2 Участок отражателя с падающим, отраженным и вспомогательными лучами

плоскости. Профильная кривая цилиндрического зеркального отражателя рассчитывается путем заполнения ее зональными кривыми. Зеркальная поверхность для зеркального люминесцентного светильника выбирается только гладкой и имеет параболический профиль. В этом случае возникает вопрос о профиле отражателя, от которого прежде всего зависит коэффициент усиления (отношение силы света СП к силе света ИС), связанный с углом охвата отражателя.

Часто профиль отражателя подбирается под требуемую КСС экспериментально. Поскольку он зависит от нескольких параметров, а изготовление отражателей разных форм и размеров затруднительно, то целесообразно моделирование отражателя. Этому служит математическое моделирование, выполняемое с помощью ЭВМ.

4 Среда программирования

При выборе среды основной акцент был сделан не на какие-либо математические и алгоритмические возможности языка, а на удобства объектно-ориентированного программирования. В качестве такой среды была выбрана CodeGear RAD Studio 2009 (язык Object Pascal), которая является одной из наиболее продуктивных для быстрой разработки приложений и сочетает высокопроизводительный компилятор, визуальные механизмы двунаправленного проектирования и методику масштабируемого доступа к базам данных [11, 12].

Подобными возможностями обладают такие среды программирования, как Java Eclipse, Microsoft Visual Studio. Однако значительным недостатком языка Java при разработке больших проектов является такой принцип распределения динамической памяти, как «сборщик мусора», что, как правило, приводит к излишней загрузке системной памяти. В связи с тем, что большую часть кода занимает реализация математических моделей, отпадает необходимость в низкоуровневых возможностях языка C++.

С другой стороны, обладая несомненными преимуществами в разработке визуальной и алгоритмической частей приложения, язык среды CodeGear RAD Studio не позволяет указывать принцип наследования классов так, как это реализовано в C++. Это было бы крайне удобно при проектировании классов с большим числом «предков».

5 Программные методы реализации задачи

Принцип построения автоматизированной системы моделирования и проектирования световых приборов (далее — система) имеет модульную структуру, где каждый из программных блоков (модулей) независимо от остальных должен решать свою детерминированную задачу.

Математические аспекты системы включают в себя математические модели проектируемых СП, методы и алгоритмы расчетов, используемых при автоматизированном проектировании.

Для моделирования отражения луча от элементарной части отражателя используются стандартные алгебраические и геометрические приемы расчета, которые занимают незначительную долю процессорного времени. При сложном отражателе, состоящем из большого числа элементарных частей, возникает необходимость быстрого нахождения конкретной части по заданным начальным координатам хода луча.

Простейшим вариантом решения этой задачи является линейный перебор всех частей, который имеет сложность $O(N)$, что существенно снижает общую скорость вычислений. В связи с этим была разработана система параллельной индексации с помощью двух самобалансирующихся красно-черных деревьев, у которых максимальная сложность поиска $O(\ln N)$. Одно дерево упорядочивает части отражателя в горизонтальном направлении, другое — в вертикальном. Таким образом значительно увеличивается скорость поиска нужной части для начала расчета хода луча по двум заданным начальным координатам.

В программе во время вычислений используются такие разделяемые «ресурсы», как таблицы ИС, частей отражателя, значений силы света, база данных материалов и др. Для выполнения определенной операции необходимы лишь некоторые из этих разделяемых ресурсов. При мультипоточном программировании во избежание конфликта доступа к отдельным ресурсам необходима их последовательная блокировка. Это может привести к нерациональному распределению по потокам выполняемых параллельно операций. В связи с этим была реализована транзакционная система, которая позволяет быстро и эффективно организовать выборку очередной задачи на выполнение так, чтобы уменьшить общее время ожидания ресурсов. Основной принцип работы этой системы заключается в том, что каждой задаче (транзакции) задается список используемых ею ресурсов, который размещается в памяти таким образом, что образуется некий двудольный граф. Одна часть графа содержит вершины, обозначающие ресурсы, другая — вершины, обозначающие активные на данный момент транзакции. Анализируя связи между этими частями, можно легко вычислить наиболее подходящую задачу на выполнение в текущий момент времени, тем самым уменьшив время ожидания.

6 Автоматизированная система моделирования и проектирования световых приборов

На основании методов расчета и проектирования СП разработано программное обеспечение для автоматизированной системы, осуществляющей моделирование, расчет и проектирование СП [13].

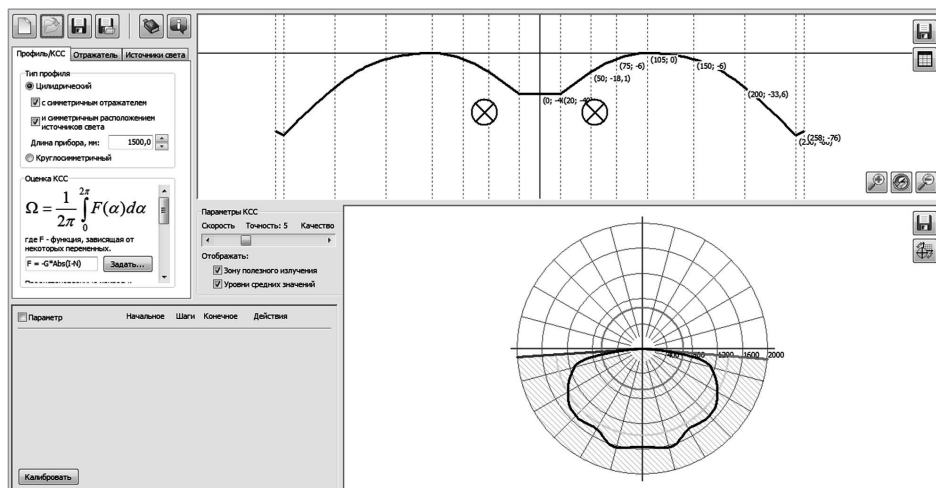


Рис. 3 Рабочее окно программы LighTooLux 1.1

Использование данной системы позволяет:

- в короткие сроки эффективно осуществлять разработку СП различного конструктивного исполнения и назначения;
- оценивать технико-экономическую эффективность светооптической системы СП;
- определять основные параметры светооптической системы СП и увязывать их с конструктивным решением светового прибора в целом с учетом специфики его эксплуатации.

Исходными данными для работы с программой LighTooLux 1.1 являются:

- требуемое светораспределение;
- тип ИС;
- количество и расположение ИС;
- оптические коэффициенты материалов светоперераспределяющей системы;
- полезный угол излучения прибора.

Работу с программой, главное окно которой представлено на рис. 3, можно разделить на несколько этапов.

На первом этапе необходимо смоделировать оптическую систему разрабатываемого СП и выбрать ИС. Поскольку ИС также участвуют в перераспределении световых лучей, то обязательно должны быть заданы их точные размеры.

Тип профиля может быть круглосимметричным или цилиндрическим. При задании последнего можно выбрать такие условия его исполнения, как с симметричным отражателем, с симметричным расположением ИС.

Оценка КСС осуществляется путем вычисления интеграла

$$\Omega = \frac{1}{2\pi} \int_0^{2\pi} F(\alpha) d\alpha,$$

где α — плоский угол, указывающий текущее направление; F — функция, которая задается пользователем.

В программе заложены следующие стандартные (предустановленные) способы оценки КСС:

- максимизация светового потока (выбирается профиль с максимальным значением светового потока во всех направлениях);
- максимизация светового потока внутри полезного угла излучения (выбирается профиль с максимальным значением светового потока внутри полезного угла);
- выравнивание КСС внутри полезного угла излучения;
- максимизация разности средних значений силы света во всем телесном угле и силы света в пределах полезного угла излучения.

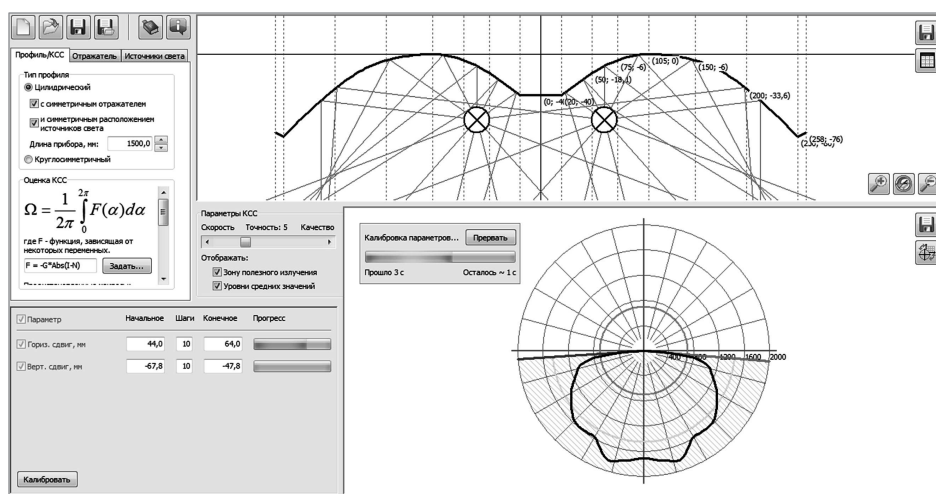


Рис. 4 Отображение главного окна в процессе калибровки

Отражатель формируется из независимых друг от друга частей, неразрывно связанных между собой. Добавлять можно фактически неограниченное количество частей, из которых планируется составить отражатель. Пример отражателя с прорисовкой падающих и отраженных световых лучей показан на рис. 4.

На втором этапе осуществляется уточнение, детальная проработка эскиза смоделированного отражателя путем качественной оценки КСС, калибровки необходимых параметров и окончательного построения КСС (рис. 4). Калибровка — это процесс подбора выбранных параметров для обеспечения максимальной оценки КСС. Кривая силы света отрисовывается в виде графика в полярной или декартовой системе координат.

На третьем этапе полученные результаты могут быть представлены для последующего анализа и интерпретации данных в виде схем, графиков и таблиц, которые сохраняются в соответствующем формате.

Результаты моделирования профиля отражателя можно представить в виде отчета, в котором будут представлены таблицы параметров ИС и трассировки отражателя.

7 Заключение

Ввиду массового производства СП различного назначения и исполнения, а также большой сложности процесса их расчета и проектирования, назрела необходимость автоматизации этой задачи. С этой целью в среде объектно-ориентированного программирования Delphi была разработана программа LighTooLux 1.1, позволяющая осуществлять расчет и построение профиля отражателя на основании информации об источнике света, коэффициенте отражения материала отражателя, типе кривой силы света и других данных.

Использованная для этого система программирования Delphi, обладающая такими преимуществами, как наиболее полная объектная библиотека VCL (Visual Component Library), полное описание функций WinAPI, удобные и надежные средства отладки (вплоть до пошагового выполнения команд процессора), позволила решить все вопросы применения алгоритмов и математических методов расчета, использованных в программе.

Литература

1. LightTools: Официальный сайт. <http://lighttools.software.informer.com>.
2. Solidworks: Официальный сайт. <http://www.solidworks.com>.
3. Lambda Research: Официальный сайт. <http://lambdares.com>.
4. Autodesk: Официальный сайт. <http://www.cad.ru>.
5. АСКОН: Официальный сайт. <http://ascon.ru>.
6. КОМПАС-3D: Официальный сайт. <http://kompas.ru>.

7. DIALux: Официальный сайт. <http://dialux.ru>.
8. Light-in-Night Road: Официальный сайт. www.l-i-n.ru.
9. Справочная книга по светотехнике / Под. ред. Ю. Б. Айзенберга. — М.: Знак, 2006.
10. *Трембач В. В.* Световые приборы. — М.: Высшая школа, 1990.
11. *Осипов Д.* Delphi. Профессиональное программирование. — СПб.: Символ-Плюс, 2006.
12. *Бакнелл Д.* Фундаментальные алгоритмы и структуры данных в Delphi. — СПб.: Питер, 2006.
13. *Байнева И. И., Байнев В. В.* Программа для моделирования и проектирования световых приборов различного исполнения и назначения. Свидетельство о государственной регистрации программы для ЭВМ № 2012612870 от 23.03.2012.

ЦИТИРОВАНИЕ ДОКУМЕНТОВ В ПАТЕНТАХ КАК ИНДИКАТОР ВЗАИМОСВЯЗИ ОБЛАСТЕЙ НАУКИ И ТЕХНОЛОГИЙ

*О. С. Кожунова*¹

Аннотация: Работа посвящена анализу взаимосвязей науки и технологий на основе данных о запатентованных изобретениях Нидерландов, Бельгии, Чили и других стран, а также оценке степени значимости приведенных в патентах цитирований научных работ соответствующих областей, в том числе технологических.

Ключевые слова: цитирование в патентах; индикатор взаимосвязи областей науки и технологий; информационный мониторинг науки

1 Введение

Цитирование научных публикаций в патентах позволяет проводить анализ и оценку доли и значимости научных и технических знаний в технологических достижениях. Это дает возможность отслеживать и применять на практике взаимосвязь между наукой и технологиями. Как правило, именно патенты в силу особенности своей сферы и жестких требований как к самим изобретениям, так и к их описаниям предоставляют исчерпывающую информацию об уровне технологического развития отдельных стран, о степени актуальности и новизны некоторых областей науки и техники, о вкладе стран, коллективов, отдельных ученых в развитие определенных областей и т. д.

В данной работе проводится анализ взаимосвязей науки и технологий на основе данных о запатентованных изобретениях Нидерландов, Бельгии, Чили и других стран, а также оценка степени значимости приведенных в патентах цитирований научных работ соответствующих областей, в том числе технологических.

Кроме того, проведенное в работе исследование позволило рассмотреть круг вопросов и проблем, затрагивающих патентную и научную сферы, в частности: влияние фактора цитируемости научных работ в патентах на их международный импакт-фактор; соотношение цитирований в описаниях изобретений, запатентованных в местных и заграничных патентных бюро; доля самоцитирований авторов-изобретателей в их же патентах; увеличение значимости отдельных

¹Институт проблем информатики Российской академии наук, okozhunova@ipiran.ru

областей наук; повышение значения научного сотрудничества как на уровне институтов и коллективов, так и на уровне стран и т. п.

Таким образом, анализ патентных цитирований представляет собой актуальную на сегодняшний день задачу и может быть использован как при проведении мониторинга и оптимизации отдельно взятых областей науки, так и при выявлении факторов научно-технологического взаимодействия.

2 Взаимодействие науки и технологий

Системы регистрации инноваций в передовых индустриальных обществах очень напоминают сложные сети взаимозависимостей между научно-техническими исследованиями и технологическим прогрессом. Лежащие в их основе процессы порождения, распространения и использования знаний обычно описываются в терминах сложных и интерактивных моделей инноваций. Технические изобретения подразумевают передачу и использование идей, практических навыков, «кодируемых» знаний и прочих необходимых составляющих. Научные исследования являются одной из таких составляющих, а именно источником информации [1–3].

Исследования такого рода предоставляют не только необходимые знания и продуктивное взаимодействие различных коллективов, но и обеспечивают осуществление важных технических достижений. Поэтому основными преимуществами знаний и навыков, полученных при проведении фундаментальных исследований в общественном секторе, являются следующие: (1) прослойка обученных научно-методу высококвалифицированных исследователей и инженеров — для решения технических задач и (2) возможность получения оптимальных решений с точки зрения достижения цели, экономии средств и ресурсов. Эти косвенные преимущества научной деятельности во многом являются одной из главных движущих сил технического прогресса [1, 2].

Одна из основных проблем в исследованиях взаимодействия технологий и науки состоит в том, чтобы оценить результаты переноса знаний из области фундаментальных исследований в сферу технологических промышленных инноваций [4]. Даже при том, что мы знаем о значительной доле вклада науки и технологий в разработку технических изобретений, измерить его масштабы и характеристики достаточно непросто. Причины этого заключаются в следующем: потоки имплицитных знаний и кодируемой информации так многочисленны, что, как правило, их трудно идентифицировать и описать с необходимой степенью подробности. Кроме того, перенос научного прогресса на новые технологические приложения часто происходит через определенный интервал времени.

Mansfieldth проводил исследования в этой области, в качестве исходных данных выбрав данные о работе американских фирм [5–7]. Его результаты оказались среди немногих работ, содержащих всесторонние эмпирические данные

по вопросу соотношения научных исследований и технологических инноваций. Результаты его работы указывают на то, что научные исследования имеют действительно большое значение для индустриальных инноваций: согласно Mansfieldth, 10% технологических новшеств не были бы изобретены или же были бы изобретены с большой задержкой, если бы были выполнены без использования соответствующих научных исследований [5, 6].

Патенты предоставляют другой богатый источник информации о технологической релевантности фундаментальных исследований [2]. Хотя патентование и сопровождается многими известными недостатками, связанными с мерами доверия к полученным инновациям, патентные данные уже давно ассоциируются с индикаторами инновационной деятельности и уровня развития технологий [2, 3]. Некоторые исследователи используют этот источник информации для выявления взаимосвязей между наукой и технологиями на уровне изобретателей: выстраивая соответствия между патентными изобретателями и авторами научно-исследовательских работ, они раскрыли несколько интересных особенностей научной составляющей патентов и научного сотрудничества в соответствующих технологических областях исследований. Другие исследователи полностью сосредоточились на содержании патентов, особенно на цитируемых релевантных документах, в которых приведены явные ссылки на научные публикации [2, 4]. Эти непатентные цитирования позволяют явным образом выявить эмпирическую связь между исследованиями и техническими изобретениями [8]. Частотные характеристики непатентных цитирований и производных количественных индикаторов могут использоваться для измерения особенностей отношений технологий и науки и для описания специфических черт опубликованных исследований, релевантных технологическим разработкам в определенной научной области.

3 Что характеризуют непатентные цитирования: голландский опыт

За последнее десятилетие было проведено несколько исследований непатентных цитирований, на которые широко ссылались в связи с их релевантностью для научно-технических разработок, особенно в США [1, 8]. Одним из ключевых вопросов таких исследований является следующий: насколько эти непатентные цитирования отражают степень «научности» патентованных инноваций? По этому поводу Carpenter и Narin было проведено крупномасштабное исследование, которое выявило тесную связь между экспертным ранжированием цитирований научной литературы в расчете на один патент и соответствующим ранжированием зависимостей науки и рассматриваемых технологий [9].

При попытке исследовать природу непатентного цитирования возникают также и другие вопросы: (1) до какой степени патенты, изобилующие непатентными ссылками, связаны с наукой и (2) действительно ли изобретатели являются основными авторами непатентных источников? Экспериментальный массив из

50 патентов был получен из выборки голландских патентов, зарегистрированных в USPTO (United States Patent and Trademark Office — Бюро по регистрации патентов и торговых марок США) [1]. Среди этих патентов были отобраны изданные за период с 1993 по 1996 гг. и содержащие, по крайней мере, одну непатентную ссылку. Результаты этого исследования подтвердили общее представление о том, что патенты USPTO с непатентными ссылками основаны на наукоемких изобретениях (94% всех патентов). В ходе этой работы также был сделан вывод о том, что авторство большинства непатентных цитирований в отобранном массиве принадлежит самим изобретателям. Это единственный источник данных, который дает более подробную информацию о наукоемкости этих патентов и степени использования научного задела в той или иной области [1, 9].

Другое интересное открытие, полученное в ходе вышеописанного исследования, имеет отношение к склонности изобретателей к самоцитированию. В общей сложности 8% всех непатентных цитирований в патентах голландских изобретателей связаны с голландскими научно-исследовательскими публикациями, т. е. в списке институциональной принадлежности их авторов содержалось, по крайней мере, одно голландское учреждение из Академии наук. Интересно и то, что авторами и/или соавторами не менее 81% непатентных цитирований из отобранного массива патентов являются сами изобретатели. Оказывается, многие из этих непатентных цитирований относятся к научным работам, непосредственно связанным с разработкой патентуемого изобретения [1].

В целом, результаты этого исследования массива патентов подтверждают, что непатентные цитирования, релевантные научной области изобретения, образуют явные связи между научно-исследовательскими и технологическими инновациями. Поэтому они могут использоваться в качестве основного источника информации для описания особенностей взаимосвязей науки и технологий [1, 4, 9].

Кроме того, голландские ученые провели исследование нескольких существенных характеристик непатентных документов. Во-первых, их заинтересовал вопрос качества цитируемых непатентных документов, то есть являются ли они высокоцитируемыми научными работами высокого уровня или представляют собой публикации по вопросам некоторых инновационных приложений. Во-вторых, ученые анализировали непатентные документы с точки зрения авторства: являются ли публикации результатом коллективного научного сотрудничества или у них единственный автор? В-третьих, голландские специалисты задались вопросом, сколько рассмотренных ими публикаций написали (или выступили в качестве соавторов) исследователи из промышленных областей [1]?

В табл. 1 приведены некоторые из вышеперечисленных ключевых характеристик непатентных документов (с высоким импакт-фактором). Наиболее часто цитируемая работа в приведенном ранжировании была опубликована достаточно давно (1980 г.), но при этом была процитирована 32 раза в патентах за период с 1993 по 1996 гг., относящихся к производству удобрений. Несколько публикаций

Таблица 1 Наиболее часто цитируемые в патентах научные голландские публикации (фрагмент)

Патентные цитирования			Процитированные научные публикации				
Цитаты в 1993–1996 гг.	Технологические области цитирования	Источники цитирования	Год публикации	Институциональная принадлежность автора изобретения в Голландии	Научное сотрудничество	Научные области	Цитирования 1993–1996 гг.
32	Не научно-направленный сектор	Голландские и иностранные	1980	Foundation Ital. Revecu BV Middelburg BV	NL/Publ.-priv. NL/Intrasec.	Сельское хозяйство, растениеводство	2

являются результатом научного сотрудничества (на национальном или международном уровне, в пределах одного института или между институтами). Некоторые из относительно недавних работ имеют чрезвычайно высокий индекс цитирования в научной сфере (923 цитаты) и при первом приближении описывают крупные научные достижения, тогда как публикации с высоким импакт-фактором насчитывали только два научных цитирования за период с 1993 по 1996 гг. Все эти разнородные публикации, рассмотренные голландскими учеными, имеют лишь одну общую особенность: все публикации характеризуются относительными индексами цитирования, которые выше среднего у соответствующего научного журнала (JCS), как и в среднем по научной области. Голландские научные статьи, процитированные в патентах, являются не просто случайным набором релевантных публикаций, отобранных патентными экспертами, исследователями и инженерами, но, как правило, имеют некоторый вес на международном уровне.

Кроме того, исследователи из Голландии подняли и другие не менее интересные вопросы. Например, насколько количество непатентных ссылок на те самые часто цитируемые голландские публикации зависит от основных характеристик патентов и статей? Для прояснения этого вопроса они произвели регрессионный анализ 280 голландских исследовательских статей из отобранного массива публикаций, изданных между 1980–1992 гг., причем эти работы были процитированы, по крайней мере, в трех патентах в 1993–1996 гг. [1]. В табл. 2 приведены суммарные статистические данные этого анализа и список независимых (прогнозирующих) переменных. Результаты показывают, что этот набор переменных не очень-то объясняет причины, определяющие число непатентных ссылок в каждом патенте: квадратичная корреляция Pearson указывает, что только 12% дисперсии имеют соответствия. Лучшими прогнозирующими переменными числа непатентных ссылок оказались: (1) научная область и (2) принадлежность патента к промышленной сфере [1]. Дата издания публикаций и абсолютное число цитат в статье на другую статью также существенны для объяснения различий в количестве непатентных ссылок.

Таблица 2 Результаты регрессионного анализа — обоснование числа цитирований голландских научных публикаций с высоким индексом цитирования в американских патентах за 1993–1996 гг.

Прогнозирующие переменные	Отбор цитирующих патентов	
	Все	Растениеводство, медицина и фармакология
1. Основная научная область публикации	0,20	-0,25
2. Дата создания публикации	0,14	0,17
3. Абсолютное число цитирований статьи на статью за 1993–1996 гг.	0,13	0,09
4. Нормализованное число цитирований статьи на статью за 1993–1996 гг.	0,05	0,10
Сектора цитирующих патентов		
5. Растениеводство	0,17	
6. Фармакология и медицина	0,10	
7. Химия материалов	0,08	
8. Электроника и электромеханика	-0,07	
9. Техническое обеспечение	0,03	
10. Транспортные задачи	-0,03	
Дисперсия (Γ^2)	0,12	0,08
Число цитируемых публикаций	280	193

Голландские исследователи пошли дальше в своем анализе патентной сферы и ее взаимосвязи с научной областью в национальном масштабе [1]. Они рассмотрели голландские научные публикации и их цитируемость на международном уровне. Оказалось, что эти публикации были процитированы приблизительно в 2 000 патентов из USOTO, выданных в 1995–1996 гг. Только 4% тех патентов были голландскими (т. е. изобретатель — резидент Нидерландов). Это лишь небольшая часть всех ссылок, однако она значительно больше, чем доля Нидерландов в суммарном выходе всех патентов в американском бюро (приблизительно 1%) или доля Нидерландов в глобальном исследовательском «выхлопе» в международных научно-технических журналах (приблизительно 2%) [1]. Ученые пришли к выводу, что в данном случае большое количество ссылок голландских ученых в изобретенных голландцами патентах связано с национальной релевантностью исследований. Другие исследования непатентных ссылок подтверждают это положение и для других индустриально развитых стран.

В табл. 3 приведена итоговая статистика тенденций патентования голландских изобретений в USPTO и соответствующие данные о непатентных ссылках. Почти две трети голландских патентов относятся к двум ключевым индустриальным секторам в Нидерландах: (1) электроника и (2) химия материалов. Наиболее

Таблица 3 Тенденции в распределении американских патентов, выданных голландским изобретателям, по технологическим секторам

Сектор	Доля в голландских патентах		Доля голландских патентов, цитирующих голландские научные публикации	
	1995/96	Изменения за период 1987/88–1995/96	1995/96	Изменения за период 1987/88–1995/96
Электроника и электромеханика	33,5%	+1,1	9,8%	+0,4
Химия материалов	29,0%	–0,6	3,3%	–6,3
Техническая поддержка	15,5%	+3,8	2,4%	+2,4
Транспортные задачи	13,1%	–6,0	9,8%	+2,4
Фармакология и медицина	5,1%	+1,7	2,3%	–6,2

наукоемкий сектор — фармацевтика и медицина (включая биотехнологии) — содержит только 5% всех патентов USPTO [1].

Наблюдения голландских ученых подтверждают их гипотезу о специфичности взаимосвязей выдачи патентов и соответствующих им непатентных ссылок в ракурсе определенных секторов деятельности. Сравнительный анализ доли секторов, к которым принадлежат выданные голландцам патенты, с одной стороны, и их доли в этих секторах по патентам, цитирующим голландские научно-исследовательские работы, с другой стороны, демонстрирует как сходство, так и некоторые отличительные особенности. Что касается сходства, голландские ученые обнаружили тесное соответствие между обеими долями в случае сектора электроники [1].

Результаты голландского исследования представлены в табл. 4 и 5, в которых продемонстрирована сложность взаимосвязей между пятью основными патентными секторами и соответствующими областями науки. Четыре «главных» научных области вносят свой существенный вклад в несколько технологических секторов: биомедицинские исследования, клиническую медицину, биохимию и биологию и химию. В целом эти данные о взаимосвязях иллюстрируют убывающую долю естественных наук (физика, химия) и повышение роли наук о жизни, что связано с данными недавних библиометрических исследований американских патентных цитирований и их взаимосвязей. Интересно, что каждый сектор связан с определенной областью науки. Так, в случае с фармацевтикой и медициной можно обнаружить существенный вклад из области биомедицинских исследований, биохимии и биологии. Напротив, сектор химии материалов выделяется существенной долей цитат на исследования из области химии [1].

Таблица 4 Тенденции в образовании научных связей между голландскими и иностранными патентами, цитирующими голландские научные публикации

Сектор	Среднее число непатентных ссылок на голландские научные публикации в патентах 1995/96 гг.		Изменения за периоды с 1987/88 и с 1995/96 гг.	
	Иностранные	Голландские	Иностранные	Голландские
Фармакология и медицина	2,9	3,8	42%	11%
Техническое обеспечение	2,8	3,9	59%	—
Химия материалов	2,2	3,7	8%	43%
Транспортные задачи	1,9	2,0	10%	100%
Электроника и электромеханика	0,3	2,6	-1%	14%
Суммарно	2,8	3,3	46%	23%

Таблица 5 Цитирования голландских научных публикаций по секторам и научным областям: цитирования за 1995/96 гг. и тенденции в периоды с 1987/88 и с 1995/96 гг.

Научная область	Фармакология и медицина	Химия материалов	Техническая поддержка	Электроника и электромеханика	Транспортные задачи
Биохимические исследования	29% (↓)	17% (—)		19% (↓)	7% (—)
Клиническая медицина	19% (—)	17% (↑)	22% (↑↑)	26% (↑↑)	6% (↑)
Биомедицинские технологии			9% (↑↑)	5% (↓)	6% (↑↑)
Биохимия и биология	28% (↑)	19% (↑)		20% (↑)	14% (—)
Биотехнологии					20% (↑↑)
Фармакология	7% (↓)	9% (—)			
Химия	4% (↓↓)	17% (↓↓)		5% (↓↓)	23% (↓)
Химические технологии					6% (↓↓)
Физика			18% (↓↓)	7% (↑)	9% (↓↓)
Материаловедение			2% (↓↓)		1% (↓↓)
Электротехника			24% (↑)	5% (↓)	
Инструментология			3% (↓↓)		
Математика и компьютерные науки			7% (—)		

Замечание. Изменение в долях за периоды с 1987/88 и с 1995/96 гг.: (↑↑) — повышение более чем на 25%; (↑) — повышение на 10%-25%; (—) — стабильное между -10% и +10%; (↓) — снижение на 10%-25%; (↓↓) — снижение более чем на 25%.

4 Анализ взаимосвязей науки и технологий: опыт Бельгии

Понимание и интерпретация взаимодействия науки и технологий требует четкого понимания сути и научной, и технологической систем [2]. Первым шагом в этом направлении является измерение. Одна из главных проблем в этом отношении: описание научно-технологической деятельности как в качественных мерах, так и в количественных, с тем чтобы индикаторы, характеризующие их, охватывали и индивидуальные, и общие аспекты [2]. Общая проблема, с которой обычно сталкиваются при осуществлении такого рода усилий, заключается в том, что результаты науки и технологий могут быть измерены лишь косвенно. При этом используются индикаторы непосредственных результатов (output), целевых результатов (outcome) и индикаторы взаимосвязей и влияния результатов (impact) [2, 10]. Патенты и публикации, как результаты достижений науки и технологий, называют мерами доверия (proxy measures) [2]. Можно ли базировать анализ научно-технологического взаимодействия на таких мерах? В настоящее время такие меры представляются наилучшими доступными индикаторами, поскольку они дают достаточно широкий спектр аналитических возможностей [11].

Патенты как детальный источник информации об изобретательской деятельности предлагают интересный инструмент мониторинга для выявления главных тенденций в развитии технологий, а при соблюдении определенных условий — и возможность более детального анализа специфических научно-исследовательских процессов. Кроме того, патенты характеризуют некоторый минимальный уровень требований к изобретениям, которые проходят полную экспертизу и отбор патентными бюро по принципу их новизны и потенциальной применимости [11]. Патент обычно является следствием успешной научно-исследовательской деятельности, таким образом детально характеризуя саму деятельность.

Что касается научных публикаций, они представляют собой индикатор непосредственных результатов научно-исследовательской деятельности [2]. Одна из целей научной публикации состоит в распространении научных результатов в пределах и вне научного сообщества. Следовательно, публикация в научных журналах серийных изданий играет ведущую роль в распространении научных результатов. Патенты, как и научные публикации, позволяют проводить подробный анализ соответствующих структур в сферах технологической и научной деятельности. На основе доступной информации о (со)авторах, (со)изобретателях и (со)патентовладельцах, их институциональной принадлежности, их адресах, ссылках и цитатах на предшествующие научные и технологические разработки и т. д. можно обнаружить и проанализировать социальные и когнитивные сети [2, 11].

Одними из ведущих специалистов, определивших понятие качественного анализа научно-технологического взаимодействия, являются de Solla Price [12] и Rosenberg и Birdzell, Jr. [13]. До начала 1990-х гг. исследования такого вза-

имодействия все же были ориентированы в большей степени на количественные показатели. Однако далее в центре внимания исследователей оказались именно качественные индикаторы (см., например, [14]).

Сегодня различают два подхода к изучению отношений между наукой и технологиями: подход «косвенные взаимосвязи» и подход «непосредственные взаимосвязи» [2]. Подход «косвенные взаимосвязи» направлен на изучение взаимосвязей науки и технологий на основе данных о мобильности ученых и инженеров, образовательных процессов и т. д. Подход «непосредственные взаимосвязи» обращен к возможности изучения взаимодействий науки и технологий через библиографические ссылки, представленные в патентных документах. В рамках данного подхода интересны в основном непатентные ссылки, поскольку они демонстрируют непосредственные связи между наукой и технологиями [2].

В работе бельгийских ученых Verbeek и др. отдается предпочтение второму подходу. На его основе разработана методология, описанная в [2].

Методология исследования взаимосвязей науки и технологий, описанная в их работе, в качестве единиц анализа использует непатентные ссылки, а именно цитирования публикаций из научных журналов [2]. Выявление непатентных ссылок и последующая идентификация «исходной» публикации в Научном индексе цитирования (SCI) позволяет более широко охарактеризовать взаимосвязь областей науки и технологий, а также посредством международной классификации патентов и классификационной системы журнала SCI-ISI (ISI — Институт научной информации). В исследовании также были использованы патентные данные Бюро патентов и регистрации товарных знаков Соединенных Штатов и данные Европейского патентного бюро [2].

В предложенной методологии исследователи различают две основных стадии. На первой стадии была разработана модель науки и технологий (рис. 1). Предложенная методология может быть применена неоднократно для обновления схемы взаимодействия науки и технологий. Вторая стадия включает много шагов, на которых результаты первой стадии переносятся на более узконаправленную задачу. На данном этапе применяется, главным образом, количественный подход с рядом качественных аспектов (в основном путем привлечения экспертной деятельности) [2].

На первом шаге разработанной методологии происходит определение предмета дальнейшего анализа. В работе бельгийских ученых был выбран временной период в 17 лет (1980–1996 гг.), разделенный на 3 аналитических показательных периода (1984–1986, 1987–1991 и 1992–1996 гг.) [2].

При определении эффективности взаимодействия науки и технологий во времени появляется возможность анализа соэволюции их взаимодействия. В работе бельгийцев основным периодом обсуждения является период 1992–1996 гг., названный ими «справочным» [2]. В результате моделирования взаимодействия науки и технологий (S&T) за указанный временной период ученые из Бельгии по-



Рис. 1 Структура методологии (I — создание схем научно-технологического взаимодействия, II — введение в эксплуатацию)

лучили «фактическую» схему взаимосвязей S&T, которая может использоваться в качестве исходных данных для дальнейших исследований. Кроме того, выбранный временной период характеризуется достаточно большим числом непатентных цитирований (68% всех непатентных ссылок за период 1980–1996 гг.) [2].

На втором шаге описываемой методологии происходил отбор и выборка патентов и соответствующих непатентных ссылок. Причем критерием отбора как самих патентов, так и ссылок внутри них было соответствие набору критериев, определенных на первом шаге. В своей работе бельгийские ученые сообщают об отборе 656 695 патентов из USPTO. Из них для дальнейшей обработки были отобраны 1 147 160 непатентных ссылок. В ходе подготовки информационных ресурсов из этого диапазона были отобраны цитирования из журналов. Публикации в научных журналах являются основной коммуникационной средой в

пределах научного сообщества, и в качестве таковой их можно считать мерой доверия к научной деятельности. Заключительная цель второго шага состояла в том, чтобы идентифицировать «исходные» публикации из данных, предоставляемых SCI посредством применения базового подхода методологии [2].

На третьем шаге предложенной методологии была реализована стандартизация и унификация журнальных ссылок. Для идентификации и разбора ссылок научных журналов на несколько полей, таких как {имя автора} и {год публикации}, был разработан сложный алгоритм парсинга, основанный на подходе к анализу текста. Грамматические отклонения типа неправильного написания, неуместной пунктуации, заглавных или строчных букв делают эту операцию сложной и трудоемкой. Из каждой журнальной ссылки бельгийские разработчики извлекли следующие поля: {имя автора}, {название публикации}, {название журнала}, {том}, {номер издания}, {год публикации}, и {первая страница}. Каждый текстовый фрагмент был приписан одному из этих типов данных. Например, текстовый фрагмент «том 55» должен быть преобразован в «55», а запись «12-05-1986» должна быть заменена на «1986». По итогам шага 3 на период 1992–1996 гг. было идентифицировано, разобрано и успешно стандартизировано 296 679 ссылок научных журналов (26% всех непатентных ссылок) [2].

На четвертом шаге было выстроено соответствие цитирований публикациям, включенным в SCI. Подход, разработанный для отслеживания «исходной» публикации, включенной в SCI, основан на ключевых соответствиях [15]. Ключевое соответствие составляется из комбинации следующих полей: {имя автора}, {год публикации}, {том} и {первая страница} (рис. 2). Таким образом удалось избежать использования названия журнала для выстраивания этих соответствий, поскольку в нем могут быть орфографические ошибки, синонимы и даже акронимы. Как только «исходная» публикация была идентифицирована в SCI, была обнаружена и соответствующая область науки по ее связи с классификацией журналов SCI. Процесс соответствия проводился в несколько итераций. Первоначально для составления ключевого соответствия были использованы все четыре поля. В последующих итерациях было попеременно использовано одно «изменяемое» поле, за исключением поля {имя автора}, которое присутствовало на всех этапах. Имя автора было сокращено до первых шести букв с тем, чтобы предотвратить несоответствия между цитированием и «исходной» публикацией [15]. Когда все поля ключевых соответствий совпадали с соответствующими областями в SCI, предполагалось, что журнальная цитата была уникальным образом связана с публикацией из списка SCI (это имело место в 106 636 журнальных цитатах) [2].

На пятом, заключительном шаге первой стадии бельгийской методологии были созданы и проанализированы схемы взаимосвязей / взаимодействия науки и технологий [2]. Названия технологий и научных областей были идентифицированы и обработаны в пределах классов Международной патентной классификации

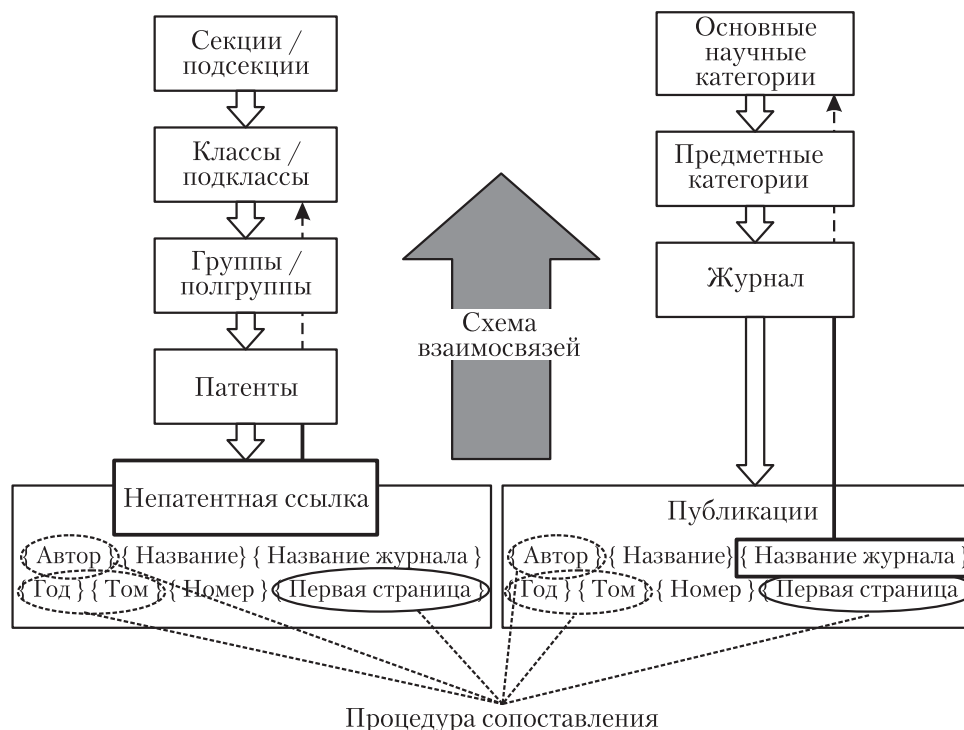


Рис. 2 Процедура сопоставления и построения научно-технологических взаимосвязей

(первые четыре цифры) и распределены в соответствии с классификацией журналов Института научной информации (SCI-ISI) (рис. 2).

Исходя из соответствия между отдельно взятой журнальной ссылкой, заданной в патенте, и «исходной» публикацией, внесенной в перечень SCI, бельгийские ученые затем проследили класс Международной патентной классификации рассматриваемого патента и классификацию журнала, в котором появилась «исходная» публикация. Результаты полученных взаимосвязей были спроецированы в матрицу, содержащую классы технологий, научные области и частоту кросс-цитирования между обеими сферами деятельности. Была применена «нормальная» схема подсчета. Все патенты были повторно классифицированы согласно Международной патентной классификации. Помимо описания взаимодействия науки и технологий и их анализа первого порядка, было рассчитано много дополнительных индикаторов. Такую процедуру бельгийские исследователи назвали «статическим анализом взаимосвязей». Для ряда областей науки и технологий может быть также проведен «динамический анализ взаимосвязей». Он дает возможность проанализировать как соэволюцию патентной и публи-

кационной деятельности, так и всеобщую эволюцию вариантов взаимодействия S&T в течение долгого времени. Динамический анализ взаимосвязей является частью стадии II методологии бельгийских ученых. Методологические шаги 4 и 5 проиллюстрированы на рис. 2 [2].

Предложенная методология и первые результаты ее применения были проверены экспертным комитетом (группой международных экспертов, состав которой был утвержден ЕС). На второй стадии исследования бельгийские ученые сосредоточились на актуальных сегодня областях, в которых они провели детальный анализ взаимодействия науки и технологий. Критерием выбора рассматриваемых областей стали научно-исследовательские приоритеты 6-й Рамочной программы ЕС. В списке областей особенно были выделены биотехнологии, информационные технологии и телекоммуникации, нанотехнологии, интеллектуальные материалы и новые производственные процессы, авионавтика и космические исследования, нутрицевтика и фармакология, новые технологии и глобальные процессы [2].

В опубликованной работе бельгийские исследователи привели результаты разработки методологии на стадии I [2]. Они обратили внимание на то, что соэволюция патентов, зарегистрированных в USPTO, и соответствующих непатентных ссылок особенно характерна начиная с 1988 г. и далее, поскольку число непатентных ссылок в какой-то момент начинает превышать число патентов (рис. 3). Однако это вызвано не общим увеличением числа патентов, в которых цитируются непатентные документы, а, согласно выводам бельгийских ученых, происходит за счет увеличения числа непатентных ссылок на один патент в опре-

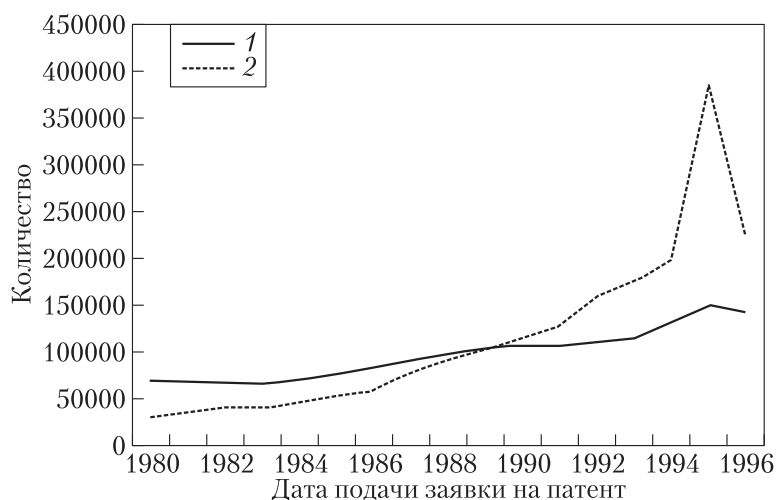


Рис. 3 Эволюция числа патентов (1) и непатентных ссылок (2)

деленных технологических областях. Кроме того, выявленная закономерность объясняется также особенностями патентования в США.

Процедура прохождения экспертизы в USPTO занимает значительное время, поэтому в ряде областей (таких как, например, биотехнологии, сельское хозяйство) выдается ограниченное число патентов. Как следствие, непатентные ссылки цитируются вместо не доступных пока патентных документов, на которые обычно ссылаются в первую очередь как на первоисточники. Бельгийские исследователи предполагают, что из-за таких рассогласований во времени патентные эксперты сегодня имеют значительно больше стимулов заниматься поиском релевантных патентам исследований. Кроме того, они полагают, что увеличение количества непатентных ссылок связано с повышением роли научных исследований в технологических разработках. На этом предположении основана методология, которую разработали бельгийские ученые, и построенная ими модель взаимодействия науки и технологий [2].

Несимметричное распределение непатентных ссылок (причем большинство патентов не содержит ссылок, и лишь небольшая часть из рассмотренных патентов содержит многочисленные ссылки) является важным аспектом обоснованности «прямого» анализа взаимодействия науки и технологий [16, 17]. За период с 1992 по 1996 гг. исследователями отмечено следующее распределение: 65% всех патентов содержали 0 непатентных ссылок; 8% только 1 ссылку такого рода; 19% от 2 до 4 ссылок; 1% точно 5 ссылок; наконец, 7% всех патентов содержало больше 5 таких ссылок. В целом 35% патентов содержат 1 или более ссылок на непатентный документ. Отмечено также, что существуют области, в которых взаимодействие между наукой и технологиями достаточно тесное, поэтому распределение ссылок по патентам зависит от области исследования. К числу таких областей можно отнести биотехнологии, органическую химию, физику полупроводников и информационные технологии [2].

5 Связь между наукой и технологическими инновациями: ракурс Чили

Ученые из Университета Сантьяго, Чили рассматривают патентование как информетрический инструмент, позволяющий оценивать инновации и анализировать многомерные взаимосвязи между наукой, технологиями и инновациями. В своем исследовании они рассматривают влияние чилийских научных статей на развитие технологий на международном уровне на примере связей между научными работами, опубликованными в Чили, и американскими патентами [18].

Например, результаты исследования показали, что с 1987 по 2003 год в 509 патентах было приведено 562 цитаты на 273 статьи, написанные, по крайней мере, одним автором-сотрудником какого-либо чилийского института [18]. В целом эти и другие результаты подтверждают, что в Чили превалирует непатентная

культура оформления инноваций, в которой задействованы исследователи и научные институты. Поэтому при изучении взаимосвязей науки и технологий посредством анализа отношений между патентами и научными публикациями чилийские ученые преследовали свои цели, а именно показать необходимость внедрения в научную сферу развивающихся стран подхода, который позволил бы поддерживать устойчивое и продуктивное взаимодействие между фундаментальной и прикладной наукой, а также вывел бы академическую науку на конкурентный мировой уровень [18].

Итак, на основе данных на 2005 г. была проведена дискуссия о необходимости в защите интеллектуальной собственности в странах Латинской Америки, однако она не привела эти страны к какому-либо продуктивному решению [19]. По мнению чилийских исследователей, такая позиция во многом обусловлена социокультурными предпосылками и сложившимся отношением к научным исследованиям и инновациям в странах Латинской Америки. В частности, согласно данным Чилийской академии наук, в 2002 г. на научные исследования в Чили потрачено лишь 0,7% ВВП (валового национального продукта) [18]. Кроме того, по ряду показателей Чили отстает от развитых стран, хотя темпы его собственного развития, включая и наличие фундаментальных и прикладных исследований, приближаются к последним. Также немаловажен тот факт, что, несмотря на весьма скромные масштабы патентования чилийских инноваций, научные статьи того же происхождения регулярно цитируются в американских патентах. Таким образом, согласно выводам чилийских ученых суммарная значимость производимых в Чили инноваций явно недооценивается, на что указывает индикатор взаимосвязи науки и технологий — ссылки на чилийские научные статьи в американских патентах [18].

В своем исследовании чилийские специалисты рассматривали Чилийский национальный корпус цитирований за период 1981–2003 гг. (38 862 документа). Найденные документы были стандартизованы на предмет принадлежности авторов цитирований к институциональной и/или научно-исследовательской среде. Дальнейший отбор происходил по нескольким атрибутам цитируемых в американских патентах статей: дата публикации статьи, полное название журнала, фамилия первого автора и номера страниц [18].

В результате были получены следующие распределения по числу авторов из Чили (табл. 6). Оказалось, что число статей уровня «mainstream», опубликованных авторами из Чили, одно из самых невысоких в сравнении с аналогичным показателем для авторов из других стран (что также связано и с небольшой плотностью населения в Чили) [18].

Среди американских патентов, которые были выданы изобретателям из других государств, патенты Чили занимают одну из самых низких позиций: 11 патентов с изобретателями-чилийцами, выданные в 2003 г., соответствуют 0,7 патента на миллион жителей в год [18].

Таблица 6 Количественные индикаторы результатов науки и технологий для стран с разным уровнем экономического развития (фрагмент)

Страна происхождения	Публикации, изданные авторами из соответствующей страны		Американские патенты, выданные по первому автору из соответствующей страны		Соотношение патентов и публикаций на страну происхождения
	За 2003 г.	За 2003 г. на 1 млн жителей	За 2003 г.	За 2003 г. на 1 млн жителей	Американские патенты 2003 г. на 100 публикаций 2003 г. по стране происхождения
Финляндия	7851	1510	865	166,3	110
Ирландия	3061	785	163	41,8	53
Австралия	23620	1199	900	45,7	38
Новая Зеландия	4666	1167	135	33,8	29
Мексика	5902	56	84	0,8	14
Аргентина	4679	121	63	1,6	13
Бразилия	12699	70	130	0,7	10
Чили	2550	162	11	0,7	4

В последней колонке табл. 6 приведен так называемый индекс перехода: соотношение патентов и публикаций. Это число патентов, выданных USPTO в 2003 г. изобретателям из определенной страны, на 1000 статей, опубликованных изобретателями из той же страны. И здесь Чили занимает последнее место [18].

Когда чилийские специалисты изучали вопрос о том, имеют ли научные статьи, написанные чилийскими учеными, какое-либо влияние на технологии международного уровня, они сопоставляли свою базу данных за 1981–2003 гг. с 38 862 научными статьями с базой данных Thompson, содержащей ссылки из американских патентов за период с 1984 по 2003 гг. Это сопоставление дало следующие результаты: в 509 патентах было обнаружено 562 цитирования, из них 273 — ссылки на статьи. Итоги исследования показали, что именно американские, а не чилийские компании, являются патентообладателями патентов, которые чаще всего цитируют чилийские научные статьи (лишь четвертая часть таких патентов принадлежала Чили). Это лишь подтвердило прежнее допущение чилийских исследователей о том, что научные исследования не используются при разработке местных технологий [18].

Если рассматривать области, к которым относятся статьи чилийских авторов, процитированные в американских патентах, то получается следующая картина: в области биомедицинских исследований приведено 211 цитат, клинической меди-

пины — 163 цитаты, химии — 119, биологии — 25, инженерных технологий — 24, физики — 9, наук о Земле и космосе — 5, математики — 3 и психологии — 3 [18].

Хотя цитирование статей, написанных чилийцами, в американских патентах началось в 1987 г., устойчивый характер оно приобрело лишь к 1997 г. Как показано на рис. 4, темп роста числа американских патентов, ссылающихся на чилийские статьи, выше, чем темп роста числа статей американских авторов. Средняя доля цитирований за периоды с 1994 по 1998 гг. и с 1999 по 2003 гг. чилийских статей в областях биомедицины и химии значительно возросла, например соответствующая доля для биохимии и биофизики возросла с 3,92 до 5,29, иммунологии — с 2,33 до 6,05 [18]. Чилийские исследователи полагают, что доступность статей в эру Интернета дает возможность ознакомиться с исследованиями, проводимыми в развивающихся странах, что, в свою очередь, повышает вероятность и создает условия для совместных исследований вместе с учеными из других стран. В большинстве чилийских статей, процитированных в американских патентах, в качестве соавторов указаны исследователи из США, Италии, Аргентины, Франции и других стран.

Поэтому растущее число цитирований научных статей уровня «mainstream», написанных чилийцами, в американских патентах, вероятно, отражает растущее число международных научных объединений и число публикаций такого уровня (рис. 4).

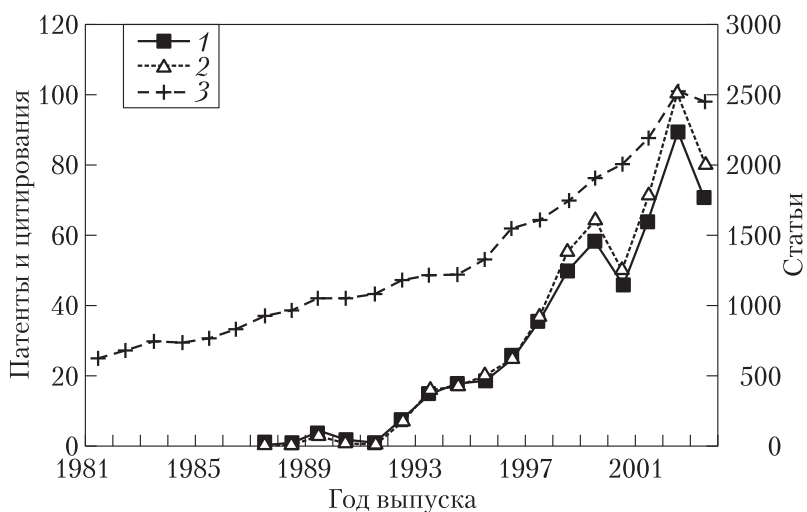


Рис. 4 Ссылки в американских патентах на чилийские публикации по году выпуска цитирующего патента: 1 — число патентов, цитируемые публикации; 2 — число ссылок на публикации; 3 — общее число статей

Хотя сложно определить тот вклад, который чилийцы внесли в патентование изобретений, публикации, процитированные в американских патентах, в исследовании чилийских специалистов представляют в основном чилийскую науку и ее результаты. Это, вероятно, обусловлено теми специфичными областями, в рамках которых авторы этих статей проводят исследования в своей стране [18].

6 Другие попытки изучить связи между наукой и технологиями

6.1 Индикаторы передачи знаний

Испанские ученые провели исследование непатентных ссылок, процитированных в качестве источников в американских патентах в области биотехнологий, с тем чтобы оценить способность ученых из европейских научных центров CNRS (Франция), CNR (Италия) и CSIC (Испания) проводить исследования и

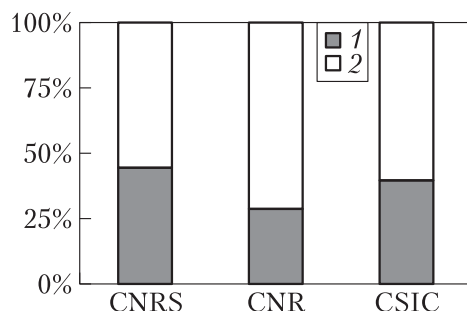


Рис. 5 Авторы, цитируемые (1) и нецитируемые (2) в американских патентах

публиковать соответствующие научные статьи с технологическим уклоном [20]. Анализ научной литературы, процитированной в патентных документах, — это процедура, обычно используемая для исследования связей между наукой и технологиями [21–25]. Индикаторы передачи знаний, полученные в результате анализа ряда цитирований в американских патентах на публикации с

таким же числом исследователей (63) из CNR, CSIC и CNRS показал, что в течение одного и того же периода

процент процитированных авторов составляет 28,5%, 39,7% и 44,4% соответственно (рис. 5).
Из массива американских патентов, цитирующих исследователей из CNR, CNRS или CSIC, 56% цитируют ученых CNRS, в то время как 22% цитируют ученых CSIC и 22% — ученых CNR. Распределение цитирований весьма отличается от распределения патентов, цитирующих публикации. Если рассматривать общее количество найденных цитирований (453) для всех ученых из вышеперечисленных организаций, 62% из них принадлежат ученым из CNRS (281 цитата), в то время как ученые из CNR и CSIC снова находятся в похожем диапазоне цитирований (81 и 91). Таким образом, среди рассмотренных трех учреждений CNRS является не только наиболее цитируемым в американских патентных документах, но также единственным институтом, который получил максимальное количество цитирований [20].

Изучение патентов с цитированиями также позволяет выявить самые важные/успешные области биотехнологий, представленные в этих трех институтах. Так, ученые из CNR, главным образом, цитируются в американских патентах по экологии, ученые из CSIC — в патентах по сельскохозяйственным вопросам и патентах по промышленным способам производства, в то время как их коллеги из CNRS, главным образом, специализируются на патентах из области здравоохранения [20].

6.2 Патенты, процитированные в научной литературе: исследование обратных цитирований

В работе ученых из Бельгии, Венгрии и Финляндии описывается новый подход к изучению связей между наукой и технологиями [1]. В отличие от большинства исследований в этой области исследователи из Евросоюза не изучают цитирование научной литературы в патентах. Они исследуют цитирование патентов в научной литературе. Исследователи проанализировали статьи, зафиксированные в SCI из ежегодных выпусков 1996–2000 гг. ISI (на компакт-дисках) и патентные данные, предоставляемые USPTO [26].

Почти 30 000 американских патентов были процитированы в научных статьях. Ученые проанализировали «цитатные» связи по научным областям и технологическим секторам. Оказалось, что из всех рассмотренных документов патенты из областей, связанных с химией, цитируются в статьях чаще всего. Среди технологических секторов химический сектор четко доминирует в плане цитирования патентов, за ним по частоте ссылок следуют патенты из фармакологических и медицинских секторов. Далее исследователи ранжировали патенты по странам, беря за основу географическую принадлежность изобретателей процитированных патентов, а также более детально просматривали десять наиболее часто цитируемых патентов и анализировали смещения между классами патентов и областями наук [26].

Ученые получили результат, что в среднем приблизительно в 13 500 публикациях ежегодно цитируются патенты. Это составляет около 1,7% всех публикаций, проиндексированных в базе данных SCI. Среди них ежегодно 7800 публикаций цитируют американские патенты. Как и ожидалось, большинство публикаций, цитирующих патенты, — статьи. Исследователи выбрали в качестве периода анализа 1998 г. Из рассматриваемого массива 92% публикаций, которые цитировали патенты, оказались статьями и примечаниями, на долю обзоров пришлось 6,8% и письма 0,8%. Остальные 2,2% представляют собой редакционный материал. В другие годы распределение аналогичное [26].

Картина сильно меняется, если проанализировать распределение по предметным классификациям. На рис. 6 представлено распределение по областям знаний для всех публикаций, зафиксированных в SCI, для публикаций, цитирующих патенты, в целом и для публикаций, цитирующих американские патенты,

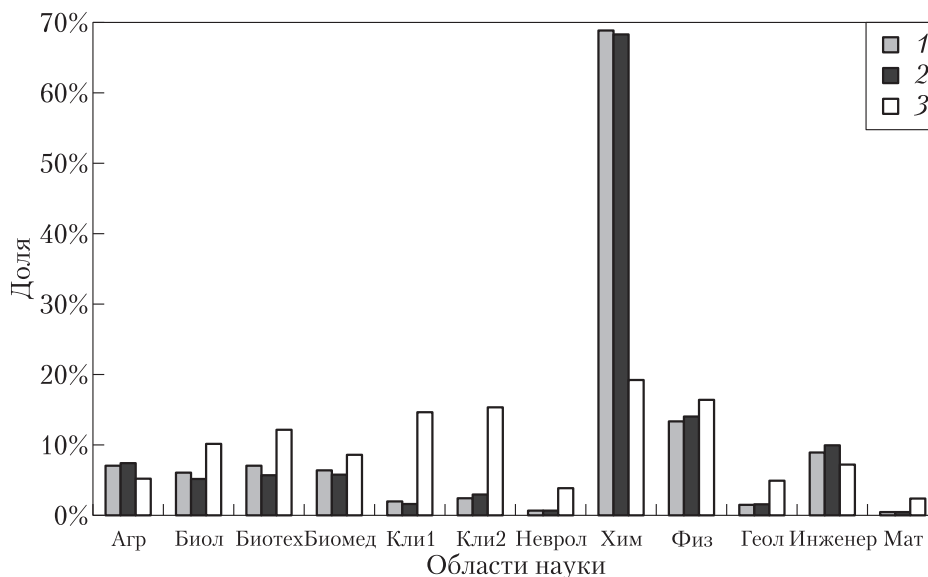


Рис. 6 Распределение публикаций из SCI, цитирующих патенты: 1 — цитируют все патенты; 2 — цитируют американские патенты; 3 — публикации из SCI

в частности. Показатели по американским патентам практически не отличаются от общих показателей такого рода. Поэтому в том, что касается рассматриваемого цитирования, американские патенты можно считать репрезентативными. В ракурсе технологических секторов смещения следующие: почти 70% публикаций, цитирующих патенты, относятся к сектору химии, далее следуют сельское хозяйство и экология. Кроме того, часто цитируются патенты по клинической и экспериментальной медицине и по математике [26].

7 Заключение

Хотя при разработке технологий предполагается некоторая их связь с научной деятельностью, выявить закономерности их взаимодействия не так просто. Механизмы мониторинга и анализа результатов в этих сферах человеческой деятельности разные в силу разности мотиваций и целей. Однако именно наукоемкие технологии, основанные на фундаментальных исследованиях, демонстрируют преимущества взаимодействия с наукой.

Поэтому фокусом данной работы стали именно взаимосвязи областей науки и технологий. Они рассматривались в ракурсе цитирования документов, как правило научных статей, в патентах. Такой ракурс был выбран неслучайно: патенты являются, по сути, квинтэссенцией технологического прогресса, непо-

средственными результатами инновационной деятельности в сфере технологий. Научные публикации, в свою очередь, достаточно разносторонне характеризуют состояние некоторой научной области в целом, дают представление о коллективах ученых и институтах, о характере проводимых ими исследований и т. д. Когда в патентах цитируются научные публикации, возникает множество вопросов о связи научных областей и сферы технологий, о мерах их влияния друг на друга, о необходимости такого взаимодействия, а также, например, о влиянии фактора цитируемости научных работ в патентах на их международный импакт-фактор, о соотношении цитирований в описаниях изобретений, запатентованных в местных и заграничных патентных бюро, о доле самоцитирований авторов-изобретателей в их же патентах, об увеличении значимости отдельных областей наук, о повышении значения научного сотрудничества как на уровне институтов и коллективов, так и на уровне стран и т. п.

В работе была предпринята попытка осветить некоторые из этих вопросов и проанализировать взаимосвязи науки и технологий на основе данных о запатентованных изобретениях Нидерландов, Бельгии, Чили и других стран. Кроме того, была проведена оценка степени значимости приведенных в патентах цитирований научных работ соответствующих областей, в том числе технологических.

Литература

1. *Tissen R. J. W., Buter R. K., Leeuwen Th. N.* Technological relevance of science: An assessment of citation linkages between patents and research papers // *Scientometrics*, 2000. Vol. 47. No. 2. P. 389–412.
2. *Verbeek A., Debackere K., Luwel M., et al.* Linking science to technology: Using bibliographic references in patents to build linkage schemes // *Scientometrics*, 2002. Vol. 54. No. 3. P. 399–420.
3. *Looy van B., Magerman T., Debackere K.* Developing technology in the vicinity of science: An examination of the relationship between science intensity (of patents) and technological productivity within the field of biotechnology // *Scientometrics*, 2007. Vol. 70. No. 2. P. 441–458.
4. *Looy van B., Zimmermann E., Veugelers R., et al.* Do science–technology interactions pay on when developing technology? An exploratory investigation of 10 science-intensive technology domains // *Scientometrics*, 2003. Vol. 57. No. 3. P. 355–367.
5. *Mansfieldth E.* Academic research and innovation // *Research Policy*, 1991. Vol. 20. P. 1–12.
6. *Mansfieldth E.* Academic research underlying industrial innovations: Sources, characteristics and financing // *Rev. Economic Statistics*, 1995. Vol. 77. P. 55–62.
7. *Mansfieldth E.* Academic research and industrial innovation: An update of empirical findings // *Research Policy*, 1998. Vol. 26. P. 773–776.
8. *Callaert J., Looy van B., Verbeek A., et al.* Traces of prior art: An analysis of non-patent references found in patent documents // *Scientometrics*, 2006. Vol. 69. No. 1. P. 3–20.

9. *Carpenter M. P., Cooper M., Narin F.* Linkage between basic research and patents // *Research Management*, 1980. Vol. 23. P. 30–35.
10. The measurement of scientific and technological activities: Using patent data as science and technology indicators // *Patent Manual: Proceedings of OECD*. — Paris, 1994.
11. *Griliches Z.* Patent statistics as economic indicators: A survey // *J. Economic Literature*, 1990. Vol. 28. P. 1661–1707.
12. *De Solla Price D. J.* Is technology historically independent of science? A study in statistical historiography // *J. Technology Culture*, 1965. Vol. 6. P. 553–568.
13. *Rosenberg N., Birdzell L. E., Jr.* Science, technology and the Western miracle // *Scientific American*, 1990. No. 263(5). P. 42–54.
14. *Schmoch U., Strauss E., Grupp H., Reiss T.* Indicators of the scientific base of European patents // EUR 15 330 EN: Report to the European Commission. — Brussels/Luxemburg, 1993.
15. *Luwel M., Noyons E. C. M., Moed H. F.* Bibliometric assessment of research performance in Flanders: Policy background and implications // *R&D Management*, 1999. Vol. 29. Is. 2. P. 133–142.
16. *Van Vianen B. G., Moed H. F., Van Raan A. F. J.* An exploration of the science base of recent technology // *Research Policy*, 1990. Vol. 19. P. 61–81.
17. *Schmoch U.* Tracing the knowledge transfer from science to technology as reflected in patent indicators // *Scientometrics*, 1993. Vol. 26. P. 193–211.
18. *Krauskopf M., Krauskopf E., Mendez B.* Low awareness of the link between science and innovation affects public policies in developing countries: The Chilean case // *Scientometrics*, 2007. Vol. 72. No. 1. P. 93–103.
19. *Chen Y., Puttitanun T.* Intellectual property rights and innovation in developing countries // *J. Development Economics*, 2005. Vol. 79. P. 474–493.
20. *Albert A., Granadino B., Plaza L. M.* Scientific and technological performance evaluation of the Spanish Council for Scientific Research (CSIC) in the field of Biotechnology // *Scientometrics*, 2007. Vol. 70. No. 1. P. 41–51.
21. *Mcmillan G. S., Narin F., Deeds D. L.* An analysis of the critical role of public science in innovation: The case of Biotechnology // *Research Policy*, 2000. Vol. 29. P. 1–8.
22. *Tijssen R. J. W., Buter R. K., Van Leeuwen Th. N.* Technological relevance of science: An assessment of citation linkages between patents and research papers // *Scientometrics*, 2000. Vol. 47. P. 389–412.
23. *Raan van A.* Advanced bibliometric methods for the evaluation of universities // *Scientometrics*, 1999. Vol. 45. No. 3. P. 417–423.
24. *Boyack K. W., Klavans R., Boerner K.* Mapping the backbone of science // *Scientometrics*, 2005. Vol. 64. No. 3. P. 351–374.
25. *Börner K., Penumathy S., Meiss M., Ke W.* Mapping the diffusion of scholarly knowledge among major U.S. research institutions // *Scientometrics*, 2006. Vol. 68. No. 3. P. 415–426.
26. *Glaenzel W., Meyer M.* Patents cited in the scientific literature: An exploratory study of ‘reverse’ relations // *Scientometrics*, 2003. Vol. 58. No. 2. P. 415–428.

ИНТЕЛЛЕКТУАЛЬНЫЕ МЕХАНИЗМЫ СЕМАНТИЧЕСКОГО ПОИСКА В СЕТИ ИНТЕРНЕТ

И. П. Кузнецов¹, М. М. Шарнин², А. Г. Мацкевич³

Аннотация: Рассматриваются системы поиска и сбора информации в сети Интернет, основанные на использовании лингвистических процессоров (ЛП), обеспечивающих автоматическое выделение информационных объектов (именованных сущностей) и их связей из текстов естественного языка (ЕЯ). В результате формируются структуры знаний. Аналогичным образом формируется структура запроса. Поиск, называемый семантическим, обеспечивается за счет сопоставления таких структур, где учитываются связи объектов, а также их участие в событиях, действиях. Рассматриваются различные виды семантического поиска и их особенности.

Ключевые слова: семантико-ориентированный лингвистический процессор; извлечение знаний из текстов; семантический поиск; база знаний

1 Введение

Одно из важных направлений совершенствования систем поиска информации в сети Интернет связано с повышением их полноты и точности. Имеется в виду информация, представленная в виде текстов на ЕЯ. По запросу пользователя должны выдаваться сведения, которые ему требуются. В существующих поисковых системах (Яндекс, Google и др.) используется так называемый **текстовый поиск**, т. е. поиск по ключевым словам. При этом учитывается степень близости слов в плане их расположения в предложении (тексте). Для каждого слова находятся близкие по смыслу слова или словоформы, которые также участвуют в поиске. Это позволяет расширить поисковый запрос и найти большее число «адекватных» документов. Последние ранжируются: выделяются их компоненты с большей плотностью ключевых слов, которые и выдаются пользователю.

Здесь возникает ряд трудностей. Во-первых, часто выдаются предложения, в которых ключевые слова оказываются не связанными, т. е. относятся к различным «сущностям». Например, при поиске человека по ФИО сравнительно часто выдаются предложения, где идет перечисление лиц, но запрашиваемые фамилия, имя, отчество относятся к разным лицам, которые расположены рядом.

¹Институт проблем информатики Российской академии наук, igor-kuz@mtu-net.ru

²Институт проблем информатики Российской академии наук, keywen1@mail.ru

³Институт проблем информатики Российской академии наук, xmag@mail.ru

Во-вторых, как правило, не учитываются данные, которые уточняют запрос, например профессия лица, его место жительства и др. Такие данные могут находиться в документе на значительном расстоянии от самого лица.

В связи с этим возникло новое направление — **семантический поиск**, когда при сравнении запроса и документов учитываются семантические связи. Многие крупные компании инициируют работы в этом направлении. Следует отметить тематическую базу Powerset (купленную Microsoft), которая работает с предварительно формализованной и структурированной информацией. Система SearchMonkey компании Yahoo использует семантические аннотации для формирования более полного ответа. Компании Яндекс и Google постоянно вводят в поисковые механизмы расширения, позволяющие учитывать семантику слов, а в ряде случаев — фраз. Патентуются технологии, в которых учитываются свойства и действия запрашиваемых объектов [1, 2].

Все большее развитие получают семантические технологии: язык онтологий OWL (Web Ontology Language), язык описания ресурсов RDF (Resource Description Framework), язык запросов SPARQL, — позволяющие расширить пространство поиска и учитывать связи. Такие технологии планируются как основа нового типа поисковых машин — Семантического WEB (WEB-3), где предлагается учитывать, какого типа объекты необходимы пользователю и каким должно быть их содержимое. Для этого разрабатываются новые технологии встраивания семантики в существующие веб-страницы.

Из сказанного следует, что семантический поиск — это технология будущего. Однако его использование рационально только для сложных запросов, где важную роль играют связи. С простыми запросами легко справляются обычные поисковые системы.

Направление, связанное с семантическим поиском, интенсивно развивается. Следует отметить информационно-аналитическую систему «Медиалогия», учитывающую лингвистическое окружение при поиске объектов, «интеллектуальную» поисковую систему Нигма.РФ, которая структурирует найденную информацию и способна осуществлять поиск не только по текстам, но и по математическим формулам, химическим соединениям и др.

Дальнейшее развитие семантического поиска связано, во-первых, с развитием методик выделения из запросов и документов семантических компонентов и связей. Для этого требуется глубинный анализ текстов — построение соответствующих ЛП. А во-вторых, с техникой сопоставления запросов при выборе адекватных документов.

2 Особенности поиска информационных объектов

В то же время в ИПИ РАН в рамках плановых тем («Криминал», «Аналитик», «Поток») уже более 15 лет развивается направление, связанное с автома-

тической формализацией текстов на ЕЯ, выявлением семантической информации (структур знаний) и реализацией на этой основе различных видов семантических поисков. Учитывается тот факт, что большинство пользователей — специалистов в своей области — интересуются лишь конкретными вещами для своих задач. Например, следователям важны фигуранты, их местожительства, телефоны, приметы, действия лиц (с указанием места, времени), связи и др. Такая информация является основой для оперативно-розыскных действий, поиска по связям, различных видов анализа [3–5]. Других интересуют памятные места, их местонахождение, автор, архитектор и т. д. [6, 7]. Подобную информацию будем называть **информационными объектами** (другое название — «именованные сущности»), которые различаются по типам. Например, лица и фигуранты — это объекты одного типа, адреса — другого и т. д.

Отметим, что связи между объектами, которые интересуют пользователей, могут иметь высокую степень разнообразия. Например, интерес может представлять не только связь лиц между собой или с анкетными данными, но и действия или события, в которых эти лица участвуют. Такие события привязаны к времени, месту. Более того, одни события могут быть составной частью других и т. д. Их тоже нужно искать и обрабатывать. Поэтому следует считать, что действия и события — это тоже информационные объекты, связанные между собой и с другими информационными объектами. Возникают сложные структуры знаний.

Для использования таких структур в различных приложениях в рамках проектов ИПИ РАН разработаны:

- язык для представления структур знаний (объектов и их связей, в том числе фактов участия объектов в действиях или событиях);
- семантико-ориентированные ЛП, выделяющие из запросов и документов на ЕЯ информационные объекты и их связи;
- процедуры (поисковые машины), реализующие различные виды поиска объектов на основе структур знаний.

В процессе разработок создана научная база: расширенные семантические сети (РСС), методики представления сложных видов знаний, инструментальная среда ДЕKL обработки структур знаний, семантико-ориентированные ЛП для извлечения знаний из текстов [6, 8, 9]. Представляется перспективным использование этой базы и накопленного опыта в качестве основы развития новых семантических технологий при обработке и поиске в больших массивах неформализованной информации (в сети Интернет), представленной в виде текстов на ЕЯ.

В данной статье рассматривается методика семантического поиска для случая, когда из запроса выделен информационный объект и его связи. В результате запрос преобразуется в структуру на РСС. Необходимо найти в сети Интернет аналогичные или похожие объекты.

На первом этапе идет поиск текстов, в которых может находиться запрашиваемый объект. Для этого из запроса выделяются ключевые слова, которым присваиваются веса по степени их значимости или уникальности в плане уточнения объекта. По ним идет обращение к стандартным поисковикам (например, Google, Яндекс и др.), которые на основе ключевых слов находят соответствующие документы. Из них выделяются связанные предложения с наибольшей плотностью ключевых слов. Такие предложения образуют исходный материал анализа. На последующих этапах предложения преобразуются ЛП в структуры знаний, которые накапливаются и образуют **оперативную базу знаний** (БЗ). Она служит основой для семантических поисков, которые сводятся к сопоставлению структуры запроса (на РСС) и содержимого БЗ. В результате находятся информационные объекты — аналогичные запрашиваемому или похожие на него. На их основе формируется ответ.

Подобная методика (в своей основной части) использована в системе «Криминал», где на предварительном этапе криминальные документы (сводки происшествий, обвинительные заключения и др.) обрабатывались ЛП. Полученные структуры запоминались в долговременной БЗ, из которой по ключевым словам выбирались структуры, которые составляли оперативную БЗ. Она служила основой для различных видов поиска, в которых учитывались всевозможные виды связей. В данной статье рассматривается развитие данной методики применительно к задаче сбора сведений в сети Интернет об интересующем пользователя объекте. Конечно, речь идет об объектах, которые выделяются ЛП, т. е. методика работает в предметной области, на которую настроен данный процессор.

3 Семантико-ориентированный лингвистический процессор

Извлечение знаний из текстов осуществляется семантико-ориентированным ЛП, который является основой предлагаемой методики семантического поиска в сети Интернет. Такие ЛП использовались в системах «Криминал» и «Аналитик» для формирования семантических структур (СС) документов и реализации объектных поисков.

Рассмотрим основные компоненты семантико-ориентированного ЛП.

3.1 Блок лексико-морфологического анализа

Выделяет из документа на ЕЯ слова и предложения и выдает в виде семантической сети (РСС), представляющей последовательность компонентов (слов в нормальной форме, чисел, знаков) и их основные признаки. Такую сеть будем называть **пространственной структурой** (ПС) **документа**. Блок использует набор тематических словарей (словарь стран, регионов России, имен, видов оружия и др.) для придания словам и словосочетаниям дополнительных семантических признаков [7].

3.2 Блок синтактико-семантического анализа

Путем анализа ПС документа блок выделяет объекты и связи. На их основе строит другую семантическую сеть (РСС), представляющую **семантическую структуру документа**, которую также называют **содержательным портретом документа** [6, 8, 10]. Блок управляется **лингвистическими знаниями** (ЛЗ), за счет которых обеспечивается:

- извлечение информационных объектов (лиц, организаций, событий, их места и времени и многое другое), их идентификация с учетом анафорических ссылок, сокращений;
- выявление связей объектов (например, как лица связаны с организациями, адресами и др.);
- анализ глагольных форм, причастных и деепричастных оборотов с выявлением фактов участия объектов в соответствующих действиях;
- выявление связей действий с объектами типа место или время (где и когда имело данное действие или событие);
- анализ причинно-следственных и временных связей между действиями и событиями.

Лингвистические знания также записываются на языке РСС — в виде специальных фрагментов и правил.

3.3 Экспертные системы, логический вывод

На основе СС документа формируют новые знания, которые дополняют СС документа [3, 7, 10]. Например, по действиям лица определяют его роль в происшествии (потерпевший, обвиняемый) и т. д.; относят криминальные происшествия к определенным типам (в соответствии с классификаторами криминальной полиции).

3.4 База лингвистических и экспертных знаний

Содержит правила анализа текста и экспертных решений во внутреннем представлении (РСС). Они определяют работу ЛП.

3.5 Обратный лингвистический процессор

Преобразует СС документа или вновь сформированные знания в текст на ЕЯ, который выдается пользователю в качестве ответа.

4 Семантическая структура документа и запроса

Семантическая структура документа состоит из фрагментов языка РСС, представляющих выделенные объекты, их признаки и связи, а также участие в действиях. Такие структуры, с одной стороны, должны представлять сложные виды информации, а с другой стороны — быть удобными с точки зрения обработки и поиска. Поэтому для их представления выбран язык РСС, а для обработки — инструментальная среда ДЕKL. Множество СС документов образует БЗ.

Пример 1. Документ (с номером 26) взят из сводок происшествий ГУВД г. Москвы:

01.02.98 года в 22.30 час заявление гр-ки Ивановой Л. А., 1979 года рожд., прож.: ул. Яна Райниса, дом 5-5-14. О привлечении к уголовной ответственности мужа Иванова М. И., 1964 года рожд., который по месту жительства учинил скандал и угрожал убийством. На месте уч. insp. Голубевым был задержан.

Семантическая структура данного документа имеет вид:

ДОК_(26, "1-02-98", "СВОДКА"/0+) 0-(RUS)
ФИО(ИВАНОВА, Л., А., 1979/1+)
АДР_(УЛ., ЯНА, РАЙНИСА, ДОМ, 5, 5, 14/2+)
ПРОЖ.(1-, 2-)
ФИО(ИВАНОВ, М., И., 1964/3+)
МУЖ(1-, 3-/4+)
СООБЩИТЬ(1-/5+) 5-(26, АСТ_-)
ДАТА_(1998, 02, ~01, 22.30/6+)
КОГДА(5-, 6-)
"ПРИВЛЕЧЕНИЕ К УГОЛОВНОЙ ОТВЕТСТВЕННОСТИ"(3-/7+) 7-(26, АСТ_-)
СКАНДАЛ(3-/8+) 8-(26, АСТ_-)
"УГРОЖАТЬ УБИЙСТВОМ"(3-/9+) 9-(26, АСТ_-)
МИЛ_(уч. insp., ГОЛУБЕВ/10+)
ЗАДЕРЖАТЬ(МЕСТО, 10-/11+) 11-(26, АСТ_-)
ПРЕДЛ_(26, 5-, 2-/12+) 12-(1, 14, 128)
ПРЕДЛ_(26, 7-, 8-, 9-/13+) 13-(3, 129, 283)
ПРЕДЛ_(26, 11-/14+) 14-(5, 284, 324)
ANAL_(“Преступные действия”, ХУЛИГАНСТВО, СКАНДАЛ/15+)
15-(26, ANAL_-)

Семантическая структура документа состоит из элементарных фрагментов, аргументами которых являются слова в нормальной форме (например, для существительных — в единственном числе, именительном падеже, для прилагательных — дополнительно мужской род и т. д.). Это необходимо для поиска и обработки. Каждый элементарный фрагмент имеет свой уникальный код, который записывается в виде числа со знаком «+» и отделяется косой линией.

Так, во фрагменте ФИО(ИВАНОВА,Л.,А.,1979/1+) знак «1+» есть его код, а «1-» — ссылки на него). Например, во фрагменте ПРОЖ.(1-,2-) знаки «1-» и «2-» означают, что лицо (Иванова Л. А.) проживает по адресу, представленному в АДР_(УЛ.,ЯНА,РАЙНИСА,ДОМ,5,5,14/2+). Такие знаки (1+, 2+, . . .) играют роль идентификаторов объектов во внутреннем представлении.

Фрагменты ДОК_(26, “1-02-98.ТХТ”, “СВОДКА” /0+) 0-(RUS) указывают, что СС построена на основе русскоязычного текста документа с номером 26 из файла 1-02-98.ТХТ”, который обрабатывался как сводка происшествий (от этого зависят ЛЗ). Знаки 1+, 1-, 2+, 2-, 3+, 3- . . . — это коды фрагментов, с помощью которых задаются их связи и отношения. Например, фрагмент МУЖ(1-,3-/4+) представляет отношение «быть мужем» двух лиц (представленных как ФИО с кодами 1+ и 3+). Действия также представляются в виде фрагментов типа СКАНДАЛ(3-/8+) 8-(26, АСТ_), где представлено, что «лицо (ФИО с кодом 3+), *учинило скандал*». С помощью кода (8+,8-) указывается, что фрагмент представляет действие (АСТ_) и относится к документу с номером 26. Такие коды также служат для представления времени, места действия и фактов их комбинирования — когда одно действие включено в состав другого. Следующие фрагменты ДАТА_(. . . /6+) и КОГДА (5-,6-) представляют время (ДАТА_) и что оно относится к действию «*сообщить*» (код 5+).

Особую роль играют фрагменты ПРЕДЛ_(. . .), которые соответствуют предложениям. Они заполняются словами, не вошедшими в информационные объекты, а также кодами самих объектов. К этим фрагментам добавляются указатели их местоположения в тексте. Например, фрагменты ПРЕДЛ_(26, 5-,2-/12+) и 12-(1,14,128) представляют тот факт, что объекты с кодами «5-» (соответствует действию «*сообщить*» с его субъектом и временем), «2-» (соответствует адресу) и др. находятся в предложении, которое начинается с 1-й строки текста документа и занимают место от 14-го байта до 128-го.

Фрагмент ANAL_(. . .) формируется экспертной системой, которая оценивает документ и относит его к определенному пункту в соответствии с классификаторами криминальной полиции. В данном примере система определила «преступные действия» как ХУЛИГАНСТВО с уточнением — СКАНДАЛ. Такие фрагменты крайне необходимы при поиске похожих происшествий (см. разд. 9).

Помимо сказанного, к СС документа добавляются фрагменты DESC_(. . .), представляющие описание объектов и действий:

DESC_(1-, “гр-ки Ивановой Л. А. 1979 года рожд.”)

DESC_(2-, “прож.: ул. Яна Райниса дом 5-5-14”) . . .

В первом фрагменте содержится описание лица (представленного фрагментом с кодом 1+), взятого из текста и т. д. Такие фрагменты необходимы для работы обратного ЛП.

Аналогичным образом представляется **семантическая структура запроса**, что упрощает его сопоставление с СС документов при поиске и оценке степени близости. Например, вопрос «Кто скандалил и угрожал убийством?» представляется следующим образом:

ФИО(' ',' ',1/15+) СКАНДАЛ(15-) “УГРОЖАТЬ УБИЙСТВОМ”(15-)?(15-),

где запрашивается (?) неизвестный (ФИО), а словосочетания «*учинить скандал*» и «*скандалил*» преобразуются в СКАНДАЛ в процессе нормализации [7, 11].

Использование СС документов позволяет учитывать при поиске и идентификации объектов гораздо большее число факторов, чем в других поисковых системах, например участие лица в действиях, время этого действия, связи лица и др.

5 Оперативная база знаний

Под **оперативной базой знаний** понимается множество отобранных СС документов, которые используются для поиска ответа на запрос. Рассмотрим, как формируется такая база.

Методика основана на использовании ЛП и стандартных поисковых систем (типа Google, Яндекс). Для полноты поиска используется несколько поисковых систем, на которые подается один и тот же запрос.

Вначале запрос обрабатывается ЛП. Все его слова нормализуются. Формируется СС запроса. Каждому нормализованному слову присваивается вес в зависимости от степени его информативности в плане поиска объектов. Будем называть такие **слова ключевыми**. Некоторым словам присваивается **статус обязательности**. Такой статус может быть задан в запросе, или же система сама отбирает для этого слова с учетом степени их уникальности. Ключевые слова подаются на вход упомянутых поисковых систем, которые находят документы.

Из этих документов специальной программой выделяются связанные предложения с наибольшей плотностью ключевых слов. Далее отсеиваются предложения, в которых нет обязательных слов. Подсчитывается вес каждого оставшегося предложения — как сумма весов находящихся в них ключевых слов. Документы (с их URL) ранжируются по весам и выделяются те из них, вес которых превышает заданный порог.

Выделенные предложения образуют исходный материал анализа. Предложения формализуются с помощью ЛП. В результате формируется **оперативная БЗ**. Ответ находится путем сопоставления СС запроса и структур в оперативной БЗ, которая после выдачи ответа архивируется или уничтожается.

При использовании предлагаемой методики возникает проблема выделения из документов **связанных предложений**, на основе которых формируется оперативная БЗ (такое выделение осуществляется до их синтактико-семантического анализа). Если брать только предложения с наибольшей плотностью ключевых

слов или с наибольшим их суммарным весом, то может быть потеряна значимая информация, относящаяся к описанию связей. В текстах на ЕЯ многие связи даются с помощью анафорических ссылок (указательных местоимений), различных форм и оборотов (причастных, деепричастных), кратких описаний. Признаки и связи запрашиваемого объекта могут быть рассредоточены по множеству предложений. Важно выбрать и формализовать именно эти предложения. Простейший способ решения этой проблемы — анализировать наличие в соседних предложениях анафорических ссылок (указательных местоимений, связок «*тот, который*»).

6 Особенности поиска объектов

Ниже будут рассматриваться различные виды **семантического поиска** для случая, когда запрашивается информационный объект, который формализован (преобразован в СС запроса) и по которому уже сформирована оперативная БЗ. Итак, роль запроса играет информационный объект — его компоненты, свойства и связи. Будем называть такой **объект исходным**. Требуется найти в оперативной БЗ объекты с такими же компонентами, свойствами и связями. Поиск идет на уровне СС документов, где учитываются различные виды связей. Поэтому такой поиск относится к классу семантических. Степень близости или «похожести» найденных или отобранных объектов (по отношению к исходному объекту) оценивается с помощью веса объекта, который подсчитывается с учетом весов совпавших слов. Отметим, что многие виды семантического поиска нашли свое применение в системах «Криминал» и «Аналитик» [3, 4, 6, 11].

6.1 Элементы исходного объекта

Исходный объект (СС запроса) содержит, во-первых, упорядоченное **множество элементов**, каждый из которых есть или слово в нормальной форме, или число, или какой-либо символ, и, во-вторых, связи, представленные в виде фрагментов РСС. Элементы различаются по их значимости.

Имеются слова, которые не учитываются при анализе текста. Они удаляются сразу и не входят в объект (СС запроса). К ним относятся частицы, междометия и слова с минимальной смысловой нагрузкой (ПРИМЕРНО, НЕКОТОРЫЙ, ЛЮБОЙ, САМЫЙ, НАВЕК и др.)

Для всех информационных объектов задается множество **незначимых слов**, которые встречаются в объектах, но не должны учитываться при объектном поиске. К ним относятся указательные местоимения, которые не были идентифицированы, предлоги и общезначимые слова ВРЕМЯ, ПРЕДМЕТ, СОСТАВ и др.

Для каждого типа информационных объектов задаются **малозначимые слова**, которые в малой степени (с малыми весами) должны учитываться при

поиске по исходному объекту определенного типа. Это слова, которые часто встречаются в объектах или которые играют роль **типизаторов** — указателей категории следующего слова.

Примеры типизаторов для дат: ГОД, МЕСЯЦ, ЧАС, МИН., СЕК., СУТКИ.

Примеры типизаторов для адресов: ГОР., УЛ., ДОМ, КВ., ОБЛАСТЬ.

Многие информационные объекты имеют стандартные компоненты. Например, для лиц — это фамилия, имя, отчество, год рождения. Для адресов — название города, улицы, номер дома, квартиры. Это **значимые элементы**, которые будем называть **типизированными** (перед ними могут стоять слова-типизаторы). Помимо этого, имеются **сокращения** типа «г.», «пер.», «след.» и др. Многие из них уточняются в процессе анализа текста и заменяются обычными словами. Но для некоторых сокращений это невозможно, например для инициалов.

Значимым называется **элемент** исходного объекта, который не относится к множеству незначимых и малозначимых слов. Степень значимости элемента зависит от его уникальности и характеристических свойств. Такая степень задается весами.

Некоторым элементам присваивается **статус обязательности**. Обычно это типизированные элементы (слова), например фамилии или некоторые (редко встречающиеся) значимые слова. Статус зависит от типа объекта. В этом случае производится поиск объектов только с такими элементами.

6.2 Критерий полноты

Отметим, что семантический поиск (если нет связей) не эффективен, когда исходный объект состоит из небольшого числа часто встречающихся элементов — слов. Например, поиск адресов по слову «Москва» или «Россия, Москва». Или поиск лиц по слову «Иван». Тогда лучше работают обычные поиски. Поэтому вводится понятие **полноты объекта**, который определяет, когда следует включать семантический поиск.

Критерий полноты объекта зависит от его типа и задается шаблонами. Например, шаблон <слово><слово><число> означает, что объект должен содержать два значимых элемента-слова и один элемент-число. Порядок не играет роли (в простейшем случае).

Для объектов каждого типа задается свой шаблон или несколько шаблонов. Если исходный объект определенного типа удовлетворяет одному из его шаблонов, то это значит, что он обладает критерием полноты и для него следует включать объектный поиск.

Приведем примеры шаблонов, определяющих полноту объектов (в системе «Криминал»).

Шаблоны для адресов:

<число><число><слово> — наличие в адресе двух чисел и значимого слова.
<УЛ.,ДОМ,ПРОСПЕКТ,ПРОЕЗД,ДЕР.><слово><число>>.

Шаблон для автотранспорта: <число><число><слово> и т. д.

Критерии полноты могут изменяться в зависимости от предметной области.

6.3 Виды семантического поиска

Рассмотрим виды семантического поиска, реализованные в логико-аналитической системе «Криминал», предназначенной для решения задач оперативных работников [3, 4, 12].

Точный поиск объектов. Требуется полноты исходного объекта и наличия всех обязательных элементов у найденных объектов. При точном поиске лиц требуется наличие полных ФИО (или инициалов вместо имени и отчества) и совпадение всех элементов у найденных объектов. При этом допускается совпадение имен, отчеств и инициалов — по первым буквам. Такое *совпадение* будем называть *частичным*. При поиске учитывается год рождения, который считается факультативным элементом ФИО.

Поиск похожих объектов. Не требуется наличия всех значимых элементов запроса у найденных объектов в оперативной БЗ. По совпавшим элементам подсчитывается вес найденного объекта (он определяет степень его близости к исходному объекту). Если не совпадают типизированные элементы, то из общей суммы вычитается определенный вес, например, когда при сравнении лиц не совпали отчества или инициалы.

Поиск похожих объектов с учетом онтологий. Онтологии используются при сопоставлении объекта в СС запроса с объектами оперативной БЗ. Допускается частичное совпадение элементов (с учетом связей типа «род–вид», сокращений, или степени близости слов). По совпавшим элементам подсчитывается вес найденного объекта.

Поиск похожих объектов с учетом связей. Признаки и связи являются источником ключевых слов и значимых элементов, которые дополняют множество значимых элементов исходного объекта. Например, это может быть кличка, имя отношения или действия, в котором участвует объект и др. Все элементы полученного множества участвуют в поиске и подсчете весов. По совпавшим элементам подсчитывается вес. Пример — поиск лиц без ФИО по приметам или лиц с частично заданными ФИО.

Навигация по связям (см. разд. 10). По исходному объекту (на основе фрагментов СС запроса) производится поиск связанных с ним объектов. Для каждого из них ищутся аналогичные объекты (из других документов) и связанные с ними объекты и т. д. В результате строится граф связей.

Отметим, что тип исходного объекта во многом определяет эффективность поиска. Например, при поиске телефонов, сайтов, дат, документов, номерных вещей, статей УК (где числовые данные определяют объект) работают только точные поиски. Онтологии в основном требуются при поиске похожих действий, поиске по приметам (реже — для автотранспорта, оружия). Поиск похожих объектов с учетом связей — в основном для лиц без ФИО.

Будем называть **объекты**, полученные в результате точного поиска, **аналогичными**, иначе — **похожими**.

7 Онтологии в виде расширенной семантической сети

Онтологии служат при формировании оперативной БЗ (за счет размножения значимых элементов исходного объекта), а также при сопоставлении СС запроса и оперативной БЗ.

Все элементы полученного множества участвуют в поиске и подсчете весов. Это необходимо для расширения пространства поиска с целью учета различных способов и средств описания того, что есть в исходном объекте. В результате повышается точность и надежность результатов, обеспечивается достаточная свобода использования слов и терминов в запросах и документах.

В системах «Криминал» и «Аналитик» онтологии имеют вид терминологического словаря, который определяет семантическое пространство терминов и признаков, используемых при поиске и оценке степени близости. Рассмотрим основные компоненты такого словаря.

7.1 Близкие по смыслу слова

Представляются с помощью фрагментов NEAR.

Например, фрагмент NEAR(ПРОДАЖА,ПРОДАВЕЦ), во-первых, служит для размножения ключевых слов запроса, а во-вторых, означает возможность сопоставления этих двух слов: ПРОДАЖА <-> ПРОДАВЕЦ.

Аналогично

NEAR(АЛКОГОЛЬ,ПЬЯНЫЙ,АЛКОГЛИК,ВЫПИВКА,БУТЫЛКА)

означает:

ПЬЯНЫЙ <-> АЛКОГОЛИК, ПЬЯНЫЙ <-> ВЫПИВКА. . .

Другой пример:

NEAR(«УГРОЖАТЬ НОЖОМ», «УГРОЖАТЬ УБИЙСТВОМ». . .).

Отметим, что многие такие слова приводятся к одному виду в процессе нормализации. Для них фрагментов NEAR не требуется.

7.2 Отношение класс–подкласс (род–вид)

Представляются с помощью фрагментов SUB. Образуются так называемые SUB-деревья. Например:

SUB(ПРОНИКАТЬ,ПРОЛЕЗТЬ)

SUB(ПРОНИКАТЬ,ПРОЙТИ)

SUB(ПРОНИКАТЬ,ВОЙТИ)

означает

ПРОНИКАТЬ <-> ПРОЛЕЗТЬ, ПРОЙТИ. . .

Аналогично

SUB(ВЕСНА,МАРТ)

SUB(ВЕСНА,АПРЕЛЬ)

SUB(ВЕСНА,МАЙ)

означают:

ВЕСНА <-> МАРТ, ВЕСНА <-> АПРЕЛЬ. . .

7.3 Несовместимые элементы

Представляются фрагментами типа OR_OR (разделительное ИЛИ). Например:

OR_OR(МУЖЧИНА,ЖЕНЩИНА,РЕБЕНОК)

OR_OR(АТЛЕТИЧЕСКИЙ,ХИЛЫЙ).

При сопоставлении учитываются со знаком минус (к общему весу).

В реальности при поиске похожих объектов система учитывает значительно большее число несовместимых элементов, которые не вводятся в словарь. Например, при поиске по ФИО учитывается несовпадение имен, отчеств и инициалов. При поиске похожих адресов учитывается несовпадение номеров домов, квартир, названий улиц и др. При этом величина, которая вычитается из общего веса (или добавляется к нему), зависит от типа не совпавших элементов и степени близости слов или чисел.

7.4 Особенности сопоставления элементов-чисел

Имеется в виду сопоставление фрагментов типа ВОЗРАСТ, РОСТ (для лиц), номеров домов, квартир, времени действия и др., представленных в СС запроса и в оперативной БЗ. При этом рассматриваются следующие варианты:

- равенство чисел;
- близость числовых значений;
- число входит в интервал;
- пересечение интервалов.

Близость числовых значений зависит от признаков или отношений с числовыми значениями и, соответственно, определяет вес, добавляемый к объекту. Таким способом нивелируются статистические ошибки, которые могут быть в запросе или документах.

8 Этапы семантического поиска объектов

Первый этап. Выделение значимых элементов исходного объекта с присвоением некоторым из них статуса обязательности. Значимым элементам присваиваются веса: типизированным элементам — наибольший вес, малозначимым элементам — наименьший. При этом учитываются связи объекта, которым тоже присваивается статус и вес.

Второй этап. В оперативной БЗ ищутся объекты такого же типа, содержащие все обязательные элементы. У найденных объектов может быть больше элементов, чем у исходного объекта. При этом учитываются связи исходного объекта и выделенные по ним элементы.

Третий этап. Подсчет весов найденных объектов. При совпадении элементов к весу объекта добавляется вес элемента. При этом учитываются онтологии, определяющие близость элементов или их противоречивость. Элементы, выделенные за счет связей, сравниваются с элементами, имеющими аналогичные связи. Например, клички сравниваются с кличками, наличие родственников с родственниками и т. д. Таким способом идет подсчет весов с учетом структуры связей.

Вес объекта равен сумме весов совпавших элементов.

Четвертый этап. Объекты упорядочиваются в соответствии с критерием сходства. В простейшем случае — это ранжирование по весам и выдача в порядке весов. Ясно, что у всех этих объектов должны присутствовать обязательные элементы исходного объекта.

При **поиске лиц** по ФИО имеются некоторые особенности. Учитывается частичное совпадение: имен (отчеств) с инициалами, инициалов с именами (отчествами) и инициалы с инициалами. Тогда к общему весу найденного объекта добавляется усредненное число.

При **поиске похожих объектов** не требуется совпадения всех элементов. При несовпадении элементов (в том числе типизированных) из общей суммы вычитается определенный вес. Например, при поиске похожих лиц ФИО могут быть не полными, могут не совпадать (или отсутствовать) имена, отчества или год рождения. Это приведет к уменьшению веса найденного лица.

При **поиске объектов с учетом связей** различаются два случая. Первый — когда в исходном объекте есть значимые элементы, по которым найдены объекты в оперативной БЗ. Тогда идет сравнение исходного и найденного объектов по их элементам, а также по признакам и связям. Например, если исходный

и найденный объекты связаны одним и тем же отношением с двумя другими объектами, то вес, добавляемый к найденному объекту, определяется степенью близости (с учетом онтологий) последних объектов.

Второй случай — когда в исходном объекте нет значимых элементов, достаточных для поиска. Например, поиск неизвестного лица (без ФИО) по приметам. Тогда осуществляется поиск похожих объектов на основе признаков и связей. Выделяются признаки исходного объекта и связи, например приметы неизвестного. Для них ищутся похожие, например такие же приметы (с учетом онтологий), которые связаны с каким-либо лицом. Для данного лица и подсчитывается вес — как сумма весов элементов, выделенных по связям.

9 Поиск похожих действий и событий

Поиск похожих действий (это тоже объекты) проходит те же этапы (см. разд. 8). Особенность — в выделении значимых элементов. Если в исходном действии принимает участие какой-либо объект, то в найденном действии (с тем же именем) ищется объект такого же типа. Для этого используется типология объектов, определяемая ЛП, а также фрагменты NEAR и SUB, задающие онтологию.

Например, если в исходном действии участвует лицо, то в найденном действии с тем же именем тоже ищется лицо. Эти лица сравниваются по степени похожести. В результате вес действия подсчитывается как сумма весов участвующих в них объектов.

Если происшествие или событие состоит из множества действий, то для каждого действия находятся похожие (в анализируемом документе). При этом учитывается участие одних и тех же объектов (лиц) в различных действиях, т. е. сопоставляются структуры. В результате вычисляются веса действий по их степени сходства при наличии одинаковых связей. Найденные происшествия ранжируются. Для каждого из них выдается объяснение причины сходства — в виде текстов (взятых из фрагментов DESC-), сопоставленных совпавшим объектам и действиям.

Например, на происшествие с номером 74 «. . . *Неизвестный скандалил и угрожал ножом. . . был задержан. . .*» будут выданы похожие происшествия:

Похожие происшествия на док-т 74:

Происш. 26: Угроза убийством (вес 47).

. . .

Происш. 59: Злостное хулиганство (вес 27).

. . .

При выборе 1-го пункта этого меню (происшествия 26 из раздела 3) дается объяснение:

В документе 26 встретились те же признаки:

Преступные действия: ХУЛИГАНСТВО, СКАНДАЛ.

Те же деяния: «неизвестный скандал».

Те же деяния: «и угрожал ножом» (близкое — «угрожать убийством»).

...

Общие сведения: «. . . был задержан».

...

<Текст документа 26>.

Особую роль при поиске похожих событий и действий играют аналитические фрагменты (ANAL_). С их помощью осуществляется проверка того, что события (происшествия), для которых оценивается степень похожести, по категориям «преступные действия», «способ совершения преступления», «орудие преступления» и др. относятся к одному и тому же типу. При их совпадении анализируемым документам присваивается наибольший вес.

10 Навигация по связям

По описанию объекта ЛП строится его формализованное представление в виде СС запроса. Далее формируется оперативная БЗ, в которой ищутся аналогичные объекты — точный поиск. Они образуют один или множество исходных объектов. Рассмотрим простейший вариант, когда исходный объект — один.

Для исходного объекта строится граф связей — прямых и косвенных.

Прямые связи — связи с другими объектами в рамках одного документа (происшествия). Например, для фигуранта — это адрес проживания, владение оружием с определенными характеристиками и др. Это **связи 1-го уровня**.

Косвенные связи — это связи через другие объекты, находящиеся в разных документах. Например, связанными считаются два фигуранта (из разных документов), если они входят в происшествия, в которых встретился один и тот же телефон или какой-либо другой объект, принадлежащий и тому и другому фигуранту. Это **связи 2-го уровня**.

Объекты, представляющие связи 1-го уровня, выделяются по фрагментам СС запроса, в которых представлены отношения или действия исходного объекта. Выделенные объекты должны удовлетворять критерию полноты. Далее по выделенным объектам и фрагментам, представляющим их косвенные связи, ищутся объекты второго уровня. Для найденных объектов также может осуществляться поиск связей. Это связи 3-го и более высоких уровней.

Более подробную информацию об извлечении знаний и объектных поисках в рамках предлагаемой концепции можно получить на сайте [11].

11 Заключение

В связи с развитием сети Интернет научное направление, имеющее целью построение объектно-ориентированных поисковых машин, приобретает особую степень важности. Предлагаемые в данной статье подходы могут быть основой новых поколений Семантических WEB, где семантические связи выявляются автоматически и обеспечивают более точный и качественный поиск.

Литература

1. *Charnine M.* System and method for updating facts in a fact repository. U.S. Patent 7 739 212, 2010.
2. *Charnine M.* Fact-based object merging, U.S. Patent 7 966 291, 2011.
3. *Кузнецов И. П., Козеренко Е. Б., Шарнин М. М.* Семантико-ориентированная система фактографического поиска с входом на русском и английском языках // Диалог-1998: Тр. Междунар. семинара по компьютерной лингвистике и ее приложениям. Т. 2. — Казань: Хэтер, 1998. С. 821–830.
4. *Кузнецов И. П., Мацкевич А. Г.* Методы поиска похожих объектов и событий на основе признаков и связей // Диалог-1999: Тр. Междунар. семинара по компьютерной лингвистике и ее приложениям. Т. 2. — Таруса, 1999. С. 164–171.
5. *Кузнецов И. П.* Методы обработки сводок с выделением особенностей фигурантов и происшествий // Диалог-1999: Тр. Междунар. семинара по компьютерной лингвистике и ее приложениям. Т. 2. — Таруса, 1999. С. 691–700.
6. *Кузнецов И. П., Козеренко Е. Б., Мацкевич А. Г.* Принципы организации объектно-ориентированных систем обработки неформализованной информации // Искусственный интеллект: Журнал НАН Украины, 2010. Вып. 3. С. 227–237.
7. *Кузнецов И. П., Сомин Н. В.* Особенности настройки объектно-ориентированного лингвистического процессора на тексты предметной области // Системы и средства информатики. Вып. 20. № 2. Методы и технологии, применяемые в научных исследованиях информатики. — М.: ИПИ РАН, 2010. С. 124–148.
8. *Kuznetsov I., Kozerenko E.* The system for extracting semantic information from natural language texts // Conference (International) on Machine Learning (MLMTA-03) Proceedings. — Las Vegas, 2003. P. 75–80.
9. *Кузнецов И. П.* Особенности обработки текстов естественного языка на основе технологии баз знаний // Системы и средства информатики. — М.: Наука, 2003. Вып. 13. С. 241–250.
10. *Kuznetsov I.* Discovery of role functions of persons on the basis of knowledge structures // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам Ежегодной междунар. конф. «Диалог 2011». — М.: РГГУ, 2011. Вып. 10(17). С. 391–402.
11. Лаборатория компьютерной лингвистики ИПИ РАН: Официальный сайт. [www.piranLogos.com](http://piranLogos.com).
12. *Кузнецов И. П., Рабинович Б. И.* Модель базы знаний с возможностью интеграции внешних источников // Системы и средства информатики. — М.: Наука, 2007. Вып. 17. С. 254–272.

МЕТОД КЛАССИФИКАЦИИ ИНФОРМАЦИИ НА ОСНОВЕ ИЕРАРХИЧЕСКИХ ТЕГОВ И ЕГО РЕАЛИЗАЦИЯ НА ПРИМЕРЕ СЕМЕЙНОГО АРХИВНОГО ФОНДА

И. М. Адамович¹, О. И. Волков², Н. А. Маркова³

Аннотация: Рассмотрены вопросы, связанные с современными методами классификации информации, проанализированы их достоинства и недостатки. Описаны предложения по классификации гетерогенной информации на основе иерархических тегов, проанализирован класс задач, для которых такая классификация возможна и целесообразна. Рассмотрена реализация предложенного подхода на примере программы «Семейный архивный фонд» (САФ).

Ключевые слова: классификация; иерархические теги; семейный архивный фонд

1 Введение

При разработке современных информационных систем важную роль играет организация классификации информации.

Классификация объектов, их разнесение по различным классам на основе общих свойств облегчает восприятие общей картины, позволяет ориентироваться в многообразии объектов, способствует установлению связей между ними, определению места того или иного объекта во всей их совокупности, помогает лучше их различать, сопоставлять друг с другом, уяснять их особенности [1].

В настоящее время чаще всего применяются два метода классификации: иерархический и многоаспектный. Характерными особенностями иерархического метода являются:

- наличие неограниченного количества признаков классификации;
- соподчиненность признаков классификации, что выражается в разбиении каждой классификационной группировки, образованной по одному признаку, на множество классификационных группировок по нижестоящему (подчиненному) признаку.

¹Институт проблем информатики Российской академии наук, Adam@amsd.com

²Институт проблем информатики Российской академии наук, Volkov@amsd.com

³Институт проблем информатики Российской академии наук, nMarkova@ipiran.ru

Многоаспектный метод в качестве основания классификации использует параллельно несколько независимых признаков (аспектов). Существует два типа многоаспектной классификации: фасетная и дескрипторная. Фасет — это аспект классификации, который используется для образования независимых классификационных группировок. Дескриптор — ключевое слово, определяющее некоторое понятие, которое формирует описание объекта и указывает на принадлежность этого объекта к классу, группе и т. д. [2].

Фасетная классификация обладает высокой степенью гибкости, простотой модификации без изменения структуры группировок и позволяет использовать для их создания большое число признаков классификации и их значений.

Однако по причине сложности структуры построения и низкой степени заполнения фасетный метод имеет довольно ограниченное распространение: в основном это классификация товаров и продукции. В качестве примера можно привести классификацию вин:

- по срокам выдержки: молодые, марочные, коллекционные;
- по цвету: белые, розовые, красные;
- по технологии: тихие, игристые.

Дескрипторный (описательный) метод классификации, основанный на классификации объектов на естественном языке с помощью ключевых слов, до последнего времени использовался в основном в библиотечных системах поиска (алфавитно-предметные указатели систематических каталогов, каталоги ключевых слов).

В отличие от фасетной и дескрипторной классификации, сфера применения иерархической классификации весьма обширна. Ее достоинства:

- простота построения;
- использование независимых классификационных признаков в различных ветвях иерархической структуры.

Благодаря этому наиболее широкое распространение получила именно иерархическая классификация. Примерами тому могут служить общероссийские классификаторы (ОКУД, ОКАТО, ОКПО и т. п.) [3] и Универсальная десятичная классификация (УДК) [4].

Рассмотрим альтернативный способ классификации информации, объединяющий достоинства дескрипторного и иерархического методов, и оценим перспективы его применения в контексте современных тенденций в данной области.

2 Современные тенденции в области классификации

В связи со все большим развитием сети Интернет, появлением и широким распространением социальных сетей и сетевых сервисов возникли новые задачи классификации.

В частности, относительно новым направлением исследований, которое в настоящее время активно развивается за рубежом и в России, является разработка подходов и алгоритмов решения задачи классификации многотемных документов [5–11]. При этом большинство существующих подходов является альтернативой непосредственного сведения задачи классификации многотемных документов к традиционной задаче классификации, характеризующейся тем, что классифицируемый объект может принадлежать только к одному классу (multiclass classification) [12].

В новых условиях, когда разнородная информация добавляется в систему одновременно из многих не связанных между собою источников, весьма существенными становятся недостатки иерархической системы классификации:

- жесткая структура, которая приводит к сложности внесения изменений, так как приходится перераспределять все классификационные группировки;
- невозможность группировать объекты по заранее не предусмотренным сочетаниям признаков.

Поэтому приобрела популярность классификация на основе тегов, являющаяся, по сути, дескрипторной классификацией, но используемая в условиях динамически расширяемых распределенных информационных систем.

Теги, не обладая недостатками иерархической классификации, позволяют классифицировать информационные объекты по мере их добавления в систему без изначального продумывания структуры и состава классификаторов, т. е. они применимы для гетерогенных динамических систем и, по сути, не имеют альтернатив в ряде применений, например в практике совместной категоризации информации интернет-ресурсов пользователями посредством произвольно выбираемых ключевых слов (фолксономии) [13].

Но по мере роста информационных объемов пользователи сталкиваются с негативными последствиями этой практики. Система тегов линейна, и по мере роста числа понятий она становится все более и более громоздкой. В связи с этим в системе тегов неизбежно возникают синонимы, что препятствует поиску информации по тегу. Кроме того, эффективному поиску препятствует неоднозначность (мультизначность) понятий, принятых в качестве тегов. «Облако тегов», исходно предназначенное для удобства визуализации структуры, чрезмерно разрастается, и его наглядность сводится на нет.

Таким образом, неизбежно стали производиться попытки совместить достоинства тегов и иерархических классификаторов [13–15].

3 Система иерархических тегов и область ее применения

Представим новый метод классификации, который будем называть «система иерархических тегов» (СИТ). Система иерархических тегов предполагает

свободно расширяемый набор тегов (понятий), каждый из которых может быть динамически расширен своей иерархической древовидной структурой подпонятий. Таким образом, «облако тегов» в СИТ представляет собой «лес», где «корнем» каждого «дерева» является одно понятие из набора тегов, а иерархическая структура «дерева» соответствует иерархии подпонятий, уточняющих корневое понятие.

Принципиальными отличиями СИТ от фасетной классификации на основе комплекса независимых иерархических классификаторов являются:

- необязательность для любой пары тегов быть независимыми понятиями;
- необязательность соотнесения классифицируемого объекта с каждым тегом;
- отсутствие требования соотнесения классифицируемого объекта с единственным элементом в древовидной структуре подпонятий тега;
- отсутствие требования соотнесения классифицируемого объекта только с окончательным элементом («листом») в древовидной структуре подпонятий тега.

Сформулируем основные признаки информационных систем, для которых возможно и целесообразно применение СИТ.

1. Целесообразность:

- невозможность заранее определить структуру классификатора;
- значительная гетерогенность информации;
- наличие информации, относящейся одновременно к нескольким классам (мультитемность информации);
- значительная динамика информационного потока, необходимость классификации информации по мере ее поступления.

2. Возможность:

- наличие единого (не распределенного) центра классификации поступающей информации;
- ограниченный объем информации (как следствие необходимости обработки ее в едином центре).

Про необходимость классификации поступающей информации в едином, не распределенном центре следует сказать особо: классификация тегами сама по себе не требует такой централизации. Более того, распределенная классификация (категоризация) информации пользователями того или иного «облачного сервиса» позволяет эффективно обрабатывать огромные объемы информации, одновременно поступающие в систему из многочисленных источников. Любая фолксномия обязана своим существованием именно распределенности процесса классификации.

Но как только возникает необходимость введения в классификацию иерархии, распределенность процесса становится невозможной. Как известно, при

составлении иерархического классификатора деление класса на подклассы требует полноты и чистоты деления [16]. Требование полноты к СИТ, конечно, применяться не может, поскольку состав классов и подклассов в них в каждый момент времени отражает только текущее состояние информационного хранилища, а не информационную структуру предметной области в целом. Но требование чистоты деления (т. е. чтобы члены деления не были пересекающимися понятиями [16]) очевидным образом является актуальным и для СИТ. А выполнение этого требования, учитывая субъективность в той или иной степени любой классификации, невозможно обеспечить без управления процессом деления из единого центра.

Полагаем, что множество систем, для которых возможно и целесообразно использование классификации на базе СИТ весьма обширно. Рассмотрим пример такой системы — программу САФ.

4 Программа «Семейный архивный фонд»

«Семейный архивный фонд» [17] является программным средством ведения семейного архива и предназначен для хранения, систематизации, поиска и просмотра копий документов и фотографий. От аналогов (генеалогические программы и электронные фотоальбомы) САФ отличает возможность гибкого формирования групп документов для просмотра (альбомов документов) в соответствии со сложными критериями, увязанными с объектной моделью (объект–событие–документ), возможность пособытийной группировки документов при просмотре и поддержка многостраничных документов.

Программа разработана авторами в рамках работ ИПИ РАН по теме «Методы и средства справочно-поисковой поддержки научных и социально-культурных проектов на основе интеграции данных разнородных биографических источников» в 2011–2012 гг.

В целом характеристики САФ и концептуальная модель информационной технологии подробно описаны в [18].

Специфическими чертами САФ с точки зрения проблем классификации являются:

- значительная гетерогенность данных;
- наличие всевозможных «невозможных» сочетаний характеристик документов;
- сложность для неподготовленного пользователя заранее разработать взвешенную, непротиворечивую и избыточную систему классификации;
- постепенность наполнения САФ по мере появления новых документов;
- число документов, хранимых в САФ, — до 100 000.

Поскольку САФ удовлетворяет условиям как возможности, так и целесообразности, для него была разработана система, обеспечивающая возможность множественной расширяемой классификации.

5 Система иерархических тегов, реализованная в «Семейном архивном фонде»

Элементы САФ классифицируются на базе СИТ. Три независимых классификатора соответствуют типам элементов САФ: документам, событиям и объектам.

Документ как материальный объект, содержащий информацию в зафиксированном виде, привязан в САФ к двум классификаторам:

- (1) классификатор типа документа отражает материальную сущность носителя информации (например: «фотография», «портрет», «письмо», «видеозапись»). Целесообразность применения именно СИТ, а не иерархического классификатора для классификации типа документа следует из факта, что документ может иметь сложную структуру. Например, письмо может содержать рисунок, а на фотографии может быть зафиксирован человек на фоне портрета с грамотой в руках;
- (2) классификатор событий/содержания документов отражает информационную составляющую документа. Этот же классификатор служит для классификации событий, поскольку содержательно документ и событие имеют одну природу (документ, как правило, отражает одно или несколько событий).

В интерфейсе программы соотносённость элемента САФ с тем или иным понятием или подпонятием обозначается отметкой в соответствующем узле дерева и визуально отображается как галочка напротив имени узла (аналогично стандартному элементу управления TreeView [18]).

При этом в «лесе» тегов, соответствующих одному элементу САФ, могут присутствовать как деревья без единой отметки, так и деревья со многими отметками. Единственным естественным ограничением в положении отметок является невозможность поставить две галочки так, что одна из них будет находиться непосредственно под другой. То есть если объект соотносён с каким-либо понятием, то он не может быть соотносён и с каким-либо подпонятием этого понятия, что следует из самой сути иерархичности.

На рис. 1 приведен внешний вид интерфейса программы, позволяющий посмотреть теги объекта на конкретном примере. Объект Мария Иванова (в девичестве — Петрова) имеет отмеченными тег «Петровы» и подтег «московские» тега «Ивановы». Пример демонстрирует иерархичность и мультисмысленность СИТ.

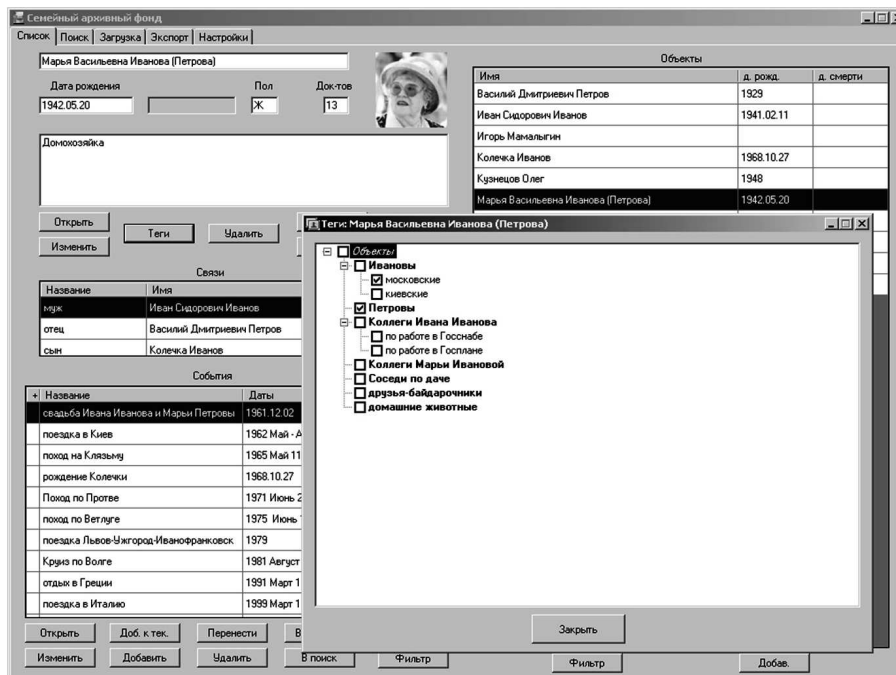


Рис. 1 Пример тегов объекта

Программный интерфейс позволяет при редактировании элемента САФ:

- поставить отметку у любого тега или его подпонятия. При этом все нижележащие отметки будут сняты;
- убрать отметку у любого тега или его подпонятия;
- добавить новый тег или подпонятие к существующему тегу на любом уровне иерархии и указать его имя;
- удалить тег или его подпонятие. При этом будут удалены и все нижележащие подпонятия;
- переименовать любой тег или его подпонятие.

Для отбора или поиска элементов САФ используется интерфейс задания фильтра, позволяющий:

- поставить отметку у любого тега или его подпонятия. При этом у всех нижележащих подпонятий отметки будут также установлены, т. е. при поиске элементов САФ, соотнесенных с некоторым понятием, будут также найдены элементы САФ, соотнесенные с любым его подпонятием;

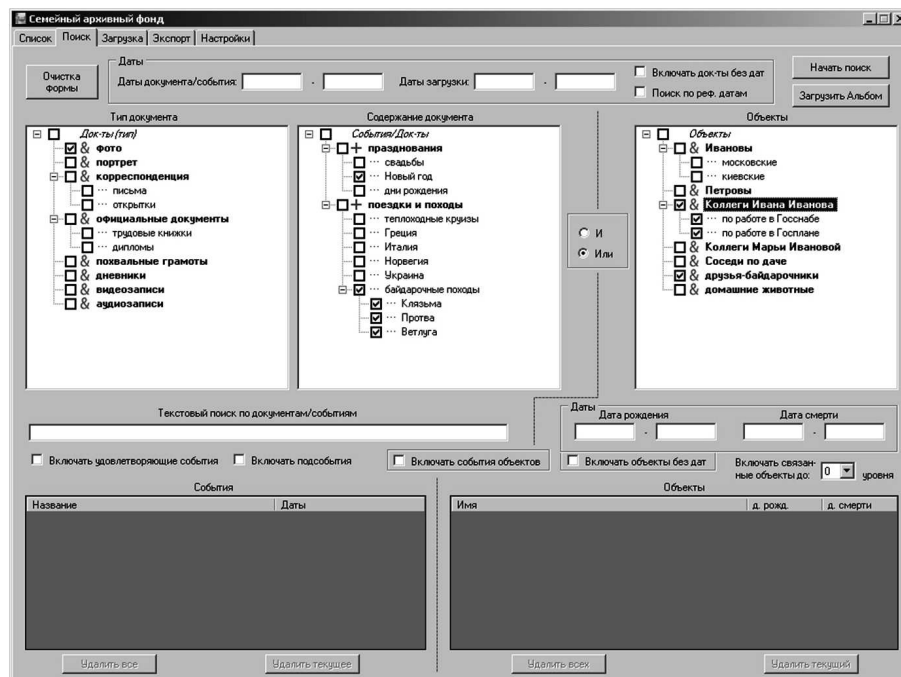


Рис. 2 Пример фильтра для формирования альбома документов

- убрать отметку у любого тега или его подпонятия. При этом у всех вышележащих и нижележащих подпонятий отметки будут также убраны;
- установить у любого тега (корня соответствующего дерева) знак «И» или «Или». (По умолчанию исходно у каждого тега установлен знак «И».)

На рис. 2 приведен внешний вид интерфейса программы, служащего для задания условия выбора документов с целью их просмотра (формирования альбома).

В примере условие задано исключительно с помощью фильтров (программа позволяет задавать и иные условия). В соответствии с выбранными условиями в альбом будут отобраны только фотографии, относящиеся к празднованию Нового года или к байдарочным походам, а также фотографии всех коллег Ивана Ивановича, являющихся также его друзьями по байдарочным походам.

Алгоритм поиска (отбора) элементов САФ по фильтру описан ниже.

При поиске для каждого элемента САФ осуществляется проверка соответствия условию фильтра согласно следующим шагам:

1. Теги, в дереве которых в фильтре не установлено ни одной отметки, игнорируются. Если у всех тегов не установлено ни одной отметки, то в результате отбора попадут абсолютно все элементы САФ данного типа. Далее в описании алгоритма под тегами подразумеваются только теги, в дереве которых в фильтре установлена хоть одна отметка.
2. Для каждого тега, имеющего установленные отметки в фильтре:
 - (а) для каждой отметки одного тега фильтра проверяется наличие отметки в том же месте дерева, соотнесенного с элементом САФ. Результатом проверки будет ИСТИНА, если отметка имеется, и, соответственно, ЛОЖЬ, если отметка отсутствует;
 - (б) по результатам проверки всех отметок одного тега вычисляется агрегирующий результат проверки для тега в целом как дизъюнкция (логическое «Или») результатов проверки отметок.
3. Для всех тегов фильтра, помеченных знаком «Или», вычисляется агрегирующий результат проверки как дизъюнкция (логическое «Или») результатов проверки тегов. Если таких тегов нет, то результат принимается как ИСТИНА.
4. Для всех тегов фильтра, помеченных знаком «И», вычисляется агрегирующий результат проверки как конъюнкция результатов проверки тегов. Если таких тегов нет, то результат принимается как ИСТИНА.
5. Агрегирующий результат проверки по фильтру в целом вычисляется как конъюнкция агрегирующего результата проверки тегов, помеченных знаком «Или», и агрегирующего результата проверки тегов, помеченных знаком «И».
6. Элемент САФ будет отобран в результаты поиска, только если агрегирующий результат проверки по фильтру в целом есть ИСТИНА.

Данный подход позволил совместить в САФ гибкость и открытость классификации с помощью тегов со структурностью и упорядоченностью иерархической классификации.

6 Предложения по дальнейшему развитию системы иерархических тегов в «Семейном архивном фонде»

Опытная эксплуатация САФ показала как удобство данного метода классификации, так и желательность некоторых усовершенствований, а именно:

1. В связи с тем, что пользователь в качестве тега может выбрать абсолютно любое понятие, при наполнении САФ может возникнуть следующая ситуация: в структуру тегов и их подпонятий потребуется ввести новое понятие,

являющееся не независимым тегом и не уточняющим подпонятием существующего тега, а понятием более высокого уровня по отношению к одному из существующих тегов. Для разрешения подобной коллизии в программе должна быть предусмотрена возможность не только добавления новой ветви в древовидную структуру подпонятий, но и вставки нового узла в любое место дерева, включая вставку в корень дерева, с тем чтобы ранее введенный тег становился уточняющим подпонятием нового тега.

2. Время от времени возникает ситуация, когда результат поиска по фильтру оказывается слишком обширным и возникает потребность исключить из него какую-либо группу элементов. Для этого необходимо для тегов фильтра иметь возможность, помимо указания способа агрегирования посредством значков «И» и «Или», задать признак отрицания «Не», обеспечив тем самым четыре способа агрегирования: «И», «Или», «И Не» и «Или Не». Результат проверки для тега в целом, помеченного признаком «Не», должен инвертироваться.

Данные усовершенствования предполагается реализовать в следующей версии программы.

Литература

1. Классификация // Философский словарь / Под ред. И. Т. Фролова. — 4-е изд. — М.: Политиздат, 1981.
2. Воробович Н. П. Проектирование информационных систем. — Красноярский аграрный государственный университет, кафедра «Информационные системы и технологии в экономике», 2007–2008. <http://www.kgau.ru/istiki>.
3. Библиографическая информация об общероссийских классификаторах. http://www.gks.ru/metod/OK_BIB.htm.
4. UDC Consortium Site. <http://www.udcc.org>.
5. Comite F. D., Gilleron R., Tommasi M. Learning multi-label alternating decision tree from texts and data // Machine Learning and Data Mining in Pattern Recognition (MLDM'03): 3rd Conference (International) Proceedings. Lecture Notes in Computer Science 2734. — Berlin: Springer, 2003. P. 35–49.
6. Boutell M. R., Luo J., Shen X., Brown C. M. Learning multi-label scene classification // Pattern Recognition, 2004. Vol. 37. P. 1757–1771.
7. Zhang M.-L., Zhou Z.-H. A k -nearest neighbor based algorithm for multi-label classification // GrC'05: 1st IEEE Conference (International) on Granular Computing Proceedings. — Beijing, China, 2005. P. 718–721.
8. Пескова О. В. Классификация документов в электронных библиотеках. <http://www.gpntb.ru/win/inter-events/crimea2007/cd/63.pdf>.
9. Петровский М. И., Глазкова В. В. Алгоритмы машинного обучения для задачи анализа и рубрикации электронных документов // Вычислительные методы и программирование, 2007. № 8. С. 57–69.

10. Анализ методов и предлагаемое решение для классификации больших объемов многотемных документов: Реферат // Реферативный сборник научно-исследовательских и опытно-конструкторских работ, выполненных НИИ СП в 2009 г. — М: НИИ СП, 2010. 229 с.
11. *Bo L., Yanshan X., Longbing C., Philip S. Yu.* Orientation distance-based discriminative feature extraction for multi-class classification // CIKM'10: 19th ACM Conference (International) on Information and Knowledge Management Proceedings. — NY, USA, 2010. P. 909–918.
12. *Глазкова В. В.* Исследование и разработка методов построения программных средств классификации многотемных гипертекстовых документов: Автореф. дис. . . . канд. физ.-мат. наук. — М.: МГУ, 2008.
13. *Витязев Я. М.* Фолксномия и/или таксономия // Персональный блог. <http://blog.vityasev.ru/2007/02/06/folksonomy-and-or-taxonomy>.
14. *Калитин Е.* Построение иерархических классификаторов на основе тщательно спроектированной системы тегов // Блог-платформа Ehh.ru. <http://ehhu.ru/people/user/6/blog/30511>.
15. *Бирман И.* Использование тегов. Древовидная классификация // Blogengine.ru. <http://blogengine.ru/help/tags>.
16. Классификация // Энциклопедический словарь Брокгауза и Ефрона: В 86 т. (82 т. и 4 доп.). — СПб., 1890–1907.
17. *Адамович И. М., Волков О. И., Маркова Н. А.* «Семейный архивный фонд» // Свидетельство о государственной регистрации программы ЭВМ № 2012617921.5.6.2012.
18. *Адамович И. М., Волков О. И., Маркова Н. А.* Организация семейного архивного фонда // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL'2011): Тр. XIII Всеросс. науч. конф. — Воронеж: ВГУ, 2011. С. 139–145.
19. TreeView Class / MSDN. <http://msdn.microsoft.com/en-us/library/system.web.ui.webcontrols.treeview.aspx>.

МЕТОДЫ ОБРАБОТКИ СИГНАЛОВ ДЛЯ ЛОКАЛИЗАЦИИ НЕВОСПОЛНИМЫХ ОБЛАСТЕЙ ГОЛОВНОГО МОЗГА*

*Т. В. Захарова¹, С. Ю. Никифоров², М. Б. Гончаренко³, М. А. Драницына⁴,
Г. А. Климов⁵, М. Ш. Хазиахметов⁶, Н. В. Чайнов⁷*

Аннотация: Настоящая работа посвящена проблеме синхронизации сигналов миограммы и магнитоэнцефалограммы. Методика вызванных потенциалов, которая применяется для локализации источников мозговой активности, основана на усреднении магнитоэнцефалограммы относительно опорных точек. В качестве опорных точек выбраны точки начала произвольного движения пациента. Основная проблема при усреднении заключается в неопределенности точек начала движения. В работе представлен новый высокоточный метод определения опорных точек, основанный на вероятностных свойствах сигналов. В разработанном методе точки начала движения ищутся в два этапа. Сначала они определяются по электромиограмме (ЭМГ) и/или актограмме, далее их местоположение уточняется по сигналу магнитоэнцефалограммы. Результаты могут быть использованы при решении задач локализации в смежных областях.

Ключевые слова: проверка гипотез; выборочная оконная дисперсия; выборочная квантиль; магнитоэнцефалография; магнитоэнцефалограмма; актограмма; миограмма

*Работа выполнена при поддержке РФФИ (проекты 11-01-00515 и 11-01-12026-офи-м), а также Министерства образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы».

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, lsa@cs.msu.ru

²Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, nikisimonmsu@gmail.com

³Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, goncharenko.mir@yandex.ru

⁴Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, margarita1Zapril@mail.ru

⁵Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, gregklimov@yandex.ru

⁶Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, khaziakhmetov@yandex.ru

⁷Центр нейрокогнитивных исследований (МЭГ-центр) при Московском городском психолого-педагогическом университете, oia.mgppu@gmail.com

1 Введение

Точная дооперационная локализация невосполнимых зон мозга и соотношение их местоположения с пораженными областями являются основой для планирования нейрохирургических вмешательств и минимизации послеоперационных осложнений для пациента, страдающего различными поражениями головного мозга [1].

В клинической практике одной из самых главных и трудных является задача локализации первичной моторной коры (М1) [2], особенно области представительства руки в зоне М1. Различные индивидуальные изменения, возникающие как последствия поражений центральной нервной системы [3], могут приводить к смещению функциональных зон коры относительно их «стандартного» расположения [4], что ведет к недостаточности информации о местонахождении этих зон только по анатомическим данным.

Одним из методов предоперационной локализации различных зон мозга выступает магнитоэнцефалография (МЭГ), которая, обладая высоким пространственно-временным разрешением, в сочетании с магнитно-резонансной томографией может использоваться для локализации источников мозговой активности [5–7].

Подробное описание методики получения экспериментальных данных, используемых в данной работе, изложено в статье [8].

Для точного определения местоположения моторных зон коры используют общепринятую методику вызванных потенциалов [9–11]. Классический метод вызванных потенциалов предполагает усреднение мозговой активности (энцефалограммы) относительно опорных точек [11] по всему сигналу. В качестве событий могут быть использованы как внешние стимулы, так и произвольная активность испытуемого (например, движение). Этот метод приводит к увеличению соотношения сигнал/шум, что позволяет выделить из фоновой высокоамплитудной активности мозга фазово-связанную с событием активность. Шум в данном случае представляет собой суперпозицию физических шумов (шумы сенсоров, усилителей, аналого-цифрового преобразования, внешних источников сигнала, сетевые помехи, развертка монитора и т. п.) и физиологических (фоновая активность мозга, фазово не связанная с событием).

Основная проблема метода усреднения ответа мозга относительно начала движения заключается во временной неопределенности опорных точек, которые требуется вычислить по параметрам мышечной активности или актограммы (сигналов акселерометра, позволяющих отслеживать движение).

Настоящая работа посвящена проблеме устойчивого определения начала движения.

В разработанном методе точки начала движения ищутся в два этапа: сначала они определяются по ЭМГ и/или актограмме, далее их местоположение уточняется по сигналу магнитоэнцефалограммы.

2 Исследование вероятностных характеристик сигналов

2.1 Спектральные характеристики

Исследуемые сигналы миограмм и магнитоэнцефалограмм являются нестационарными, что будет показано ниже, но несмотря на это, специальным образом примененный Фурье-анализ позволил исследовать их стационарную составляющую.

Исследование ЭМГ и электроэнцефалограмм началось с анализа различных спектральных и вероятностных характеристик. Спектр активного МЭГ-канала при движении представлен на рис. 1.

К примеру, на рис. 1 видны пики на частотах, кратных 50 Гц, которые дают сетевые наводки. Характерные воронки в этих частях спектра остались от аппаратных фильтров, использованных при снятии сигнала. Такие наводки подавлялись на первом этапе цифровой обработки режекторными фильтрами.

На рис. 2 кроме самих сигналов выведена кнопка (штриховые линии), которая регистрирует моменты, когда палец поднимается с поверхности или возвращается на нее (два положения: кнопка нажата или отжата).

Интересные результаты дало исследование спектральной плотности сигнала. Пусть

$$f(\lambda) = \sum_{m=-\infty}^{\infty} C_x(m) e^{-2\pi i \lambda m} \quad (1)$$

есть спектральная плотность случайного процесса $X(t)$.

Формула (1) есть преобразование Фурье функции автоковариации этого процесса. Здесь

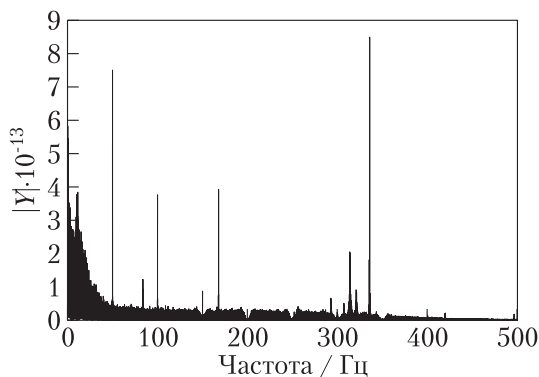


Рис. 1 Спектр активного МЭГ-канала

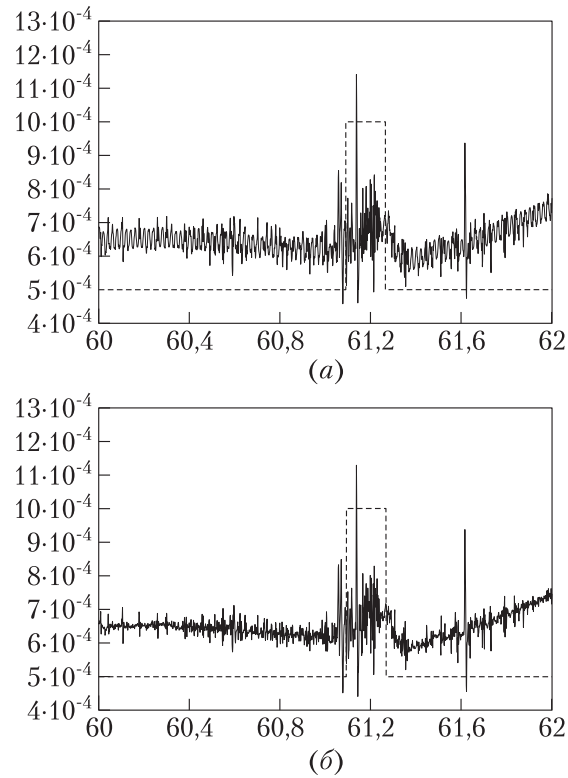


Рис. 2 Участок миограммы с движением до (а) и после (б) обработки режекторным фильтром

$$C_x(m) = \text{cov}(x_{t+m}, x_t) = \mathbb{E}[(X_{t+m} - \mathbb{E}X_{t+m})(X_t - \mathbb{E}X_t)]$$

есть автоковариационная функция случайного процесса $X(t)$.

Используемые выборочные автоковариационные функции:

$$\hat{C}'_x(m) = \frac{1}{N} \sum_{t=1}^{N-m} [X_{t+m} - \bar{X}] [X_t - \bar{X}] ; \quad (2)$$

$$\hat{C}''_x(m) = \frac{1}{N-m} \sum_{t=1}^{N-m} [X_{t+m} - \bar{X}] [X_t - \bar{X}] , \quad (3)$$

где $\bar{X} = (1/N) \sum_{t=1}^N X_t$ (см. [12]).

Оценка (2), в отличие от (3), является смещенной, т. е. $\mathbb{E}\hat{C}_x'(m) \neq C_x(m)$, но является положительно определенной. В результате исследования спектральной плотности энцефалограммы была выявлена особенность в частотной области сигнала. В начале движения главный пик спектральной плотности (обычно расположенный в пределах 5–30 Гц) либо пропадал совсем, либо смещался, т. е. происходила смена основной частоты. Определить точки начала движения по этому признаку не удалось из-за его нестабильности, поэтому от этого подхода пришлось отказаться (подробнее о применении спектрального анализа к подобным задачам см. [13–19]).

Все представленные сигналы являются нестационарными, а эпохи (отрезки сигнала, содержащие одно движение) неоднородны. Шум в записях сильно зависит от испытуемого человека, на протяжении одной записи длины эпох и самих движений могут сильно различаться. Типичный пример: испытуемый начинает с быстрых частых движений, затем устает и движения становятся длинными и более редкими, а интервалы между движениями могут возрасти от 750 мс до 30 с.

Полезным сигналом здесь считается отклик на движение пальца, а под шумом понимается физиологическая активность мозга, не связанная с движением, шум от аппаратуры и внешние наводки. Поэтому характер шумов очень сильно зависит от испытуемого и его состояния, а распределение шума не является нормальным. Подробные обоснования приведены ниже.

2.2 Исследование асимптотических свойств сигналов

При исследовании сигналов пришлось столкнуться с некоторыми их особенностями, которые сильно затрудняли обработку. Длительность записи исходного сигнала составляла около 1 часа и проходила с частотой оцифровки 1 кГц (т. е. съем параметров сигнала велся каждую миллисекунду). Таким образом, было необходимо обрабатывать массивы большой размерности, содержащие более чем 3 000 000 точек отсчетов сигналов всех 306 МEG-сенсоров, миограммы и актограммы.

Вследствие этого встал вопрос о том, можно ли изучить свойства всего сигнала только на некоторой его части? И правомерно ли применять операцию усреднения каких-либо характеристик, вычисленных по разбиениям сигнала?

Исследуем поведение выборочных характеристик. На каждой эпохе было вычислено выборочное среднее МЭГ-сигнала. Если сигнал обладает свойством стационарности, то разброс средних должен стремиться к нулю. На рис. 3, а представлен график реализаций выборочного среднего эпох магнитоэнцефалограммы.

На рис. 3, б показано изменение значения выборочной дисперсии выборочного среднего. Видно, что дисперсия не только не уменьшается, но имеет выраженную тенденцию к увеличению, что говорит о нестационарности сигнала.

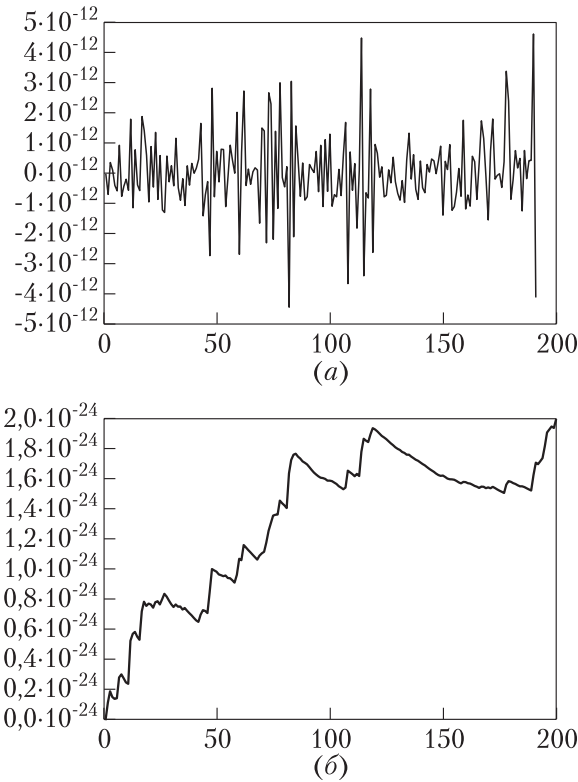


Рис. 3 Выборочное среднее МЭГ сигнала на эпохах (а) и изменение выборочной дисперсии выборочного среднего МЭГ с ростом числа эпох (б)

Аналогично исследовалось поведение выборочного среднего на эпохах сигнала миограммы. На рис. 4, а представлен график реализаций выборочного среднего на эпохах миограммы, а на рис. 4, б показано изменение выборочной дисперсии средних миограммы.

Здесь также показана нестационарность миограммы, так как выборочная дисперсия средних, вычисленных на эпохах миограммы, не стремится к нулю.

Во многих работах, посвященных цифровой обработке сигналов, изучают соответствующие математические модели [20, 21]. В них предполагается нормальное распределение шума, что зачастую адекватно отображает реальную задачу и подтверждается сравнением теоретических результатов и экспериментальных данных. В задачах фильтрации нестационарных сигналов такое предположение может либо сильно упростить модель, либо исказить ее. Допустимо ли предположение о нормальном распределении шума для такого сложного сигнала, как магнитоэнцефалограмма?

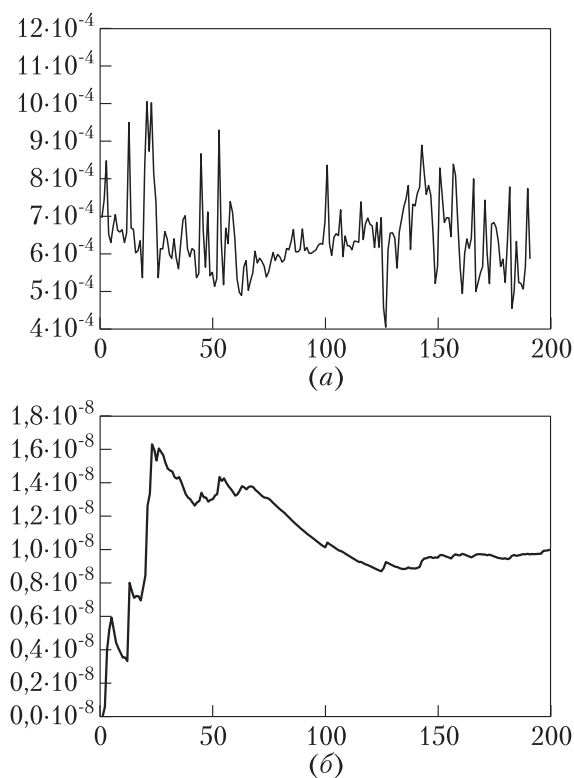


Рис. 4 Выборочное среднее сигнала миограммы на эпохах (а) и изменение выборочной дисперсии миограммы с ростом числа эпох (б)

Выше было показано, что магнитоэнцефалограмму можно рассматривать как нестационарный процесс, поэтому для обоснования предлагаемого метода обработки и анализа необходимо статистическое исследование эмпирического распределения шума. В ходе работы была проведена проверка нормальности распределения по критерию согласия хи-квадрат К. Пирсона. Уровень значимости α был выбран равным 0,05, объем выборки равнялся 76 000. Значение критерия составило 1275 при критической границе 14,1. Отсюда следует вывод о том, что гипотеза о нормальном распределении не согласуется с выбранным наблюдением.

Аналогичным путем была проверена гипотеза о нормальном распределении шума миограммы. В этом случае значение критерия составило 35 855, что привело к отклонению гипотезы.

Для проверки гипотезы о равенстве медиан нескольких выборок с неизвестным распределением был использован критерий Краскела–Уоллиса. Данный

Таблица 1 Результаты однофакторного дисперсионного анализа (SS — сумма квадратов разностей между средним арифметическим и значениями выборки по каждому виду дисперсии; df — число степеней свободы по каждому виду дисперсии; MS — среднее значение суммы квадратов разностей по каждому виду дисперсии, определяемое как отношение SS/df)

Вид дисперсии	SS	df	MS	χ^2	$p > \chi^2$
Внутригрупповая	$3,90307 \cdot 10^{14}$	191	$2,04349 \cdot 10^{12}$	126799,02	0
Межгрупповая	$2,01286 \cdot 10^{14}$	192 000	$1,04837 \cdot 10^9$	—	—
Общая	$5,91593 \cdot 10^{14}$	192 191	—	—	—

критерий является многомерным обобщением критерия Уилкоксона–Манна–Уитни. Выборки были образованы массивами значений миограммы длиной в 1000 отсчетов. Для анализа взяты участки сигнала, где в качестве опорных точек выступают моменты нажатия кнопки, являющиеся серединами рассматриваемых интервалов. В табл. 1 представлены результаты тестирования 192 подобных выборок, которые ведут к отклонению гипотезы о равенстве выборочных медиан.

Полученные результаты еще раз подтверждают исключительную сложность и особенность исследуемых сигналов.

3 Методы нахождения опорных точек

В ходе исследования найти опорные точки по МЭГ-каналам не представлялось возможным из-за высокого уровня шума в сигналах. В данной статье рассматриваются методы нахождения опорных точек по сигналам миограммы и актограммы. Обязательным условием является одновременное снятие сигналов миограммы, актограммы и магнитоэнцефалограммы. Далее при применении метода усреднения магнитоэнцефалограмм по опорным точкам улучшается соотношение сигнал/шум для МЭГ-каналов.

Также в статье приводится описание метода уточнения опорных точек по максимальному ответу в МЭГ-сигнале на движение.

3.1 Расстановка опорных точек по выборочной оконной дисперсии миограммы

Метод основан на физиологических особенностях строения мышечной системы человека, который при совершении движения не может задействовать все мышечные волокна одновременно, а использует их по очереди. В результате движением будем считать ту часть сигнала, на которой оконная выборочная дисперсия миограммы превышает порог. Далее рассмотрим этапы работы алгоритма.

3.1.1 Вычисление выборочной оконной дисперсии миограммы

На первом этапе работы алгоритма вычислим оконную выборочную дисперсию миограммы:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (4)$$

Размер окна дисперсии зависит от качества исследуемой миограммы. На рис. 5 представлены примеры сигнала миограммы и ее выборочной оконной дисперсии.

Для решения задачи нахождения точек привязки для точного усреднения каждого МЭГ-канала, в этом методе и во всех последующих, в качестве опорных точек будем выбирать точки начала движения пальца пациента.

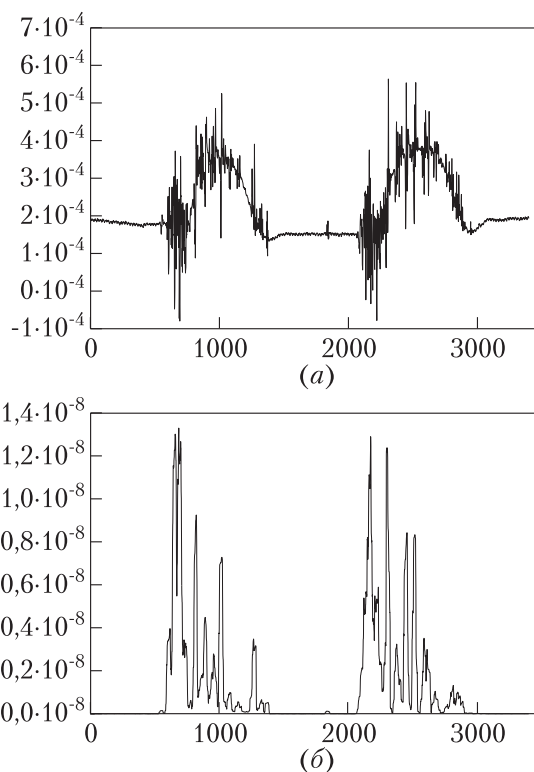


Рис. 5 Участок миограммы здорового пациента (два движения) (а) и оконная выборочная дисперсия миограммы (окно равно 30 точкам) (б)

3.1.2 Вычисление оконной выборочной квантили дисперсии миограммы

Выборочная квантиль порядка α вычисляется по вариационному ряду $x^{(1)} \leq x^{(2)} \leq \dots \leq x^{(m)}$, построенному путем упорядочивания выборки как $x^{[m\alpha+1]}$, где квадратные скобки обозначают целую часть. В результате значение квантили изменяется со временем, что помогает справиться с проблемой нестационарности сигнала.

3.1.3 Расстановка и уточнение опорных точек

Точки пересечения порога (равного квантили) с дисперсией миограммы назовем предварительными точками начала движения, так как они не являются точками начала движения руки пациента и нуждаются в уточнении. Для этого применяется специальный алгоритм, который вычисляет шум сигнала дисперсии миограммы перед движением и «спускает» точки до определенного уровня в зависимости от параметров данного шума.

3.1.4 Удаление ложных срабатываний

На заключительном этапе работы алгоритма каждой опорной точке сопоставляется некий признак, чаще всего — медиана значений точек в ее окрестности (квантиль порядка $1/2$). Далее на гистограмме значений данных признаков удаляются движения, медиана которых на некоторую величину отличается от большинства медиан. Пример гистограммы значений медиан окрестности точек начала движения приведен на рис. 6. Все движения, расположенные справа от пунктирной линии, удаляются, так как считаются ложными срабатываниями алгоритма.

Результатом работы алгоритма является массив опорных точек. В дальнейшем точная расстановка опорных точек может быть использована при исследованиях

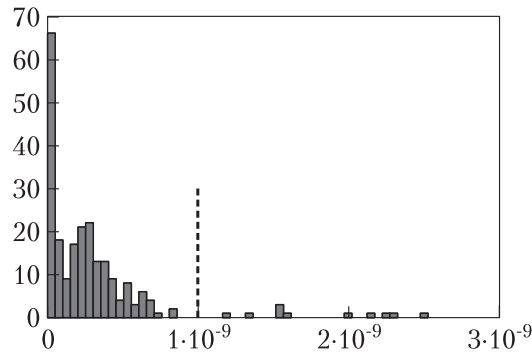


Рис. 6 Гистограмма значений медиан

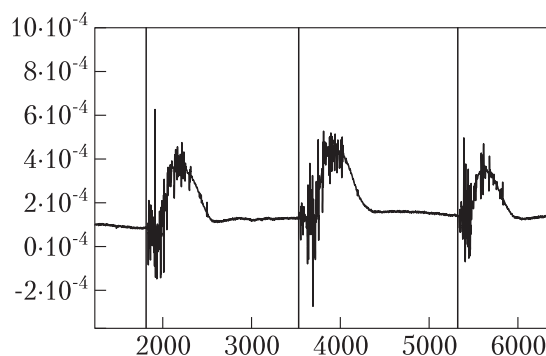


Рис. 7 Миограмма с расставленными опорными точками (точки начала движения)

зависимости между МЭГ-каналами и сигналом миограммы. На рис. 7 приведены несколько эпох миограммы с расставленными на ней опорными точками.

Данный алгоритм нахождения опорных точек применим только к пациентам с миограммой хорошего качества. По некоторым миограммам распознать движения не представлялось возможным в принципе.

3.2 Поиск по шаблону для расстановки опорных точек по сигналу акселерометра

Вне зависимости от того, что движения пациента не похожи друг на друга и сигнал не является стационарным, поиск по сформированному специальным образом шаблону (по сигналу акселерометра) дал хорошие результаты. К тому же использование сигнала акселерометра очень удобно, так как зачастую он является менее шумным, чем миограмма, и определить точки начала движения можно с большей точностью, что соответствует ранее полученным данным [8].

Рассмотрим этапы работы алгоритма, использующего поиск по шаблону на сигнале акселерометра.

3.2.1 Расстановка точек начала движения

С помощью ранее приведенного алгоритма (см. п. 3.1) расставляются точки начала движения по оконной выборочной дисперсии акселерометра (4). На примере нескольких эпох движения акселерометра (рис. 8, а) и их выборочной оконной дисперсии (рис. 8, б) видно, что приведенный алгоритм успешно работает.

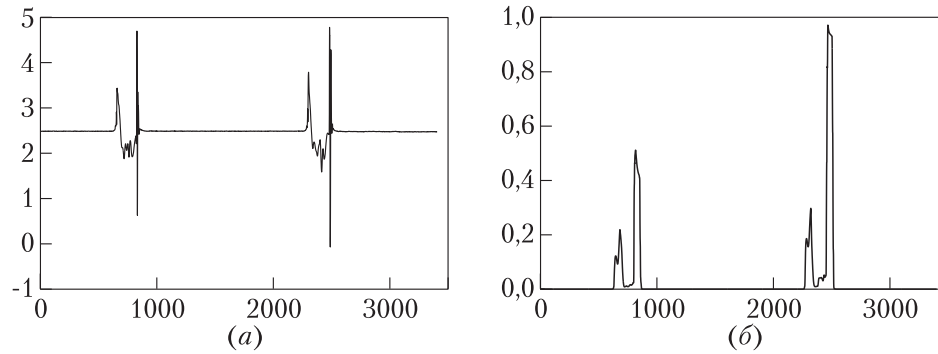


Рис. 8 Участок сигнала акселерометра (два движения) (а) и оконная выборочная дисперсия данного участка акселерометра (б)

3.2.2 Формирование динамического шаблона и построение выборочной корреляции шаблона с сигналом

Сформируем шаблон, который изменяет свое значение по ходу сигнала для точного нахождения опорных точек.

Пусть имеется движение (исследуемое движение), для которого нужно сформировать шаблон. Для этого используем движения в некоторой окрестности исследуемого. Чем движение расположено ближе к исследуемому, тем с большим коэффициентом оно войдет в шаблон. Данное решение основано на нестационарности сигнала и на том, что чем ближе движение расположено к исследуемому, тем больше оно похоже на него.

После формирования специальной системы шаблонов необходимо найти выборочную корреляцию каждого из шаблонов с соответствующим ему движением. Формула корреляции для выборок X и Y

$$\text{Corr}(X, Y) = \frac{\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]}{\sqrt{\mathbb{D}X\mathbb{D}Y}}. \quad (5)$$

3.2.3 Уточнение точек начала движения с помощью корреляционной картины

На рис. 9 приведен пример корреляционной картины (с точками начала движения), рассчитанной на предыдущем этапе. Видно, что построенные на первом этапе точки начала движения можно уточнить, если найти локальные максимумы корреляционной картины вблизи этих точек. Данный метод является итерационным, т. е. уточнение точек начала движения по шаблону может быть применено к сигналу неоднократно до получения необходимой точности.

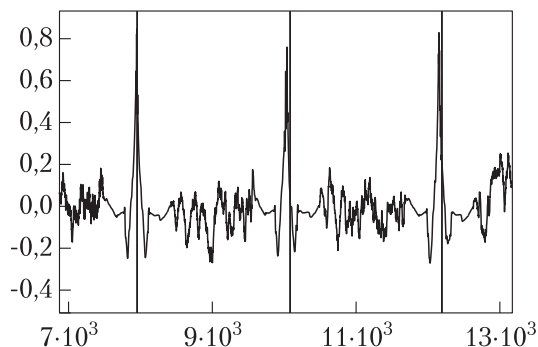


Рис. 9 Корреляционная картина с расставленными точками начала движения

3.3 Поиск по шаблону на активном МЭГ-канале для уточнения опорных точек

МЭГ-канал, имеющий наибольшую амплитуду отклика на движение, назовем активным.

Метод привязывает опорные точки к максимальному ответу на движение в магнитоэнцефалограмме пациента. Ответ на активных МЭГ-каналах отчетливо виден сразу за найденным началом движения. Он является суперпозицией откликов различных зон головного мозга, активных во время совершения движения (моторная кора, сенсорная, премоторная и др.) [8].

Метод является уточняющим и может быть применим к любому алгоритму, результатом которого является массив опорных точек.

3.3.1 Поиск активного МЭГ-сигнала

Поиск активного МЭГ-канала осуществляется с помощью усреднения магнитоэнцефалограмм относительно опорных точек, найденных любым из предложенных ранее способов. Пример усредненной магнитоэнцефалограммы на активном канале приведен на рис. 10. В дальнейшем такой сигнал используется для построения шаблона.

3.3.2 Использование поиска по шаблону на МЭГ-сигнале

На первом этапе аналогично п. 3.2.2 формируется специальный шаблон для каждого движения и строится корреляционная картина шаблонов и активного МЭГ-канала. Вблизи точек начала движения расположены локальные максимумы корреляционной картины (рис. 11), которые и будем считать новыми опорными точками.

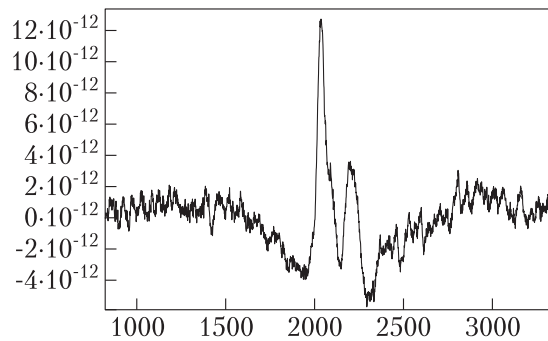


Рис. 10 Усредненный МЭГ-сигнал на активном канале

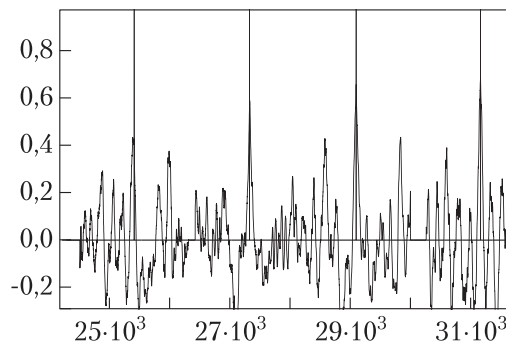


Рис. 11 Корреляционная картина МЭГ-канала с шаблоном

3.4 Использование вейвлет-анализа для нахождения опорных точек по сигналу миограммы

В настоящее время существует множество методов для анализа цифровых сигналов. Классическим примером такого метода является анализ Фурье, он же является одним из наиболее часто применимых. Преобразование Фурье раскладывает сигнал на сумму гармонических компонент с разными частотами. Однако такой метод хорошо работает лишь со стационарными сигналами, когда удастся найти периодические составляющие. Кроме того, преобразование Фурье показывает форму сигнала в частотной области, но оно не дает информации о том, в какой момент времени в сигнале присутствовали колебания с данной частотой, тем самым теряется информация о локализации явлений во временной области.

Поэтому было принято решение использовать вейвлет-анализ, являющийся идеальным инструментом для обработки нестационарных сигналов. Подробный материал об основах вейвлет-анализа содержится в [20, 22].

3.4.1 Выбор вейвлета и его применение

В Matlab Wavelet Toolbox позволяет анализировать одномерные массивы с применением и дискретного, и непрерывного вейвлет-преобразований. Вместе с тем, очевидно, второй вариант предпочтительнее, так как можно исследовать детали произвольной длины, а не только степени двойки.

В ходе экспериментов с данными были исследованы многие семейства вейвлетов, для некоторых из них были получены интересные результаты. Наиболее значимые результаты были получены при использовании вейвлета Морле (в обозначениях Matlab — “cmor1-1.5”).

Вейвлет Морле обладает наилучшим спектральным разрешением среди известных вейвлетов, и этим определяется его выбор для исследования нестационарных сигналов, для которых желательно получить высокую точность частотной характеристики, меняющейся со временем.

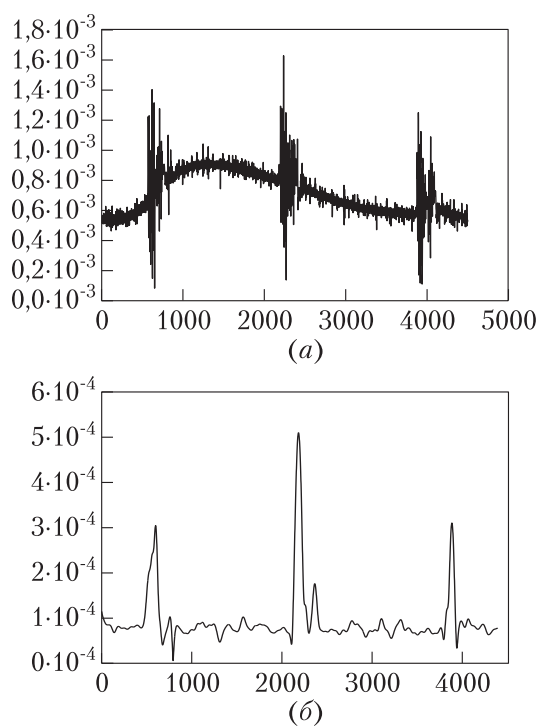


Рис. 12 Участок миограммы с движением (а) и коэффициенты одного уровня после применения вейвлета Морле (б)

Данный вейвлет — комплексная функция следующего вида:

$$\psi(t) = e^{2\pi it - t^2/2}.$$

После применения вейвлет-преобразования к сигналу миограммы найдем уровни вейвлет-коэффициентов наибольшей амплитуды. Пример миограммы с движением и ее вейвлет-коэффициентов приведен на рис. 12.

По уровню с наибольшей амплитудой коэффициентов находим точки начала движения (по аналогии с подразд. 3.1).

4 Сравнение алгоритмов

В статье представлены три алгоритма нахождения опорных точек и один алгоритм уточнения, основанный на использовании МЭГ-канала. Сравним алгоритмы нахождения опорных точек, чтобы найти оптимальный для каждого конкретного пациента.

4.1 Сравнение алгоритмов на основе максимальной амплитуды усредненного МЭГ-канала

Для 7 пациентов в результате применения каждого из алгоритмов нахождения опорных точек определены МЭГ-каналы с наибольшей амплитудой. На координатной плоскости были нанесены точки: амплитуда откладывалась по оси OY , номер пациента по оси OX (рис. 13). Метод уточнения точек начала движения применялся только к алгоритму, показавшему наилучший результат для пациента.

Данная метрика качества алгоритмов направлена на максимизацию ответа на движение.

Метод уточнения (1 на рис. 13) повышает точность найденных опорных точек для всех пациентов, кроме пациента 3. У пациента 3 ответ на движение найти не удалось.

Метод нахождения опорных точек по оконной выборочной дисперсии миограммы (2 на рис. 13) дал хорошие результаты только на пациентах 1 и 5. Это пациенты с качественными сигналами миограмм. На других пациентах, кроме пациента 6, метод дал приемлемые результаты. Для пациента 6 метод не работает, так как его миограмма содержит высокий уровень шума.

Итерационный метод нахождения опорных точек по динамическому шаблону с использованием акселерометра (3 на рис. 13) дал отличные результаты на всех пациентах, кроме пациента 3. Удовлетворительный результат данного метода на пациенте 3 обусловлен особенностями его актограммы.

Метод нахождения опорных точек по вейвлету Морле (4 на рис. 13) сравним по качеству с методом, использующим дисперсию миограммы. Однако из-за

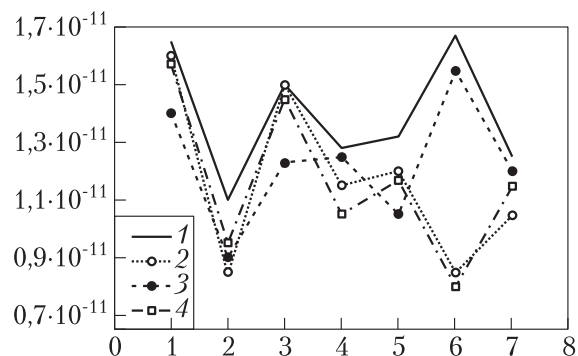


Рис. 13 Сравнение методов: 1 — метод уточнения; 2 — метод нахождения опорных точек по оконной выборочной дисперсии миограммы; 3 — итерационный метод нахождения опорных точек по динамическому шаблону с использованием акселерометра; 4 — метод нахождения опорных точек по вейвлету Морле

особенностей сигнала миограммы для каждого пациента необходимо использовать коэффициенты разложения нескольких уровней. Поэтому в целом метод работает хуже, хотя является более устойчивым к некоторым видам шума.

4.2 Сравнение числа ложных срабатываний алгоритмов

Важным критерием качества приведенных алгоритмов является число ошибок I и II рода, т. е. количество не найденных движений пациента и количество ложно найденных движений. Несколько неверно определенных опорных точек могут полностью испортить детали усредненной картины, что является недопустимым. Поэтому во всех методах важно было минимизировать число ложных срабатываний. Количество ошибок I и II рода напрямую зависит от качества сигналов. Для сигнала миограммы значения обычно больше, чем для сигнала акселерометра. Но во всех алгоритмах входные параметры могут быть выбраны таким образом, чтобы количество ошибок не превышало 5%.

5 Основные результаты и направления дальнейшей работы

В настоящем исследовании разработаны и протестированы методы нахождения опорных точек в сигнале, способствующие решению задачи по локализации областей мозга, отвечающих за движения. Алгоритмы позволяют устойчиво находить точки начала движения пальца пациента. Также были детально исследованы вероятностные характеристики и асимптотические свойства магнитоэнцефалограмм, обоснована их нестационарность и приведены спектральные и энергетические характеристики. В дальнейшем необходимо совершенствовать

разработанные алгоритмы для применения к сигналам пациентов с различными заболеваниями, что является существенно более сложной задачей. При этом практическая ценность результатов возрастет многократно. Планируется интегрировать созданные алгоритмы с системами управления электронными устройствами посредством сигналов мозга человека.

Результаты могут быть использованы при решении задач синхронизации в смежных областях.

6 Благодарность

Выражаем благодарность за экспериментальные материалы, которые были любезно предоставлены Центром нейрочувствительных исследований (МЭГ-центр) при Московском городском психолого-педагогическом университете.

Литература

1. *De Tiege X., Connelly A., Liegeois F., Harkness W., Clark C. A., Chong W. K., Gadian D. G., Cross J. H.* Influence of motor functional magnetic resonance imaging on the surgical management of children and adolescents with symptomatic focal epilepsy // *Neurosurgery*, 2009. Vol. 64. No. 5. P. 856–864.
2. *Chouinard P. A., Paus T.* The primary motor and premotor areas of the human cerebral cortex // *Neuroscientist*, 2006. Vol. 12. No. 2. P. 143–152.
3. *Haseeb A., Asano E., Juhasz C., Shah A., Sood S., Chugani H. T.* Young patients with focal seizures may have the primary motor area for the hand in the postcentral gyrus // *Epilepsy Res.*, 2007. Vol. 76. No. 2–3. P. 131–139.
4. *Flor H.* Cortical reorganisation and chronic pain: Implications for rehabilitation // *J. Rehabil. Med.*, 2003. Suppl. 41. P. 66–72.
5. *Hamalainen M., Hari R., Ilmoniemi R. J., Knuutila J., Lounasmaa O. V.* Magnetoencephalography — theory, instrumentation, and applications to noninvasive studies of the working human brain // *Rev. Mod. Phys.*, 1993. Vol. 65. P. 413–497.
6. *Bardoule T., Ross B.* MEG imaging of sensorimotor areas using inter-trial coherence in vibrotactile steady-state responses // *NeuroImage*, 2008. Vol. 42. P. 323–331.
7. *Roger A., Thornton D.* Evaluation of a technique to measure latency jitter in event-related potentials // *J. Neuroscience Methods*, 2008. Vol. 168. P. 248–255.
8. *Чаянов Н. В., Прокофьев А. О., Морозов А. А., Строганова Т. А.* Локализация моторных зон коры мозга человека методом магнитоэнцефалографии // *Журнал высшей нервной деятельности им. И. П. Павлова*, 2012 (в печати).
9. *McGillem C. D., Aunon J. I.* Analysis of event-related potentials // *Methods of Analysis of Brain Electrical and Magnetic Signals: EEG Handbook*. — Elsevier Science Publs., 1987.
10. *Kilner J. M., Baker S. N., Salenius S., Hari R., Lemon R. N.* Human cortical muscle coherence is directly related to specific motor parameters // *J. Neuroscience*, 2000. Vol. 20. No. 23. P. 8838–8845.

11. *Fabiani M., Gratton G., Federmeier K. D.* Event-related brain potentials: Methods, theory, and applications // *Handbook of Psychophysiology* / Eds. J. T. Cacioppo, L. G. Tassinary, G. G. Berntson. — Cambridge: Cambridge University Press, 2007. P. 85–119.
12. *Баласанов Ю. Г., Дойников А. Н., Королева М. Ф., Юровский А. Ю.* Прикладной анализ временных рядов с программой Эвриста. — М.: Диалог-МГУ, 1991. 270 с.
13. *Crone N. E., Miglioretti D. L., Gordon B., Sieracki J. M., Wilson M. T., Uematsu S., Lesser R. P.* Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis // *Brain*, 1998. Vol. 121. P. 2271–2299.
14. *Farmer S. F.* Rhythmicity, synchronization and binding in human and primate motor systems // *J. Physiology*, 1998. Vol. 509.1. P. 3–14.
15. *Lachaux J.-P., Rodriguez E., Martinerie J., Valera F. J.* Measuring phase synchrony in brain signals // *Human Brain Mapping*, 1999. Vol. 8. P. 194–208.
16. *Grosse P., Cassidy M. J., Brown P.* EEG-EMG, MEG-EMG and EMG-EMG frequency analysis: Physiological principles and clinical applications // *Clinical Neurophysiology*, 2002. Vol. 113. P. 1523–1531.
17. *June S. K., Chun K. Ch.* Robust source analysis of oscillatory motor cortex activity with inherently variable phase delay // *NeuroImage*, 2007. Vol. 37. P. 518–529.
18. *Cheyne D., Bells S., Ferrari P., Gaetz W., Bostan A. C.* Self-paced movements induce high-frequency gamma oscillations in primary motor cortex // *NeuroImage*, 2008. Vol. 42. P. 332–342.
19. *Xiaolin H., Jing X., Yingying W., Kirtman E. G., Kotecha R., Fujiwara H., Hemasilpin'N., Rose D. F., Degrauw T.* Gamma oscillations in the primary motor cortex studied with MEG // *Brain Development*, 2009. Vol. 32. No. 8. P. 619–624.
20. *Малла С.* Вэйвлеты в обработке сигналов / Пер. с англ. — М.: Мир, 2005.
21. *Gonzales R. C., Woods R. E.* Digital image processing. — Prentice Hall, 2005.
22. *Захарова Т. В., Шестаков О. В.* Вейвлет-анализ и его приложения: Уч. пособие. — 2-е изд. — М: ИНФРА-М, 2012.

АРХИТЕКТУРА ГИБРИДНОЙ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ GEFEST (GENERATION–EXPANSION–FILTERING–SORTING–TRUNCATION)*

*Ю. С. Неведова*¹

Аннотация: Рекомендательная система представляет сравнительно новый класс программного обеспечения, в задачу которого входит предсказание того, какие объекты (фильмы, музыка, книги, новости, веб-сайты) будут интересны пользователю, с помощью анализа его действий и оценок. В данной работе описано построение такой системы. Решается задача формирования подходящего для пользователя списка рекомендуемых товаров. Упор сделан на возможность использовать всю доступную информацию, содержащуюся в данных о товаре (описание, характеристика), данных о пользователе (профиль пользователя) и в данных о поведении пользователя (сведения об использовании или покупке товара и/или явная оценка товара).

Ключевые слова: рекомендательные системы; коллаборативная фильтрация; метод ближайших соседей; алгоритмы классификации

1 Введение

Рекомендательные системы — это программные средства, технологии и методы выработки полезных для пользователя предложений, советов и рекомендаций. Предоставленные предложения направлены на поддержку пользователя в различных процессах принятия решений, а именно: какие товары купить, какую музыку послушать или какие новости прочитать. Рекомендательные системы помогают пользователю ориентироваться в ситуации стремительно нарастающих объемов контента, они оказались ценным средством для борьбы с информационной перегрузкой и стали одним из самых мощных и популярных инструментов в электронной коммерции.

Рекомендательные системы в первую очередь предназначены для лиц, которые не имеют достаточного личного опыта или компетентности для оценки

*Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-12026-офи-м, 11-01-00515а), а также Министерством образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы» (проект МК–2256.2012.1).

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики; Институт проблем информатики Российской академии наук, y.nefedova@gmail.com

подавляющего количества альтернативных объектов, которые веб-сайт или, например, интернет-магазин может предложить.

«Объект» является общим термином, используемым для обозначения того, что система рекомендует пользователям. Рекомендательные системы обычно фокусируются на определенном типе объекта, например: книгах, DVD-дисках, новостях, товарах.

В своей простейшей форме персональные рекомендации предоставляются в виде упорядоченных по рейтингу списков объектов. Для расчета рейтинга рекомендательные системы, основываясь на предпочтениях пользователя, пытаются предсказать наиболее подходящие товары или услуги. Для того чтобы реализовать такую вычислительную задачу, рекомендательные системы собирают от пользователей информацию об их предпочтениях, которые могут быть явно выражены, например как рейтинг товаров или оценка фильмов, или могут быть выведены путем интерпретации действий пользователя. Так, рекомендательная система может рассмотреть переход на определенную веб-страницу как неявный признак предпочтения для объектов, показанных на этой странице.

Рекомендательные системы являются относительно новым предметом исследования по сравнению с разработками многочисленных средств и методов в других классических информационных системах (например, в системах баз данных или поисковых системах). Рекомендательные системы выделилась в самостоятельную область исследований в середине 1990-х годов, однако мощнейшим толчком к развитию этой темы стал конкурс Netflix Prize [1], объявленный в 2006 г. компанией Netflix, одним из лидеров на американском рынке проката DVD. Компания предоставила участникам данные о 100 млн рейтингов, предоставленных 480 тысячами пользователей 18 тысячам фильмов в течение 6 лет. Участник, показавший результаты на 10% лучше, чем их собственный движок рекомендаций, получал миллион долларов. Благодаря этому конкурсу начали бурно развиваться методы построения рекомендательных систем, однако задача оказалась сложной — несмотря на ажиотаж и жесточайшую конкуренцию, на выполнение условий у участников ушло три года.

В последние годы интерес к рекомендательным системам резко возрос, о чем свидетельствуют следующие факты.

1. Важную роль рекомендательных систем высоко оценили такие популярные интернет-сайты, как Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Last.fm, IMDb, Imhonet. Более того, многие медиа-компании в настоящее время разрабатывают и внедряют рекомендательные системы как часть услуг, которые они предоставляют своим абонентам.
2. Регулярно проводятся многочисленные конференции и семинары, относящиеся к данной области. В частности, специально основанная в 2007 г. ACM Recommender Systems (RecSys), по сей день являющаяся главным ежегодным событием в области исследований по рекомендательным технологиям. Кро-

ме того, сессии, посвященные рекомендательным системам, часто бывают включены в состав более традиционных конференций в области баз данных, информационных и адаптивных систем. Среди этих конференций стоит отметить ACM Special Interest Group on Information Retrieval (SIGIR), User Modeling, Adaptation and Personalization (UMAP) и ACM Special Interest Group on Management Of Data (SIGMOD).

3. В высших учебных заведениях по всему миру читаются специальные курсы, посвященные рекомендательным системам; учебники по рекомендательным системам являются крайне популярными; а также следует упомянуть недавно вышедшие книги [2, 3], вобравшие в себя уже накопившийся к этому времени немалый опыт построения рекомендательных систем.
4. Многие академические журналы издавали специальные выпуски, охватывающие исследования и разработки в области рекомендательных систем. Среди журналов, которые посвятили такие выпуски вопросам, связанным с рекомендательными системами, следует отметить: AI Communications (2008); IEEE Intelligent Systems (2007); International Journal of Electronic Commerce (2006); International Journal of Computer Science and Applications (2006); ACM Transactions on Computer-Human Interaction (2005) и ACM Transactions on Information Systems (2004).

Развитие рекомендательных систем началось с довольно простого наблюдения: люди часто полагаются на рекомендации, предоставленные другими при принятии повседневных решений. Например, при выборе книги для чтения мы обычно рассчитываем на рекомендации друзей, работодатели учитывают рекомендательные письма при принятии решений о трудоустройстве, при выборе фильма для просмотра люди склонны ориентироваться на обзоры фильмов, которые написал кинокритик.

Стремясь подражать такому поведению, первые рекомендательные системы применяли алгоритмы, использующие рекомендации, выработанные сообществом пользователей, чтобы предоставить рекомендацию активному (прогнозируемому) пользователю. Эти рекомендации состояли из объектов, которые высоко оценили сходно мыслящие пользователи (с похожими вкусами). Такой подход стал называться *коллаборативной фильтрацией*. Его обоснование заключается в том, что если активный пользователь соглашался в прошлом с некоторыми пользователями, то новые рекомендации, поступающие от этих схожих пользователей, должны также представлять интерес для активного пользователя.

Современные же рекомендательные системы генерируют рекомендации, используя различные типы знаний и данных о пользователях, доступных объектах и предыдущих операциях пользователей над объектами. Затем пользователь может просматривать список рекомендаций. Он может принимать их или нет, может дать сразу или на следующем этапе явную или неявную реакцию, обеспечивая обратную связь. Все эти действия пользователя и обратная связь могут быть

сохранены в базе данных рекомендательной системы и могут быть использованы для создания новых рекомендаций при следующем взаимодействии пользователя и системы.

К настоящему времени можно выделить три основных подхода к построению рекомендательных систем:

1. **Collaborative filtering** (коллаборативная фильтрация). Как уже упоминалось выше, простая и оригинальная реализация этого подхода [4] рекомендует активным пользователям элементы, которые другим пользователям со схожими вкусами понравились в прошлом. Сходство во вкусе двух пользователей рассчитывается на основе сходства в рейтингах пользователей.

Традиционный алгоритм коллаборативной фильтрации представляет пользователя как N -мерный вектор объектов, где N — число различных предлагаемых объектов (количество товаров в каталоге, фильмов, книг). Компоненты вектора положительны при покупке или явной положительной оценке пользователем соответствующего объекта и отрицательны для негативного по рейтингу объекта. Алгоритм генерирует рекомендации на основе рейтингов нескольких клиентов, которые наиболее близки к активному пользователю. Сходство двух пользователей A и B обычно измеряется косинусом угла между двумя пользовательскими векторами:

$$\text{similarity}(\vec{A}, \vec{B}) = \cos(\vec{A}, \vec{B}) = \frac{\vec{A} \bullet \vec{B}}{\|\vec{A}\| * \|\vec{B}\|}.$$

Коллаборативная фильтрация считается самым популярным и широко применяемым подходом к построению рекомендательных систем.

2. **Content-based** [5]. Данный подход основан на использовании данных профилей пользователей и данных об объектах. Содержание профиля пользователя может состоять не только из истории его покупок и оценок объектов, но и из множества других параметров: возраста, места жительства и т. п. То же относится и к профилю (описанию) объектов. Данные в этом случае можно разделить по двум важным параметрам:
 - (а) способ получения: явный (голосование, заполнение профиля и т. п.) или неявный (история покупок, географическое положение по IP-адресу, история посещений сайтов или же другая информация, собранная при помощи специализированных программ);
 - (б) структурированность. Структурированными данными являются стандартные поля профиля, истории покупок, цена товара, тип объекта и т. п. К неструктурированным можно отнести свободные поля профиля (например, «о себе»), свободные комментарии об объектах, текстовые описания товаров, тексты новостей и т. п.

Основная сложность работы с неструктурированными данными — трудность «оцифровки» (например, приведение к векторному виду). Для работы с такими данными требуются специальные алгоритмы, например, нормализации слов и алгоритмы семантического анализа текстов (Latent Semantic Analysis, случайное индексирование [6] и т. п.). Обработка неструктурированных данных является очень сложной и не решенной до конца задачей, что говорит о перспективности данного направления, но в то же время и о необходимости крайне аккуратной работы с ними. Структурированные же данные довольно легко представляются в необходимом виде.

Следует отметить, что в классическом content-based подходе используются исключительно данные о товарах, причем рекомендации пользователю даются только на основе его личных покупок. Затем с помощью некоей меры близости, часто — коэффициента Пирсона, определяются объекты, близкие к профилю пользователя.

3. **Гибридная рекомендательная система.** Различные подходы к созданию рекомендательных систем обладают своими достоинствами и недостатками, в связи с чем разумным представляется совмещение нескольких подходов в одной системе. Такие рекомендательные системы, основанные на комбинации описанных выше методов, называются **гибридными**. Гибридная система, сочетающая методы *A* и *B*, пытается использовать преимущества *A*, чтобы исправить недостатки *B*. Например, метод коллаборативной фильтрации страдает от проблемы «холодного старта», то есть этот метод не может рекомендовать объекты, которые не имеют рейтингов. Это не ограничивает content-based подход, у которого рекомендация новых объектов основана на их описании (характеристиках), что, как правило, легко доступно.

В заключение отметим, что разработка рекомендательных систем является многопрофильной задачей, требующей привлечения экспертов из различных областей, таких как искусственный интеллект, информационные технологии, интеллектуальный анализ данных, статистика, маркетинг.

2 Оценка качества работы системы

Важным пунктом для разработки рекомендательных систем является оценка качества работы таких систем. В этом разделе приводятся некоторые общепринятые метрики, такие как точность, полнота и т. п.

Следует отметить, что некоторые оценки хорошо применять лишь для систем, предсказывающих рейтинг, а некоторые — лишь для систем, дающих рекомендации типа Top N. Обычно область применения метрики видна из определения.

2.1 Точность (accuracy)

Существует множество различных метрик для оценки точности.

- **mean absolute error (MAE):**

$$\text{MAE} = \frac{\sum_{u \in U} \sum_{i \in \text{testset}_u} |\text{rec}(u, i) - r_{u,i}|}{\sum_{u \in U} |\text{testset}_u|},$$

где u — пользователь; $i \in \text{testset}_u$ — товар из тестовой выборки; $\text{rec}(u, i)$ и $r_{u,i}$ — соответственно оцененный системой и реальный рейтинги товара i у пользователя u ;

- **root mean square error (RMSE):**

$$\text{RMSE} = \sqrt{\frac{\sum_{u \in U} \sum_{i \in \text{testset}_u} (\text{rec}(u, i) - r_{u,i})^2}{\sum_{u \in U} |\text{testset}_u|}};$$

- **normalized MAE and RMSE.** Значения MAE и RMSE делятся на $(r_{\max} - r_{\min})$ — разность между минимальным и максимальным рейтингом;
- **усредненные MAE и RMSE.** Если выборка неравномерна по товарам (пользователям), то можно рассчитывать MAE и RMSE для каждого товара (пользователя) отдельно, а затем брать среднее этих значений.

Метрики MAE и RMSE, как видно из определения, лучше работают с рейтингами (хотя унарный рейтинг (предлагать товар) тоже может быть использован).

В задачах, где рекомендательная система предлагает отсортированный список рекомендаций, можно использовать параметры, стандартные для различных систем классификации.

Можно выделить 4 типа исходов (ошибок или их отсутствия) для рекомендаций:

	Recommended	Not recommended
Used	true positive (tp)	false negative (fn)
Not Used	false positive (fp)	true negative (tn)

На множестве таких исходов можно построить следующие стандартные метрики:

– **точность (precision)**

$$P = \frac{\#tp}{\#tp + \#fp};$$

– **полнота (recall)**

$$R = \frac{\#tp}{\#tp + \#fn};$$

– **уровень ложноположительных (fp) ответов**

$$FP = \frac{\#fp}{\#fp + \#tn};$$

– **аккуратность (accuracy)**

$$A = \frac{\#tp + \#tn}{\#tp + \#tn + \#fp + \#fn}.$$

Значения этих метрик, очевидно, зависят от длины списка выдаваемых рекомендаций. Например, при увеличении длины списка полнота увеличится, а точность упадет. В связи с этим можно в качестве оценок использовать не только сами значения параметров, но и их отношения при изменении длины списка, так называемые Receiving Operating Characteristic: например, график соотношения P к R или R к FP .

Также существуют общепринятые объединенные метрики:

$$F_1 = \frac{2PR}{P + R};$$

$$F_\beta = (1 + \beta^2) \frac{2PR}{\beta^2 P + R}.$$

Часто используются параметры $\beta = 0,5$ и 2 .

2.2 Метрики ранжирования

Если пользователю рекомендуется несколько товаров, важно учитывать, в каком порядке данные товары были рекомендованы. Можно реализовать множество таких метрик. Приведем пример одной из них.

Пусть $hits_u$ — множество товаров, правильно рекомендованных пользователю u . Тогда в качестве метрики для оценки ранжирования можно взять, например,

$$rankscore_u = \sum_{i \in hits_u} \frac{1}{2^{(\text{rank}(i)-1)/\alpha}},$$

где α — параметр, $\text{rank}(i)$ — каким номером был выдан товар в рекомендации.
 Можно нормализовать такую оценку разделив на

$$\text{rankscore}_u^{\max} = \sum_{i \in \text{testset}_u} \frac{1}{2^{(\text{idx}(i)-1)/\alpha}},$$

где $\text{idx}(i)$ — функция перечисления всех товаров из тестовой выборки для пользователя u .

2.3 Покрытие

Рассмотрим метрики покрытия на примере товаров. Для пользователей такие метрики могут быть введены аналогично.

Возможная выдача: $\#(\text{Items_which_can_be_recommended})/\#(\text{All_items})$.

Также для оценки равномерности покрытия можно учитывать вероятность $p(i)$ рекомендации (или выбора) товара i . В качестве метрик можно взять:

- **Gini Index** $1/(n_{\text{items}} - 1) \sum_{j=1}^{n_{\text{items}}} (2j - n_{\text{items}} - 1)p(i_j)$, последовательность номеров товаров i_j упорядочена по неубыванию $p(i_j)$;
- **энтропию Шеннона** $\sum_{i=1}^{n_{\text{items}}} p(i) \log p(i)$.

2.4 Диверсифицированность выдачи

Численно можно оценить, насколько диверсифицирована выдача рекомендаций для пользователя:

$$\text{ILS}_u = \sum_{i \in \text{recset}_u} \sum_{j \in \text{recset}_u, i > j} \text{sim}(i, j),$$

где $\text{sim}(i, j)$ — некоторая функция, показывающая близость товаров i и j .

3 Постановка задачи

В данной работе описано построение архитектуры гибридной рекомендательной системы для решения следующей задачи: для пользователя необходимо выдавать список заданного числа наиболее подходящих к рекомендации объектов (например, товаров из интернет-магазина, книг, фильмов, мобильных приложений). При этом для каждой рекомендации желательно получать вещественный

параметр, характеризующий важность этой рекомендации. Также крайне желательно по каждой рекомендации пользователю давать объяснение, почему эта рекомендация была дана (например, функция «why?» на Amazon.com).

Основная важная особенность поставленной задачи — это необходимость использовать сразу все доступные данные:

- данные профиля пользователя (пол, возраст, местоположение, если есть возможность определить по GPS, интернет-активность, частота и длительность использования тех или иных мобильных приложений и т. п.);
- данные о товаре (описание и характеристики);
- данные покупок (рейтингов) — либо явные оценки пользователя, либо собранная информация об использовании или покупке товаров.

4 Архитектура гибрида. Общая идея

В ходе разработки архитектуры рекомендательной системы удалось совместить различные идеи в один общий гибрид. Данный гибрид имеет несколько этапов работы, в связи с чем на каждом этапе можно использовать подходы, методы и алгоритмы, наиболее подходящие для решения поставленной задачи.

4.1 Основные этапы

Предложенная архитектура системы представляет собой шестиступенчатый гибрид, который было предложено назвать GEFEST в соответствии с его основными этапами:

- **Generation** — генерация начального списка возможных рекомендаций;
- **Expansion** — расширение этого списка схожими товарами;
- **Filtering** — фильтрация расширенного списка;
- **Estimation** — оценка, прогнозирование рейтингов для товаров;
- **Sorting** — сортировка по рейтингу и/или вероятности рейтинга (точности прогноза);
- **Truncation** — усечение, укорачивание отсортированного списка.

Данное разделение на этапы видится очень удачным по следующим причинам:

1. На первом этапе формируется лишь список рекомендаций, соответственно, не требуется от рекомендательной системы выставления рейтингов. Таким образом, могут использоваться системы типа Top N, которые сложно использовать в общем случае. Кроме того, отпадает необходимость строго следить за числом выдаваемых рекомендаций: верхняя граница может быть очень высока и выбирается лишь за счет вычислительной сложности системы.

2. Хорошее прогнозирование рейтингов, особенно если необходимо еще и оценивать достоверность прогноза, является самой вычислительно трудной задачей, т. к. могут быть использованы сложные алгоритмы и классификаторы. Часто используются отдельные классификаторы для каждого товара. Соответственно, прогнозировать рейтинг желательно для как можно меньшего числа товаров. В простых подходах к построению рекомендательных систем приходится для формирования рекомендации вычислять рейтинг для каждого товара, что с учетом большого и постоянно увеличивающегося количества товаров в решаемой задаче недопустимо. Наличие этапов 1–3, не требующих прогнозирования рейтинга, решает эту проблему.
3. Для каждого этапа можно вводить свои метрики оценки качества.

4.2 Требования к данным

Рассматриваемая ниже реализация предложенной идеи требует наличия определенных данных.

Во-первых, это, конечно же, **матрица рейтингов** $R = \{r_{ij}\}$ размера $n_{\text{users}} \times n_{\text{items}}$, где n_{users} — число пользователей, n_{items} — количество объектов-товаров, r_{ij} — вещественнозначный рейтинг, оценка, выставленная явным способом пользователем i товару j либо подсчитанная путем учета частоты и длительности использования данного приложения, покупки или возврата товара в интернет-магазине, посещения веб-страниц с описанием товара и т. п.

Этап 1 (Generation) требует **функции сравнения профилей пользователей** $\text{sim}_{\text{user}}(i, j)$, где i, j — номера пользователей.

Этап 2 (Expansion) требует **функции сравнения профилей товаров** $\text{sim}_{\text{item}}(i, j)$.

На этапе 4 (Estimation) уже требуются векторные представления профилей пользователей и товаров. В этом снова наблюдается преимущество предлагаемого подхода: некоторые неструктурированные данные легко сравнивать, но сложно кодировать в виде векторов. Такие данные можно использовать на этапах 1 и 2, а на этапе 4 не использовать.

Рассмотрим работу каждого этапа более подробно.

5 Generation. Формирование начального списка возможных рекомендаций

На данном этапе должен сформироваться как можно больший список возможных рекомендаций. Планируется использовать данные рейтингов и профилей пользователей.

Результатом работы данного этапа должно стать множество товаров $I_1 = \{i_1, i_2, \dots, i_n\}$, которые хотя бы потенциально могут быть интересны пользователю.

Для обработки данных матрицы рейтингов R планируется использовать item-based collaborative filtering (CF) — модификацию классического подхода к коллаборативной фильтрации, предложенную компанией Amazon [7]. В данном методе, в отличие от классической коллаборативной фильтрации, сравниваются не пользователи, а товары. Метрика и составление векторов товаров могут быть использованы те же, что и в случае сравнения пользователей.

5.1 Item-based collaborative filtering

Первый шаг — формирование списка I_{i-bCF} при помощи Item-based collaborative filtering. Особенность его использования в данном случае — очень высокий порог схожести товаров, которые надо оставлять в рекомендации. В общем-то, этот порог можно сделать неограниченным (т. е. оставлять все товары, хоть как-то схожие по матрице R). Item-based CF, пожалуй, идеально подходит для этого этапа. Причины этого в следующем:

- простота;
- большинство вычислений происходит в фоновом режиме;
- очень высокая скорость работы в реальном времени;
- используются все данные матрицы R ;
- формируются неожиданные рекомендации;
- частично решается проблема холодного старта по пользователям;
- используется разреженность данных;
- метод отлично проверен и, пожалуй, является одним из лучших среди CF.

Недостатки метода: холодный старт по товарам и то, что не используются данные о товарах и пользователях, — будут решены другими частями гибрида.

Итак, после работы метода получается список I_{i-bCF} потенциально рекомендованных товаров на основе данных матрицы R .

5.2 Ближайшие соседи по профилям пользователей

Здесь формируется список товаров I_{upNN} . Пусть требуется выработать рекомендацию для пользователя под номером i . Тогда среди всех остальных пользователей выбирается n (параметр алгоритма) таких, у которых функция $\text{sim}_{\text{user}}(i, j)$ максимальна. Вычисление $\text{sim}_{\text{user}}(i, j)$ для всех пользователей довольно ресурсоемко, в связи с чем предлагается вынесение части вычислений в фоновый режим. А именно будем хранить для каждого пользователя i список n ближайших к нему

пользователей и соответствующие значения функции $\text{sim}_{\text{user}}(i, j)$. Обновление списков можно делать следующим образом:

- с некоторой регулярностью обновлять все списки;
- при добавлении нового пользователя или изменении профиля пользователя. В данном случае, пожалуй, разумно обновлять только список нового или измененного пользователя.

Далее для n ближайших пользователей формируется множество всех товаров $I_{\text{up}NN}$, которые эти пользователи положительно оценили.

Плюсы:

- простота;
- возможность выносить вычисления в фоновый режим;
- очень высокая скорость работы в реальном времени;
- используются данные профилей пользователей;
- в паре с Item-based CF максимально решается проблема холодного старта для пользователей;
- не обязательно наличие векторной модели профиля пользователя;
- можно не особо аккуратно подбирать параметр n и брать его достаточно большим.

Минусы метода: холодный старт по товарам и то, что не используются данные о товарах, — устраняются другими частями гибрида.

5.3 Добавление товаров пользователя

Важно, что к множеству I_1 необходимо добавить и все товары, положительно оцененные пользователем — список I_{user} .

5.4 Формирование общего множества рекомендаций

Множество I_1 формируется объединением $I_{i-bCF} \cup I_{\text{up}NN} \cup I_u$.

6 Expansion. Расширение списка возможных рекомендаций

Так как данные рейтингов и профилей пользователей уже задействованы, расширение будет проводиться на основе данных профилей товаров. Будет использоваться метод ближайших соседей по профилям товаров. Для каждого товара $i \in I_1$ формируется множество Sim_i из n (параметр алгоритма) товаров с наиболее большим значением $\text{sim}_{\text{item}}(i, j)$ (сам товар i тоже принадлежит этому множеству, т.е. $i \in \text{Sim}_i$). Как и в случае с пользователями, вычисление

$\text{sim}_{\text{item}}(i, j)$ для всех товаров довольно ресурсоемко и его можно также вынести в фоновый режим. А именно будем хранить для каждого товара i список n ближайших к нему товаров и соответствующие значения $\text{sim}_{\text{item}}(i, j)$. Обновление списков можно делать следующим образом:

- с некоторой регулярностью обновлять все списки;
- при добавлении нового товара или изменении описания какого-либо товара.

Итогом работы этапа будет множество товаров $I_2 = \bigcup_{i \in I_1} \text{Sim}_i$.

Плюсы:

- простота;
- возможность вынесения вычислений в фоновый режим;
- вкупе с первым этапом максимально решается проблема холодного старта для товаров;
- используются данные профиля;
- можно не особо аккуратно подбирать параметр n и брать его достаточно большим, не оптимизируя его значение на этом этапе.

Минус — не используются данные рейтингов и профилей пользователей. Эта проблема решается другими частями гибрида.

7 Общие выводы для этапов 1–2

7.1 Плюсы и минусы

- + Все используемые методы просты.
- + Решается проблема холодного старта как для пользователей, так и для товаров.
- + Используются все возможные данные, кроме контекстных.
- + Все методы масштабируемы по пользователям и товарам.
- + Большинство вычислений могут быть вынесены в фон.
- + Не обязательны векторные модели профилей пользователей и товаров.
- + Список I_2 может быть на порядки меньше n_{items} , что существенно упрощает работу следующих этапов.
- + Возможность объяснения рекомендаций (см. ниже).
- Большинство товаров в списке I_2 являются плохими рекомендациями (проблема решается на последующих этапах).

7.2 Оценка качества работы

Для I_2 основной и, пожалуй, единственной важной метрикой будет полнота. Пусть I_i — все товары, положительно оцененные пользователем i . Разделим их на обучающие I_i^{train} и тестовые I_i^{test} . Тестовые рейтинги будем считать нулевыми. После этого проведем шаги 1–2 и получим I_2^{test} . Необходимо добиться максимального значения $|I_2^{\text{test}} \cap I_i|/|I_i|$.

7.3 Объяснение рекомендаций

Существенным плюсом использования метода ближайших соседей является возможность объяснения рекомендаций. Если товар из множества I_2 в итоге попадает в выдачу рекомендаций, причину его появления можно легко объяснить пользователю исходя из того, как он попал в список I_2 . Объяснения могут быть следующих типов:

- товар похож на товар i , купленный Вами;
- товар похож на товар i , купленный пользователем u , похожим на Вас;
- товар купил (оценил) пользователь u , похожий на Вас;
- товар покупают вместе с товаром i , купленным Вами, и т. д.

8 Filtering. Фильтрация списка рекомендаций

На данном этапе происходит фильтрация множества I_2 . Из I_2 должны быть исключены заведомо неподходящие рекомендации. На выходе должно получиться $I_3 \subseteq I_2$, причем $I_2 \setminus I_3$ — заведомо плохие рекомендации.

8.1 Простейшая фильтрация

Убираем из множества I_2 товары I_u , уже купленные пользователем. Также можно убрать дубликаты товаров (если имеется возможность их поиска).

8.2 Knowledge-based фильтрация

Knowledge-based системы рекомендуют объект, основываясь на специальных знаниях о том, какие определенные функции объекта отвечают потребностям пользователя и его предпочтениям, и, в конечном счете, насколько объект полезен для пользователя. Их разработка является трудоемкой деятельностью ввиду необходимости создания правил. В связи с этим, не останавливаясь подробно, обозначим лишь некоторые факты.

Желательно реализовать две системы Knowledge-based фильтрации:

- (1) **интерактивную**, когда пользователь сам указывает, что он ищет. Например, тип товара (электротехника, игры, художественные фильмы и т. п.), ограничения по стоимости товара и т. п.;
- (2) **неинтерактивную**, где фильтрация происходит по параметрам пользователя.

8.3 Context Aware фильтрация

Context Aware фильтрация является усовершенствованием Knowledge-based фильтрации и вводит в правила и данные Knowledge-based еще и контекст. Например: время, когда был куплен товар, местоположение пользователя, когда была совершена покупка и т. п. Разработка такой фильтрации еще более сложна, нежели KB, поэтому, пожалуй, нецелесообразна на начальном этапе.

Плюсы и минусы:

- + убираются заведомо неправильные рекомендации;
- + возможность использования разных фильтров;
- + увеличена скорость работы фильтрации за счет этапов 1–2;
- + присутствие фильтрации до прогнозирования рейтингов, а не после формирования списка рекомендаций позволяет еще больше упростить работу классификаторам, задействованным в оценке рейтинга;
- очень высока сложность реализации систем Knowledge-based.

9 Estimation. Прогнозирование рейтингов

Данный этап, пожалуй, является наиболее интересным и дает возможность использовать наиболее современные методы обработки данных и прогнозирования. На вход на данном шаге подается множество $I_3 = \{i_1, i_2, \dots, i_n\}$, на выходе должно получиться множество пар товар–рейтинг: $RI = \{(i_1, r_1), (i_2, r_2), \dots, (i_{|I_3|}, r_{|I_3|})\}$. Или, что более предпочтительно, для каждого товара $i \in I_3$ получить список (если, например, рейтингов пять) $(p_1, p_2, p_3, p_4, p_5)$, где p_i — прогноз вероятности того, что рейтинг равен i .

Стоит снова упомянуть, что важнейшим плюсом предлагаемой системы является то, что множество I_3 относительно небольшое, к тому же его размер можно регулировать параметрами предыдущих шагов. В связи с этим для прогнозирования рейтингов можно использовать достаточно сложные системы, что не скажется критически на времени формирования рекомендации.

Главной сложностью при разработке этапа Estimation является необходимость использовать все три типа данных: профили пользователей, профили товаров и

рейтинги. Как уже отмечалось ранее, профили пользователей и описания товаров должны быть представлены в векторном виде.

Введем обозначения:

$R = \{r_{ij}\}$ — как и раньше, матрица рейтингов;

u_i — вектор рейтингов пользователя $u_i = (r_{i1}, r_{i2}, \dots, r_{in_{\text{items}}})$;

a_i — вектор рейтингов товара $a_i = (r_{1i}, r_{2i}, \dots, r_{n_{\text{users}}i})$;

pr_i — вектор профиля i -го пользователя $pr_i = (pr_{i1}, pr_{i2}, \dots, pr_{in_{pr}})$, где n_{pr} — длина вектора профиля пользователя;

d_i — вектор описания i -го товара $d_i = (d_{i1}, d_{i2}, \dots, d_{in_d})$, где n_d — длина вектора описания товара.

9.1 Выбор классификатора

Для хорошего прогнозирования необходим хороший современный классификатор, обладающий следующими свойствами:

- быстрая скорость классификации (как минимум необходимо наличие обучения);
- высокое качество прогнозирования;
- возможность оценки достоверности прогнозирования;
- возможность легкого дообучения (не обязательно, но крайне желательно).

И если метод ближайших соседей отлично подходил для формирования предварительного списка, то он уже не подходит ни по одному из вышеперечисленных пунктов, кроме последнего.

9.2 Проблемы применения классификаторов на используемых данных

Допустим в системе имеется какой-либо классификатор, на вход которого подается информация о пользователях. Вся информация о пользователе i — это векторы u_i и p_i . Логично записать компоненты этих векторов в один вектор («склейка» векторов) $(pr_i, u_i) = (pr_{i1}, pr_{i2}, \dots, pr_{in_{pr}}, r_{i1}, r_{i2}, \dots, r_{in_d})$ и подавать его на классификатор. В таком подходе возникают следующие проблемы:

Разреженность и большая размерность. Получившийся вектор будет очень сильно разрежен, а далеко не все классификаторы могут это использовать. Соответственно, за счет огромного числа товаров не использующие разреженность классификаторы будут работать непозволительно для рекомендательной системы медленно.

Нефиксированная размерность (масштабируемость). Практически все (а возможно, и все) классификаторы с обучением требуют фиксированной размерности подаваемых векторов. Длина вектора профиля пользователя будет меняться крайне редко, а вот длина u_i — очень часто. Получается, что при добавлении в систему нового товара необходимо полностью перестраивать классификаторы, что непозволительно.

Несбалансированность. Части u_i и pr_i в общем и целом должны иметь одинаковый или, по крайней мере, регулируемый приоритет. Однако в итоговом векторе полностью отсутствует баланс: часть u_i на порядки длиннее pr_i , pr_i — полностью заполненный вектор, а u_i — очень сильно разреженный. Для некоторых классификаторов такой дисбаланс может критически сказываться на качестве работы.

Следует отметить, что данные проблемы являются чисто техническими, а не относящимися к качеству рекомендательной системы. Для их решения предлагается использовать один простой и очень хорошо апробированный метод.

9.3 Снижение размерности случайным кодированием

Рассмотрим простейший случай бинарных векторов $B^n = 0, 1^n$, где n очень большое. Пусть также на этих векторах есть какая-то метрика близости (x, y) , например $1 - \cos(x, y)$. Имеется некоторое множество таких векторов $M \subset B^n$, причем векторы в этом множестве разрежены. Задача состоит в следующем: необходимо задать некоторое отображение $f : B^n \rightarrow B^m$, где $m < n$, так, чтобы для векторов из M норма примерно сохранилась, т. е. $E(x, y) > E(u, v) \Rightarrow E(f(x), f(y)) > E(f(u), f(v))$ для большинства векторов x, y, u, v из M . Существует хороший и апробированный метод, позволяющий это сделать. Данный метод показал отличные результаты в обработке текстов, причем такое снижение размерности не только ускоряло, но и улучшало качество работы классификаторов, видимо, за счет добавления некоторой небольшой «нечеткости».

Сопоставим каждому числу $i = \overline{1, n}$ случайный разреженный вектор $y(i) \in B^m$ с k единицами и $(m - k)$ нулями. Тогда для $x \in B^n$

$$f(x) = \bigvee_{i: x_i=1} y(i) = \bigvee_i x_i \cdot y(i),$$

т. е. для векторов $y(i)$, соответствующих единицам в векторе x , делается побитовое OR.

Параметрами такого алгоритма являются m и k , причем m надо взять как можно меньшим. Эти параметры элементарно подбираются по множеству M и тесту $E(x, y) > E(u, v) \Rightarrow E(f(x), f(y)) > E(f(u), f(v))$.

9.3.1 Обобщение на целочисленные векторы

На целочисленные векторы метод можно обобщить естественным образом: формула

$$f(x) = \bigvee_i x_i \cdot y(i)$$

остаётся неизменной, а вместо \bigvee используется покомпонентный максимум. Конечно же, m в таком случае заметно увеличится.

9.3.2 Пример использования в рекомендательных системах

Рассмотрим использование случайного кодирования на примере, описанном в пункте 9.2 Будем уменьшать размерность вектора u_i . Т. е. каждому товару при его появлении в базе будет приписываться случайный разреженный бинарный вектор размерности m , где m заметно меньше n_{items} . Теперь вектор u_i будет формироваться «сложением» (в смысле покомпонентного максимума) товаров, рейтинги которых для пользователя нам известны. То же самое можно сделать и для вектора a_i .

Такой подход решает **все** перечисленные в подразд. 9.2 проблемы, кроме проблемы сбалансированности, так как полученный вектор u_i все равно будет разреженным в большинстве случаев, а rg_i — нет. Для многих классификаторов это не страшно, однако для наиболее быстрых (например, линейного или нейронных сетей с малым числом слоев) неразрезанные данные вообще плохо подходят. Для них необходима некоторая «средняя» разреженность (чтобы классы могли разделиться, но чтобы размерность оставалась в удобных для работы пределах). Векторы u_i можно сделать очень хорошо подходящими для таких классификаторов. Векторы rg_i также можно превратить в такие же векторы ровно таким же способом расширения размерности, вся разница лишь в том, что m в данном случае больше n . Такой метод используется в зрительной системе человека и отлично проявил себя, например, в задачах распознавания образов.

Кроме того, что очень важно и отличает данный метод от других методов снижения размерности, при новой оценке пользователя все, что необходимо сделать, — это прибавить к его уже сформированному вектору u_i соответствующий вектор товара, умноженный на рейтинг.

9.4 Прогнозирование рейтингов

Планируется использовать один большой классификатор, имеющий n_{ratings} классов, где n_{ratings} — количество возможных рейтингов, т. е. классами будут $r = 1, r = 2, \dots, r = n_{\text{ratings}}$. На вход классификатора будет подаваться вся информация о паре (рейтинг, товар). Например, если пользователь i поставил

товару j оценку «4», то на обучение, как элемент класса $r = 4$, подается вектор (pr_i, u_i, a_j, d_j) , где векторы u_i и a_j получены методом снижения размерностей. На обучение подаются все известные рейтинги.

Соответственно, для получения предполагаемого рейтинга на классификатор подаются четыре «склеенных» вектора (pr_i, u_i, a_j, d_j) .

9.5 Возможность использования Context Aware

Именно на этом этапе можно наиболее активно использовать контекст, например, подавать на классификатор еще и вектор контекстных данных (время, местоположение и т. п.), т. е. в вектор, попадающий на обучение или на классификацию, добавится вектор контекста, например:

$c(\text{year, season, month, day, day_or_night, hour, country, city, network_type, } \dots)$.

Тогда вектор, подаваемый на классификатор для обучения или классификации, будет иметь вид (pr_i, u_i, a_j, d_j, c) . Соответственно, необходимо хранить c для каждой покупки и/или использования. Для вектора c может быть актуально расширение размерности.

В данном случае можно использовать встроенные возможности OLAP-систем (в общем-то, те же классификаторы). Реализация СА и уж тем более эффективная работа с OLAP-системами является непростой задачей, однако включение какого-либо элементарного и очевидного контекста, несомненно, повысит точность рекомендаций.

9.6 Взвешенный гибрид и адаптация

В связи с наличием огромного числа различных классификаторов, разумно на этапе 4 использовать их взвешенный гибрид.

Пусть имеется n_c классификаторов. Для пользователя i и товара j классификатором k были предсказаны вероятности рейтинга p_1^k, p_2^k, \dots . Тогда итоговыми вероятностями будут

$$p_r^k = \frac{\sum_{k=1}^{n_c} w_k p_r^k}{n_c},$$

где w_k — веса классификаторов, изменяемые адаптацией.

Кроме того, именно в классификаторах четвертого этапа скрыто большинство управляемых параметров, по которым может производиться адаптация системы.

10 Sorting и Truncation. Формирование окончательной рекомендации

На данном этапе формируется окончательная рекомендация. Простейший способ — отсортировать все товары из I_3 по рейтингу с максимальной вероятностью или среднему значению рейтинга $\sum p_i i$ и выдать N лучших. Кроме того, можно разбить выдачу на категории или интерактивно просить пользователя задать число выводимых рекомендаций.

11 Общие замечания

11.1 Плюсы и минусы

- + Разбиение на этапы имеет простую и естественную структуру.
- + Используются **все возможные данные** как для формирования начального списка, так и при формировании рейтингов.
- + На этапах 1–2 используются простые и апробированные методы.
- + Небольшое относительно n_{items} число элементов, попадающих на классификатор, позволяет использовать сложные современные классификаторы.
- + Решена проблема холодного старта как для товаров, так и для пользователей.
- + На всех этапах основные вычисления можно вынести в фоновый режим.
- + Все этапы легко масштабируемы как по пользователям, так и по товарам.
- + Все этапы должны быстро работать в онлайн-режиме.
- + Мгновенная реакция на проставление новых рейтингов.
- + Отсутствие необходимости наличия векторной модели профиля пользователя и описания товара на этапах 1–2.
- + Достаточно быстрая реакция на изменение профиля пользователя (рейтинги меняются сразу, а вот формирование начального списка требует вычислений близости с остальными пользователями, хотя это не очень затратно).
- Отсутствие данных об апробированности склейки векторов, производимой на этапе Estimation.

Литература

1. *Bennett J., Lanning S.* The Netflix Prize // KDD Cup Workshop at SIGKDD-07, 13th ACM Conference (International) on Knowledge Discovery and Data Mining Proceedings. — San Jose, California, USA, 2007. P. 3–6.
2. *Jannach D., Zanker M., Felfernig A., Friedrich G.* Recommender systems: An introduction. — Cambridge: Cambridge University Press, 2010.

3. *Ricci F., Rokach L., Shapira B., Kantor P. B.* Recommender systems handbook. — New York: Springer-Verlag, 2010.
4. *Schafer J. B., Frankowski D., Herlocker J., Sen S.* Collaborative filtering recommender systems // *The Adaptive Web*. — Berlin/Heidelberg: Springer, 2007. P. 291–324.
5. *Pazzani M. J., Billsus D.* Content-based recommendation systems // *The Adaptive Web*. — Springer Verlag, 2007. P. 325–341.
6. *Sahlgren M.* An introduction to random indexing // *Methods and Applications of Semantic Indexing Workshop at the 7th Conference (International) on Terminology and Knowledge Engineering*. — Citeseer: TKE, 2005.
7. *Linden G., Smith B., York J.* Amazon.com recommendations: Item-to-item collaborative filtering // *Internet Computing, IEEE*, 2003. Vol. 7. P. 76–80.

НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ ФУНКЦИИ ПЛОТНОСТИ СМЕСЕЙ ВЕРОЯТНОСТНЫХ ЗАКОНОВ С ПОМОЩЬЮ EM-АЛГОРИТМА*

В. Ю. Королев¹, А. Ю. Корчагин², О. А. Морева³

Аннотация: Проведено эмпирическое сравнение ядерных оценок плотности вероятности и оценок плотности, получаемых различными версиями EM-алгоритма на выборках из различных классов смесей вероятностных законов, сгенерированных с помощью датчика псевдослучайных чисел. Также рассмотрен модифицированный сеточный аналог EM-алгоритма.

Ключевые слова: ядерная оценка плотности; EM-алгоритм; сеточный метод разделения плотности

1 Введение

Задача оценивания плотности распределения случайных величин является традиционной задачей статистического анализа данных, которая представляет не только теоретический интерес, но и имеет большое практическое значение. Обычно выделяют два основных типа методов статистического оценивания: параметрические и непараметрические.

При использовании непараметрических методов не требуется полная спецификация модели, вместо этого данные сами формируют ее определенным образом. По этой причине непараметрические методы удобно использовать, когда не удастся построить параметрическую модель, хорошо согласующуюся с исходными данными. В данной работе анализируются результаты работы EM-алгоритма с информационным критерием выбора количества компонент смеси и ядерных методов на модельных выборках. Результаты работы сеточного аналога EM-алгоритма сопоставляются с ядерными оценками. Обсуждаются модификации сеточного аналога EM-алгоритма.

*Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 11-01-12026-офи-м, 12-01-00115а), Министерством образования и науки (госконтракт 16.740.11.0133).

¹Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики РАН, vkorolev@cs.msu.su

²Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, sasha.korchagin@gmail.com

³Университет г. Ульм, Германия, o.moreva@gmail.com

Впервые итерационная процедура типа EM-алгоритма была опубликована в работе [1]. Затем после довольно большого перерыва эта идея вновь возникла в работе [2]. Возникновение электронно-вычислительных машин позволило реализовывать довольно сложные итерационные процедуры. Это стимулировало развитие идей, лежащих в основе EM-алгоритма и нашедших впоследствии свое отражение в работах [3–6]. Одним из основных применений EM-алгоритма является разделение конечных смесей нормальных законов, которые, как известно, представляют собой довольно гибкую модель исходного распределения. Подробное описание EM-алгоритма для статистического разделения смесей вероятностных законов можно найти в книге [7].

Идея, лежащая в основе методов восстановления плотности распределения с помощью ядерных оценок, была предложена в техническом докладе Военно-воздушных сил США (USAF), а первая опубликованная статья Rosenblatt [8] вышла в 1956 г. С того времени ядерные оценки привлекли большое внимание статистиков, что подтверждается обширной статистической литературой на эту тему, в частности работами [9–15]. Подробный обзор ядерных методов можно найти в работе [16].

Как правило, в качестве ядра в непараметрических оценках плотности распределения выбираются плотности вероятностей. При этом ядерные оценки являются конечными смесями распределений, соответствующих таким плотностям. В ядерных оценках присутствует параметр масштаба, характеризующий гладкость получаемой оценки. Методы выбора этого параметра на практике во многом эвристичны и фактически оставляют решение о конкретном значении этого параметра на произвол исследователя. В то же время при разделении смесей с помощью алгоритмов EM-типа, во-первых, вместо единственного параметра масштаба возникают несколько параметров масштаба компонент и, во-вторых, они не выбираются произвольно, но оцениваются автоматически. Это обстоятельство делает возможным применение алгоритмов EM-типа для построения непараметрических оценок плотности.

В статье проводится эмпирическое сравнение упомянутых выше методов построения непараметрических оценок плотности на выборках, полученных при генерации реализаций случайных величин из различных классов смесей вероятностных законов с помощью датчика псевдослучайных чисел.

Помимо этого, в статье рассматривается сеточный аналог EM-алгоритма. В данной работе он обозначается аббревиатурой GCG (Grid Conjugate Gradient). Сеточные методы разделения смесей вероятностных законов предложены в работе [17]. Идея, лежащая в основе сеточных методов, заключается в замене интервалов, на которых оцениваются параметры, некими сетками с целью значительного уменьшения мощности параметрического множества. В статье предложены варианты улучшения сеточного алгоритма в части выбора сетки и введения второго прохода.

2 Смеси вероятностных законов

Смеси вероятностных законов находят самое широкое применение как модели распределения многих случайных величин, наблюдаемых на практике. Так, распределения приращений процессов эволюции финансовых индексов и процессов плазменной турбулентности хорошо описываются конечными сдвиг-масштабными смесями нормальных законов. При исследовании стохастической структуры хаотических информационных потоков в сложных телекоммуникационных сетях находят свое применение смеси гамма-распределений [18].

Дадим формальное определение смеси. Рассмотрим функцию $F(x, y)$, определенную на множестве $\mathbb{R} \times \mathbb{Y}$. Для простоты будем предполагать, что \mathbb{Y} — это некоторое подмножество m -мерного евклидова пространства, $\mathbb{Y} \in \mathbb{R}^m$ при некотором $m \geq 1$, причем множество \mathbb{Y} снабжено борелевской σ -алгеброй Σ . Более того, предположим, что при каждом фиксированном y функция $F(x, y)$ является функцией распределения по x , а при каждом фиксированном x функция $F(x, y)$ измерима по y , т. е. для любых $x \in \mathbb{R}$ и $c \in \mathbb{R}$ выполнено условие $\{y : F(x, y) < c\} \in \Sigma$. Пусть Q — вероятностная мера, определенная на измеримом пространстве (\mathbb{Y}, Σ) . Функция распределения

$$H(x) = \int_{\mathbb{Y}} F(x, y) Q(dy), \quad x \in \mathbb{R},$$

называется смесью функции распределения $F(x, y)$ по y относительно Q . Распределение $F(x, y)$ называется смешиваемым, в то время как мера Q задает смешивающее распределение.

Если случайный вектор Y имеет дискретное распределение и принимает значения y_1, y_2, \dots с вероятностями p_1, p_2, \dots соответственно, то получается смесь вида

$$H(x) = EF(x, Y) = \sum_{j \geq 1} p_j F(x, y_j), \quad x \in \mathbb{R},$$

называемая дискретной. В таком случае функции распределения $F(x, y_j)$ называются компонентами смеси $H(x)$, а числа p_j , $j \geq 1$, называются весами соответствующих компонент. Если в дискретной смеси число ненулевых весов конечно (т. е. случайный вектор Y принимает k (конечное число) значений), то дискретная смесь называется конечной, а k называют числом компонент смеси

$$H(x) = EF(x, Y) = \sum_{j=1}^k p_j F(x, y_j), \quad x \in \mathbb{R}.$$

Если функции распределения $F(x, y)$ соответствует плотность $f(x, y)$, то дискретной смеси $H(x)$ соответствует плотность

$$h(x) = \sum_{j \geq 1} p_j f(x, y_j), \quad x \in R.$$

Далее будем рассматривать конечные смеси нормальных законов, плотность которых имеет вид:

$$h(x) = \sum_{i=1}^k p_i f_i(x, y_i), \quad (1)$$

где

$$f(x, y_j) = f_j(x) = \frac{1}{\sigma_j} \phi\left(\frac{x - a_j}{\sigma_j}\right), \quad \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \quad x \in R.$$

3 ЭМ-алгоритм для разделения смесей

В случае конечных смесей нормальных законов задачей разделения смеси

$$f(x) = \sum_{j=1}^k \frac{p_j}{\sigma_j} \phi\left(\frac{x - a_j}{\sigma_j}\right), \quad x \in R,$$

принято называть задачу статистического оценивания параметров $\theta = (p_1, \dots, p_k, a_1, \dots, a_k, \sigma_1, \dots, \sigma_k)$ по известным реализациям случайной величины X .

Необходимым условием существования решения задачи разделения смеси вероятностных распределений вида (1) является идентифицируемость смеси, т. е. совпадение неизвестных параметров при тождественном совпадении смесей вида (1) как функций аргумента x . Конечные смеси нормальных законов обладают свойством идентифицируемости (см., например, [7]).

Для решения задачи разделения смесей традиционно используется ЭМ-алгоритм. ЭМ-алгоритм — итерационный алгоритм, используемый в математической статистике для нахождения оценок максимального правдоподобия параметров вероятностных моделей в случае, когда модель зависит от некоторых скрытых переменных. Пусть $X \in \mathbb{R}^n$ — некоторые значения наблюдаемых переменных, а $Y \in \mathbb{R}^m$ — скрытые переменные. Вместе X и Y образуют полный набор данных. Будем считать, что совместная плотность случайного вектора (X, Y) абсолютно непрерывна относительно меры Лебега. Положим f — плотность вероятности полного набора данных с параметром $\theta \in \Theta$: $f(X, Y|\Theta) = f_\theta(x, y) = P_\theta(X = x, Y = y)$. Эту функцию можно понимать как функцию правдоподобия всей модели, если рассматривать ее как функцию параметра θ .

Итерация алгоритма состоит из двух шагов. На E-шаге (expectation) вычисляется ожидаемое значение функции правдоподобия, при этом скрытые переменные рассматриваются как наблюдаемые. На M-шаге (maximization) вычисляется оценка максимального правдоподобия, таким образом увеличивается ожидаемое правдоподобие, вычисляемое на E-шаге. Затем это значение используется для E-шага на следующей итерации. Алгоритм выполняется до сходимости. Опишем подробно этапы EM-алгоритма.

1. Этап вычисления математического ожидания (E-этап). Определим функцию $Q(\theta; \theta^{(m)})$ как условное математическое ожидание логарифма полной функции правдоподобия при известном значении наблюдаемой компоненты X :

$$Q(\theta; \theta^{(m)}) = E_{\theta^{(m)}} [\log f_{\theta}(X, Y) | X] . \quad (2)$$

В этом определении θ является аргументом функции $Q(\theta; \theta^{(m)})$, X и $\theta^{(m)}$ — параметры, так что в соотношении (4) символ $E_{\theta^{(m)}}$ означает усреднение по Y относительно меры $P_{\theta^{(m)}}$. При известном значении $X = x$ функцию $Q(\theta; \theta^{(m)})$ можно вычислить по формуле

$$Q(\theta; \theta^{(m)}) = \int_{\mathbb{R}^m} [\log f_{\theta}(x, y)] f_{\theta^{(m)}}(y|x) dy .$$

2. Этап максимизации (M-этап). На этом этапе вычисляется

$$\theta^{(m+1)} = \arg \max_{\theta} Q(\theta; \theta^{(m)}) .$$

Итерационный процесс останавливается в соответствии с заранее согласованным критерием остановки. Например, заранее выбирается какая-нибудь метрика $\rho(\theta_1, \theta_2)$ и фиксируется малое положительное число ϵ . Процесс останавливается на m -м шаге, если $\rho(\theta^{(m)}, \theta^{(m-1)}) < \epsilon$.

Для конечных смесей нормальных законов EM-алгоритм сводится к одношаговому итерационному процессу. Для данного случая апостериорная вероятность того, что наблюдение x_j было сгенерировано в соответствии с распределением, задаваемым i -ой компонентой смеси (1), имеет вид:

$$f_{\theta}(i|x_j) = \frac{(p_i/\sigma_i)\phi((x_j - a_i)/\sigma_i)}{\sum_{r=1}^k (p_r/\sigma_r)\phi((x_j - a_r)/\sigma_r)} .$$

Пусть значение $\theta^{(m)} = (p_1^{(m)}, \dots, p_k^{(m)}, a_1^{(m)}, \dots, a_k^{(m)}, \sigma_1^{(m)}, \dots, \sigma_k^{(m)})$ параметра θ на m -й итерации EM-алгоритма известно. Обозначим

$$g_{ij}^{(m)} = f_{\theta^{(m)}}(i|x_j) = \frac{\left(\frac{p_i^{(m)}}{\sigma_i^{(m)}}\right) \phi\left(\frac{(x_j - a_i^{(m)})}{\sigma_i^{(m)}}\right)}{\sum_{r=1}^k \left(\frac{p_r^{(m)}}{\sigma_r^{(m)}}\right) \phi\left(\frac{(x_j - a_r^{(m)})}{\sigma_r^{(m)}}\right)} =$$

$$= \frac{\left(\frac{p_i^{(m)}}{\sigma_i^{(m)}}\right) \exp\left\{\left(\frac{(x_j - a_i^{(m)})}{\sigma_i^{(m)}}\right)^2 / 2\right\}}{\sum_{r=1}^k \left(\frac{p_r^{(m)}}{\sigma_r^{(m)}}\right) \exp\left\{\left(\frac{(x_j - a_r^{(m)})}{\sigma_r^{(m)}}\right)^2 / 2\right\}}.$$

Тогда уточненные значения параметров p_i , a_i и σ_i на $(m + 1)$ -й итерации EM-алгоритма имеют вид:

$$p_i^{(m+1)} = \frac{1}{n} \sum_{j=1}^n g_{ij}^{(m)}; \quad a_i^{m+1} = \frac{1}{\sum_{j=1}^n g_{ij}^{(m)}} \sum_{j=1}^n g_{ij}^{(m)} x_j;$$

$$\sigma_i^{m+1} = \left[\frac{1}{\sum_{j=1}^n g_{ij}^{(m)}} \sum_{j=1}^n g_{ij}^{(m)} (x_j - a_i^{(m+1)})^2 \right]^{1/2}, \quad i = 1, \dots, k.$$

Эти соотношения определяют EM-алгоритм для разделения конечных смесей нормальных законов. Фактически указанные рекуррентные соотношения реализуют оба этапа (E-этап и M-этап) EM-алгоритма одновременно.

3.1 Выбор числа компонент смеси

При использовании EM-алгоритма для разделения конечных смесей считается, что число компонент смеси известно заранее, однако на практике это не всегда так. Иногда число компонент удается оценить визуально, однако такой подход препятствует автоматизации процесса анализа данных. Можно решать задачу несколько раз при последовательном увеличении числа компонент смеси k , построить график зависимости функции правдоподобия выборки от k и выбрать наименьшее k , при котором график претерпевает резкий скачок правдоподобия.

Наличие резкого скачка свидетельствует о том, что модель была выбрана удачно; отсутствие же не дает никакой информации о выборе k .

Понятно, что чем больше параметров в модели, тем лучше она описывает исходные данные. Однако большое число компонент смеси, во-первых, затрудняет интерпретацию модели, а во-вторых, существенно увеличивает время работы алгоритма.

Для определения числа компонент смеси будем использовать информационный критерий Акаике (AIC), предложенный Хиротугу Акаике в 1971 г. В качестве показателя качества модели X . Акаике предложил рассматривать расстояние Кульбака–Лейблера между истинной и наиболее правдоподобной модельной плотностями. Модель тем лучше, чем меньше указанное расстояние. Подробно об этом критерии можно прочитать в [7].

Величина AIC определяется как

$$\text{AIC} = -2 \log f_M(X; \hat{\theta}(X)) + 2d,$$

где $f_M(X; \hat{\theta}(X))$ — функция правдоподобия модели, d — размерность параметра θ . Статистика AIC служит критерием качества модели $f_M(x, \theta)$, точнее, мерой несогласия модели и реальных данных: чем меньше AIC, тем лучше модель. При этом ясно, что включение в модель дополнительных параметров может только увеличить правдоподобие модели и, стало быть, уменьшить первое слагаемое в AIC. Однако при этом увеличивается второе слагаемое, играющее роль штрафа за использование дополнительных параметров.

Пусть теперь $x = (x_1, \dots, x_n)$ — исходная наблюдаемая выборка, элементы которой являются независимыми реализациями случайной величины, плотность распределения которой имеет вид:

$$f(x) = \sum_{i=1}^k \frac{p_i}{\sigma_i} \phi\left(\frac{x - a_i}{\sigma_i}\right), \quad x \in R, \quad (3)$$

где $\phi(x)$ — стандартная нормальная плотность, $a_i \in R$, $\sigma_i > 0$, $p_i \geq 0$, $p_1 + \dots + p_k = 1$, $i = 1, \dots, k$. Соответствующая этой модели функция правдоподобия имеет вид:

$$L_k(\theta; x) = \prod_{j=1}^n \left[\sum_{i=1}^k \frac{p_i}{\sigma_i} \phi\left(\frac{x_j - a_i}{\sigma_i}\right) \right], \quad (4)$$

где $\theta = (p_1, \dots, p_k, a_1, \dots, a_k, \sigma_1, \dots, \sigma_k)$. Размерность параметра θ равна $d = 3k - 1$. Поэтому с использованием критерия Акаике ответ на вопрос о том, какое значение числа k компонент смеси является оптимальным, имеет вид:

$$k_{opt} = \arg \min_k \left\{ -\log L_k(\hat{\theta}(k); x) + 3k - 1 \right\},$$

где $\hat{\theta}(k)$ — оценка максимального правдоподобия параметра θ , построенная в соответствии с моделью (3).

3.2 Выбор начальных приближений

Функция правдоподобия (4) модели, вообще говоря, не является выпуклой, поэтому EM-алгоритм может находить как локальные, так и глобальные максимумы. В связи с этим EM-алгоритм не всегда находит «правильные» оценки параметров. Так как не существует критерия, позволяющего достоверно распознать природу достигнутого максимума — является ли он глобальным или локальным, — то результат работы EM-алгоритма очень сильно зависит от начальных приближений.

В данной работе начальные приближения выбирались с помощью датчика псевдослучайных чисел следующим образом:

$$p = (p_1, \dots, p_k) = \frac{1}{\sum_{i=1}^k u_i} (u_1, \dots, u_k),$$

где $u_i \sim U[0, 1]$ (равномерное распределение на $[0, 1]$) и независимы,

$$a = (a_1, \dots, a_k) = (u_1, \dots, u_k),$$

где $u_i \sim U[X_{(1)}, X_{(n)}]$ и независимы, $X_{(1)}$ и $X_{(n)}$ — минимальный и максимальный элементы выборки

$$\sigma = (\sigma_1, \dots, \sigma_k) = \bar{\sigma}(v_1, \dots, v_k),$$

где $v_i \sim U[0, 25, 1, 75]$ и независимы, $\bar{\sigma}$ — оцененное стандартное отклонение всей выборки.

4 Сеточный аналог EM-алгоритма

Рассмотрим конечную смесь нормальных законов вида

$$f(x) = \sum_{i=1}^k p_i \phi\left(\frac{x - a_i}{\sigma_i}\right),$$

где $\phi(x)$ — стандартная нормальная плотность распределения, а a_i , σ_i , p_i ($p_1 + \dots + p_k = 1$) — параметры, подлежащие оцениванию.

Если известны диапазоны изменения оцениваемых параметров (например, что $\underline{a} \leq a_i \leq \bar{a}$, $\forall i$), задача сводится к отысканию оценок из заданных интервалов. Идея, лежащая в основе сеточных методов, заключается в замене интервалов $[\underline{a}; \bar{a}]$ и $(0; \bar{\sigma}]$ возможных значений параметров масштаба σ_i и сдвига a_i дискретными множествами известных точек:

$$f(x) = \sum_{i=1}^k p_i \phi\left(\frac{x - a_i}{\sigma_i}\right) \approx \sum_{i=1}^K \tilde{p}_i \phi\left(\frac{x - \tilde{a}_i}{\tilde{\sigma}_i}\right) \equiv \tilde{f}(x),$$

где K — число узлов полученной сетки; p_i , a_i , σ_i и \tilde{p}_i — неизвестные параметры; a_i и σ_i — известные параметры.

Задача свелась к поиску весов ограниченного числа фиксированных компонент, где веса заданной компоненты могут быть близки к нулю. Принципы работы сеточных методов подробно описаны в работе [17].

Применительно к сеточным методам логарифмическая «сеточная» функция правдоподобия имеет вид

$$L(p, x) = \log \prod_{j=1}^n \sum_{i=1}^K \tilde{p}_i \phi_{ij} = \sum_{j=1}^n \log \left(\sum_{i=1}^K \tilde{p}_i \phi_{ij} \right),$$

где $p = (\tilde{p}_1, \dots, \tilde{p}_K)^T$.

Задачу поиска МП-оценок можно свести к задаче условной оптимизации $J(p) = L(p; x) \rightarrow \sup, p \in \mathcal{P} \subset \mathbb{R}^K$. Оптимизация осуществляется методом условного градиента и детально описана в работах [19, 20].

На рис. 1 приведен пример равномерной сетки (по математическим ожиданиям), построенной автоматически по исходным данным и наложенной на график функции плотности.

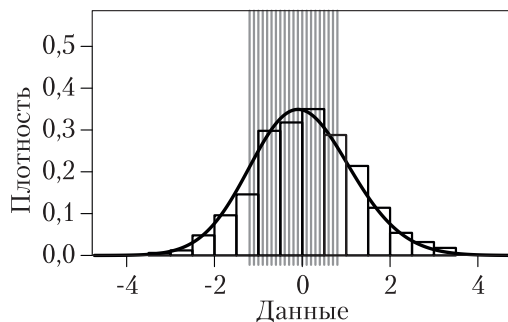


Рис. 1 Пример равномерной сетки математических ожиданий (серые линии)

Основными преимуществами GCG-метода являются устойчивость к входным данным и достаточное быстродействие по сравнению с классической версией EM-алгоритма. Именно быстродействие алгоритма позволяет за разумное время прогонять его несколько раз и получать оценки на разных начальных приближениях, которые генерируются случайно. Эта процедура помогает устранить влияние выбора начального приближения на итоговые результаты. Далее пойдет речь о другой особенности при работе с сеточными алгоритмами — собственно выборе сетки, по которой ведется приближение.

При реализации алгоритма компоненты с незначительными весами отбрасывались по одному из двух критериев: критерию Акаике или пороговому критерию. Пороговый метод предполагает введение некоторого порога значимости таким образом, что все компоненты, полученные GCG-методом, меньшие этого порога, исключаются из рассмотрения. Примером может служить $p_{div} = 0.05$. Для любого из двух методов при отбрасывании («обнулении») компонент вес отброшенных компонент распределяется пропорционально по оставшимся ненулевым компонентам, чтобы соблюсти условие равенства единице суммы вероятностей.

4.1 Проблемы выбора сетки, двухпроходный алгоритм

Перед любым исследователем, который решает воспользоваться сеточными методами, непременно встает вопрос выбора сетки. С одной стороны, сетку нужно выбирать достаточно частой, чтобы можно было оценить параметры компонент (математическое ожидание и дисперсию) достаточно точно. С другой стороны, она должна быть достаточно редкой и состоять из небольшого числа узлов, чтобы можно было достаточно точно оценить вес данного конкретного узла на основании относительно небольшой исходной выборки.

По умолчанию будем рассматривать *прямоугольную* сетку, которая получается пересечением сетки по математическим ожиданиям с сеткой по дисперсиям, т. е. сначала выбирается разбиение математических ожиданий $a_1 < a_2 < \dots < a_M$ размерности M и разбиение дисперсий $\sigma_1 < \sigma_2 < \dots < \sigma_N$. Итоговая сетка получается как декартово произведение выбранных разбиений, т. е. множество пар (a_i, σ_j) , где $1 \leq i \leq M$, $1 \leq j \leq N$ размерности MN .

Сразу отметим, что это оставляет пространство для дальнейшей оптимизации: можно разумным образом выбирать, какие из узлов этой прямоугольной сетки действительно нужны, а какие можно не рассматривать. Вопрос выбора *непрямоугольных* сеток выходит за рамки этой работы. На рис. 2 показаны примеры прямоугольной и непрямоугольной сеток.

Разумно начать решать вопрос выбора сетки с выделения двух следующих фактов. Во-первых, число узлов итоговой сетки можно выбирать на основании размера исходной выборки, т. е. чем больше выборка, тем больше можно выбирать узлов в итоговой сетке. Во-вторых, границы сетки (минимальные и максимальные значения математического ожидания и дисперсии) нужно задавать аккуратно —

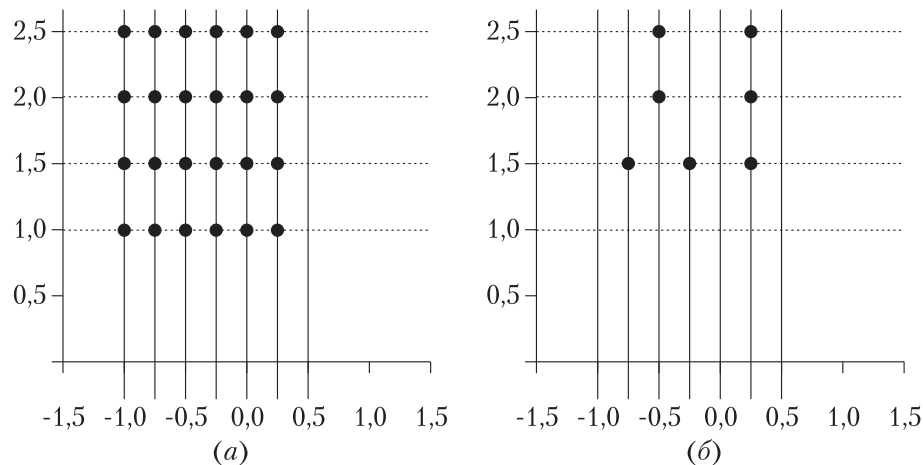


Рис. 2 Сравнение прямоугольной (а) и непрямоугольной (б) сеток

так, чтобы не войти в зону, где весомых компонент точно не будет, тем самым «израсходовав» узлы, но и так, чтобы не пропустить возможные «весомые» компоненты из-за отсутствия узлов рядом.

При первоначальных прогонах сеточного алгоритма сетка выбиралась равномерной ($a_i - a_{i-1} = h_1$, $\sigma_i - \sigma_{i-1} = h_2$). При этом в качестве середины сетки математических ожиданий бралось выборочное среднее, а границы определялись на основе выборочной дисперсии, умноженной на некоторый коэффициент. Итоговое число компонент сетки выбиралось в зависимости от $n = 100, 1000, 5000$ и $10\,000$; например, при $n = 5000$ оно равнялось 21×11 .

После некоторых экспериментов возникла идея прогонять алгоритм второй раз с использованием данных первого прогона для уточнения сетки. После первого прогона уже доступна некоторая оценка плотности и, естественно, доступна исходная равномерная сетка с весами в некоторых ее узлах (не всех).

Следующим логичным шагом является переход к неравномерной сетке. Понятно, что шаг сетки должен быть меньше там, где наиболее вероятно встретится весомый компонент. Это позволяет увеличить точность оценки математических ожиданий и дисперсий, в то же время оставив общее число узлов сетки на разумном уровне, чтобы не уменьшить точность оценки веса компоненты.

Рассматривались два метода построения сетки на втором шаге. В первом методе за основу берется $\hat{f}_1(x)$ — плотность распределения, построенная при первом шаге алгоритма (результат его работы), и $\hat{g}_1(x)$ — дополнительная искусственная плотность, построенная по сетке дисперсий первого шага. Второй алгоритм использует по отдельности только веса математических ожиданий и

веса дисперсий, полученных в результате первого шага. Рассмотрим детально эти два метода.

I метод. Рассмотрим сначала построение сетки для мат. ожиданий. Пусть $\hat{f}_1(x)$ — оценка плотности распределения, построенная при первом проходе алгоритма, M — число узлов в сетке математических ожиданий. Задача — построить неравномерную сетку с числом узлов M на основе $\hat{f}_1(x)$. Для этого решаются уравнения:

$$\left. \begin{aligned} \int_{-\infty}^{m_1} \hat{f}_1(x) &= \frac{1}{M+1}; \\ \int_{m_{i+1}}^{m_i} \hat{f}_1(x) &= \frac{1}{M+1} \quad (i = 1, \dots, M-1); \\ \int_{m_M}^{+\infty} \hat{f}_1(x) &= \frac{1}{M+1}. \end{aligned} \right\} \quad (5)$$

В результате решения этих уравнений (по порядку) получается сетка (m_1, m_2, \dots, m_M) с числом узлов M . Физический смысл данного подхода следующий: площадь под графиком $\hat{f}_1(x)$ разделяется на $M+1$ равных по площади кусков (самый левый и самый правый куски бесконечны по длине). При этом точки границ этих кусков (абсциссы) используются как сетка (рис. 3). Чем больше значение плотности в заданной области, тем чаще там будут ставиться точки сетки, тем самым точки сетки расставляются недалеко от пиков.

Аналогичная процедура производится для сетки дисперсии с числом узлов N , только для этой стадии надо подготовить вспомогательную функцию

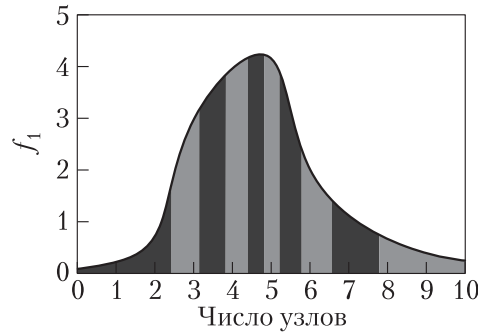


Рис. 3 Построение сетки на втором проходе первым методом

$$\hat{g}_1(x) = \sum_{i=1}^K p_i \phi(\sigma_i, 1),$$

где p_i — вес компоненты из первого прохода, а σ_i — ее дисперсия. Далее необходимо решить уравнения, аналогичные уравнениям (5):

$$\int_0^{n_1} \hat{g}_1(x) = \frac{1}{N+1}; \quad \int_{n_i}^{n_{i+1}} \hat{g}_1(x) = \frac{1}{N+1} \quad (i = 1, \dots, N-1); \quad \int_{n_N}^{+\infty} \hat{g}_1(x) = \frac{1}{N+1}.$$

Одним из плюсов данного метода является то, что его можно запускать даже на первом шаге исходного процесса, используя гистограмму как $\hat{f}_1(x)$ (по крайней мере, для оценки сетки математических ожиданий). Из минусов данного подхода можно отметить нечувствительность к компонентам в случае, если они находятся не на пике (рис. 4). Для устранения этой особенности был предложен второй метод, который обсуждается далее.

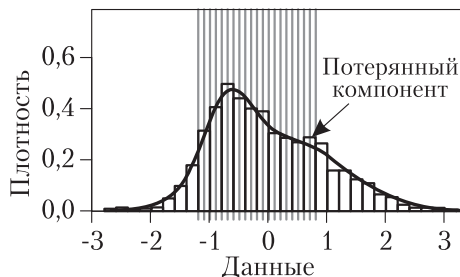


Рис. 4 Особенность первого метода — нечувствительность к компонентам, находящимся не на пиках плотности

II метод. Основная идея этого алгоритма заключается в следующем: если первый проход не разместил весомых компонент на соседних узлах сетки a_i, a_{i+1} , то весь этот интервал исключается из рассмотрения, т.е. никакие точки не

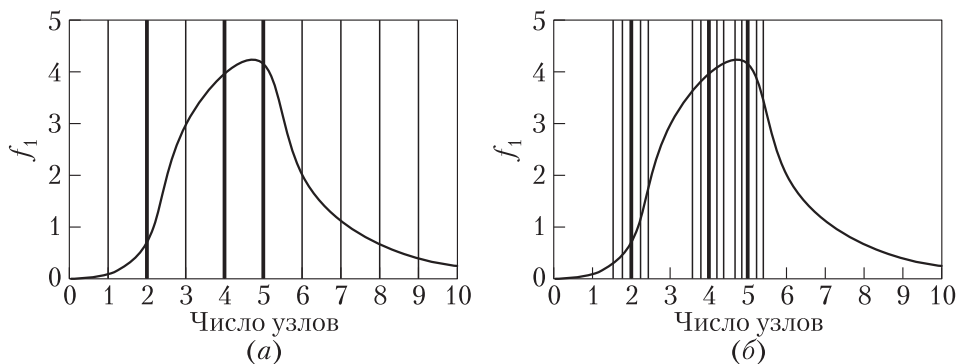


Рис. 5 Построение сетки на втором проходе вторым методом: (а) исходная сетка; (б) полученная сетка

добавляются между a_i, a_{i+1} и сами границы исключаются из сетки. Назовем эту процедуру *чисткой*. В результате чистки удаляются несколько узлов сетки и получается некоторый запас из T вершин для дальнейшего неравномерного размещения.

Вторую процедуру в этом методе назовем *размещением*. Она заключается в том, что если первый шаг разместил весовую компоненту на узле a_k , то в промежутки (a_{k-1}, a_k) и (a_k, a_{k+1}) добавляются узлы, равномерно делящие эти промежутки. Если этот промежуток уже детализирован из-за наличия весовой компоненты на a_{k-1} или a_{k+1} , то детализация не требуется. При этом необходимо следить, чтобы общее число добавленных весов равнялось T .

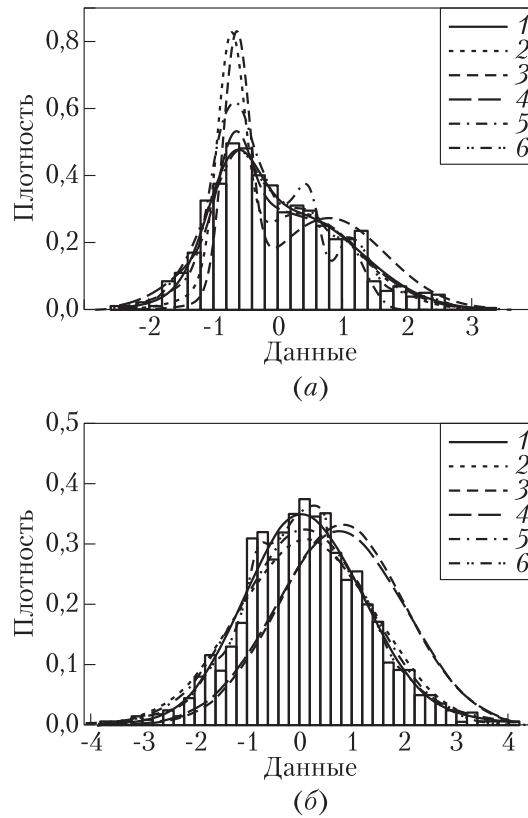


Рис. 6 Пример работы алгоритмов на модельных данных при втором (а) и первом (б) методах (двухпроходных): 1 — реальная плотность; 2 — классический EM-алгоритм; 3 — GCG метод; 4 — GCG метод, II проход, сетка Корчагина; 5 — GCG метод, II проход, сетка Королева; 6 — оценка плотности

Абсолютно аналогичная процедура проводится для сетки дисперсий. Схематичное изображение работы алгоритма можно увидеть на рис. 5, *a*, где изображена исходная сетка, жирными линиями помечены компоненты, на которых в результате первого прохода были размещены веса. На рис. 5, *б* изображена полученная сетка.

Как уже отмечалось, у каждого из алгоритмов есть свои плюсы и минусы. В случае если первый проход алгоритма сработал точно и хорошо оценил примерное нахождение компонент, второй метод оказывается точнее на втором проходе. Это видно на рис. 6, *a*. В случае же если первый проход сработал недостаточно точно, то универсальный в некотором смысле первый метод оказывается лучше, что показано на рис. 6, *б*.

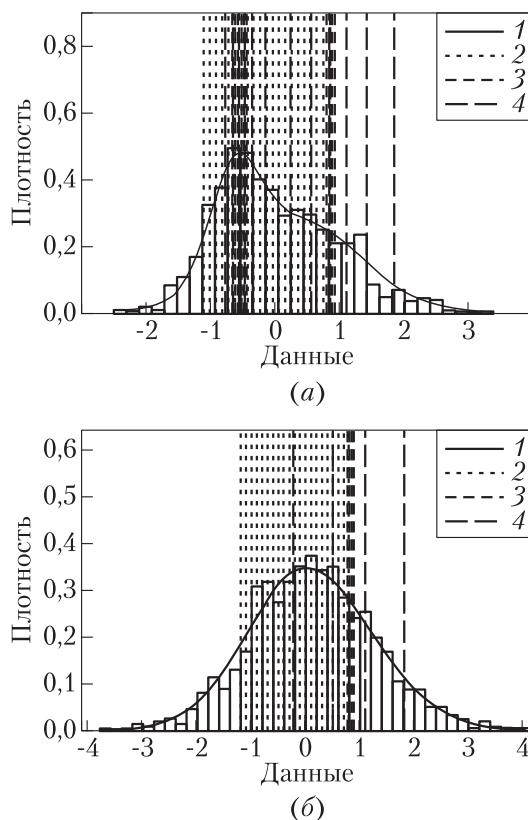


Рис. 7 Пример сеток при работе двухпроходного алгоритма при втором (*a*) и первом (*б*) методах: 1 — реальная плотность; 2 — GCG метод, I проход; 3 — GCG метод, II проход, сетка Корчагина; 4 — GCG метод, II проход, сетка Королева

На рис. 7 представлены примеры сеток при работе двухпроходного алгоритма при первом и втором методах.

5 Ядерные оценки плотности

5.1 Основные определения

Пусть X_1, X_2, \dots, X_n — выборка размера n . Тогда ядерная оценка функции плотности $f(x)$ в точке x определяется следующим образом:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right), \quad (6)$$

где $h > 0$ — это сглаживающий параметр, или *ширина окна*, а K — ядро, некоторая неотрицательная функция, обладающая следующими свойствами:

$$\int_{-\infty}^{+\infty} K(y) dy = 1; \quad \int_{-\infty}^{+\infty} yK(y) dy = 0; \quad \int_{-\infty}^{+\infty} y^2 K(y) dy = \mu_2(K) < \infty.$$

5.2 Выбор ширины окна

Ядерная оценка функции плотности (6), как видно из формулы, зависит от ширины окна h и ядра K . Однако выбор функция ядра K практически не влияет на точность восстановления плотности. Решающим образом на качество восстановления плотности влияет ширина окна h . При слишком широком окне функция плотности чрезмерно сглаживается, при слишком узком — претерпевает слишком частые колебания. Часто h выбирают как значение, минимизирующее некоторый функционал ошибки.

5.2.1 Эвристические правила

Интегральная среднеквадратичная ошибка (MISE) агрегирует среднеквадратичные ошибки по всей области определения функции плотности, являясь глобальной мерой ошибки, и при подстановке выражений для асимптотического смещения и дисперсии принимает вид:

$$\begin{aligned} \text{MISE} \{ \hat{f}_h \} &= E \left\{ \int \left(\hat{f}_h(y) - f(y) \right)^2 dy \right\} = \\ &= \int \text{Bias} \left(\hat{f}_h(y) \right)^2 dy + \int \text{Var} \left(\hat{f}_h(y) \right) dy, \quad (7) \end{aligned}$$

где

$$\begin{aligned} \text{Bias} \left\{ \hat{f}_h(x) \right\} &= \frac{h^2}{2} \mu_2(K) f''(x) + o(h^2); \\ \text{Var} \left\{ \hat{f}_h(x) \right\} &= \frac{1}{nh} R(K) f(x) + o\left(\frac{1}{nh}\right); \\ R(f'') &= \int_{-\infty}^{+\infty} [f''(y)]^2 dy. \end{aligned}$$

Устремим n к бесконечности в (7) и получим асимптотическую интегральную среднеквадратичную ошибку:

$$\text{AMISE} \left\{ \hat{f}_h \right\} = \frac{1}{nh} R(K) + \frac{h^4}{4} \mu_2(K)^2 R(f''), \quad (8)$$

где

$$R(f'') = \int_{-\infty}^{+\infty} [f''(y)]^2 dy.$$

Оптимальная ширина окна в данном случае зависит от функции плотности f , которая неизвестна, поэтому неизвестную величину $R(f'')$ оценивают эвристическими методами и уже потом минимизируют $\text{AMISE} \left\{ \hat{f}_h \right\}$. Если в качестве ядра взять $N(0, 1)$ и предположить, что исходная выборка имеет нормальное распределение, то $R(f'') = 3/(8\sqrt{\pi}\sigma^5)$ [16], где σ — стандартное отклонение выборки. Исходя из этого и некоторых эвристических соображений, выражение для h можно записать как

$$h = \frac{0,9}{n^{1/5}} \min \left\{ \sigma, \frac{\text{IQR}}{1,34} \right\},$$

где σ — стандартное отклонение выборки, а IQR — интерквартильный размах. Недостатком этого метода является чересчур сглаженная оценка функции распределения.

5.2.2 Кросс-валидация на основе наименьших квадратов

Кросс-валидация на основе наименьших квадратов (least squares cross-validation) — это полностью автоматический и диктуемый данными метод выбора сглаживающего параметра. Этот метод основан на принципе выбора ширины ок-

на, минимизирующей интегральную среднеквадратичную ошибку получающейся оценки. Интеграл квадрата разности $\hat{f}(x)$ и $f(x)$ имеет вид:

$$\int \left[\hat{f}_h(x) - f(x) \right]^2 dx = \int \hat{f}_h(x)^2 - 2 \int \hat{f}_h(x) f(x) dx + \int f(x)^2 dx.$$

Легко заметить, что последнее слагаемое не зависит от h . Второе слагаемое зависит от неизвестной плотности $f(x)$. В работе [21] показано, что при больших n величину $\int \hat{f}_h(x) f(x) dx$ можно оценить как $(1/n) \sum_{i=1}^n \int (\hat{f}_{-i}(y))^2 dy$, где $\hat{f}_{-i}(y)$ — ядерная оценка, построенная по исходной выборке без наблюдения X_i . Таким образом, нужно минимизировать функцию:

$$\text{LSCV} \left\{ \hat{f}_h \right\} = \int_{-\infty}^{+\infty} \left(\hat{f}_h(y) \right)^2 dy - \frac{2}{n} \sum_{i=1}^n \hat{f}_{-i}(X_i).$$

Одна из проблем данного способа выбора ширины окна — его чувствительность к наличию округленных или дискретизированных данных, а также к мелкомасштабным эффектам в данных.

5.2.3 Смещенная кросс-валидация

Смещенная кросс-валидация (biased cross-validation) основана на замене $R(f'')$ в (8) выражением $R(\hat{f}_h'') - (1/(nh^5)) R(K'')$, где \hat{f}_h'' — вторая производная ядерной оценки плотности. Такая замена обусловлена тем, что $R(\hat{f}_h'')$ является смещенной оценкой $R(f'')$ [12]. В этом случае минимизируемый функционал будет выглядеть так:

$$\text{BCV} \left\{ \hat{f}_h \right\} = \frac{1}{nh} R(K) + \frac{h^4}{4} \mu_2(K)^2 \left[R(\hat{f}_h'') - \frac{1}{nh^5} R(K'') \right],$$

где

$$R(K'') = \int_{-\infty}^{+\infty} [K''(y)]^2 dy.$$

6 Результаты применения алгоритмов

Для тестов использовались выборки размером 100, 300, 1000, 5000 и 10 000. Они генерировались несколько раз для устранения случайности. Максимальное

число компонент смеси в классическом EM-алгоритме — 15. Расстояния от полученной оценки функции плотности до истинной считались в метриках C , L_1 и L_2 :

$$C(\hat{f}_h(x), f(x)) = \max_{x \in X} |f(x) - \hat{f}_h(x)|;$$

$$L_1(\hat{f}_h(x), f(x)) = \int_{x \in X} |f(x) - \hat{f}_h(x)| dx;$$

$$L_2(\hat{f}_h(x), f(x)) = \left(\int_{x \in X} |f(x) - \hat{f}_h(x)|^2 \right)^{1/2}.$$

6.1 Сравнение EM-алгоритма и ядерных оценок

Ниже приведены примеры использования EM-алгоритма и ядерных оценок плотности с различными h . В качестве данных для исследования были взяты модельные данные, функции плотности которых имеют вид:

$$p_1 * U[0, 3] + p_2 * U[0, 5] + p_3 * U[2, 4];$$

$$p_1 * U[0, 1] + p_2 * U[2, 4] + p_3 * U[5, 8];$$

$$p_1 * \text{Exp}(0.1) + p_2 * \text{Exp}(p_1) + p_3 * \text{Exp}(p_3);$$

$$p_1 * \text{Exp}(7) + p_2 * \text{Exp}(2) + p_3 * \text{Exp}(5);$$

$$p_1 * \text{Exp}(1) + p_2 * U[0, 10] + p_3 * \text{Laplace}(-2, 3);$$

$$p_1 * \text{Exp}(1) + p_2 * U[-2, 10] + p_3 * \text{Laplace}(2, 3);$$

$$p_1 * \text{Exp}(1) + p_2 * U[-10, 10] + p_3 * \text{Laplace}(-10, 1);$$

$$p_1 * \text{Laplace}(-2, p_3) + p_2 * \text{Laplace}(0, 1) + p_3 * \text{Laplace}(3, 3);$$

$$p_1 * \text{Laplace}(-5, p_3) + p_2 * \text{Laplace}(0, 1) + p_3 * \text{Laplace}(7, 5);$$

$$p_1 * \text{Norm}(1, 0.5) + p_2 * \text{Norm}(2, 2) + p_3 * \text{Norm}(3, 0.1);$$

$$p_1 * \text{Norm}(0, 0.5) + p_2 * \text{Norm}(0, 2) + p_3 * \text{Norm}(0, 0.1);$$

$$p_1 * \text{Norm}(0, 0.5) + p_2 * \text{Norm}(0, 2) + p_3 * \text{Norm}(3, 4);$$

$$p_1 * \text{Norm}(-2, 0.5) + p_2 * \text{Norm}(0, 0.5) + p_3 * \text{Norm}(3, 0.5);$$

где $U[x_1, x_2]$ — равномерное распределение на отрезке $[x_1, x_2]$, $\text{Laplace}(a, b)$ — распределение Лапласа с параметрами ab , $\text{Exp}(a)$ — экспоненциальное распределение с параметром a , $\text{Norm}(a, \sigma)$ — нормальное распределение с параметрами a, σ , веса $p = (p_1, p_2, p_3)$ принимали значения

$$(1/3, 1/3, 1/3);$$

$$(0.2, 0.3, 0.5);$$

$$(0.7, 0.2, 0.1);$$

$$(0.15, 0.7, 0.15).$$

На рис. 8–10 представлены результаты применения EM-алгоритма и ядерных методов для оценки функции плотности вероятности: диаграммы и графики.

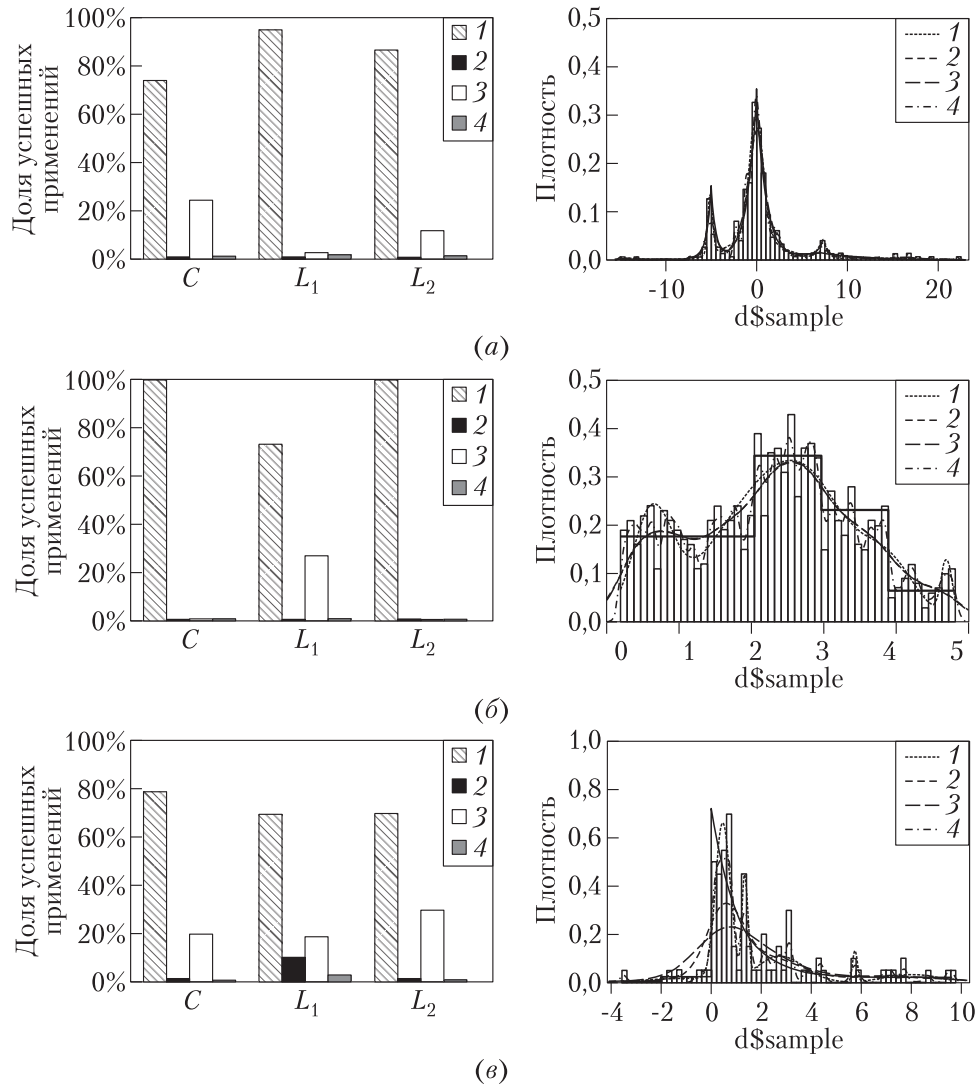


Рис. 8 Результаты применения EM-алгоритма (1) и ядерных методов (2 — NRD; 3 — UCV; 4 — BCV) для оценки функции плотности вероятности: (а) $0,15 * \text{Laplace}(-5, 0,5) + 0,7 * \text{Laplace}(0, 1) + 0,15 * \text{Laplace}(7, 5)$, Size = 300; (б) $(1/3) * U[0, 3] + (1/3) * U[0, 5] + (1/3) * U[2, 4]$, Size = 1000; (в) $0,7 * \text{Exp}(1) + 0,2 * \text{Uniform}[-2, 10] + 0,1 * \text{Laplace}(2, 3)$, Size = 100

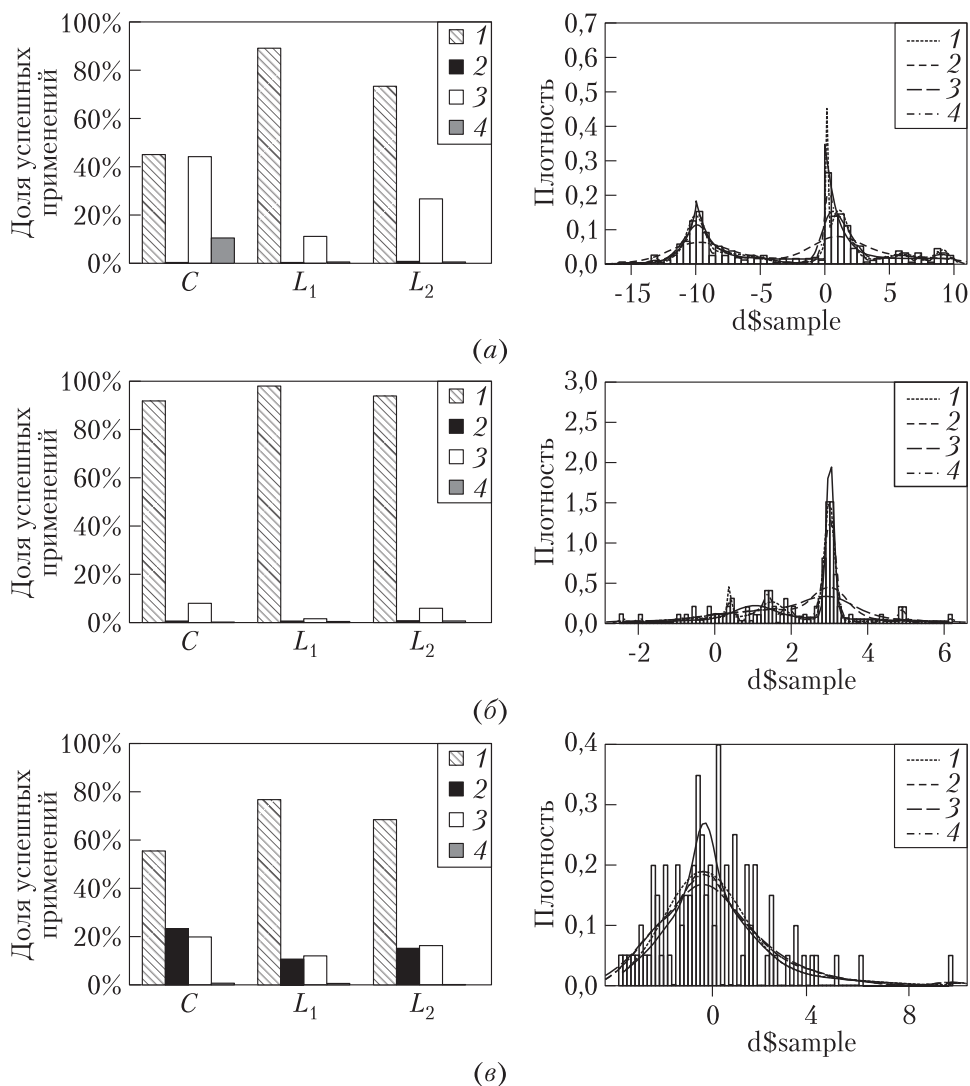


Рис. 9 Результаты применения EM-алгоритма (1) и ядерных методов (2 — NRD; 3 — UCV; 4 — BCV) для оценки функции плотности вероятности: (а) $(1/3) * \text{Exp}(1) + (1/3) * \text{Uniform}[-10, 10] + (1/3) * \text{Laplace}(-10, 1)$, Size = 300; (б) $0,2 * \text{Norm}(1, 0,5) + 0,3 * \text{Norm}(2, 2) + 0,5 * \text{Norm}(3, 0,1)$, Size = 100; (в) $0,15 * \text{Norm}(0, 0,5) + 0,7 * \text{Norm}(0, 2) + 0,15 * \text{Norm}(3, 4)$, Size = 100

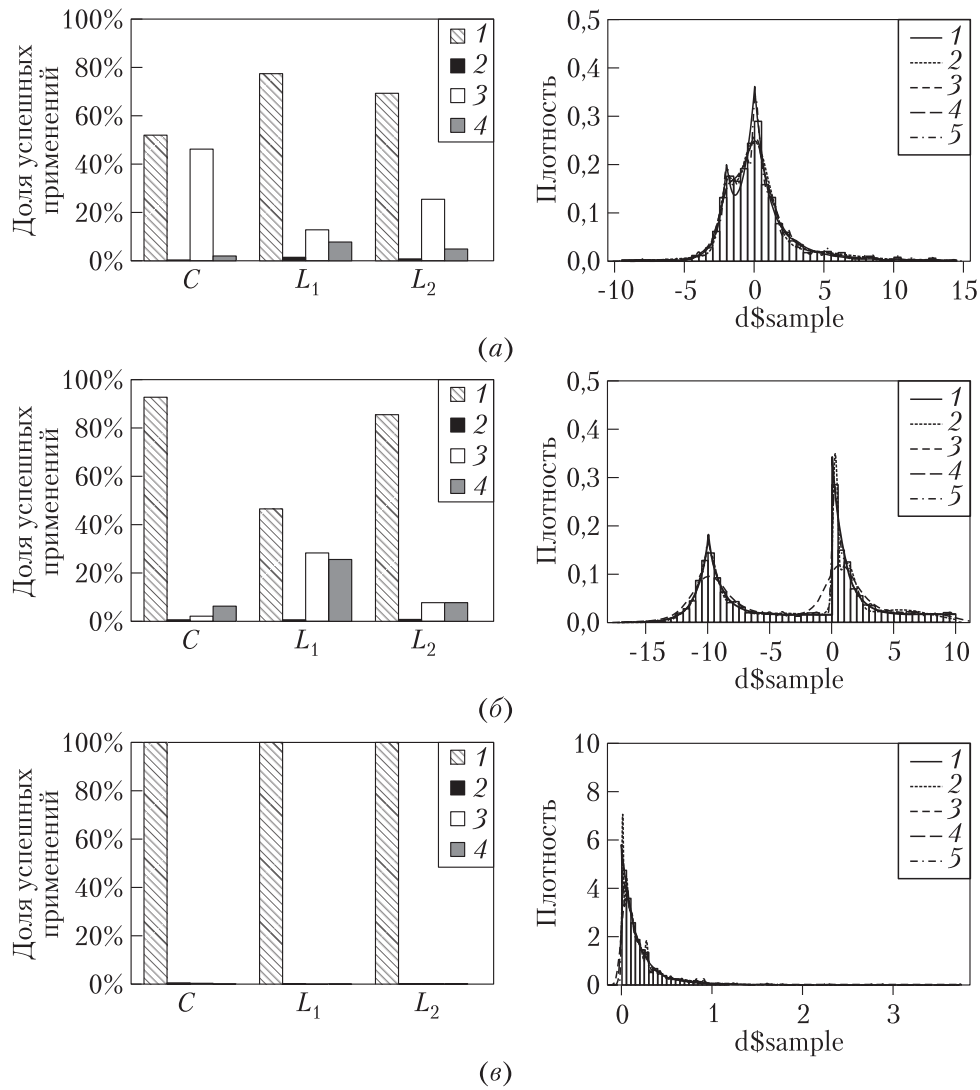


Рис. 10 Результаты применения EM-алгоритма (1) и ядерных методов (2 — NRD; 3 — UCV; 4 — BCV) для оценки функции плотности вероятности: (а) $0,15 * \text{Laplace}(-2, 0,5) + 0,7 * \text{Laplace}(0, 1) + 0,15 * \text{Laplace}(3, 3)$, Size = 1000; (б) $(1/3) * \text{Exp}(1) + (1/3) * \text{Uniform}(-10, 10) + (1/3) * \text{Laplace}(-10, 1)$, Size = 5000; (в) $0,7 * \text{Exp}(7) + 0,2 * \text{Exp}(2) + 0,1 * \text{Exp}(5)$, Size = 1000

На диаграммах по оси y отложена доля «успешных» применений каждого алгоритма. Под «успешностью» подразумевается наибольшая близость оценки функции плотности (NRD соответствует эвристическим правилам, UCV — кросс-валидации на основе наименьших квадратов, BCV — смещенной кросс-валидации) к истинной функции плотности в какой-либо метрике. На графиках изображены истинная функция плотности и ее полученные оценки.

6.2 Сравнение сеточного аналога EM-алгоритма и ядерных оценок

Ниже представлена общая сводная таблица победителей на тех или иных распределениях (табл. 1). В этих экспериментах менялся критерий выбора ширины окна и критерий оценки компонент смеси.

Как видно из табл. 1, смеси нормальных законов лучше оценивает сеточный метод. Если есть уверенность в правильной спецификации модели (т. е. что данные — это смесь нормальных законов), нужно использовать EM-алгоритмы. Когда важна скорость выполнения алгоритма (например, при разделении прямых потоков данных финансовых показателей) — стоит использовать GCG, точность которого была повышена введением второго прохода и уточнением сетки. Если уверенности в модели нет, нужно использовать непараметрический подход. Ядерные методы приходят на помощь исследователям, когда конкретную модель подобрать пока не удастся, а гистограммы уже недостаточно.

В табл. 2 приведены расширенные таблицы результатов в метриках L_1 и C .

На рис. 11–16 приведены гистограммы выигрышей алгоритмов. Серым цветом отмечены выигрыши сеточного метода, черным — ядерных оценок. Чем больше в столбце серого цвета, тем лучше при данных параметрах работает сеточный метод. В левой части расположены результаты в метрике L_1 , а в правой — в метрике C . Каждый столбец соответствует одному фиксированному размеру выборки.

Таблица 1 Сводная таблица результатов

Распределение	Лучший метод
$N[-0,3; 1,2]$	Сеточный метод
$0,3N[-0,7; 0,4] + 0,7N[0,3; 1]$	Сеточный метод
$0,4N[-1,2; 0,2] + 0,6N[0,8; 0,4]$	Сеточный метод
$0,45N[-0,2; 1] + 0,55N[0,4; 1,2]$	Сеточный метод
$0,3R[-0,6; 0,6] + 0,7R[0,2; 1,2]$	Ядерное оценивание
$R[-0,6; 0,6]$	Ядерное оценивание
$G[3; 2]$	Почти ничья, сеточный лучше
$\exp[2], \exp[5]$	Сеточный метод
$0,5N[-0,6; 1,2] + 0,5R[0,3; 0,7]$	Ядерное оценивание
$0,6N[1,3; 1,2] + 0,4G[2; 3]$	Ядерное оценивание

Таблица 2 Расширенная таблица результатов (метрика L_1 и C)

Распределение	Выборка	GCG		Ядерный		Всего
		L_1	C	L_1	C	
$N[-0,3; 1,2]$	100	48	50	4	2	52
	1000	49	52	3	0	52
	5000	52	52	0	0	52
	10000	52	52	0	0	52
$0,3N[-0,7; 0,4] + 0,7N[0,3; 1]$	100	29	24	23	28	52
	1000	49	50	3	2	52
	5000	52	52	0	0	52
	10000	51	51	1	1	52
$0,4N[-1, 2; 0,2] + 0,6N[0,8; 0,4]$	100	47	44	5	8	52
	1000	49	50	3	2	52
	5000	52	52	0	0	52
	10000	52	52	0	0	52
$0,45N[-0,2; 1] + 0,55N[0,4; 1, 2]$	100	47	47	5	5	52
	1000	50	50	2	2	52
	5000	50	51	2	1	52
	10000	50	51	2	1	52
$0,3R[-0,6; 0,6] + 0,7R[0,2; 1,2]$	100	10	89	90	11	100
	1000	0	67	100	33	100
	5000	0	28	100	72	100
	10000	0	14	100	86	100
$R[-0,6; 0,6]$	100	2	47	50	5	52
	1000	0	31	52	21	52
	5000	0	7	52	45	52
	10000	0	4	52	48	52
$R[-1,2; 0,7]$	100	16	4	84	96	100
	1000	1	0	99	100	100
	5000	0	0	100	100	100
	10000	0	0	100	100	100
$G[3; 2]$	100	18	22	34	30	52
	1000	33	24	19	28	52
	5000	25	18	27	34	52
	10000	22	18	30	34	52
$\exp[2]$	100	23	14	29	38	52
	1000	41	20	11	32	52
	5000	43	37	9	15	52
	10000	42	42	10	10	52
$\exp[5]$	100	21	9	31	43	52
	1000	42	18	10	34	52
	5000	35	34	17	18	52
	10000	28	37	24	15	52

Окончание табл. 2 на с. 221

Таблица 2 (окончание) Расширенная таблица результатов (метрика L_1 и C)

Распределение	Выборка	GCG		Ядерный		Всего
		L_1	C	L_1	C	
$0,5N[-0,6; 1,2] + 0,5R[0,3; 0,7]$	100	97	10	3	90	100
	1000	40	1	60	99	100
	5000	17	0	83	100	100
	10000	18	0	82	100	100
$0,6N[1,3; 1,2] + 0,4G[2; 3]$	100	37	30	15	22	52
	1000	34	29	18	23	52
	5000	20	23	32	29	52
	10000	15	19	37	33	52

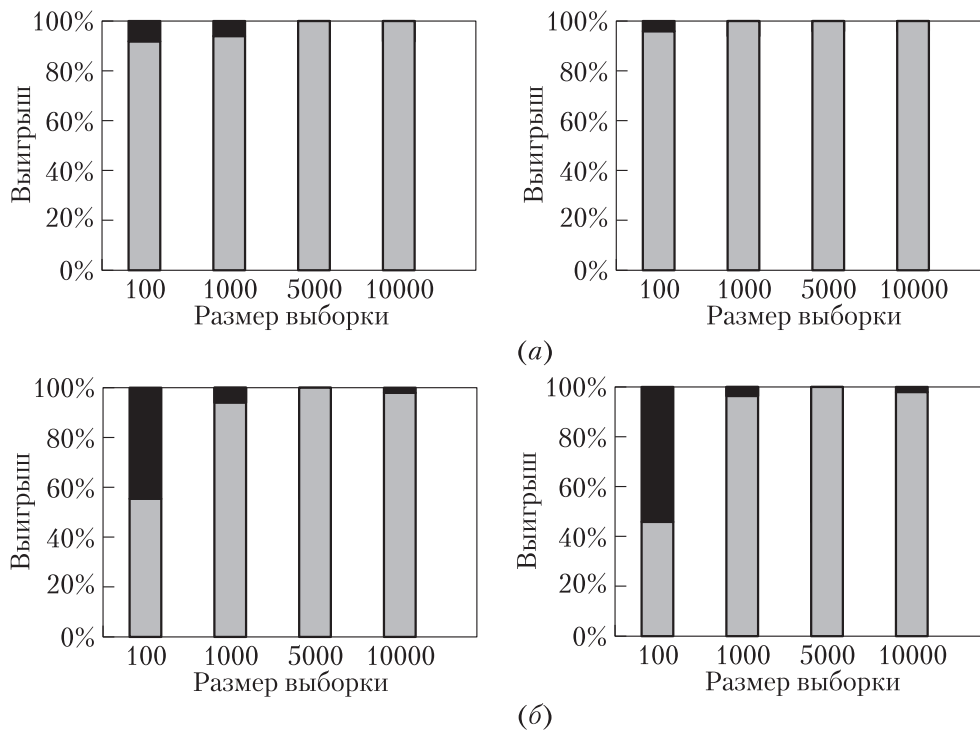


Рис. 11 Гистограммы выигрышей алгоритмов: (а) $N[-0,3; 1,2]$; (б) $0,3N[-0,7; 0,4] + 0,7N[0,3; 1]$;

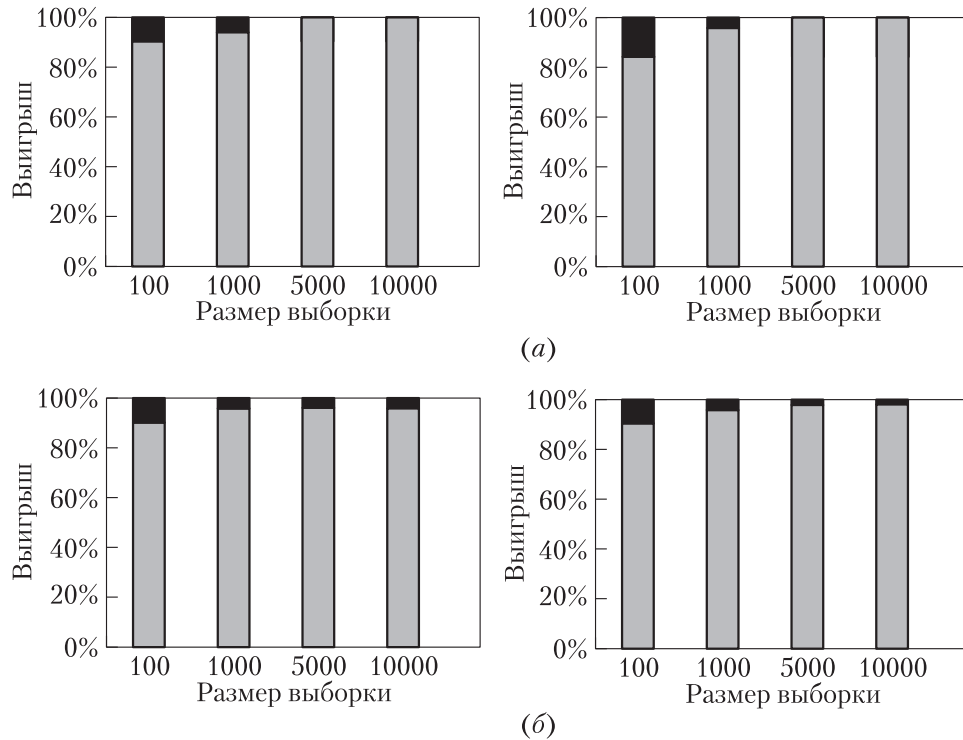


Рис. 12 Гистограммы выигрышей алгоритмов: (а) $0,4N[-1,2; 0,2] + 0,6N[0,8; 0,4]$; (б) $0,45N[-0,2; 1] + 0,55N[0,4; 1,2]$

7 Выводы

В работе было рассмотрено решение задачи оценивания плотности вероятности смесей вероятностных распределений с помощью ядерных оценок, стандартного EM-алгоритма и сеточных модификаций EM-алгоритма.

В подавляющем большинстве (80% всех выборок) случаев EM сработал лучше ядерных методов в метриках C , L_1 и L_2 . Однако, по сравнению с алгоритмом ядерного оценивания, EM-алгоритм работает гораздо дольше, поэтому имеет смысл использовать ядерные оценки, когда критична скорость работы. Помимо близости оценки функции плотности к истинной, преимуществом EM-алгоритма является еще и то, что для хранения оценки функции плотности, построенной по выборке, не требуется хранить всю выборку (как в ядерных методах), а только вектор параметров размерности $3k - 1$. Следовательно, если важно качество оценки функции плотности вероятности и стоит задача сократить объем

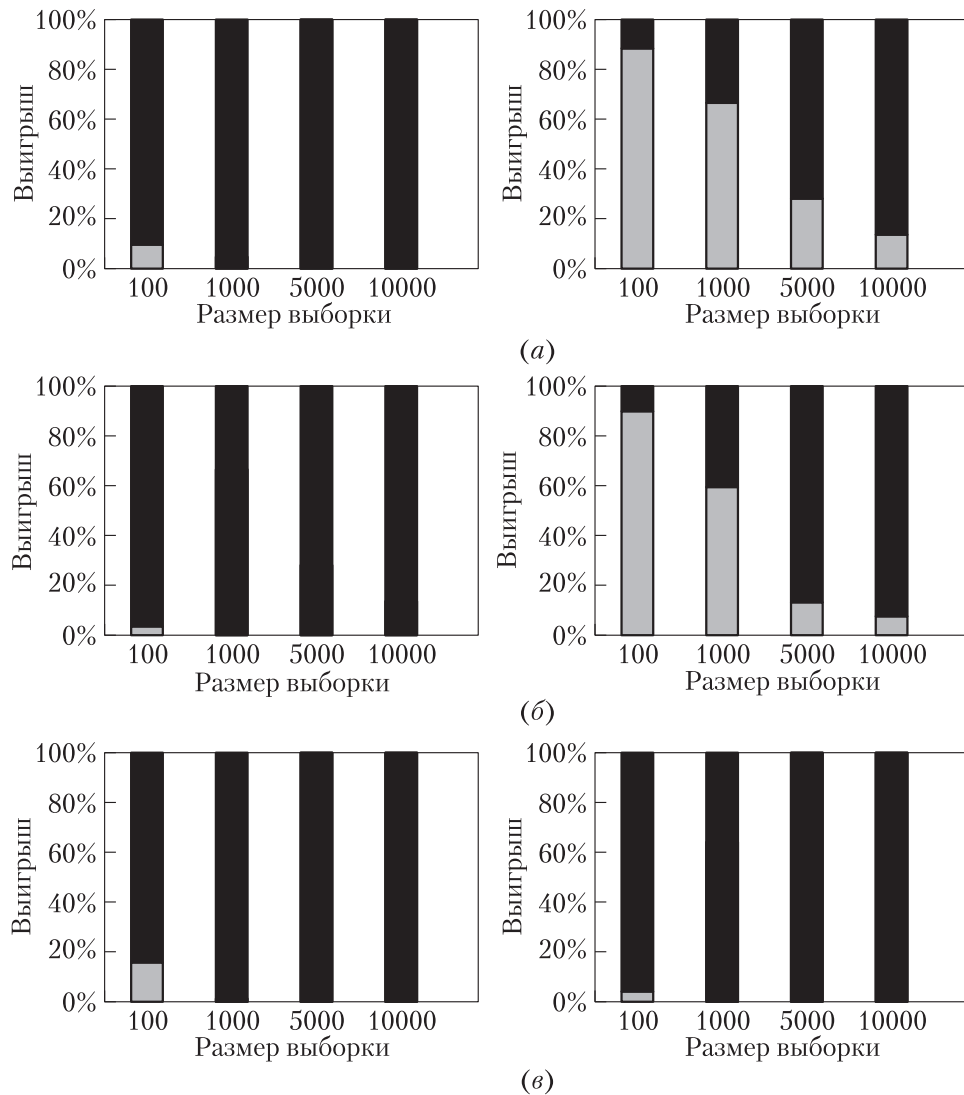


Рис. 13 Гистограммы выигрышей алгоритмов: (а) $0,3R[-0,6; 0,6] + 0,7R[0,2; 1,2]$; (б) $R[-0,6; 0,6]$; (e) $R[-1,2; 0,7]$

хранимых данных, то целесообразно использовать непараметрические оценки, полученные с помощью EM-алгоритма.

Помимо этого, в результате экспериментов на искусственно сгенерированных данных было показано, что для разделения смесей нормальных распределений

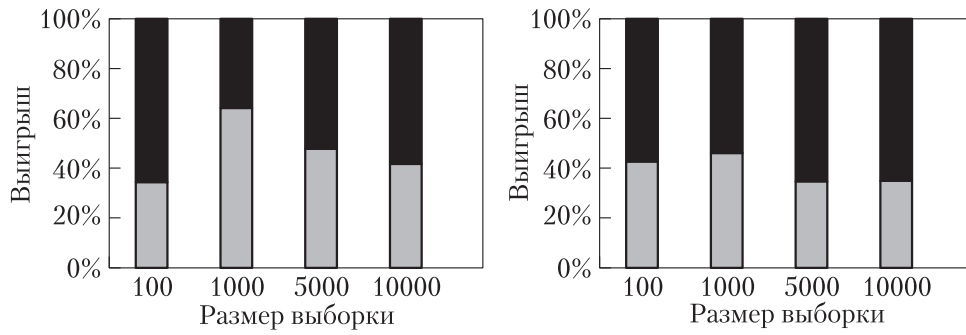


Рис. 14 Гистограммы выборки для алгоритма $G[3; 2]$

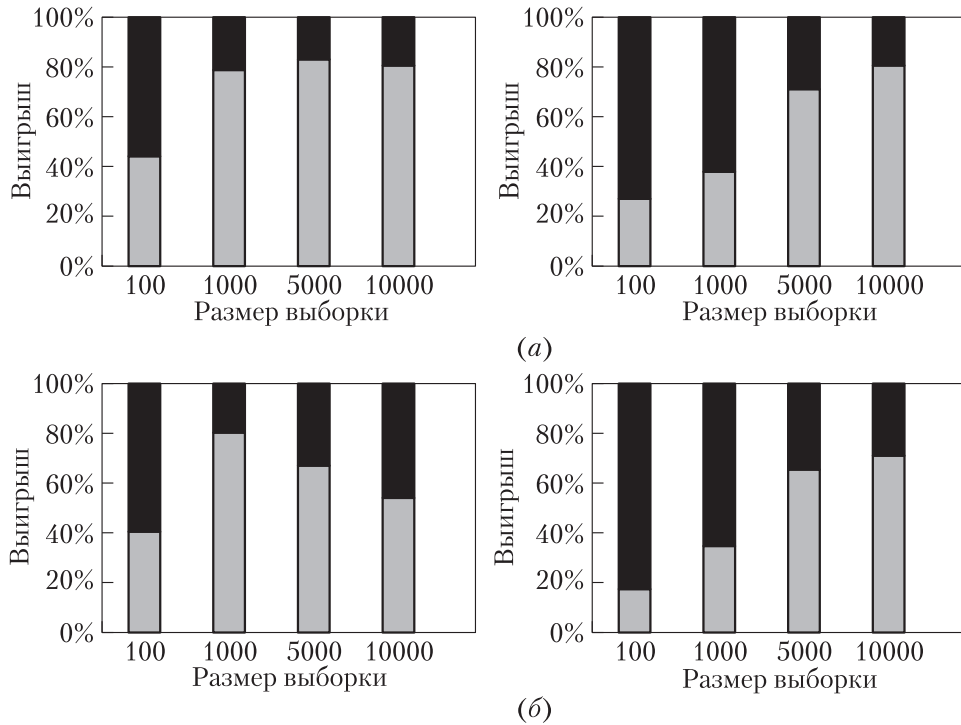


Рис. 15 Гистограммы выигрышей алгоритмов: (а) $\exp[2]$; (б) $\exp[5]$

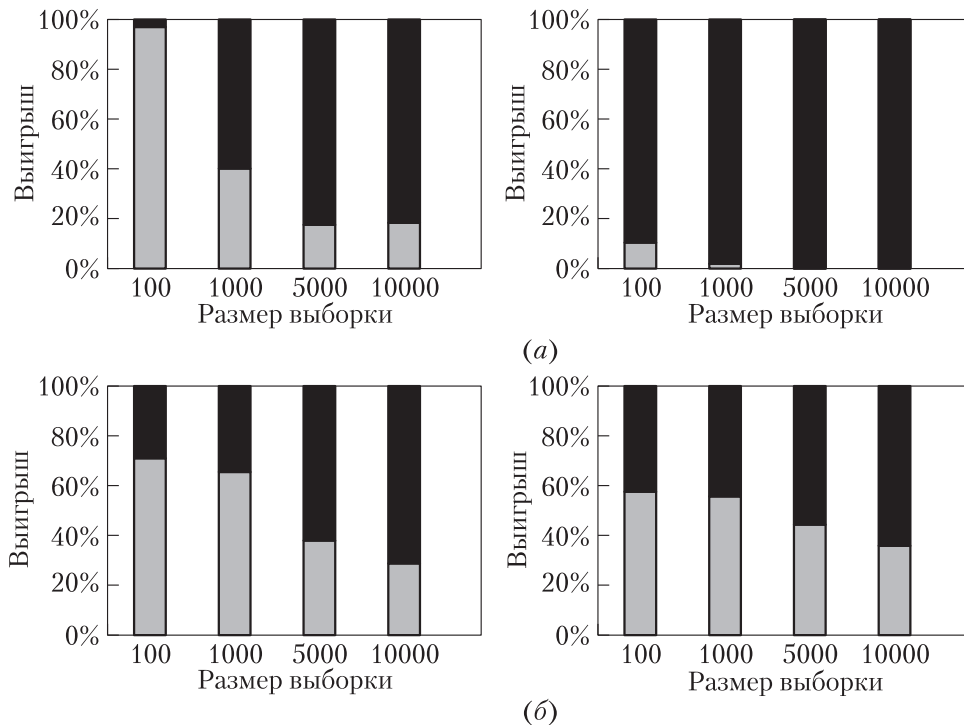


Рис. 16 Гистограммы выигрышей алгоритмов: (а) $0,5N[-0,6; 1,2] + 0,5R[0,3; 0,7]$; (б) $0,6N[1,3; 1,2] + 0,4G[2; 3]$

и некоторых других видов распределений сеточные алгоритмы работают лучше ядерных оценок. Авторами была произведена доработка сеточного алгоритма путем введения второго прохода для построения неравномерной сетки. Предложенный второй проход позволяет значительно повысить точность приближения к реальным данным, при этом время работы алгоритма увеличивается не более чем в два раза.

Литература

1. *Newcomb S.* A generalized theory of the combination of observations so as to obtain the best result // Amer. J. Math., 1886. Vol. 8. P. 343–366.
2. *McKendrick A. G.* Applications of mathematics to medical problems // Proceedings Edinburgh Math. Soc., 1926. Vol. 44. P. 98–130.
3. *Healy M. J. R., Westmacott M. H.* Missing values in experiments analyzed on automatic computers // Appl. Stat., 1956. Vol. 5. P. 203–206.

4. *Hartley H. O.* Maximum likelihood estimation from incomplete data // *Biometrics*, 1958. Vol. 14. P. 174–194.
5. *Buck S. F.* A method of estimation of missing values in multivariate data suitable for use with an electronic computer // *J. R. Stat. Soc., Ser. B*, 1960. Vol. 22. P. 302–306.
6. *Dempster A., Laird N., Rubin D.* Maximum likelihood estimation from incomplete data // *J. R. Stat. Soc., Ser. B*, 1977. Vol. 39. P. 1–38.
7. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011.
8. *Rosenblatt M.* Remarks on some nonparametric estimates of a density function // *Ann. Math. Stat.*, 1956. Vol. 27. P. 832–837.
9. *Надарая Э. А.* Некоторые новые оценки функций распределения // *Теория вероятн. и ее примен.*, 1964. Т. 9. Вып. 3. С. 550–554.
10. *Надарая Э. А.* О непараметрических оценках плотности вероятности и регрессии // *Теория вероятн. и ее примен.*, 1965. Т. 10. Вып. 1. С. 199–203.
11. *Silverman B. W.* Density estimation for statistics and data analysis. — New York: Chapman & Hall, 1986.
12. *Scott D. W., Terrell G. R.* Biased and unbiased cross-validation in density estimation // *J. Amer. Stat. Association*, 1987. Vol. 82. P. 1131–1146.
13. *Prakasa Rao B. L. S.* Identifiability in stochastic models. — New York–London: Academic Press, 1992.
14. *Wand M. P., Jones M. C.* Kernel smoothing. — London: Chapman and Hall, 1995.
15. *Simonoff J. S.* Smoothing methods in statistics. — New York: Springer, 1996.
16. *Расин Дж.* Непараметрическая эконометрика: вводный курс // *Квантиль*, 2008. Вып. 4. С. 7–56.
17. *Королев В. Ю.* Вероятностно-статистический анализ хаотических процессов с помощью смешанных гауссовских моделей. Декомпозиция волатильности финансовых индексов и турбулентной плазмы. — М.: ИПИ РАН, 2007.
18. *Королев В. Ю., Шоргин С. Я.* Математические методы анализа стохастической структуры информационных потоков. — М.: ИПИ РАН, 2011.
19. *Назаров А. Л.* Разделение смесей вероятностных распределений сеточным методом максимального правдоподобия при помощи алгоритма условного градиента // *Сб. статей молодых ученых факультета ВМиК МГУ*, 2009. Вып. 6. С. 128–135.
20. *Королев В. Ю., Назаров А. Л.* Разделение смесей вероятностных распределений при помощи сеточных методов моментов и максимального правдоподобия // *Автоматика и телемеханика*, 2010. Вып. 3. С. 98–116.
21. *Stone M.* An asymptotically optimal window selection rule for kernel density estimates // *Ann. Stat.*, 1984. Vol. 12. P. 1285–1297.

О СОСТОЯТЕЛЬНОСТИ ОЦЕНОК ПАРАМЕТРОВ МАСШТАБНЫХ СМЕСЕЙ НОРМАЛЬНЫХ РАСПРЕДЕЛЕНИЙ, ПОЛУЧАЕМЫХ С ПОМОЩЬЮ СЕТОЧНЫХ МЕТОДОВ*

*А. Л. Назаров*¹

Аннотация: В задаче разделения масштабных смесей нормальных законов исследуются асимптотические свойства оценок смешивающего распределения, полученных с помощью сеточного метода максимального правдоподобия. Выписаны достаточные условия состоятельности оценок, описывающие алгоритм согласованного изменения структуры параметрического множества (шага сетки) и числа элементов выборки.

Ключевые слова: смесь нормальных распределений; сеточный метод разделения смесей; состоятельность; расстояние Кульбака–Лейблера; пространство Скорохода

1 Введение

Смеси нормальных законов хорошо зарекомендовали себя в качестве «тяжелохвостых» моделей статистических закономерностей в различных прикладных задачах, в частности в физике при анализе данных, полученных в экспериментах с турбулентной плазмой, в финансовом анализе при анализе эволюции цен разнообразных финансовых активов и индексов, а также в практической информатике при анализе данных о трафике в информационных и телекоммуникационных системах. Задача оценки смешивающего распределения по наблюдаемым реализациям смешанных данных часто решается с помощью EM (expectation-maximization) алгоритма и его модификаций [1]. При использовании этого алгоритма семейство допустимых смешивающих распределений ограничивается классом распределений, сосредоточенных в фиксированном числе точек. Однако классический EM-алгоритм обладает некоторыми существенными недостатками. Например, при старте итерационного процесса поиска оценок из различных начальных точек на одной и той же выборке наблюдений могут быть получены заметно отличающиеся результаты. Также необходимо явно задавать число точек, в которых сосредоточенно смешивающее распределение.

*Работа поддержана Российским фондом фундаментальных исследований (проекты 11-01-00515а, 11-07-00112а, 12-07-00115-а, 11-01-12026-офи-м), Министерством образования и науки (госконтракт 16.740.11.0133).

¹Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, nazarov.vmik@gmail.com

Для преодоления описанных выше трудностей были предложены сеточные методы разделения смесей вероятностных распределений. Они достаточно подробно рассмотрены в работах [1–4]. В данных методах используется предположение о том, что оценку смешивающего распределения можно искать в классе распределений, сосредоточенных в конечном числе фиксированных точек. При этом точки расположены в области, содержащей носитель оцениваемого распределения. Краткое описание этих алгоритмов приведено ниже.

С возможностью использования сеточных методов на практике связаны проблемы стохастической устойчивости нормальных смесей. Исследованию устойчивости различных классов смесей вероятностных распределений к возмущениям смешивающих распределений был посвящен ряд работ. Результаты для частных случаев были рассмотрены, например, в работах [5, 6]. В [7] выписаны оценки близости смесей так называемых сдвиг/масштабных нормальных законов в метрике Леви через близость смешивающих распределений в метрике Леви–Прохорова.

Указанные результаты позволили обосновать замену исходного распределения приближенным при использовании сеточных методов. Благодаря им были получены первые результаты исследования свойств оценок, полученных с помощью сеточных методов [8]. В данной работе представлены асимптотические свойства оценок смешивающего распределения, полученных с помощью сеточного метода максимального правдоподобия для масштабных смесей нормальных законов. Выписаны достаточные условия состоятельности оценок, описывающие алгоритм согласованного изменения структуры параметрического множества и числа элементов выборки. Изложенные ниже факты в основном опираются на результаты монографии [9] и источников, приведенных в ней (см. также [10]).

2 Сеточный метод максимального правдоподобия для разделения смесей вероятностных распределений

Идея, которая лежит в основе сеточных методов разделения смесей, рассматриваемых в данной работе, близка к идее гармонического анализа. Периодическая функция раскладывается в ряд Фурье, т. е. представляется взвешенной комбинацией (рядом), возможно, бесконечно большого числа синусов и косинусов с различными (но кратными) периодами (т. е. с различными параметрами масштаба — частотами). Аналогично при использовании данных алгоритмов оценка смешивающего распределения ищется в классе распределений, сосредоточенных в конечном множестве фиксированных точек. Причем эти точки равномерно покрывают область, содержащую носитель истинного распределения.

Подробнее работу алгоритма рассмотрим на следующем примере. Пусть случайная величина Y может быть представлена в виде

$$Y = VX, \quad X \sim \mathcal{N}(0, 1), \quad V > 0, \quad P(Y < x) = E\Phi\left(\frac{x}{V}\right), \quad x \in \mathbb{R},$$

где Φ — функция распределения стандартного нормального закона, а случайные величины X и V стохастически независимы. Очевидно, Y является масштабной смесью нормальных законов. Задача разделения смесей в данном случае заключается в оценке распределения величины V по реализациям Y .

Исходное смешивающее распределение приближается распределением из параметризованного семейства \mathcal{P}_Ξ , сосредоточенного на сетке с известными узлами. Утверждается, что для плотности смеси Y

$$p_0(x) = E\frac{1}{V}\phi\left(\frac{x}{V}\right) \approx \sum_{i=1}^K p_i \frac{1}{\sigma_i} \phi\left(\frac{x}{\sigma_i}\right) \equiv p_\xi(x), \quad x \in \mathbb{R}, \quad \xi \in \Xi,$$

$$\Xi = \{(\sigma_j, p_j) \mid \sigma_j \in Z, p_j \geq 0, j = 1, \dots, K, p_1 + \dots + p_K = 1, K = \#Z\},$$

где p_0 — плотность истинного распределения, ϕ — плотность стандартного нормального распределения. При этом Z фиксировано и при работе алгоритма оценивается лишь набор весов p_1, \dots, p_K . Множество Z можно выбрать таким образом, чтобы оно равномерно покрывало область, содержащую носитель смешивающего распределения. Поэтому Z иногда называют сеткой, а рассматриваемое семейство алгоритмов — сеточными методами. В случае, если распределение V сосредоточено в k точках, $K \gg k$.

Для масштабных смесей нормальных законов существуют верхние оценки устойчивости к возмущениям смешивающих распределений, поэтому замена исходного смешивающего распределения приближенным допустима. Если два смешивающих распределения близки в метрике Леви–Прохорова, то соответствующие им смеси близки в метрике Леви [7]. Следовательно, если сетка Z содержит большое число элементов и равномерно покрывает область, включающую носитель распределения V , то в предложенном параметризованном семействе содержится распределение, близкое к распределению Y .

В сеточном методе максимального правдоподобия поиск оценок весов проводится с помощью итерационного процесса. В функцию правдоподобия, соответствующую распределениям из набора \mathcal{P}_Ξ , подставляются реализации случайной величины Y . Далее численно ищется точка максимума этой функции как функции параметров. Распределение, соответствующее полученным величинам, используется в качестве оценки истинного. При это максимизируемая функция выпукла вверх по неизвестным параметрам.

Заметим, что распределение выборки независимых реализаций Y может быть достаточно близко к распределению реализаций некоторой величины с распределением из \mathcal{P}_Ξ . Несмотря на это, классическая теория, описывающая свойства оценок максимального правдоподобия, не может быть применена для

изучения свойств полученных оценок. Однако последние могут быть исследованы с помощью аппарата M -оценок (M -estimators), достаточно подробно описанного в работе [9].

В настоящей статье рассматривается следующее предположение: чем меньше расстояние между узлами сетки Z , покрывающей носитель истинного распределения, тем лучше можно приблизить истинное распределение распределением из \mathcal{P}_Ξ , следовательно, может существовать алгоритм согласованного увеличения размера выборки и числа элементов Z таким образом, чтобы последовательность получаемых оценок сходилась в некоторой метрике к оцениваемому распределению.

Ниже представлены результаты, описывающие поведение оценок, полученных с помощью сеточного метода максимального правдоподобия как для случая неограниченного увеличения размера выборки, так и для случая согласованного увеличения размера выборки и количества элементов множества Z . Будем придерживаться обозначений, используемых в главах работы [9], посвященной M -оценкам и эмпирическим процессам.

3 Основные обозначения

Для формального описания процесса «измельчения» сетки, на которой сосредоточен класс распределений, приближающих исходное смешивающее распределение, введем следующие обозначения. В отличие от работы [8], далее существенно используется тот факт, что работа ведется с масштабными смесями нормальных законов.

Рассмотрим некоторое множество $Z_0 \subset \mathbb{R}$, содержащее носитель смешивающего распределения. Обозначим через $Z_1 \subset Z_2 \subset \dots$ последовательность вложенных конечных подмножеств Z_0 , исчерпывающую счетное, всюду плотное подмножество Z_0 . Сопоставим каждому множеству Z_i параметрическое множество

$$\Theta_i = \{(\sigma_j, f_j) \mid \sigma_j \in Z_i, \sigma_j < \sigma_{j+1}, f_j = \sum_1^j p_j, \\ p_j \geq 0, j = 1, \dots, k_i, p_1 + \dots + p_{k_i} = 1, k_i = \#Z_i\}, \quad i = 1, 2, \dots$$

Рассмотрим метрические пространства $(\Theta_i, d_i) \equiv (\Theta_i, d)$, $i = 1, 2, \dots$, где расстояние между двумя элементами $\theta^1, \theta^2 \in \Theta_i$ задается через

$$d \equiv d_i(\theta^1, \theta^2) = \max_{1 \leq j \leq \#Z_i} |f_j^{(1)} - f_j^{(2)}|, \quad i = 1, 2, \dots$$

Далее рассмотрим статистическую структуру $(\mathfrak{X}, \mathcal{F}, \mathcal{P})$, где $\mathfrak{X} = \mathbb{R} \times \mathbb{R} \times \dots$; $\mathcal{F} = \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \dots$, $\mathcal{B}(\mathbb{R})$ — борелевская сигма-алгебра; $\mathcal{P} = \{\mathbf{P}_\theta, \theta \in \Theta_i$,

$i = 1, 2, \dots\} \cup P_0$ — параметризованное семейство вероятностных распределений, отвечающих последовательностям независимых одинаково распределенных случайных величин (X_1, X_2, X_3, \dots) . Здесь P_0 — распределение реализаций оцениваемой смеси. Для каждого фиксированного i класс $\{P_\theta, \theta \in \Theta_i\}$ содержит распределения, приближающие истинное среди смесей, соответствующих смешивающим распределениям, сосредоточенным на множестве Z_i .

Семейству \mathcal{P} соответствует семейство $\mathcal{P}^1 = \{P_{\theta_i}^1, \theta_i \in \Theta_i, i = 1, 2, \dots\} \cup P_0^1$ одномерных распределений. Одномерные распределения данных последовательностей для $\{P_\theta, \theta \in \Theta_i, i = 1, 2, \dots\}$ являются конечными смесями нормальных законов. При этом каждому фиксированному $\theta \in \Theta_i, i = 1, 2, \dots$, соответствует плотность одномерного распределения последовательности относительно меры Лебега на \mathbb{R} :

$$p_\theta(x) = \sum_{j=1}^{k_i} p_j \frac{1}{\sigma_j} \phi\left(\frac{x}{\sigma_j}\right), \quad x \in \mathbb{R},$$

где ϕ — плотность стандартного нормального закона. Заметим, что от P_0 , вообще говоря, лишь требуется, чтобы оно соответствовало распределению некоторой масштабной смеси нормальных законов. Будем считать, что P_0^1 имеет плотность p_0 относительно меры Лебега.

Рассмотрим статистики, с помощью которых проводится оценка истинного распределения. Зафиксируем некоторый номер i сетки Θ_i . Как было описано выше, в функцию правдоподобия, соответствующую распределениям из набора $\{P_\theta, \theta \in \Theta_i\}$, подставляются реализации независимых случайных величин с распределением P_0^1 . Затем ищется точка максимума этой функции как функции параметров.

В данном случае исследуются свойства последовательности оценок $\hat{\theta}_n^{(i)}$:

$$\hat{\theta}_n^{(i)} = \arg \max_{\theta \in \Theta_i} M_n(\theta),$$

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n m_\theta(X_i), \quad m_\theta(x) = \ln \frac{p_\theta(x)}{p_{\theta_0}(x)}, \quad \theta \in \Theta_i.$$

Такие оценки принадлежат классу M -оценок. При выполнении ряда условий они обладают некоторыми важными свойствами.

В следующем разделе исследуются свойства последовательности $\theta_0^{(i)}$:

$$\theta_0^{(i)} = \arg \min_{\theta \in \Theta} D_{\text{KL}}(P_0^1, P_\theta^1),$$

где D_{KL} — расстояние Кульбака–Лейблера. Для распределений P, Q , имеющих плотность p, q относительно меры Лебега,

$$D_{\text{KL}}(P, Q) = \int_{\mathbb{R}} \ln \frac{p(x)}{q(x)} p(x) dx$$

(будем считать, что $\min_{\theta \in \Theta} D_{\text{KL}}(P_0^1, P_\theta^1) < \infty$).

Легко показать, что для каждого натурального i точка $\theta_0^{(i)}$ единственна. Действительно, для каждого фиксированного значения x подинтегральная функция является выпуклой на выпуклом множестве как функция весов. При этом из того, что плотность p_0 является строго положительной, а конечные смеси нормальных законов идентифицируемы [11], напрямую следует, что указанная функция не может достигать минимума в двух различных точках.

Последовательность параметров $\{\theta_0^{(j)}\}_{j=1}^\infty$ соответствует последовательности распределений, сосредоточенных на сетках $\{Z_j\}_{j=1}^\infty$, ближайших к истинному в смысле расстояния Кульбака–Лейблера.

4 Основные свойства целевой функции

При использовании сеточного метода максимального правдоподобия задается множество точек, в которых может быть сосредоточено распределение, оценивающее истинное смешивающее распределение. Рассмотрим поведение получаемых оценок при неограниченном увеличении размера выборки реализаций смеси.

Используя введенные выше обозначения, зафиксируем параметрическое множество $\Theta = \Theta_i$ для одного из значений i , $i \in \mathbb{N}$. Сетка $\{\sigma_j \in Z_i, j = 1, \dots, k_i\}$ для него неизменна, поэтому для простоты под элементом параметрического множества $\theta \in \Theta$ будем понимать лишь вектор значений функции смешивающего распределения. Ниже получены основные свойства целевой функции, которые будут использоваться при исследовании асимптотических свойств оценок, получаемых с помощью сеточных методов.

Лемма 1. Для $d(\theta_0, \theta) > \delta$, $\theta \in \Theta$,

$$P_0^1 m_{\theta_0} - P_0^1 m_\theta \geq C_H(\Theta) \delta^2,$$

где

$$P_0^1 f \equiv \int_{\mathbb{R}} f(x) P_0^1(dx)$$

для интегрируемой функции f , $C_H(\Theta)$ — константа, зависящая от множества Θ .

Доказательство. Рассмотрим разложение в ряд Тейлора функции $P_0^1 m_\theta$ в окрестности точки θ_0 .

$$P_0^1 m_\theta = P_0^1 m_{\theta_0} + (\nabla P_0^1 m_{\theta_0}, \theta - \theta_0) + (\theta - \theta_0)^T H(P_0^1 m_{\theta_0})(\theta - \theta_0) + \dots,$$

где $\nabla P_0^1 m_{\theta_0}$, $H(P_0^1 m_{\theta_0})$ — градиент и гессиан функции в точке θ_0 соответственно. Заметим, что из того, что $\theta_0 = \arg \max_{\theta \in \Theta} \int_{\mathbb{R}} m_{\theta}(x) P_0^1(dx)$, следует, что

$$(\nabla P_0^1 m_{\theta_0}, \theta - \theta_0) \leq 0.$$

Легко видеть, что $-H(P_0^1 m_{\theta_0})$ является матрицей Грама, образованной из линейно независимых элементов в L^2 . Следовательно, $H(P_0^1 m_{\theta_0})$ является отрицательно определенной. Учитывая, что рассматриваемое параметрическое множество — компакт, получаем, для $d(\theta_0, \theta) > \delta$

$$P_0^1 m_{\theta_0} - P_0^1 m_{\theta} \geq C_H(\Theta) \delta^2,$$

где $C_H(\Theta)$ зависит от спектра матрицы Грама и параметрического множества.

Лемма 2. Если $p_k \geq c_{\sigma}/\sqrt{k}$ для всех $\theta \in \Theta$, то для любых $\theta^1, \theta^2 \in \Theta$, $x \in \mathbb{R}$,

$$|m_{\theta^1}(x) - m_{\theta^2}(x)| \leq \dot{m}_{\Theta}(x) d(\theta^1, \theta^2),$$

где

$$\dot{m}_{\Theta}(x) \leq \frac{2\sqrt{k}}{c_{\sigma}} \sum_{j=1}^k \frac{\sigma_k}{\sigma_j}, \quad \sigma_1 < \sigma_2 < \dots < \sigma_k.$$

Доказательство. Воспользуемся свойством логарифмической функции:

$$\begin{aligned} |m_{\theta^1}(x) - m_{\theta^2}(x)| &= \left| \ln \frac{\sum_{j=1}^k p_j^{(1)}(1/\sigma_j) \phi(x/\sigma_j)}{\sum_{j=1}^k p_j^{(2)}(1/\sigma_j) \phi(x/\sigma_j)} \right| \leq \\ &\leq \frac{\left| \sum_{j=1}^k p_j^{(1)}(1/\sigma_j) \phi(x/\sigma_j) - \sum_{j=1}^k p_j^{(2)}(1/\sigma_j) \phi(x/\sigma_j) \right|}{\min \left(\sum_{j=1}^k p_j^{(1)}(1/\sigma_j) \phi(x/\sigma_j), \sum_{j=1}^k p_j^{(2)}(1/\sigma_j) \phi(x/\sigma_j) \right)} \leq \\ &\leq \frac{\sum_{j=1}^k |p_j^{(1)} - p_j^{(2)}| (1/\sigma_j) \phi(x/\sigma_j)}{(c_{\sigma}/\sqrt{k})(1/\sigma_k) \phi(x/\sigma_k)} \end{aligned}$$

для любого рационального x . Пусть $d(\theta^1, \theta^2) \leq \delta$, тогда

$$|p_j^{(1)} - p_j^{(2)}| \leq 2\delta, \quad j = 1, \dots, k,$$

следовательно,

$$|m_{\theta^1}(x) - m_{\theta^2}(x)| \leq \frac{2\sqrt{k}\delta \sum_{j=1}^k (1/\sigma_j)\phi(x/\sigma_j)}{c_\sigma (1/\sigma_k)\phi(x/\sigma_k)} \leq \frac{2\sqrt{k}\delta \sum_{j=1}^k \sigma_k}{c_\sigma \sigma_j}. \quad \square$$

Далее, обозначим набор измеримых функций, определяемых параметрическим множеством Θ и числом δ , $\mathcal{F}_{\Theta, \delta} = \{m_\theta - m_{\theta_0} : d(\theta_0, \theta) < \delta, \theta \in \Theta\}$. Для доказательства следующего свойства необходимо оценить величину $E_{P_0} \sup_{f \in \mathcal{F}_{\Theta, \delta}} |\mathbb{G}_n f|$, где \mathbb{G} — эмпирический процесс, определенный на некотором наборе функций \mathcal{F} ,

$$\mathbb{G}_n f = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n f(X_i) - P_0^1 f \right), \quad f \in \mathcal{F}.$$

Оценим размер $\mathcal{F}_{\Theta, \delta}$ в терминах энтропии. Опираясь на результаты [9], введем понятие ε -отрезка $[l, u]$ в пространстве $L_2(P_0^1)$, состоящего из всех функций $f \in L_2(P_0^1)$, таких что $l \leq f \leq u$, $P_0^1(l - u)^2 < \varepsilon^2$. Далее введем понятие ограничивающего числа (bracketing number) $N_{[\cdot]}(\varepsilon, \mathcal{F}_{\Theta, \delta}, L_2(P_0^1))$ — минимального числа ε -отрезков, необходимых, чтобы покрыть набор функций $\mathcal{F}_{\Theta, \delta}$. Логарифм ограничивающего числа будем называть ограничивающей энтропией (entropy with bracketing). Рассмотрим также понятие ограничивающего интеграла (bracketing integral)

$$J_{[\cdot]}(\delta, \mathcal{F}_{\Theta, \delta}, L_2(P_0^1)) = \int_0^\delta \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{F}_{\Theta, \delta}, L_2(P_0^1))} d\varepsilon.$$

Введенные понятия не только помогают получить некоторые полезные свойства класса \mathcal{F}_Θ , но и позволяют оценить скорость сходимости оценок, получаемых сеточным методом. Справедливо следующее утверждение.

Лемма 3. Для класса \mathcal{F} измеримых функций, таких что

$$\forall f \in \mathcal{F} \quad |f| \leq F,$$

справедливо

$$E_{P_0} \sup_{f \in \mathcal{F}} |\mathbb{G}_n f| \leq C_m J_{[\cdot]} \left(\|F\|_{L_2(P_0^1)}, \mathcal{F}, L_2(P_0^1) \right),$$

где C_m — некоторое число, не зависящее от класса \mathcal{F} .

Доказательство. Основная идея доказательства опирается на использование неравенства Бернштейна. Полное доказательство данного факта приводится в [9].

Воспользовавшись данным фактом можно доказать следующее утверждение.

Лемма 4. Пусть Θ — некоторое параметрическое множество из последовательности $\{\Theta_i, i \in \mathbb{N}\}$, $p_k \geq c_\sigma / \sqrt{k}$ для всех $\theta \in \Theta$. Для набора измеримых функций $\mathcal{F}_{\Theta, \delta} = \{m_\theta - m_{\theta_0} : d(\theta_0, \theta) < \delta, \theta \in \Theta\}$ справедливо неравенство

$$E_{P_0}^* \sup_{f \in \mathcal{F}_{\Theta, \delta}} |\mathbb{G}_n f| \leq C \frac{k\delta}{c_\sigma} \sum_{j=1}^k \frac{\sigma_k}{\sigma_j},$$

где C — число, не зависящее от Θ .

Доказательство. Для $\mathcal{F}_{\Theta, \delta}$ существует ограничивающая функция F , такая что

$$\forall f \in \mathcal{F}_{\Theta, \delta} \quad |f| \leq F.$$

В данном случае $F = m_\Theta \delta \leq (2\sqrt{k}\delta/c_\sigma) \sum_{j=1}^k \sigma_k/\sigma_j$.

Из леммы 3 следует, что задача покрытия набора $\mathcal{F}_{\Theta, \delta}$ сводится к задаче покрытия параметрического множества Θ сферами соответствующего размера. Точнее, справедливо следующее неравенство:

$$N_{[\cdot]} \left(\varepsilon \|m_\Theta\|_{L_2(P_0^1)}, \mathcal{F}_{\Theta, \delta}, L_2(P_0^1) \right) \leq \left(\frac{2\delta}{\varepsilon} \right)^k, \quad 0 \leq \varepsilon \leq 2\delta,$$

где k — число узлов сетки.

Следовательно, для $\mathcal{F}_{\Theta, \delta}$ справедлива цепочка неравенств

$$\begin{aligned}
 E_{P_0} \sup_{f \in \mathcal{F}_{\Theta, \delta}} |\mathbb{G}_n f| &\leq C_m J_{[\cdot]} \left(\delta \|\dot{m}_{\Theta}\|_{L_2(P_0^1)}, \mathcal{F}_{\Theta, \delta}, L_2(P_0^1) \right) = \\
 &= C_m \int_0^{\delta \|\dot{m}_{\Theta}\|_{L_2(P_0^1)}} \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{F}_{\Theta, \delta}, L_2(P_0^1))} d\varepsilon = \\
 &= C_m \|\dot{m}_{\Theta}\|_{L_2(P_0^1)} \int_0^{\delta} \sqrt{\log N_{[\cdot]}(\varepsilon \|\dot{m}_{\Theta}\|_{L_2(P_0^1)}, \mathcal{F}_{\Theta, \delta}, L_2(P_0^1))} d\varepsilon \leq \\
 &\leq C_m \|\dot{m}_{\Theta}\|_{L_2(P_0^1)} \int_0^{\delta} \sqrt{k \log \left(\frac{2\delta}{\varepsilon} \right)} d\varepsilon = C \frac{k\delta}{c_{\sigma}} \sum_{j=1}^k \frac{\sigma_k}{\sigma_j}. \quad \square
 \end{aligned}$$

5 Поведение оценок при фиксированном параметрическом множестве

Последовательность рассматриваемых оценок имеет предел. Покажем это.

Теорема 1. Для последовательности оценок $\{\hat{\theta}_n\}_{n=1}^{\infty}$, таких, что

$$M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_{P_0}(1)$$

и $\int_{\mathbb{R}} \sup_{\theta \in \Theta} m_{\theta}(x) P_0^1(dx) < \infty$, справедливо утверждение: для любого $\varepsilon > 0$

$$P_0(d(\hat{\theta}_n, \theta_0) \geq \varepsilon) \rightarrow 0.$$

Доказательство. Действительно,

$$\theta_0 = \arg \max_{\theta \in \Theta} \int_{\mathbb{R}} m_{\theta}(x) P_0^1(dx),$$

m_{θ} непрерывна по θ для всех значений x и множество Θ — компакт. Следовательно, выполняются условия теоремы 5.14 из [9], и сформулированное утверждение верно. Отметим, что доказательство теоремы опирается на «классический» подход к доказательству состоятельности оценок максимального правдоподобия, описанный в [12], существенно использующий компактность параметрического множества. \square

Из доказанного утверждения следует, что в сеточном методе максимального правдоподобия при достаточно большом размере выборки смешивающее распределение оценивается среди распределений, сосредоточенных на известной сетке, с помощью ближайшего к истинному в смысле расстояния Кульбака–Лейблера. Очевидно, если носитель оцениваемого распределения лежит на данной сетке, то получаемые оценки совпадают с оценками максимального правдоподобия.

Воспользовавшись выведенными выше свойствами, перейдем к оценке скорости сходимости в сеточном методе разделения смесей вероятностных распределений. Будем опираться на общий результат, приведенный в [9].

Теорема 2. Пусть Θ — некоторое параметрическое множество из последовательности $\{\Theta_i, i \in \mathbb{N}\}$, $p_k \geq c_\sigma/\sqrt{k}$ для всех $\theta \in \Theta$, $\hat{\theta}_n$ — оценка, полученная при использовании сеточного метода, такая, что

$$M_n(\hat{\theta}_n) \geq M_n(\theta_0) - \frac{C_H(\Theta)}{2nV^2(\Theta)}. \quad (1)$$

Тогда $\forall \lambda > 0$

$$P_0 \left(\sqrt{n}V(\Theta)d(\hat{\theta}_n, \theta_0) > \lambda \right) \leq \frac{C}{\lambda},$$

где

$$V(\Theta) = \frac{C_H(\Theta)}{(k/c_\sigma) \sum_{j=1}^k \sigma_k/\sigma_j};$$

C — некоторая константа.

Доказательство. Представим параметрическое множество без точки θ_0 в виде счетного объединения множеств вида

$$S_{j,n} = \{ \theta : 2^{j-1} < \sqrt{n}V(\Theta) d(\theta, \theta_0) \leq 2^j, \theta \in \Theta \}, \quad j, n \in \mathbb{N}.$$

Для любого $l \in \mathbb{N}$

$$P_0 \left(\sqrt{n}V(\Theta)d(\hat{\theta}_n, \theta_0) > 2^l \right) \leq \sum_{j \geq l} P_0 \left(\sup_{\theta \in S_{j,n}} (M_n(\theta) - M_n(\theta_0)) \geq -\frac{C_H(\Theta)}{2nV^2(\Theta)} \right).$$

Заметим, что максимизируемая функция является выпуклой на выпуклом компакте. В связи с этим, в отличие от доказательства в [9], нет необходимости рассматривать только малую окрестность точки θ_0 . Можно подобрать константы в лемме 1 таким образом, чтобы условия выполнялись на всем параметрическом множестве.

Воспользовавшись леммой 1, получаем

$$\sup_{\theta \in S_{j,n}} P_0^1(m_{\theta_0} - m_{\theta}) \geq C_H(\Theta) \frac{4^{j-1}}{nV^2(\Theta)}.$$

Далее

$$\begin{aligned} & \left\{ \sup_{\theta \in S_{j,n}} (M_n(\theta) - M_n(\theta_0)) \geq -\frac{C_H(\Theta)}{2nV^2(\Theta)} \right\} \subseteq \\ & \subseteq \left\{ \sup_{\theta \in S_{j,n}} (M_n(\theta) - M_n(\theta_0)) \geq -C_H(\Theta) \frac{4^{j-1}}{2nV^2(\Theta)} \right\} \cap \\ & \cap \left\{ \sup_{\theta \in S_{j,n}} P_0^1(m_{\theta_0} - m_{\theta}) \geq C_H(\Theta) \frac{4^{j-1}}{nV^2(\Theta)} \right\} \subseteq \\ & \subseteq \left\{ \sup_{\theta \in S_{j,n}} (M_n(\theta) - P_0^1 m_{\theta} - M_n(\theta_0) + P_0^1 m_{\theta_0}) \geq C_H(\Theta) \frac{4^{j-1}}{2nV^2(\Theta)} \right\}. \end{aligned}$$

С учетом этого

$$\begin{aligned} P_0 \left(\sup_{\theta \in S_{j,n}} (M_n(\theta) - M_n(\theta_0)) \geq -\frac{1}{n} \right) & \leq \\ & \leq P_0 \left(\sup_{\theta \in S_{j,n}} |\mathbb{G}_n(m_{\theta} - m_{\theta_0})| \geq \sqrt{n} C_H(\Theta) \frac{4^{j-1}}{2nV^2(\Theta)} \right) \leq \\ & \leq \frac{2\sqrt{n}V^2(\Theta)}{C_H(\Theta)4^{j-1}} C \frac{2^j}{\sqrt{n}V(\Theta)} \|\dot{m}_{\Theta}\|_{L_2(P_0^1)} = V(\Theta) \frac{C\|\dot{m}_{\Theta}\|_{L_2(P_0^1)}}{C_H(\Theta)} \frac{8}{2^j} = \frac{8C}{2^j}. \end{aligned}$$

Таким образом, рассматриваемый ряд сходится и для любого натурального l

$$P_0 \left(\sqrt{n}V(\Theta)d(\hat{\theta}_n, \theta_0) > 2^l \right) \leq \frac{16C}{2^l}. \quad \square$$

Отметим, что требование (1) естественно и выполнимо, так как максимизируемая функция является выпуклой на выпуклом компакте. Данный результат отличается от аналогичной теоремы, доказанной в [9]. Полученные оценки являются уточненными, так как они отражают зависимость скорости сходимости от структуры сетки параметров. Их можно использовать при доказательстве сходимости в схеме серий.

6 Состоятельность оценок, полученных с помощью сеточных методов

Воспользовавшись результатами предыдущих разделов, перейдем к исследованию сходимости последовательности полученных оценок в функциональном пространстве при согласованном увеличении размеров сетки и числа элементов выборки, по которой строится оценка.

Поставим каждому $\theta^{(i)} \in \Theta_i$ в соответствие элемент $x_{\theta^{(i)}}$ пространства Скорохода $D[0, 1]$ следующим образом. Пусть F_0 — функция оцениваемого смешиваемого распределения, тогда

$$x_{\theta^{(i)}}(t) = F_{\theta^{(i)}}(F_0^{-1}(t)), \quad t \in [0, 1]; \quad x_{\theta^{(i)}}(1) = 1;$$

где

$$F_0^{-1}(t) = \inf_{x \in \mathbb{R}} \{x | F_0(x) \geq t\}, \quad t \in [0, 1]; \quad F_{\theta^{(i)}}(y) = \min_{j=1 \dots k_i} \{f_j | \sigma_j \geq y\}, \quad y \in \mathbb{R}.$$

Напомним, что для каждого натурального i оценка, полученная с помощью сеточного метода максимального правдоподобия по выборке размера n на параметрическом множестве Θ_i , обозначается через $\hat{\theta}_n^{(i)}$. С учетом этого $\hat{\theta}_n^{(i)}$ соответствует семейство конечномерных распределений $\mu_{\hat{\theta}_n^{(i)}}$:

$$\mu_{\hat{\theta}_n^{(i)}; t_1, t_2, \dots, t_l}(A) = P_0 \left((x_{\hat{\theta}_n^{(i)}}(t_1), \dots, x_{\hat{\theta}_n^{(i)}}(t_l)) \in A \right), \\ 0 \leq t_1 < t_2 < \dots < t_l \leq 1, \quad A \in \mathcal{B}(\mathbb{R}^l).$$

Данное семейство является согласованным. Справедливо следующее утверждение.

Лемма 5. Пусть Θ_i — некоторое параметрическое множество из последовательности $\{\Theta_i, i \in \mathbb{N}\}$, $p_{k_i} \geq c_\sigma / \sqrt{k_i}$ для всех $\theta \in \Theta_i$, $\hat{\theta}_n^{(i)}$ — оценка, полученная при использовании сеточного метода,

$$M_n(\hat{\theta}_n^{(i)}) \geq M_n(\theta_0^{(i)}) - \frac{C_H(\Theta_i)}{2nV^2(\Theta_i)}.$$

Тогда в $D[0, 1]$ существует случайный элемент $X_{\hat{\theta}_n^{(i)}}$ с конечномерными распределениями $\mu_{\hat{\theta}_n^{(i)}}$.

Доказательство. Действительно,

$$\mu_{\hat{\theta}_n^{(i)}; t_1, t, t_2}((\beta_1, \beta, \beta_2) : |\beta_1 - \beta| \geq \lambda, |\beta_2 - \beta| \geq \lambda) \leq \\ \leq \mu_{\hat{\theta}_n^{(i)}; t_1, t, t_2} \left((\beta_1, \beta, \beta_2) : |\beta_1 - x_{\theta_0^{(i)}}(t_1)| \geq \frac{\lambda}{3}, |\beta - x_{\theta_0^{(i)}}(t)| \geq \frac{\lambda}{3} \right),$$

$$\begin{aligned} & \left(|x_{\theta_0^{(i)}}(t_1) - x_{\theta_0^{(i)}}(t)| \geq \frac{\lambda}{3}, |\beta_2 - x_{\theta_0^{(i)}}(t_2)| \geq \frac{\lambda}{3}, |x_{\theta_0^{(i)}}(t) - x_{\theta_0^{(i)}}(t_2)| \geq \frac{\lambda}{3} \right) \leq \\ & \leq P_0 \left(d(\hat{\theta}_n^{(i)}, \theta_0^{(i)}) \geq \frac{\lambda}{3}, |x_{\theta_0^{(i)}}(t_2) - x_{\theta_0^{(i)}}(t_1)| \geq \frac{\lambda}{3} \right) \leq \frac{K_i (t_2 - t_1)^2}{\lambda \delta_i^2}, \end{aligned}$$

где K_i — константа, зависящая от параметрического множества, δ_i — длина минимального отрезка, на котором $x_{\theta_0^{(i)}}$ постоянна, $t_1 \leq t \leq t_2$ и

$$\forall \varepsilon : \lim_{h \downarrow 0} \mu_{\hat{\theta}_n^{(i)}; t_1, t_2} \{(\beta_1, \beta_2) : |\beta_1 - \beta_2| \geq \lambda\} = 0, \quad 0 \leq t \leq 1.$$

Следовательно, выполняются условия теоремы существования случайного элемента в пространстве Скорохода с заданными конечномерными распределениями [13]. \square

Таким образом, каждому значению параметра θ соответствует некоторая траекторию x_θ , а каждой оценке $\hat{\theta}_n^{(i)}$ — случайный процесс $X_{\hat{\theta}_n^{(i)}}$ со значениями в пространстве Скорохода $D[0, 1]$ с заданными выше конечномерными распределениями. Перейдем теперь к доказательству основного утверждения.

Теорема 3. *Рассмотрим последовательность случайных элементов $\{X_{\hat{\theta}_n^{(r(n))}}\}_{n=1}^\infty$ со значениями в пространстве Скорохода $D[0, 1]$, где $r(n)$ — возрастающая натуральная последовательность.*

Пусть выполнены следующие условия:

1. *Последовательность распределений $P_{\theta_0^{(n)}}^1$ сходится в смысле расстояния Кульбака–Лейблера:*

$$D_{KL}(P_0^1, P_{\theta_0^{(n)}}^1) \rightarrow 0.$$

2. *Для каждого элемента последовательности параметрических множеств $\{\Theta_i\}_{i=1}^\infty$ справедливо неравенство $p_{k_i} \geq c_\sigma / \sqrt{k_i}$ и*

$$M_n(\hat{\theta}_n^{(i)}) \geq M_n(\theta_0^{(i)}) - \frac{C_H(\Theta_i)}{2nV^2(\Theta_i)},$$

где

$$V(\Theta_i) = \frac{C_H(\Theta_i)}{(k_i/c_\sigma) \sum_{j=1}^{k_i} \sigma_k^{(i)} / \sigma_j^{(i)}},$$

3. Последовательность чисел $\{r(n)\}_{n=1}^{\infty}$ такова, что

$$\sqrt{n}V(\Theta_{r(n)}) \rightarrow \infty$$

и существует константа K :

$$\sqrt{n}V(\Theta_{r(n)})\delta_i^2 \geq K, \quad \forall n \in \mathbb{N},$$

где δ_i — длина минимального отрезка, на котором $x_{\theta_0^{(i)}}$ постоянна.

4. Семейство логарифмов случайных величин, отвечающих $\mathcal{P}^1 = \{\mathbb{P}_{\theta_i}^1, \theta_i \in \Theta_i, i = 1, 2, \dots\} \cup \mathbb{P}_0^1$, — плотно.

Тогда

$$X_{\hat{\theta}_n^{(r(n))}} \xrightarrow{d} x_{\theta_0}.$$

Доказательство. При доказательстве данной теоремы будем опираться на теорему Прохорова и ее следствия [13]. Рассмотрим последовательность случайных элементов пространства $D[0, 1]$

$$\{Z_n = X_{\hat{\theta}_n^{(r(n))}} - x_{\theta_0^{(r(n))}}\}_{n=1}^{\infty}.$$

Конечномерные распределения данной последовательности сходятся к конечномерным распределениям элемента, вырожденного в z , $z(t) = 0$ для любого $t \in [0, 1]$.

Далее,

$$\begin{aligned} P_0(|Z_n(t) - Z_n(t_1)| \geq \lambda, |Z_n(t) - Z_n(t_2)| \geq \lambda) &\leq \\ &\leq P_0\left(|Z_n(t)| \geq \frac{\lambda}{3}, |Z_n(t_1)| \geq \frac{\lambda}{3}, |x_{\theta_0^{(r(n))}}(t_1) - x_{\theta_0^{(r(n))}}(t)| \geq \frac{\lambda}{3}, \right. \\ &\quad \left. |Z_n(t_2)| \geq \frac{\lambda}{3}, |x_{\theta_0^{(r(n))}}(t) - x_{\theta_0^{(r(n))}}(t_2)| \geq \frac{\lambda}{3}\right) \leq \\ &\leq P_0\left(d(\hat{\theta}_n^{(r(n))}, \theta_0^{(r(n))}) \geq \frac{\lambda}{3}, |x_{\theta_0^{(r(n))}}(t_2) - x_{\theta_0^{(r(n))}}(t_1)| \geq \frac{\lambda}{3}\right) \leq \\ &\leq \frac{C}{\sqrt{n}V(\Theta_{r(n)})\lambda} \frac{(t_2 - t_1)^2}{\delta_i^2} \leq \frac{C(t_2 - t_1)^2}{K\lambda}. \end{aligned}$$

Следовательно, последовательность случайных элементов

$$Z_n \xrightarrow{d} z.$$

По условию $D_{KL}(P_0^1, P_{\theta_0^{(n)}}^1) \rightarrow 0$. Из неравенства Пинскера следует, что данная последовательность сходится в смысле расстояния полной вариации. Так как расстояние полной вариации сильнее [14] метрики Леви, последовательность сходится в смысле расстояния Леви. Для масштабных смесей существуют верхние оценки устойчивости [7]. Семейство логарифмов случайных величин, отвечающих смешивающим распределениям, плотно. Следовательно, последовательность логарифмов этих случайных величин сходится в метрике Леви [15]. Отсюда напрямую следует, что

$$x_{\theta_0^{(r(n))}} \xrightarrow{d} x_{\theta_0}.$$

Далее, воспользовавшись неравенством треугольника, получаем утверждение теоремы. \square

Следовательно, при выполнении указанных условий оценки, полученные с использованием сеточных методов, являются состоятельными.

7 Заключение

Показано, что оценки смешивающих распределений, полученные с помощью сеточных методов разделения смесей вероятностных распределений, для указанных классов смесей нормальных законов являются состоятельными при согласованном увеличении размера выборки и числа узлов сетки. Данный факт является важным теоретическим обоснованием применимости сеточных методов на практике. Он согласуется с результатами, полученными при анализе работы алгоритмов на различных наборах данных. Аналогичные результаты можно получить и для других классов смесей вероятностных распределений.

Литература

1. *Королев В. Ю.* Вероятностно-статистический анализ хаотических процессов с помощью смешанных гауссовских моделей. Декомпозиция волатильности финансовых индексов и турбулентной плазмы. — М.: ИПИ РАН, 2007.
2. *Королев В. Ю., Непомнящий Е. В., Рыбальченко А. Г., Виноградова А. В.* Сеточные методы разделения смесей вероятностных распределений и их применение к декомпозиции волатильности финансовых индексов // Информатика и её применения, 2008. Т. 2. Вып. 2. С. 3–18.
3. *Назаров А. Л.* Разделение смесей вероятностных распределений сеточным методом максимального правдоподобия при помощи алгоритма условного градиента // Сб. статей молодых ученых факультета ВМиК МГУ, 2009. № 6. С. 128–135.
4. *Korolev V., Nazarov A.* Separating mixtures of probability distributions with the grid method of moments and the grid maximal likelihood method // Automation Remote Control, 2010. Vol. 71. No. 3. P. 455–472.

5. *Tukey J. W.* A survey of sampling from contaminated distributions // Contributions to probability and statistics: Essays in honor of Harold Hotelling / Eds. I. Olkin, S. G. Ghurye, W. Hoeffding, *et al.* — Stanford, CA: Stanford University Press, 1960. P. 448–485.
6. *Hall P.* On measures of the distance of a mixture from its parent distribution // Stochastic Proc. Their Applications, 1979. Vol. 8. No. 3. P. 357–365.
7. *Назаров А. Л.* Об устойчивости смесей вероятностных законов к возмущениям смешивающих распределений // Статистические методы оценивания и проверки гипотез, 2010. № 22. С. 154–172.
8. *Назаров А. Л.* Асимптотические свойства оценок, полученных с помощью сеточных методов разделения смесей вероятностных распределений // Статистические методы оценивания и проверки гипотез, 2012. № 24. С. 22–35.
9. *Van der Vaart A. W.* Asymptotic statistics. — Cambridge: Cambridge University Press, 2000. Cambridge ser. in statistical and probabilistic mathematics.
10. *Pollard D.* Empirical processes: Theory and applications. — Hayward, CA: Institute of Mathematical Statistics, 1990. NSF-CBMS Regional Conference ser. in probability and statistics. Vol. 2
11. *Teicher H.* Identifiability of finite mixtures // Ann. Math. Stat. 1963. Vol. 34. No. 4. P. 1265–1269.
12. *Wald A.* Note on the consistency of the maximum likelihood estimate // Ann. Math. Stat. 1949. Vol. 20. P. 595–601.
13. *Биллингсли П.* Сходимость вероятностных мер. — М.: Наука, 1977.
14. *Золотарев В. М.* Современная теория суммирования независимых случайных величин. — М.: Наука, 1986.
15. *Назаров А. Л.* Нижние оценки в задаче устойчивости смесей нормальных распределений к возмущениям смешивающих распределений // Информатика и её применения, 2012. Т. 6. Вып. 4.

ОПТИМИЗАЦИЯ РЕСУРСОВ ВЫЧИСЛИТЕЛЬНОГО КОМПЛЕКСА СО СЛАБОСВЯЗАННОЙ АРХИТЕКТУРОЙ*

*М. Я. Агаларов*¹

Аннотация: Рассматривается задача расчета оптимального объема вычислительных ресурсов (ВР) вычислительного комплекса (ВК) со слабосвязанной архитектурой и пуассоновскими входящими потоками заданий специального класса. Предлагается эвристический алгоритм условной оптимизации объема ресурсов ВК. Приведены результаты вычислительных экспериментов и оценки трудоемкости предлагаемого алгоритма.

Ключевые слова: вычислительный комплекс; вычислительные ресурсы; поток заданий; оптимизация объема ресурсов

1 Введение

Рассматривается ВК с множеством слабо связанных ВР, на который поступает стохастический поток заданий специального класса. Подробное описание данного ВК и его математическая модель для случая, когда не выполненные задания покидают комплекс и не возвращаются, приведены в работе [1]. В данной работе ставится задача расчета минимального требуемого объема ВР с учетом повторных поступлений не выполненных заданий. Из известных работ, в которых приведены аналогичные математические постановки задач и использованы подобные подходы к их решению, наиболее близки работы, посвященные мультисервисным телекоммуникационным сетям [2, 3]. Приведем содержательное описание рассматриваемой задачи.

На вход ВК из одинаковых ВР поступает поток заданий, представляющих собой совокупности элементарных заданий (ЭЗ). Каждое ЭЗ может независимо выполняться на любом ВР. Длина задания (количество ЭЗ, составляющих задание) является независимой дискретной случайной величиной. Время выполнения каждого задания в ВК ограничивается предельным значением (контрольным временем), задаваемым заранее. Для управления распределением вычислительных и коммуникационных ресурсов, организации и контроля процессов выполнения заданий в ВК выделен отдельный вычислительный ресурс, называемый центром управления (ЦУ), который связан с остальными ВР коммуникационной средой.

* Работа выполнена при поддержке РФФИ (грант 11-07-00112).

¹ Институт проблем информатики Российской академии наук, murad-agalarov@yandex.ru

Выполнение задания в ВК происходит следующим образом. Для каждого поступающего в ВК задания ЦУ назначает подмножество ВР, которые до завершения его выполнения не доступны другим заданиям. Если в момент поступления задания в ВК отсутствует требуемое число свободных ресурсов, задание покидает ВК и, возможно, через случайный интервал времени вновь поступает в ВК. Перед тем как отправить задание на выполнение, в ЦУ задание предварительно разбивается на пакеты из определенного числа ЭЗ. Пакеты направляются на ресурсы группами, называемыми посылками, в каждой из которых (за исключением возможно последней) число пакетов совпадает с количеством выделенных для задания ресурсов. Посылка занимает и освобождает все выделенные ресурсы одновременно, при этом один пакет занимает один ресурс. Линии связи и ВР могут подвергаться сбоям (коротким самовосстанавливающимся отказам). Выполнение пакета считается успешным, если во время его выполнения не произошел сбой в линии связи или ВР, иначе выполнение пакета считается неуспешным. Неуспешно выполненный пакет включается в ЦУ в состав какой-нибудь вновь формируемой посылки этого же задания. Выполнение посылки считается завершенным, если выполнены (успешно или неуспешно) все составляющие ее пакеты. Очередную посылку ЦУ направляет на ресурсы только после завершения выполнения предыдущей посылки.

В каждом задании, возможно, одно или несколько ЭЗ имеют отличительный признак, при обнаружении которого хотя бы в одном пакете задание сбрасывается из системы. Обнаружение отличительного признака в пакете возможно только при успешном его выполнении. Считается априори, что ЭЗ с отличительным признаком (если в задании хотя бы одно ЭЗ с отличительным признаком существует) может оказаться в любом пакете с равной вероятностью. Задание покидает ВК либо в момент достижения предельного времени выполнения, либо в момент завершения выполнения очередной посылки, если обнаружится ЭЗ с отличительным признаком, либо после успешного выполнения всех пакетов задания, освобождая одновременно все занятые им ресурсы. Время выполнения задания включает время его обработки в ЦУ, время передачи посылок по коммуникационной среде и вычисления на ВР.

Считается, что ВК выполняют задания с требуемым качеством, если значения показателя качества выполнения заданий не ниже заданной величины.

Одной из важных задач на этапе проектирования ВК является задача расчета минимального объема ресурсов, обеспечивающего требуемое качество выполнения заданий. Объем ВР, необходимый для выполнения заданий, может зависеть как от предъявляемых требований к качеству выполнения заданий, так и от процесса поступления заданий, управления ресурсами ВК, организации процесса выполнения заданий в ВК и т. д. Решение этой задачи точными аналитическими методами вызывает определенные трудности, вызванные в основном следующими двумя причинами. Одна из них — это отсутствие точных аналитических моделей

рассматриваемого ВК для случая формирования в ЦУ очереди заданий или повторных поступлений невыполненных заданий (в отличие, например, от случая с явными потерями). Другая причина — это то, что она относится к задачам дискретной оптимизации и по сложности относится к NP-полным. Решение данной задачи точными методами дискретного программирования не применимо на практике в случае большого числа вариантов распределения ВР, иначе эффект от использования оптимального распределения может быть сведен на нет трудоемкостью получения самого распределения. Поэтому решение данной задачи с использованием приближенной модели и алгоритмов непрерывной оптимизации пока остается единственным способом, который применим на практике.

2 Модель вычислительного комплекса и постановка задачи

Пусть общий поток заданий делится на M потоков, в каждом из которых задания имеют одинаковую детерминированную длину. Предположим, что суммарный входной поток заданий (первичных и повторных) пуассоновский, время пребывания задания в ЦУ до начала передачи на ВР — случайная величина с заданным средним значением, время передачи пакета по коммуникационной среде — случайная величина с заданным средним значением, ВР системы имеют заданную одинаковую производительность, вероятности сбоя ВР и вероятности повтора невыполненного задания каждого потока — заданные величины. Всюду в дальнейшем считается, что ВК находится в стационарном режиме работы.

Введем следующие обозначения:

R — число ВР системы (ВК);

ν — производительность единицы ресурса;

M — число различных длин заданий;

L — множество различных длин заданий, $L = \{l_j, j = 1, \dots, M\}$, l_j — длина задания;

j -задание — задание длины l_j ;

λ_j — интенсивность первичного потока j -заданий;

Λ_j — интенсивность суммарного потока j -заданий, поступающих в ВК;

l_p — длина пакета j -задания;

n_j — число пакетов в j -задании;

k_{\max_j} — максимальное число ресурсов, которое ЦУ может выделить для выполнения j -задания, $0 \leq k_{\max_j} \leq n_j$, $k_{\max_j} \leq R$;

k_{\min_j} — минимальное число ресурсов, требуемых для выполнения j -задания, $0 \leq k_{\min_j} \leq k_{\max_j}$;

$m_j = k_{\max_j} - k_{\min_j}$, $m_j + 1$ — число возможных вариантов выделения ресурсов для j -задания;

(j, i) -задание — это задание, число назначенных ресурсов которому равно $(i + k_{\min_j})$, (задание (j, i) -го типа);

W — множество всех возможных пар (j, i) таких, что $j = 1, \dots, M$, $i = 0, \dots, m_j$;

Λ_{ji} — интенсивность суммарного потока (j, i) -заданий;

β_j — вероятность повтора j -задания, $0 \leq \beta_j \leq 1$, $j = 1, \dots, M$;

α_j — вероятность искажения пакета j -задания (вероятность сбоя ресурса во время выполнения пакета j -задания);

$\bar{t}z_j$ — среднее времени обработки j -задания в ЦУ;

$\bar{t}o_{ji}$ — среднее время пребывания (j, i) -задания в системе;

$t_{\text{contrz},j}$ — ограничение сверху на время пребывания j -задания в системе (контрольное время);

π_{ji} — стационарная вероятность отклонения (j, i) -задания из-за отсутствия назначенного числа свободных ресурсов;

π_j — стационарная вероятность отклонения j -задания из-за отсутствия назначенного числа свободных ресурсов.

Пусть для задания назначается $i + \text{kmin}_j$ ресурсов согласно вероятностному распределению p_{ji} , $i = 0, \dots, m_j$, $\sum_0^{m_j} p_{ji} = 1$. Набор $\bar{p} = \{\bar{p}_j, j = 1, \dots, M\}$, где $\bar{p}_j = \{p_{ji}, i = 0, \dots, m_j\}$ назовем квазистатическим планом распределения ресурсов системы.

Обозначим через $\bar{k} = \{\bar{k}_1, \dots, \bar{k}_M\}$ вектор состояния системы, где $\bar{k}_j = \{k_{j0}, \dots, k_{jm_j}\}$, k_{ji} — число (j, i) -заданий в системе; Ω — множество всех возможных состояний системы; $\bar{\Omega}_{ji}$ — множество всевозможных значений вектора \bar{k} , при которых (j, i) -задание получает отказ (т. е. $\bar{\Omega}_{ji} = \{\bar{k} : R - i - \text{kmin}_j < \sum_{i=1}^{m_j} k_{ji}(i + \text{kmin}_j) \leq R\}$); $p(\bar{k})$ — стационарная вероятность того, что вектор состояния системы равен \bar{k} ; π_{ji} — вероятность того, что (j, i) -задание получит отказ.

В обозначениях, приведенных выше, для стационарных вероятностей состояний системы справедлива формула [1, 4]:

$$P(\bar{k}, \bar{p}) = [P(\bar{0}, \bar{p})]^{-1} \prod_{(j,i) \in W} \frac{\rho_{ji}^{k_{ji}}}{k_{ji}!},$$

где $\rho_{ji} = \Lambda_j p_{ji} \bar{t}o_{ji}$ — нагрузка, создаваемая (j, i) -заданиями; вероятность $P(\bar{0}, \bar{p})$ вычисляется из условия нормировки и равна

$$P(\bar{0}, \bar{p}) = \sum_{\bar{k} \in \Omega} \prod_{(j,i) \in W} \frac{\rho_{ji}^{k_{ji}}}{k_{ji}!}.$$

Вероятность отклонения (j, i) -задания равна

$$\pi_{ji} = \sum_{\bar{k} \in \bar{\Omega}_{ji}} P(\bar{k}, \bar{p}),$$

а вероятность отклонения j -задания равна

$$\pi_j = \sum_{i=1}^{m_j} p_{ji} \pi_{ji}.$$

В стационарном режиме работы ВК для потоков Λ_j , $j = 1, \dots, M$, справедливы следующие уравнения баланса:

$$\Lambda_{jr} = \left(\lambda_j + \beta_j \sum_{s=0}^{m_j} \Lambda_{js} \pi_{js} \right) p_{jr}, \quad r = 0, \dots, m_j, \quad j = 1, \dots, M. \quad (1)$$

Обозначим через S множество всех возможных квазистатических планов \bar{p} , при которых система уравнений (1) имеет решение.

Определим показатели эффективности ВК:

- $\Lambda_{\text{ВЫХ}}(\bar{p}) = \sum_{j=1}^M \sum_{r=0}^{m_j} \Lambda_{jr} (1 - \pi_{jr}) B_{jr}(t_{\text{contrz},j})$ — интенсивность потока успешно выполненных заданий, где $B_{jr}(t_{\text{contrz},j})$ — вероятность того, что выполнение (j, i) -задания завершится успешно;
- $\Lambda_{\text{ПОВТ}}(\bar{p}) = - \sum_{j=1}^M \left(\sum_{r=0}^{m_j} \Lambda_{jr} - \lambda_j \right)$ — интенсивность потока повторных заданий, поступающих в систему.

Задача оптимизации ресурсов ВК ставится в виде задачи I:

найти целое число R_0 такое, что $\max_{\bar{p} \in S} F(\bar{p}, R) \leq A_F$ при любом объеме ресурсов $R < R_0$ и $\max_{\bar{p} \in S} F(\bar{p}, R) > A_F$ при $R = R_0$, где $F(\bar{p}, R)$ — целевая функция, совпадающая с одним из показателей эффективности ВК, A_F — заданная величина ($0 < A_F \leq \sum_{j=1}^M \lambda_j$, если $F(\bar{p}, R) = \Lambda_{\text{ВЫХ}}(\bar{p}, R)$ и $A_F \leq - \sum_{j=1}^M \lambda_j$, если $F(\bar{p}, R) = \Lambda_{\text{ПОВТ}}(\bar{p}, R)$).

3 Метод и алгоритм решения

Пусть r_j^0 ($j = 1, \dots, M$) — максимально возможное значение индекса r параметра p_{jr} для всех планов $\bar{p} \in S$. Так как система рассматривается только

в стационарном режиме работы, то, очевидно, необходимо выполнение условия $m_j \leq r_j^0$, $j = 1, \dots, M$. Отметим два простых свойства задачи I, справедливых для случая $F(\bar{p}, R) = \Lambda_{\text{ВЫХ}}(\bar{p}, R)$:

1. Задача I не имеет решения, если выполняется условие:

$$\sum_{j=1}^M \lambda_j B_{jm_j}(t_{\text{contrz},j}) \leq A_F. \quad (2)$$

2. План распределения $\bar{p} \in S$ таков, что $p_{jr} = 1$ при $r = m_j$, $p_{jr} = 0$ при $r \neq m_j$, $j = 1, \dots, M$, является решением задачи $\max_{\bar{p} \in S} F(\bar{p}, R)$ при $\beta_j = 1$, $j = 1, \dots, M$.

Первое свойство следует из справедливости следующих соотношений:

$$B_{js-1}(t_{\text{contrz},j}) \leq B_{js}(t_{\text{contrz},j}), \quad s = 1, \dots, m_j;$$

$$\begin{aligned} \lambda_j B_{jm_j}(t_{\text{contrz},j}) &= \sum_{r=0}^{m_j} \Lambda_{jr}(1 - \beta_j \pi_{jr}) B_{jm_j}(t_{\text{contrz},j}) \geq \\ &\geq \sum_{r=0}^{m_j} \Lambda_{jr}(1 - \pi_{jr}) B_{jm_j}(t_{\text{contrz},j}) \geq \sum_{r=0}^{m_j} \Lambda_{jr}(1 - \pi_{jr}) B_{jr}(t_{\text{contrz},j}) \end{aligned}$$

для любых значений R и J .

Второе свойство следует из соотношений

$$\begin{aligned} \sum_{j=1}^M \Lambda_{jm_j}(1 - \pi_{jm_j}) B_{jm_j}(t_{\text{contrz},j}) &\geq \sum_{j=1}^M \lambda_j B_{jm_j}(t_{\text{contrz},j}) \geq \\ &\geq \sum_{j=1}^M \sum_{r=0}^{m_j} \Lambda_{jr}(1 - \pi_{jr}) B_{jm_j}(t_{\text{contrz},j}) \geq \sum_{j=1}^M \sum_{r=0}^{m_j} \Lambda_{jr}(1 - \pi_{jr}) B_{jr}(t_{\text{contrz},j}) = \\ &= \Lambda_{\text{ВЫХ}}(\bar{p}, R) \end{aligned}$$

для любого $\bar{p} \in S$.

Для решения оптимизационной задачи I в общем случае предлагается алгоритм, состоящий из следующих основных действий.

1. Если $F(\bar{p}, R) = \Lambda_{\text{ВЫХ}}(\bar{p}, R)$ и выполняется условие (2), то завершить работу алгоритма, иначе выбрать начальные значения параметров R и \bar{p} такие, что первичный поток заданий реализуем и выполняются условия: $F(\bar{p}, R) > A_F$, $F(\bar{p}, R - 1) \leq A_F$.

2. Выполнить последовательно шаги метода градиентного спуска (см. ниже) для решения задачи $\max_{\bar{p} \in S} F(\bar{p}, R)$ до тех пор, пока либо $F(\bar{p}[n], R) > A_F$, либо $F(\bar{p}[n], R) - F(\bar{p}[n-1], R) < \varepsilon$, где $\varepsilon > 0$ — заданная малая величина (точность решения задачи $\max_{\bar{p} \in S} F(\bar{p}, R)$), $\bar{p}[0]$ — начальное значение плана распределения \bar{p} , полученное в п. 1, $\bar{p}[n]$ — план, полученный на n -м шаге градиентного метода.
3. Если $F(\bar{p}[n], R) < A_F$ и $F(\bar{p}[n], R) - F(\bar{p}[n-1], R) \geq \varepsilon$, то увеличить R на единицу и перейти к п. 2, иначе текущее значение R считать решением задачи и завершить работу алгоритма.

На каждом шаге метода градиентного спуска, выполняемом в п. 2 алгоритма, вычисляется $\bar{p}[n+1] = C_{\delta} \bar{p}[n]$, $n \geq 0$, где C_{δ} — отображение $S \xrightarrow{B} S$. C_{δ} определено следующим образом:

$$\bar{p}' = C_{\delta} \bar{p}, \bar{p}' = \{\bar{p}'_s, s = 1, \dots, M\}, \bar{p}'_s = \{p'_{si}, i = 0, \dots, m_s\};$$

$$p'_{si} = \begin{cases} p_{si} - \delta_s \Delta_s, & \text{если } i = i_{\min}; \\ p_{si} + \delta_s \Delta_s, & \text{если } i = i_{\max}; \\ p_{si}, & \text{если } i \neq i_{\min}, i \neq i_{\max}, i = 0, \dots, m_s, \end{cases}$$

где i_{\min} — индекс i , при котором достигается $\min_{i: i \in \{0, \dots, m_j\}, p_{si} > 0} (\partial F(\bar{p}, R) / \partial p_{si})$;
 i_{\max} — индекс i , при котором достигается $\max_{i: i \in \{0, \dots, m_s\}, p_{si} < 1} (\partial F(\bar{p}, R) / \partial p_{si})$;
 $\Delta_s = \max_{i: i \in \{0, \dots, m_s\}, p_{si} < 1} (\partial F(\bar{p}, R) / \partial p_{si}) - \min_{i: i \in \{0, \dots, m_j\}, p_{si} > 0} (\partial F(\bar{p}, R) / \partial p_{si})$, $\delta_s > 0$ — масштабный параметр, $p_{si} - \delta_s \Delta_s \geq 0$ при $i = i_{\min}$ и $p_{si} + \delta_s \Delta_s \leq 1$ при $i = i_{\max}$, $s = 1, \dots, M$.

Для производных $\partial \Lambda_{\text{ВЫХ}}(\bar{p}) / \partial p_{si}$, $\partial \Lambda_{\text{ПОВТ}}(\bar{p}) / \partial p_{si}$ ($i = 0, \dots, m_s, s = 1, \dots, M$) верны равенства:

$$\frac{\partial \Lambda_{\text{ВЫХ}}(\bar{p})}{\partial p_{si}} = \sum_{j=1}^M \sum_{r=0}^{m_j} B_{jr}(t_{\text{contrz},j}) \left[\frac{\partial \Lambda_{jr}}{\partial p_{si}} (1 - \pi_{jr}) - \Lambda_{jr} \sum_{(l,n) \in W} \bar{t}_{ln} \frac{\partial \pi_{jr}}{\partial \rho_{ln}} \frac{\partial \Lambda_{ln}}{\partial p_{si}} \right]; \quad (3)$$

$$\frac{\partial \Lambda_{\text{ПОВТ}}(\bar{p})}{\partial p_{si}} = - \sum_{j=1}^M \sum_{r=0}^{m_j} \frac{\partial \Lambda_{jr}}{\partial p_{si}}. \quad (4)$$

Пронумеруем элементы (j, r) множества W соответствующими числами $N_{jr} = \sum_{s=0}^{j-1} m_s + r$. Перенесем в (1) правую часть в левую и обозначим полученные в левой части соотношения через

$$\Phi_{N_{jr}} = \Lambda_{jr} - \left(\lambda_j + \sum_{s=0}^{m_j} \Lambda_{js} \pi_{js} \right) p_{jr}, \quad r = 0, \dots, m_j, \quad j = 1, \dots, M.$$

Используя метод вычисления производных для неявных функций, изложенный в [5], для производных неявных функций (1) получим следующие равенства:

$$\frac{\partial \Lambda_{jr}}{\partial p_{si}} = \frac{[D_{(j,r)}]}{[D]}, \quad (5)$$

где $[D]$ — детерминант матрицы D порядка $|W| \times |W|$, в которой N_{jr} -я строка является последовательностью вида $\{\partial \Phi_{N_{jr}} / \partial \Lambda_{N_{si}}, N_{si} = 1, \dots, |W|\}$; $[D_{(j,r)}]$ — детерминант матрицы $D_{(j,r)}$ порядка $|W| \times |W|$, которая получается из D заменой N_{jr} -го столбца на $\{-\partial \Phi_{N_{in}} / \partial p_{N_{si}}, N_{in} = 1, \dots, |W|\}$;

$$\begin{aligned} \frac{\partial \Phi_{N_{jr}}}{\partial \Lambda_{N_{si}}} &= [\chi_{N_{si}}(N_{jr}) - p_{si} \pi_{si} \chi_j(s)] - p_{ir} \bar{t}_{0si} \sum_{l=1}^{m_j} \Lambda_{jl} \frac{\partial \pi_{jl}}{\partial \rho_{si}}; \\ \frac{\partial \Phi_{N_{ln}}}{\partial p_{N_{si}}} &= - \left(\lambda_i + \sum_{r=0}^{m_i} \Lambda_{ir} \pi_{ir} \right) \chi_{N_{si}}(N_{ln}) = \Lambda_i \chi_{N_{si}}(N_{ln}); \end{aligned}$$

$\chi_a(b)$ — функция-индикатор; a и b — целые числа;

$$\chi_a(b) = \begin{cases} 1, & \text{если } a = b; \\ 0, & \text{если } a \neq b. \end{cases}$$

При вычислении с помощью (3)–(5) производных $\partial F(\bar{p}, R) / \partial p_{si}$ ($i = 0, \dots, m_s$; $s = 1, \dots, M$), используются следующие утверждения, позволяющие существенно уменьшить объем требуемого алгоритмом машинного времени и памяти ЭВМ.

1. Справедливы формулы

$$\pi_{si} = \left\{ \sum_{d=0}^R q_d(\bar{p}) \right\}^{-1} \sum_{d=R-i-k\min_s+1}^R q_d(\bar{p});$$

$$\pi_{jr,si} = \left\{ \sum_{d=0}^{R-i-\text{kmin}_s} q_d(\bar{p}) \right\}^{-1} \sum_{d=R-i-\text{kmin}_s-r-\text{kmin}_j+1}^{R-i-\text{kmin}_s} q_s(\bar{p});$$

$$dq_d(\bar{p}) = \begin{cases} \sum_{(s,i) \in W} (i + \text{kmin}_s) \rho_{si} q_{d-i-\text{kmin}_s}(\bar{p}) & \text{при } d \geq 0; \\ 0 & \text{при } d < 0, \end{cases}$$

где

$$q_d(\bar{p}) = \begin{cases} \sum_{\bar{k} \in D_d} \prod_{(j,i) \in W} \frac{\rho_{ji}^{k_{ji}}}{k_{ji}!} & \text{при } d \geq 0, \\ 0 & \text{при } d < 0; \end{cases}$$

$$D_d = \left\{ \bar{k} \in \Omega : \sum_{j=1}^M \sum_{i=0}^{m_j} k_{ji}(i + \text{kmin}_j) = d \right\};$$

$\pi_{jr,si}$ — вероятность того, что в системе с общим числом $(R - i - \text{kmin}_s)$ ресурсов (j, r) -задание получит отказ, $i = 0, \dots, m_s$; $r = 0, \dots, m_j$; $j, s = 1, \dots, M$.

2. Для производной функции π_{jr} по ρ_{si} справедливо равенство:

$$\frac{\partial \pi_{jr}}{\partial \rho_{si}} = (1 - \pi_{si}) (\pi_{jr,si} - \pi_{jr}),$$

где $\pi_{jr,si}$ — вероятность того, что в системе с общим числом $(R - i - \text{kmin}_s)$ ресурсов (j, r) -задание получит отказ, $i = 0, \dots, m_s$; $r = 0, \dots, m_j$; $j, s = 1, \dots, M$.

3. Верны равенства:

$$\frac{\partial \pi_{jr}}{\partial \rho_{si}} = \frac{\partial \pi_{si}}{\partial \rho_{jr}}, \quad i = 0, \dots, m_s, \quad r = 0, \dots, m_j, \quad j, s = 1, \dots, M.$$

4. Верны равенства:

$$\frac{\partial q_d(\bar{p})}{\partial \rho_{si}} = q_{d-i-\text{kmin}_s}(\bar{p}), \quad s = 1, \dots, M, \quad i = 0, \dots, m_s.$$

Начальный план $\bar{p}[0]$ предлагается выбрать следующим образом: $\bar{p}[0] = \{\bar{p}_j[0], j = 1, \dots, M\}$, где $\bar{p}_j[0] = \{p_{ji}[0], i = 0, \dots, m_j\}$, $p_{ji}[0] = 1$ при $i = m_j^0 - \text{kmin}_j$, $p_{ji}[0] = 0$ при $i \neq m_j^0 - \text{kmin}_j$, где

$$m_j^0 = \min \left\{ \max \left\{ \left[lz_j / (\nu(t_{\text{contrz},j} - \bar{tz}_j)) \right]^-, k_{\min,j} \right\}, k_{\max,j} \right\}, \text{ где } m_j^0 \leq R, j = 1, \dots, M.$$

Для обеспечения сходимости алгоритма решения задачи $\max_{\bar{p} \in S} F(\bar{p}, R)$ рекомендуется использовать градиентный метод с дроблением шага.

Трудоемкость предлагаемого алгоритма решения задачи I определяется трудоемкостью вычислений производных $\partial \Lambda_{jr} / \partial p_{si}$, выполняемых в п. 2 алгоритма. Если при вычислении производных $\partial \Lambda_{jr} / \partial p_{si}$ ($i = 0, \dots, m_s, r = 0, \dots, m_j, j, s = 1, \dots, M$) с помощью выражения (5) использовать метод Гаусса, то объем элементарных операций (арифметических или операций сравнения), выполняемых алгоритмом на шаге 2 для коррекции плана $\bar{p}[n]$, в случае без повторов ($\beta_j = 0, j = 1, \dots, M$) равен величине порядка $\approx O(N^3)$, в случае повторов $\approx O(N^4)$, где $N = \sum_{j=1}^M m_j$.

4 Пример вычислительного эксперимента

Для проведения расчетов с использованием описанного в предыдущем разделе алгоритма разработана экспериментальная программа на языке Delphi 7, в которой объем используемых массивов чисел равен $(2N^2 + 12N + 3M + 3R)$, объем элементарных операций $\approx O(N^4)$. В качестве примера рассматривается ВК со следующими исходными данными: $\nu = 10/3, \varepsilon = 0,0001, M = 5, lz_j = 2j, tz_j = 0, lp_j = 2, \alpha_j = 0,0199, j = 1, \dots, 5$. В таблице приведены результаты решения задачи при $F(\bar{p}) = \Lambda_{\text{ВЫХ}}(\bar{p}), k_{\min,j} = 1, k_{\max,j} = j$ для различных значений параметров $\lambda_j, \beta_j, t_{\text{contrz},j}, j = 1, \dots, 5$.

В таблице вектор \bar{p}' — вектор \bar{p} , у которого $p_{ji} = 1$ при $(j, i) = (0, 0), (1, 1), (2, 2), (3, 3), (4, 4)$ и $p_{ji} = 0$ при остальных (j, i) ; \bar{p}'' — вектор \bar{p} , у которого $p_{ji} = 1$ при $(j, i) = (0, 0), (1, 0), (2, 1), (3, 1), (4, 2)$ и $p_{ji} = 0$ при остальных (j, i) ; \bar{p}''' — вектор \bar{p} , у которого $p_{ji} = 1$ при $(j, i) = (0, 0), (1, 1), (2, 1), (3, 2), (4, 4)$ и $p_{ji} = 0$ при остальных (j, i) .

Проведенные с использованием экспериментальной программы вычислительные эксперименты показали, что:

- решение задачи $\max_{\bar{p} \in S} F(\bar{p}, R)$ при $A_F / \Lambda_{\text{ВЫХ}}(\bar{p}) \approx 1$ слабо зависит от интенсивностей $\lambda_j, j = 1, \dots, M$;
- при $A_F / \Lambda_{\text{ВЫХ}}(\bar{p}) \approx 1$ зависимость решения задачи I от интенсивностей $\lambda_j, j = 1, \dots, M$, при фиксированных значениях других параметров ВК близка к линейной;
- решение задачи I существенно зависит от значения параметра $t_{\text{contrz},j}, j = 1, \dots, M$.

Зависимость объема ресурсов от входной нагрузки и параметра контрольного времени

λ_j	β_j	$t_{\text{contrz},j}$	$A_F/\Lambda_{\text{ВЫХ}}(\bar{p})$	R_0	$\Lambda_{\text{ВЫХ}}(\bar{p})$	\bar{p}
0,5	0	1,0	0,9	13	2,29	\bar{p}'
	0	1,8	0,9	9	2,31	\bar{p}''
	0,8	1,0	0,85	12	2,13	\bar{p}'''
1,0	0	1,0	0,9	18	4,5	\bar{p}'
	0	1,8	0,9	13	4,55	\bar{p}''
	0,8	1,0	0,85	18	4,26	\bar{p}'''
1,5	0	1,0	0,9	24	6,83	\bar{p}'
	0	1,8	0,9	17	6,8	\bar{p}''
	0,8	1,0	0,85	23	6,39	\bar{p}'''
2,0	0	1,0	0,9	29	9,09	\bar{p}'
	0	1,8	0,9	21	9,07	\bar{p}''
	0,8	1,0	0,85	28	8,52	\bar{p}'''
2,5	0	1,0	0,9	34	11,6	\bar{p}'
	0	1,8	0,9	25	11,36	\bar{p}''
	0,8	1,0	0,85	33	10,66	\bar{p}'''
3,0	0	1,0	0,9	38	13,54	\bar{p}'
	0	1,8	0,9	29	13,5	\bar{p}''
	0,8	1,0	0,85	37	12,76	\bar{p}'''

5 Заключение

Предложенный выше алгоритм решения задачи I не гарантирует получения наименьшего значения параметра R , при котором выполняются условия задачи. Это следует из того, что метод градиентного спуска не гарантирует глобального оптимума для любой функции и, как показывают вычислительные эксперименты, $F(\bar{p}, R)$ — немонотонная функция по R . Следует также заметить, что в случае больших значений N при вычислении с помощью выражения (5) производных $\partial \Lambda_{jr} / \partial p_{si}$ может накапливаться недопустимо большая вычислительная погрешность, что может сделать невозможной оптимизацию плана распределения ресурсов в п. 2 алгоритма. Поэтому в данном случае для вычисления указанных производных необходимо использовать методы, обеспечивающие требуемую точность решения соответствующих систем линейных уравнений [5].

Разработанный алгоритм рекомендуется для оценки минимального объема ВР, необходимого для обеспечения требуемого качества работы (интенсивности полностью выполненных заданий, среднего числа повторных поступлений задания) на стадии проектирования вычислительных систем со слабосвязанной архитектурой.

Литература

1. *Агаларов М. Я.* Аналитическая модель расчета эффективности плана распределения вычислительных ресурсов многопроцессорной системы при решении специального класса задач // Информатика и её применения, 2011. Т. 5. Вып. 4. С. 29–35.
2. *Степанов С. Н.* Основы телетрафика мультисервисных сетей. — М.: Эко-Трендз, 2010.
3. *Iversen V. B.* Teletraffic engineering and network planning. DTU Course 34340. — Technical University of Denmark. Revised January 24, 2011. <http://oldwww.com.dtu.dk/education/34340/telenook.pdf>.
4. *Ross K. W.* Multiservice loss models for broadband telecommunication networks. — London – Berlin – New York: Springer-Verlag, 1995. 343 p.
5. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. — М.: Наука, 1974.

ВИРТУАЛЬНЫЕ МАШИНЫ КАК СРЕДСТВО ОПТИМИЗАЦИИ УЧЕБНОГО ПРОЦЕССА

В. С. Петрухин¹, А. Ю. Шнейдер², С. А. Лучкин³

Аннотация: Рассматриваются возможные пути использования виртуальных машин (ВМ) в учебном процессе. Приведены конкретные примеры логической структуры компьютерного класса, построенного на основе ВМ.

Ключевые слова: оптимизация; виртуализация

Широко распространено мнение, что оптимизация — это максимум результата при минимуме затрат. Если трактовать более точно это понятие [1] с точки зрения образовательного процесса, то оно означает достижение максимально возможного качества обучения в конкретных условиях определенного учебного заведения в определенный период. Кроме того, оптимизируя учебный процесс, следует четко формулировать поставленную цель и определять способы ее достижения.

Решение этой непростой задачи предполагает сочетание традиционных методов обучения и информационно-коммуникационных технологий (ИКТ) [2].

В настоящее время овладение ИКТ является первостепенной задачей для учителя, стремящегося обеспечить высокое качество обучения. Многие учителя осознали достоинства ИКТ, почувствовали необходимость воплощения своих идей в виде конкретных учебных и методических пособий. Однако на пути освоения и внедрения ИКТ в учебный процесс школы обнаруживается ряд сдерживающих факторов [3]:

- для эффективного внедрения ИКТ в учебный процесс учитель должен обладать необходимыми знаниями в области компьютерных технологий;
- наличие компьютерного класса в школе предполагает организацию процесса администрирования и обслуживания технических средств, что является дополнительной нагрузкой (особенно для учителей информатики).

Для преодоления вышеизложенных сдерживающих факторов необходимо снизить нагрузку на учителя и направить его основное внимание на организацию процесса обучения. В качестве одного из путей решения предлагается использование технологии виртуализации, которая стала неотъемлемой частью

¹Институт проблем информатики Российской академии наук, cokrat2@rambler.ru

²Институт проблем информатики Российской академии наук, alexshnd@rambler.ru

³Средняя школа № 4, г. Кольчугино, info32kab@gmail.com

современной ИТ-инфраструктуры и позволяет существенно ускорить внедрение новых систем, а также уменьшить затраты на поддержку существующих [4]. Под виртуализацией здесь подразумевается технология, которая обеспечивает возможность размещения на одном физическом компьютере (физической машине) нескольких ВМ, работающих независимо друг от друга.

Известен успешный пример [5] использования ВМ в качестве средства обучения будущих учителей информатики сетевым технологиям. Отмечено, что как средство обучения ВМ могут выполнять такие дидактические функции, как информативность, компенсаторность, интегративность.

В Южно-Уральском государственном университете (ЮУрГУ) [6] завершен очередной этап проекта по созданию инфраструктуры виртуальных рабочих мест (VDI — Virtual Desktop Infrastructure) на базе Microsoft Hyper-V Server. Данный проект является частью программы развития ЮУрГУ на 2010–2019 гг.

Чем привлекательна технология ВМ для сферы образования? В широком смысле понятие виртуализации представляет собой замену физической реализации какого-либо процесса или объекта его моделью. В компьютерных технологиях под термином «виртуализация» обычно понимается абстракция вычислительных ресурсов и предоставление пользователю упрощенной стандартизированной системы (ВМ), которая видна пользователю как реальный компьютер со своими ресурсами и операционной системой (ОС). Проще говоря, ученики и учителя работают с удобным для себя представлением компьютера, свободным от массы ненужных и мешающих им специфических технических деталей.

Таким образом, учитель избавляется от необходимости изучения тонкостей установки и настройки ОС, приложений и драйверов устройств.

Общая схема (архитектура) системы ВМ следующая:

- на физическом компьютере устанавливается ОС, которая называется хостовой (например, Windows, Linux и т. п.);
- в этой ОС устанавливается специальное приложение «гипервизор», которое осуществляет управление ВМ;
- под управлением гипервизора создается одна или несколько ВМ;
- каждая ВМ представляет собой упрощенную стандартизованную модель физического компьютера.

Далее на каждую ВМ устанавливается своя (гостевая) ОС. В качестве гостевой может использоваться любая ОС.

Гипервизор обеспечивает одновременную работу ВМ как отдельных физических компьютеров. При этом пользователи могут работать с ВМ, не обращая внимания на существующий физический компьютер и хостовую ОС.

Использование ВМ значительно повышает мобильность и оперативность работы. Она может быть перемещена на другой компьютер независимо от его

физической конфигурации и установленной на нем ОС (единственным необходимым требованием является достаточный объем оперативной памяти компьютера). Не требуется создавать специальных образов для миграции, и, кроме того, ВМ не привязана к конкретной аппаратуре. На любом компьютере ВМ всегда выглядит одинаково. Учитель, один раз создав ВМ, может переносить ее оперативно со своего компьютера на компьютеры школы.

Можно создавать виртуальное окружение для конкретного урока (например, для урока физики, математики и т. п.), установив на ней все требуемое программное обеспечение (ПО), и разворачивать ВМ по мере необходимости.

Можно создать базу данных из готовых к использованию ВМ с различными гостевыми ОС и запускать их по мере необходимости в целях обучения. ВМ можно без катастрофических последствий подвергать всяческим экспериментам, поскольку в случае порчи системы ее восстановление из сохраненного состояния займет несколько минут.

При установке новой версии ОС старую версию можно продолжать поддерживать на ВМ, пока не будет полностью обкатана новая ОС. Также можно поставить новую ОС на ВМ и опробовать ее без ущерба для основной системы (хостовой).

Существенная экономия денежных средств на приобретение аппаратного обеспечения происходит при размещении нескольких виртуальных серверов на одном физическом. В зависимости от платформы виртуализации доступны возможности по балансировке рабочей нагрузки, контролю выделяемых ресурсов и восстановлению системы. Все это влечет за собой реальную экономию денежных средств на обслуживании, управлении и администрировании инфраструктуры серверов.

Если у учителя возникает сомнение в надежности устанавливаемого на компьютер программного продукта, можно использовать его на ВМ и таким образом исключить возможность повреждения жизненно важных компонентов системы.

При использовании ВМ существенно повышается оперативность в отношении создания резервных копий, создания снимков состояний ВМ и восстановления после сбоев или вирусной атаки. Это позволяет в условиях школы обеспечить восстановление работоспособности компьютерного класса за время перемены.

Вместе с тем данной технологии свойственны следующие недостатки:

- некоторое снижение производительности ВМ по сравнению с физическим компьютером (за счет работы гипервизора);
- для формирования ВМ необходим дополнительный объем оперативной памяти и дискового пространства.

Далее рассмотрим конкретные варианты реализации компьютерного класса в школе на основе приложения (гипервизора) под названием VirtualBox, которое входит в состав практически всех пакетов, предназначенных для школ.

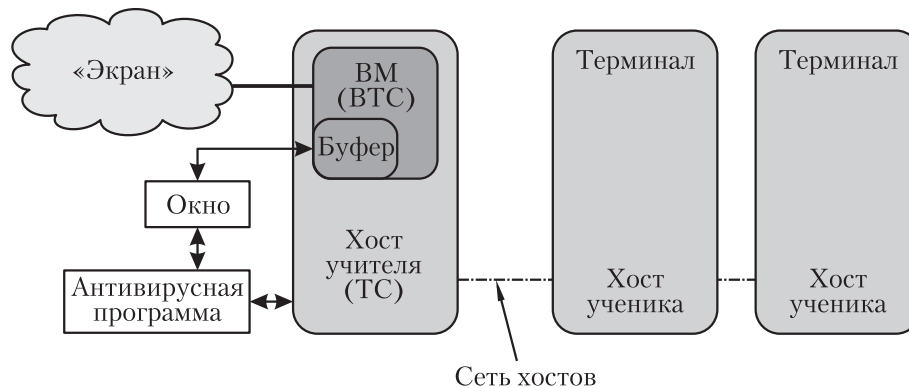


Рис. 1 Логическая схема платформы компьютерного класса на основе ТС

Внедрение технологии ВМ осуществлялось в школе № 4 г. Кольчугино и Завалинской сельской школе. К процессу обучения были подключены учителя информатики этих школ, а также группа из учеников 9 класса в количестве 4 чел. Как показала практика, для освоения данной технологии достаточно было провести 2 занятия по 45 мин, на которых рассматривались следующие вопросы:

- принципы организации ВМ;
- практические занятия по формированию ВМ в приложении VirtualBox.

Следует отметить, что ученики с интересом и увлечением работали над внедрением технологии в учебный процесс. Они под руководством учителя информатики непосредственно формировали и тестировали платформы компьютерных классов в школе.

На рис. 1 изображена логическая схема платформы компьютерного класса, построенная на основе терминального сервера (ТС) (школьного терминала). Терминальный сервер входит в состав ПО АльтЛинукс 5.0.2 и Edubuntu 11.04. Как правило, ТС устанавливается на хост компьютера учителя, выполняющего роль шлюза в сетевой структуре школы. Если Интернет не используется для проведения урока, можно осуществлять учебный процесс в компьютерном классе и администрирование непосредственно с помощью ТС.

Если урок осуществляется с использованием Интернета, то велика вероятность заражения вирусами ТС, а значит, предстоит очень трудоемкая и длительная процедура поиска и уничтожения вирусов. При этом вирусы не проникают на компьютеры учеников (в случае реализации тонких клиентов без жестких дисков).

Традиционная технология защиты от вирусов основывается на использовании специального ПО. Существенным фактором в этом подходе являются

значительные затраты вычислительных и временных ресурсов для его осуществления. Предлагается использовать подход, базирующийся на великолепных свойствах ВМ:

- ВМ в достаточной степени изолирована от хоста;
- перенос ВМ осуществляется с использованием образа жесткого диска (один файл);
- процедура восстановления состояния базируется на встроенном в VirtualBox механизме снимков ВМ.

Интернет подключается к компьютерам учеников через ВМ (в данном случае через виртуальный ТС (ВТС)). Перед загрузкой тонких клиентов учитель делает снимок состояния ВТС. Эта опция включена в состав VirtualBox и действие осуществляется в течение одной минуты. После запуска сессии загружаются тонкие клиенты на компьютерах учеников. По окончании урока учитель, используя снимок, осуществляет восстановление (уничтожение вирусов) ВТС в течение одной минуты. Если в процессе урока необходимо запомнить скачанную из Интернета информацию на внешнем носителе или в специальной папке хоста (буфере), то в процесс включается традиционная технология с использованием антивирусного ПО. Антивирусное ПО сканирует перемещаемые файлы. Далее учитель открывает совместный доступ к буферу (окно) и перемещает файлы на хост. По окончании процесса совместный доступ к буферу блокируется.

Представленная платформа компьютерного класса позволяет расширить границы для организации урока:

- (1) учитель может использовать в процессе проведения урока и загружать различные варианты реализации ВТС, которые могут отличаться, например, по составу ПО и наполнению информацией;
- (2) формирование ВТС учитель может осуществлять в домашних условиях, не загружая компьютерный класс.

Следует учитывать, что производительность данной платформы зависит от производительности компьютера учителя, на котором работает ТС, и от скорости канала обмена между ТС и компьютерами учеников.

Предложенная платформа компьютерного класса наиболее эффективно может применяться в сельских школах, которые используют в составе класса устаревшие компьютеры, а также в условиях отсутствия квалифицированных системных администраторов.

Оптимальным решением для класса, состоящего из более чем 12 компьютеров, является использование ТС, содержащего как минимум двухъядерные процессоры и канал связи с пропускной способностью в 1 Гбит/с. На рис. 2 изображена логическая схема платформы компьютерного класса, реализованная на основе виртуальной рабочей станции (например, школьный Юниор АльтЛинукс). Как

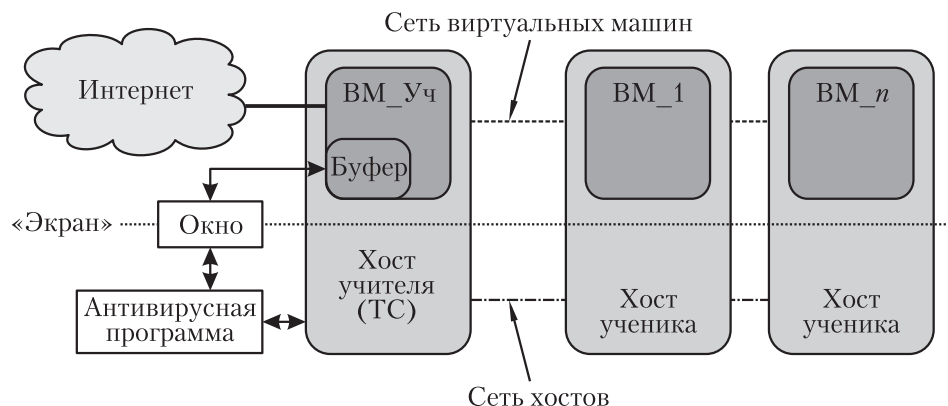


Рис. 2 Логическая схема платформы компьютерного класса на основе виртуальной рабочей станции

правило, на компьютер учителя устанавливается хостовая ОС, гипервизор и ВМ, выполненная на основе виртуальной рабочей станции (ВМ_Уч). На компьютеры учеников устанавливается хостовая ОС, гипервизор и ВМ. В логической схеме компьютерного класса присутствуют две виртуальные сети, отделенные друг от друга.

В состав первой сети входит виртуальная рабочая станция и ВМ на компьютерах учеников. К этой сети подсоединен Интернет через виртуальную рабочую станцию. Таким образом учебный процесс погружается в виртуальную среду.

Вторая сеть состоит из хоста компьютера учителя и хостов компьютеров учеников, и на ней осуществляется процесс администрирования сети и управления уроком.

Вначале запускается ОС на хостах учеников и учителя. Перед запуском ВМ учитель делает их снимок (записывает образ на жесткий диск в течение одной минуты) на всех компьютерах класса. Далее на компьютерах учеников запускаются ВМ, а на компьютере учителя виртуальная рабочая станция.

По окончании урока учитель, используя снимок ВМ, осуществляет восстановление (уничтожение вирусов) ВМ на всех компьютерах класса в течение одной минуты. Если в процессе урока необходимо запомнить скачанную из Интернета информацию на внешнем носителе или в специальной папке, то в процесс включается традиционная технология с использованием антивирусного ПО.

Данная платформа отличается от изображенной на рис. 1 более широкими возможностями в части организации урока и повышенными требованиями к ресурсам компьютеров класса.

Можно сформировать ВМ персонально для каждого ученика, что даст возможность учитывать индивидуальные особенности учащихся и классов школы.

Есть возможность существенно облегчить работу учителей, если, например, при управлении по образованию создать рабочую группу, осуществляющую формирование ВМ под заказ.

Приложение VirtualBox можно использовать в качестве инструмента для моделирования урока, что позволит отработать методику и форму учебного процесса. Для этого необходимо построить сетевую модель конкретного компьютерного класса.

Таким образом, использование ВМ в учебном процессе позволяет минимизировать затраты на подготовку и организацию уроков, обслуживание компьютерных классов, что в конечном счете оказывает существенное влияние на оптимизацию учебного процесса.

Для сельских школ наиболее востребованной оказалась платформа компьютерного класса на основе ВТС (обслуживание только компьютера учителя и минимальные требования к ресурсам компьютеров учеников).

Литература

1. *Бабанский Ю. К.* Оптимизация процесса обучения. — М.: Педагогика, 1977. 250 с.
2. *Морозов И. А.* Информационные технологии в оптимизации учебного процесса. <http://usinskaya.ucoz.ru/load/1-1-0-3>.
3. *Беляков Е. В.* Понятие информационно-коммуникационных технологий (ИКТ) и их роль в образовательном процессе. http://belyk5.narod.ru/ИКТ_new.htm.
4. *Гультяев А. К.* Виртуальные машины: несколько компьютеров в одном. — СПб.: Питер, 2006. 224 с.
5. *Рыжова Н. И., Королева Н. Ю., Ляш О. И.* Виртуальные машины как средство обучения будущих учителей информатики сетевым технологиям // Информатика и образование, 2007. № 10. С. 97–100.
6. Технологии виртуализации позволят студентам ЮУрГУ обучаться из любой точки мира с любого устройства. <http://www.npsod.ru/rus2/releases/document32173.phtml>.

ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ: ЭФФЕКТИВНОСТЬ ПРИМЕНЕНИЯ

*А. А. Федосеев*¹

Аннотация: На основе анализа процессов научения, учения и обучения выделено информационное содержание учебного процесса и построена его информационная модель. Определено место в учебном процессе желательного приложения усилий по созданию электронных образовательных ресурсов (ЭОР). Определены желательные свойства ресурсов для обеспечения их положительного влияния на эффективность учебного процесса.

Ключевые слова: учение; обучение; мотивация; восприятие; закрепление; учебный процесс; информационная модель; электронный образовательный ресурс

1 Введение

В области применения информационно-коммуникационных технологий (ИКТ) в образовании сложилась следующая картина: на протяжении десятилетий в эту область вкладываются серьезные средства, школы оснащены современной вычислительной техникой, подключены к Интернету. Существуют и доступны для применения в школах или в индивидуальном порядке десятки и сотни тысяч ЭОР. Каждый год к ним прибавляются тысячи новых. Опережающими темпами создаются методические работы по их применению. Учителя активно обмениваются опытом по использованию ЭОР. Параллельно развиваются и совершенствуются комплексы требований, которым должны удовлетворять создаваемые ЭОР, в том числе по критериям актуальности содержания, дидактической проработанности, научности, принимаются стандарты, разрабатываются механизмы обеспечения качества ЭОР. Тем не менее всего этого оказывается недостаточно: существенного положительного эффекта от использования средств ИКТ не наблюдается даже в простейшем случае — при использовании критерия лучшего освоения материала учащимися за то же время или такого же уровня освоения за меньшее время. Надо полагать, что при обнаружении заметного влияния применения ЭОР на эффективность учебного процесса в школах, об этом немедленно стало бы известно. Но таких публикаций пока не появилось. Напротив, отсутствие эффекта отмечается как у нас в стране, так и за рубежом [1, 2].

¹Институт проблем информатики Российской академии наук, A.Fedoseev@ipiran.ru

В статье сделана попытка определить, следует ли вообще ожидать положительного эффекта от существующих и продолжающих множиться в огромных количествах ЭОР. Также показано, какими свойствами должны обладать ЭОР, чтобы можно было ожидать эффекта от их применения.

2 Теоретические основы

Педагогическая психология [3] рассматривает приобретаемые в результате учебного процесса знания, умения и навыки учащихся как их индивидуальный опыт. Способность к приобретению индивидуального опыта является общим свойством биологических систем. Процесс приобретения опыта называется научением. Человеку, в отличие от других видов биологических систем, свойственно целенаправленно приобретать свой индивидуальный опыт. Этот процесс, в отличие от бессознательного научения, носит название учения. На рис. 1 [4] перечислены виды деятельности, в результате которых приобретается опыт. Учение охватывает те «деятельности», которые находятся в столбце «преднамеренное познание». Все остальные ветви схемы отображают процесс научения.

Роль учителя в процессах учения и научения (для учащихся младшего возраста) заключается в организации передачи социокультурного опыта учащимся и тех видов деятельности, которые позволят учащимся приобрести этот опыт уже в виде своего индивидуального опыта. В этих процессах учитель занимается обучением. Таким образом, учитель организует передачу опыта учащимся и в зависимости от их возрастных, интеллектуальных и психологических особенностей выбирает наиболее подходящие для этого виды деятельности. Обратим внимание, что среди видов деятельности, приводящих к присвоению опыта, особенно выделены такие виды, как труд, игра и стихийное общение, а также стихийное наблюдение, чтение, просмотр фильмов и др. Здесь важно, что, во-первых, научение есть результат активности самого учащегося и, во-вторых, научение происходит в результате деятельности учащегося. Таким образом, какие бы действия ни предпринимал учитель, пока он не организовал, не запустил, не включил собственную активность и деятельность учащегося, никакого процесса научения произойти не может. И тем более не может произойти процесса учения как осознанной деятельности.

Когда мы говорим, что учитель обучает, мы имеем в виду, что он передает учащимся учебную информацию, задает им вопросы, чтобы убедиться, что информация воспринята, задает задания, чтобы закрепить полученные учащимися знания. Но все это возможно только тогда, когда сопровождается собственной деятельностью учащихся, направленной на приобретение нового индивидуального опыта, соответствующего информации, переданной учителем. Поскольку передавать информацию, а также задавать задания учитель может не непосредственно, а с использованием средств ИКТ, смысл его деятельности перемещается

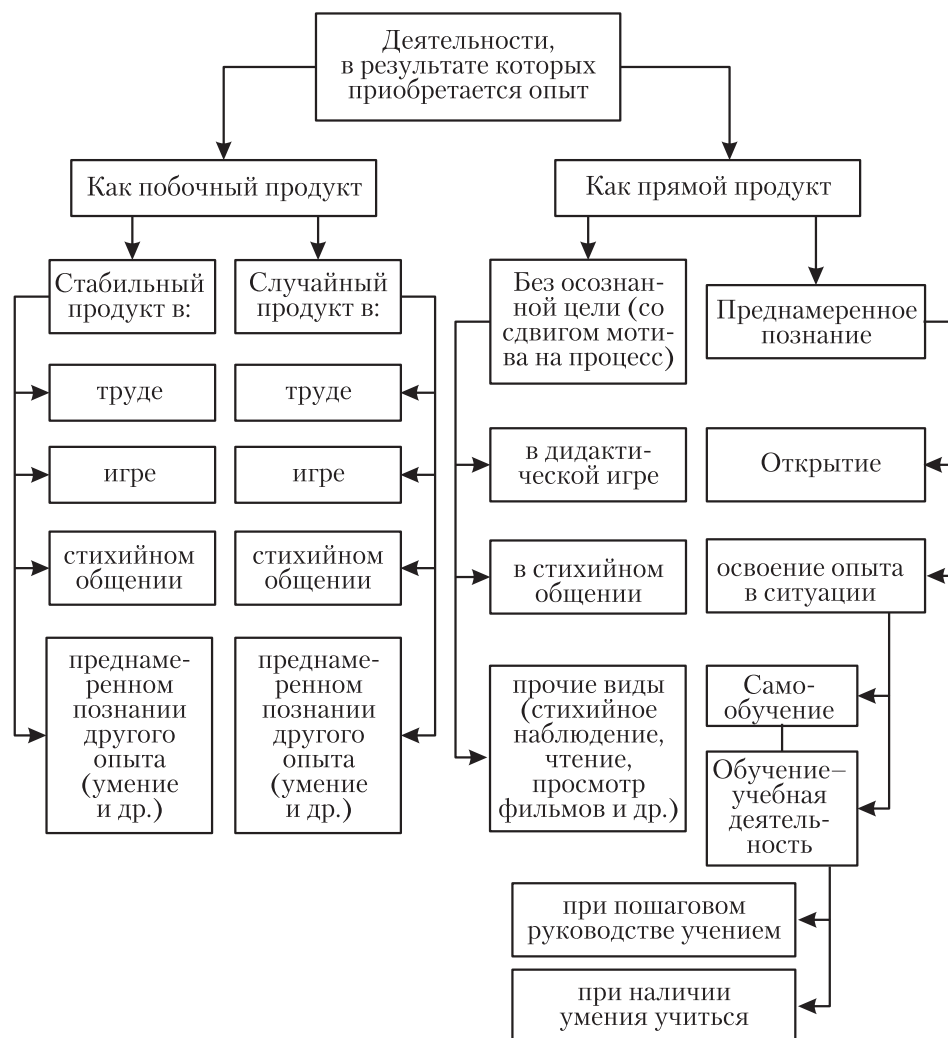


Рис. 1 Виды деятельности, в результате которых человек приобретает опыт

на обеспечение познавательной активности учащихся. Если эта активность не включена, грубо говоря, учащийся витает в облаках и не собирается ничего познавать, то учитель не в состоянии ничему такого ученика научить. Процесс учения может идти только на основе собственной деятельности учащегося и никак иначе.

Согласно [5] «принципиальная логика современного обучения состоит в том, что она неизбежно начинается с предъявления учащимся разными способами

(обязательно с помощью материальных предметов или знаков, вербально и т. д.) новой информации, продолжается организованным репродуцированием (в разных вариантах по видам и сложности) информации и связанных с ней действий». Следовательно, если говорить о структуре деятельности учащегося в процессе учения (научения), то он, во-первых, воспринимает предъявляемое и, во-вторых, воспроизводит или копирует (репродуцирует) его. При этом воспроизведение является учебной деятельностью по закреплению предъявляемого материала: как сказано в [6], «закрепление и совершенствование знаний и привитие умений и навыков». Изложенное позволяет сформулировать два элемента структуры учебной деятельности: восприятие и закрепление.

Учитывая упомянутую выше необходимость собственной деятельности для реализации процесса учения, уже выделенных элементов — восприятия и закрепления — недостаточно. В самом деле, далее в [6] сказано: «роль педагога заключается в том, чтобы вызвать у школьников собственную потребность овладеть новым понятием или законом науки, способами применения их на практике». Таким образом, в дополнение к выделенным элементам структуры учебной деятельности следует добавить особый вид деятельности: выработку учащимся потребности к учению. Другими словами, речь идет о мотивации. Можно указать работы (например, [7]), в которых мотивация выделяется как самостоятельный элемент структуры учебной деятельности. Окончательно имеем: структура учебной деятельности учащегося представляет собой последовательность следующих собственных действий:

- мотивация к предстоящему этапу (теме) процесса учения;
- восприятие учебного материала по теме;
- закрепление материала по теме.

Соответствующие этим проявлениям активности учащихся действия учителя выглядят следующим образом:

- запуск мотивации к предстоящему этапу (теме) процесса учения;
- предоставление, передача учебного материала по теме;
- руководство процессом закрепления материала по теме.

Таким образом, роль учителя заключается в организации и обеспечении всех элементов процесса. Научить без соответствующей активности самого учащегося учитель ничему не может. Однако он может (и должен) запустить процесс мотивации учащихся, передать непосредственно или опосредованно новый материал и обеспечить все необходимое для закрепления материала.

Может показаться, что автор упустил такой важный элемент структуры учебной деятельности, как контроль. В этом обвинении есть резон, поскольку, во-первых, в большинстве педагогических произведений необходимость контроля после закрепления предусмотрена и, во-вторых, всем известны такие виды учебной деятельности, как контрольная работа, тестирование, экзамен. Тем не

менее автор сознательно опускает этот элемент структуры по причинам, которые станут понятны позже. Здесь же будет уместным следующий комментарий: когда говорится об учебной деятельности, подразумеваются те виды деятельности, которые приводят к появлению нового индивидуального опыта в виде знаний, умений и навыков. Опыт появляется в результате восприятия и закрепления. Чтобы восприятие и закрепление состоялись, должна быть осуществлена мотивация собственной познавательной деятельности. Контроль не является видом деятельности, в результате которой появляется новый опыт. Контроль — это такое мероприятие, которое делает наличие (или отсутствие) нового опыта у учащихся очевидным для учителей, родителей, школы, государства. Отметим, что необходимость контроля возникает из ситуации, когда в учебной деятельности предполагается возможность (и даже неизбежность) брака, именно для обнаружения этого брака.

Рассмотрим теперь более подробно выделенные элементы структуры учебной деятельности.

3 Мотивация

У каждого человека (и учащегося в том числе) существует и действует собственная многофакторная система мотивации. В зависимости от вида предпринимаемой деятельности преобладают различные мотивационные факторы. При необходимости эти факторы могут быть усилены или ослаблены специальным воздействием. Учитель имеет возможность хорошо узнать своих учеников, чтобы индивидуально и эффективно воздействовать на них. Вместе с тем существуют общие для всех закономерности. Эти закономерности лучше всего изучены в области маркетинга и рекламы, где десятилетиями изучаются способы воздействия на человека, побуждающие его произвести определенные действия. Единственное отличие от коммерческой рекламы заключается в том, что учащийся побуждается к приобретению не товаров и услуг, а индивидуального опыта в виде знаний, умений и навыков. Следовательно, мотивационные материалы должны и могут создаваться с использованием теоретических разработок в области маркетинга и рекламы.

Вместе с тем педагогической наукой накоплены собственные методы мотивации учащихся. Одним из самых эффективных по праву считается проектный метод, позволяющий выстроить деятельность учащихся таким образом, чтобы достижение ими значимых для них целей было обусловлено приобретением определенных знаний, умений и навыков в соответствии с программами обучения. В настоящее время в педагогической литературе существует множество подробно описанных проектов.

Как минимум учащемуся должна быть передана, а им воспринята следующая информация о предмете целиком или отдельной теме (группы тем) в предмете:

- место предмета в системе знаний и культуре или темы в предмете. Как сказал Ян Коменский [8], «при образовании юношества все нужно делать как можно более отчетливо, так, чтобы не только учащий, но и учащийся понимал без всякого затруднения, где он находится и что он делает»;
- значение предмета в системе знаний и культуре, в том числе значение именно для учащихся определенного возраста, в соответствии с их интересами, или значение темы или группы тем в изучаемом предмете;
- преимущества, которыми обладает человек, овладевший предметом (темой) (и снова Коменский [8]: «Ты облегчишь ученику усвоение, если во всем, чему бы ты его ни учил, покажешь ему, какую это может принести повседневную пользу в общежитии»).

Наличие мотивации у учащихся проявляется в возникновении вопросов, осуществлении попыток найти дополнительный материал или какой-либо иной активности по изучаемой теме (предмету). Эти действия учащихся являются сигналами учителю о возникновении мотивации.

4 Восприятие

Учащийся воспринимает ту информацию, которую непосредственно или опосредованно передает учитель. При этом речь может идти об осознанном восприятии, и тогда это — процесс учения, или о восприятии неосознанном, что характеризует процесс научения. Выбор между этими двумя процессами, а следовательно, и выбор стиля подачи и содержания материала осуществляет учитель, принимая во внимание возрастные, психологические и интеллектуальные особенности учащихся.

В тех случаях, когда требуется освоение действий: движений, поз, произнесения слов и выражений (в том числе на других языках), произнесения слов и выражений с упором на интонацию или акцент, оперирования предметами, в том числе: инструментами, спортивными снарядами, средствами письма и рисования, рисования линий и изображений, написания букв, написания слов и т. п., — передача информации заключается в демонстрации действия, в том числе, при необходимости, замедленно, в деталях, с обеспечением повтора необходимое число раз. Словесные (при непосредственной передаче информации учителем) или текстовые комментарии являются в этом случае второстепенными.

В ситуациях, когда информация должна наполнить индивидуальный опыт учащихся в виде нового знания, в зависимости от того, кто является инициатором нового опыта — учитель или сам ученик, — процесс восприятия будет разным. Если информацию передает учитель в виде своей речи, предлагаемого текста или мультимедийного продукта, то учащийся воспринимает то, что передает учитель. Если же у учащегося есть инициативные вопросы, на которые следует найти

ответы, или гипотезы, которые следует проверить, то никакой заранее подготовленный материал не может решить проблему. Учащемуся следует предоставить возможность искать ответы на свои вопросы. В качестве источников знания в данном случае может выступить эрудиция учителя или других людей, библиотека и, разумеется, Интернет. Что касается проверки гипотез, то предоставление возможности для экспериментирования в виде соответствующей среды и времени позволяет обеспечить этот вид приобретения знаний. Далеко не всегда для создания среды для экспериментирования требуются серьезные затраты. Например, дети постоянно экспериментируют со своей речью и поведением, чтобы проверить гипотезы относительно допустимости произнесения некоторых слов или осуществления некоторых действий. Другим примером является работа с компьютером. Подавляющее большинство прикладных программ устроено таким образом, что позволяет экспериментировать с ними без фатальных повреждений с целью проверки ожидаемой реакции на не определенные заранее управляющие воздействия.

Необходимыми элементами этапа восприятия являются входная проверка наличия у учащихся знаний, умений и навыков, необходимых для восприятия нового материала, и обратная связь о фактическом восприятии нового материала. Очевидно, что материал не может быть воспринят и усвоен, если он опирается на отсутствующие у учащегося элементы знаний. Столь же очевидно, что выслушанные объяснения учителя не обязательно были услышаны, прочитанный текст мог не оставить о себе никаких воспоминаний, а просмотренный видеоклип не породил никаких образов и выводов в сознании учащегося. Только когда есть убежденность в том, что материал воспринят, его закрепление (следующий этап учебной деятельности) может быть эффективным.

5 Закрепление

Каким образом закрепляются новые знания, умения и навыки?

В случае, когда требуется закрепление действий, разумным механизмом является фиксация и отображение действия учащегося таким же или похожим образом, как была сделана демонстрация. Это позволит очевидным образом сопоставить действия учащегося с эталонными (демонстрируемыми) действиями. Величина отличия производимого действия от эталонного позволяет сделать заключение о достигнутой или нет правильности выполнения действия. Последующая серия повторений действия, возможно, в разбивку с другими действиями, с перерывами в осуществлении закрепляемого действия позволит убедиться в действительном его закреплении.

Новое знание, содержащееся в результатах исследований и экспериментов, закрепляется написанием соответствующих отчетов о самостоятельной работе, докладами и защитой в классе, а также, возможно, и на межшкольных меро-

приятнях. Сами по себе эти действия являются достаточными для закрепления нового знания, добытого собственными усилиями в ответ на сформулированные самостоятельно вопросы или созданные самостоятельно гипотезы.

Новое, воспринятое из материалов, переданных учителем, знание закрепляется с помощью упражнений и задач, опять-таки заданных учителем. Какими свойствами должен обладать комплект задач (упражнений), чтобы он был способен обеспечить закрепление воспринятого учащимся материала? Во-первых, этот комплект должен быть полным. Ни одна деталь переданного материала, требующая закрепления, не должна быть обойдена задачами и/или упражнениями. Во-вторых, комплект должен содержать все типы задач, связанные с закрепляемым материалом и имеющие значение для дальнейшего обучения. В-третьих, в случаях неправильного решения задач, комплект должен содержать однотипные задачи с различными числовыми данными, с тем чтобы при возникновении необходимости повторного решения задач (например, в случаях неправильного решения) учащемуся не были предъявлены в точности те же задачи. У учителя должны быть в распоряжении комплекты задач различного уровня сложности, чтобы снабжать учащихся задачами соответственно уровню их подготовки или заинтересованности. Каждый случай неправильного решения задачи должен сопровождаться анализом с целью выявления недостаточно понятых деталей материала. После прояснения материала учащемуся предъявляются такие же задачи, но с иными числовыми данными. Когда оказывается, что учащийся свободно и быстро справляется со всеми задачами комплекта, учитель может сделать обоснованный вывод о состоявшемся закреплении.

6 Информационная модель учения

В предыдущем разделе рассмотрена структура учебного процесса с точки зрения информатики: описаны сопровождающие учебный процесс информационные потоки. Оставляя в стороне вопросы, связанные с психологической сущностью процессов восприятия и закрепления, автору удалось сосредоточиться на информационной стороне дела, что позволило сформировать основную информационную модель учения в виде, представленном на рис. 2. При этом предполагается, что все материалы, необходимые для учебного процесса, у учителя имеются, что можно считать достаточно близким к реальности. Поэтому в предлагаемой модели учитель не занимается поиском соответствующего материала, а сразу передает учащимся имеющийся материал. Под материалом здесь имеется в виду любой носитель информации с данными. Это может быть речь (лекция) учителя, текст, в том числе в электронном виде, видео и прочее.

Процесс начинается с передачи учащимся мотивирующего материала. Восприятие этого материала должно ознаменоваться возникновением побуждения учащихся к овладению учебным материалом. Чтобы двигаться дальше, учитель

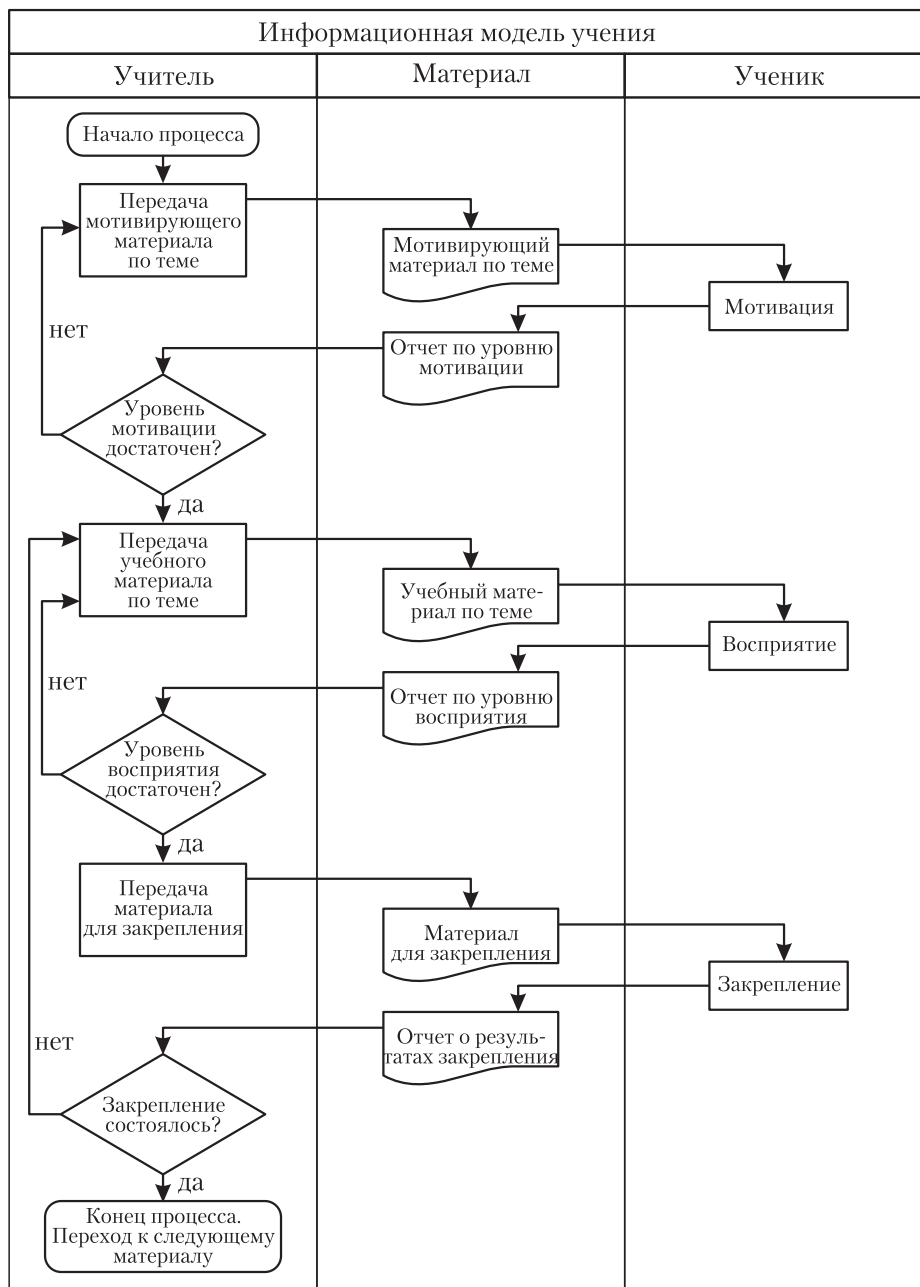


Рис. 2 Схема основной информационной модели учения

должен получить обратную связь об уровне мотивации (на схеме — отчет). Фактически на уроке, видя глаза учащихся и ощущая их возбуждение, учитель такую обратную связь имеет, особенно когда мотивирующий материал заключается в выступлении самого учителя. Учитель также может опросить некоторых учащихся класса, чтобы более точно оценить степень возникшей мотивации. После чего он принимает решение о возможности двигаться дальше. Следует отметить, что, как правило, мотивация оценивается учителем не в виде специального действия, а между делом, не задерживаясь на этом аспекте учебного процесса специально. Возникло у учителя ощущение, что все в порядке, значит можно двигаться дальше — переходить к следующему этапу учебного процесса. Тем более речь не может идти об оценке уровня мотивации каждого из учащихся класса. Следовательно, на этом этапе учебного процесса закладывается запланированный брак: не получив обратной связи о достаточном уровне мотивации всех учеников, учитель переходит к передаче нового учебного материала. Те учащиеся, которые не обрели достаточной мотивации, не будут воспринимать учебный материал с должным усердием и результативностью. Поскольку уровень мотивации всех учащихся не оценен, а недостаточная мотивация не даст возможности некоторым из учащихся воспринять в достаточной степени учебный материал, то, следовательно, неизбежно появление отстающих учеников.

Как бы то ни было, учитель принимает решение передать учащимся новый учебный материал по теме. Здесь, как и в каждом цикле учебного процесса, требуется обратная информация от учащихся. Как правило, учитель получает информацию о восприятии учебного материала методом выборочного опроса. О восприятии не опрошенных учащихся информация учителю не поступает. На этом этапе также возможен учебный брак: из-за появления не выявленных учителем учеников, которые не восприняли учебный материал в должной мере.

Тем не менее учитель принимает решение перейти к закреплению материала и выдает задания. Следует отметить, что, как правило, состав и количество заданий совершенно недостаточны для закрепления той порции знаний, умений и навыков, которая определена учебным материалом. Во-первых, в объеме заданий не предусматривается закрепление всех аспектов и нюансов знаний. Во-вторых, количество заданий определяется исходя из целей контроля, а не из целей закрепления. Для закрепления нужно выполнение большого количества однотипных, но немного отличающихся задач. Для контроля достаточно одной задачи каждого типа.

Результат выполнения заданий обычно фиксируется, что может служить соответствующим отчетом, на основании которого учитель принимает решение о необходимости выполнения дополнительных заданий или о завершении темы. Правильной стратегией выполнения этого этапа информационной модели было бы проведение анализа неверных решений с целью определения тех аспектов

нового знания, которые поняты учащимся неправильно. После чего следовало бы отправить учащегося на повторное изучение этих аспектов учебного материала, а затем снова предложить ему выполнение заданий.

Насколько сложившаяся практика соответствует детально изложенным элементам структуры учебного процесса?

Мотивационные материалы, если и существуют, то в самом зачаточном виде в предисловиях и введениях учебников. Разработчики считают, что внешняя привлекательность ЭОР и насыщенность его мультимедийным контентом повышают мотивацию. Следует, однако, отметить, что это второстепенный фактор. Учитель, когда понимает необходимость мотивационных действий, вынужден самостоятельно придумывать соответствующие материалы. Обратная связь о наличии мотивации отсутствует совершенно. Входная проверка наличия необходимых для работы с ЭОР знаний, как правило, отсутствует. Обратная связь о наличии восприятия осуществляется выборочно. Учитель не в состоянии проверить наличие восприятия всеми учениками класса точно так же, как не в состоянии убедиться в закреплении материала всеми. Учителю приходится самостоятельно собирать или изготавливать материалы для передачи учащимся на этапе восприятия. Учитель самостоятельно подбирает комплекты задач для своих учеников из различных источников. Ни один из источников не заявляет о полноте комплектов задач. Поскольку задача обеспечения полноты комплекта задач даже не ставится, то о полноте закрепления материала говорить не приходится.

Таким образом, очевидными проблемами процесса обучения являются:

- недостаточность или полное отсутствие мотивации;
- отсутствие проверки наличия необходимых знаний перед передачей нового материала;
- недостаточность или полное отсутствие обратной связи по восприятию;
- недостаточность внимания к полноте закрепления.

Позволяют ли многочисленные ЭОР устранить или хотя бы уменьшить названные проблемы? Мотивационных материалов автору найти не удалось. Входной проверки знаний перед изложением материала также не обнаружилось. Соответственно, эти действия перекладываются на учителя. Точно так же, как и в отсутствие ЭОР, учителю приходится искать, подбирать или изготавливать необходимые ему материалы. О полноте комплектов задач в ЭОР предлагается судить учителю. Подсказки, встроенные в задачи, объяснение сложностей, замена числовых данных при повторном предъявлении задач встречаются в отдельных ресурсах, но нигде не встречается все вместе, системно. Таким образом, те ЭОР, которые доступны в настоящее время, в подавляющем большинстве не способны повысить эффективность учебного процесса, что и наблюдается в действительности.

7 Какими должны быть электронные образовательные ресурсы

Попробуем сформулировать необходимые требования к ЭОР, которые помогли бы решить указанные проблемы.

ЭОР, служащие для пробуждения мотивации у учащихся, должны содержать изложение места и значения предназначенного для восприятия материала в изучаемом предмете. По возможности он должен содержать общекультурное значение материала, а также значение этого материала для учащихся соответственно их возрасту и типовым увлечениям. ЭОР должен содержать демонстрацию преимуществ человека, владеющего изучаемым материалом.

Поддача информации должна выстраиваться с использованием рекламных технологий. Для автоматизированной оценки возникшей мотивации должны быть разбросаны гиперссылки на более глубокое освещение материала, а также должны быть предложены пробные интересные и завлекательные задания, выполнить которые с легкостью может человек, освоивший предмет (тему), но представляющие известную трудность для неподготовленных учащихся. По наличию переходов по гиперссылкам и по попыткам выполнить задания можно судить о наличии мотивации. Соответствующий отчет с отображением уровня мотивации должен быть представлен учителю по завершении работы учащегося с ЭОР. На основании этого отчета учитель будет судить о том, достаточна ли возникшая мотивация учащегося и будет ли восприятие им материала успешным.

Разумным требованием было бы предварение мотивационным материалом каждого ЭОР, предназначенного для усвоения какого-либо фрагмента учебной программы. Если мотивационный материал включен в состав ЭОР, то попытки учащихся перейти от демонстрации места и важности материала к непосредственному его изучению также могут послужить сигналом о достижении достаточного уровня мотивации.

В ЭОР, предназначенном для передачи информации по какой-либо теме учащимся, должны присутствовать следующие компоненты:

- проверка наличия необходимых для освоения темы знаний. В ЭОР должны быть предусмотрены краткие справки-напоминания по забытому материалу для предъявления учащимся при непрохождении ими входной проверки по каким-либо заданиям. Должны быть также предусмотрены ссылки на неувоенный материал в более тяжелых случаях;
- сам материал по теме в объеме, необходимом и достаточном в соответствии с действующими нормативами. При этом материал должен, по возможности, подаваться в различном виде: в текстовом, голосовом, образном, схематическом и т. п., чтобы учащиеся могли выбрать предпочтительный для их индивидуальных особенностей восприятия вид подачи материала;
- расширенный и углубленный материал двух-трех уровней со ссылками, отправляющими к еще более углубленному изложению материала. Эти

дополнительные уровни подачи материала, во-первых, способствуют более глубокому, а следовательно, более качественному изучению материала, а во-вторых, служат маркерами фиксации восприятия материала;

- механизм фиксации наличия восприятия материала. Механизм может опираться на контрольные вопросы, использование гиперссылок или других интерактивных действий, анализ времени пребывания на отдельных страницах и любые иные методы измерений.

Отчет по уровню восприятия должен предоставляться учителю по завершении работы учащегося с ЭОР. По этому отчету учитель делает заключение о достаточности уровня восприятия материала и возможности перехода к этапу закрепления.

Ресурс, несущий в себе набор задач на закрепление знания, а также формирование умений и навыков по теме, должен содержать:

- генератор упражнений и задач. Назначение генератора — создавать разные задачи для различных пользователей и для последовательного предъявления однотипных (но разных) задач одному пользователю;
- статистический модуль сопоставления идентификатора учащегося и выданных и решенных задач, включая статистику попыток решения. ЭОР должен узнавать учащегося и вести список предъявленных ему задач, типов неудач с решениями, запрошенной помощи, использованных подсказок и т. п.;
- задачи всех типов, приобретение навыка решения которых необходимо для освоения последующих тем программы, а также для прохождения тестов и экзаменов;
- модуль вариативности задач помимо задач нормальной сложности должен содержать как задачи более сложные, чем это необходимо по программе, так и задачи упрощенные для тех учащихся, которые не в состоянии сразу воспринять задачи требуемого уровня сложности;
- модуль анализа встречающихся сложностей и выдачи соответствующих подсказок. В случаях, когда подсказок оказывается недостаточно, должны быть предусмотрены ссылки на теоретический материал по теме для повторного его усвоения;
- модуль выдачи результата закрепления материала по теме. Этот отчет свидетельствует о закреплении темы учащимся, и, вообще говоря, более не требуется никаких контрольных и тестовых заданий по этой теме.

Устроенный подобным образом ЭОР закрепления одновременно решает задачу контроля закрепления материала учащимся, поскольку завершающий отчет свидетельствует о том, что учащийся приобрел навык решения всех предусмотренных программой типов задач по теме.

8 Заключение

Как было показано выше, использование подавляющего большинства существующих в настоящее время ЭОР не только не приводит к повышению эффективности образования хоть по какой-то шкале, но и принципиально не может этого сделать. Дело в том, что существующие ЭОР за редким исключением практически не изменяют работы учителя. Что же касается работы (учебной деятельности) учащихся, то она также остается без изменений. Выявленные в статье компоненты информационного воздействия на учащегося, способные привести к повышению эффективности обучения, а именно: мотивационные материалы, обратная связь по мотивации, входная проверка знаний, обратная связь по уровню восприятия, полнота комплекта упражнений и задач и обратная связь по закреплению материала, — пока остаются вне пределов внимания авторов и создателей ЭОР. Тем не менее необходимость использования этих информационных компонентов непосредственно вытекает из теоретических положений педагогической психологии и педагогики.

9 Тезаурус

Восприятие — психологический процесс приема информации, получаемой при помощи органов чувств и отражения ее в сознании в виде образов субъективной картины мира. Восприятие является начальным этапом процесса познания — приобретения индивидуального опыта.

Мотивация — внутреннее побуждение к действию.

Научение — процесс и результат приобретения индивидуального опыта биологической системой (от простейших до человека как высшей формы ее организации в условиях Земли).

Обучение — целенаправленная, последовательная передача (трансляция) социокультурного (общественно-исторического) опыта другому человеку в специально созданных условиях.

Учение — научение человека в результате целенаправленного, сознательного присвоения им передаваемого (транслируемого) ему социокультурного (общественно-исторического) опыта и формируемого на этой основе индивидуального опыта. Следовательно, учение рассматривается как разновидность научения.

Литература

1. Христочевский С. А. Почему использование ИКТ в образовании до сих пор не приводит к повышению качества образования // Применение новых технологий в образовании: Мат-лы XXII Междунар. конф. — Троицк: Байтик, 2011. С. 208–209.

2. *Федосеев А. А.* К вопросу о недостаточном использовании электронных образовательных ресурсов // Электронная Казань-2011: Мат-лы III Междунар. научно-практич. конф. — Казань: Юниверсум, 2011. С. 125–128.
3. *Айсмонтас Б. Б.* Педагогическая психология: Электронный учебник. http://www.ido.edu.ru/psychology/pedagogical_psychology/index.html.
4. *Габай Т. В.* Педагогическая психология: Уч. пособие для студентов вузов. — 2-е изд., испр. — М.: Академия, 2005. 240 с.
5. *Лернер И. Я.* Теория современного процесса обучения, ее значение для практики // Советская педагогика, 1989. № 11. С. 10–17.
6. Дидактика средней школы: Некоторые проблемы современной дидактики. <http://didaktica.ru>.
7. *Бордовская Н. В., Реан А. А.* Педагогика: Уч-к для вузов. — СПб.: Питер, 2000. 304 с.
8. *Коменский Я. А., Локк Д., Руссо Ж.-Ж., Песталоцци И. Г.* Педагогическое наследие. — М.: Педагогика, 1989. 416 с.

АНАЛИЗ МЕЖДУНАРОДНЫХ ОБРАЗОВАТЕЛЬНЫХ СТАНДАРТОВ В ОБЛАСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

*В. А. Сухомлин*¹

Аннотация: Целью работы является анализ современного состояния международных стандартов программ подготовки бакалавров и магистров в области информационных технологий (ИТ) (или компьютеринга — computing), а также поиск возможностей для использования наработок мирового сообщества в методическом обеспечении отечественной высшей школы.

Ключевые слова: ИТ-образование; образовательный стандарт; kurikulum; информационные технологии; компьютеринг

1 Введение

Быстрое развитие и распространение средств вычислительной техники и программного обеспечения в 60–70-х гг. XX в. способствовало становлению индустрии ИТ как одного из ведущих секторов мировой экономики. Все это обусловило актуальность задачи массовой подготовки профессиональных кадров в области ИТ, т. е. ИТ-профессионалов. В условиях глобализации экономики большое значение для подготовки востребованных кадров имеет выработка соответствующих международных рекомендаций, обладающих высоким уровнем консенсуса в профессиональной среде и служащих ориентиром для университетов и вузов в соответствующей образовательной деятельности. Такого рода рекомендации должны систематизировать и унифицировать требования практики к выпускникам вузов и к соответствующим образовательным программам, учитывать достижения и тенденции развития предметной области, обобщать лучшую образовательную практику, служить эффективным инструментом построения актуальных образовательных программ, единого образовательного пространства.

Ответственность за решение задачи формирования таких ориентиров-рекомендаций в виде типовых учебных программ или kurikulumов (curriculum) взяли на себя ведущие международные профессиональные организации — Ассоциация компьютерной техники (Association for Computing Machinery, ACM) и

¹Московский государственный университет им. М. В. Ломоносова, sukhomlin@mail.ru

Компьютерное сообщество Института инженеров по электронике и электротехнике (Computer Society of the IEEE, или IEEE-CS), которые ведут эту работу начиная с 60-х гг. XX столетия. Вспомним основные вехи этого процесса [1, 2].

В 1965 г. комитетом по образованию организации ACM был разработан первый проект типовой программы курсов бакалавриата по компьютерным наукам (an undergraduate program in computer science) [3], который после доработки был опубликован в 1968 г. в окончательном виде, получив известность как «Curriculum 68» [4]. Через десять лет, в 1978 г., ACM выпустила новую версию этого документа, известного как «Curriculum 78» [5]. Примерно в таком же плане велась работа и в рамках IEEE-CS по разработке типовых программ подготовки бакалавров компьютерной инженерии (computer engineering).

В 1985 г. ACM и IEEE-CS объединили свои усилия, создав объединенную целевую группу под председательством профессора Питера Деннинга. В 1989 г. эта группа подготовила доклад «Computing as a discipline» [6], в котором формулировались принципы преподавания дисциплины, названной компьютерингом (computing) и объединившей в себе две дисциплины (поддисциплины) — компьютерные науки (computer science) и компьютерную инженерию (computer engineering). В 1991 г. объединенная группа опубликовала новое руководство для подготовки бакалавров по компьютерингу — «Computing Curricula 1991» (CC1991) [7], надолго ставшее, по существу, эталонной моделью для университетов в деле подготовки ИТ-кадров.

В 1998 г. вновь созданная объединенная группа специалистов под эгидой ACM и IEEE-CS приступила к разработке обновленной версии CC — «Computing Curricula 2001» (CC2001) [8]. Разработчикам этого документа уже на стадии анализа стало ясно, что за последнее десятилетие область ИТ претерпела столь значительные изменения — развитие и вширь, и вглубь, названное в документах группы драматическим, что для ее адекватного представления в академическом пространстве необходимо было разработать целую систему куррикулумов. Надо отметить, что к этому времени получили интенсивное развитие и сформировались в полноценные самостоятельные ветви ИТ-образования такие дисциплины, как информационные системы, инженерия программного обеспечения, системы ИТ, которые и были включены в состав компьютеринга в обновленной трактовке. Все это предопределило актуальность задачи разработки целостной системы куррикулумов, соответствующей современному состоянию науки и отрасли ИТ, потребностям практики в кадрах, требованиям к уровню их подготовки. Эта работа растянулась на годы. Акцент в ней делался на разработке: единого подхода к созданию учебных программ по всем векторам профильной подготовки (поддисциплинам или профилям) компьютеринга, специфических характерных требований и структурированных объемов знаний для каждого из профилей подготовки, требований к результату профессиональной подготовки выпускников (outcomes), типовых курсов для различных стратегий реализаций куррикулумов.

Следует отметить, что состав основных исполнителей этого масштабного проекта расширился вовлечением в него следующих профессиональных организаций: Ассоциации информационных систем (The Association for Information Systems — AIS), специализирующейся в области информационных систем, и Ассоциации профессионалов в области ИТ (The Association for Information Technology Professionals — АИТР), фокусирующей свой интерес в сфере использования ИТ для удовлетворения потребностей бизнеса и организаций.

К середине первого десятилетия был разработан следующий набор документов, описывающих типовые модели учебных программ, называемых стандартами куррикулумов (curriculum standards) или просто куррикулумами: «Computer Science 2001» (CS2001 или CCCS2001) [8], «Information Systems 2002» (IS2002) [9], «Computer Engineering 2004» (CE2004) [10], «Software Engineering 2004» (SE2004) [11], «Information Technology (IT2008)» [12], а также документ «Computing Curricula 2005» (CC2005) [13], имеющий общее методологическое назначение.

В последующее пятилетие (а процесс развития куррикулумов принял постоянный непрерывный характер и осуществляется на принципах консорциумной стандартизации [14, 15]) практически все указанные выше документы были переработаны и вышли в новых редакциях.

Целью настоящей работы и является анализ современного состояния системы международных стандартов образовательных программ в области ИТ (или компьютеринга — computing), а также поиск возможностей для встраивания наработок мирового сообщества в существующее методическое обеспечение отечественной высшей школы, олицетворяемое так называемыми стандартами третьего поколения.

2 Методологические основы разработки учебных программ компьютеринга

Как отмечалось во введении, процесс разработки типовых учебных программ, или куррикулумов, для подготовки профессионалов в области ИТ идет по пути создания и непрерывной актуализации системы стандартов куррикулумов. Методологической основой этой системы образовательных стандартов является документ CC2005, в котором, в частности, сформулирована современная и весьма общая трактовка понятия компьютеринга, как любой технической деятельности, вовлекающей применение компьютеров.

В качестве примеров такой деятельности определяются: проектирование и создание аппаратного и программного обеспечения; обработка, структурирование и управление различными видами информации; выполнение научных исследований с использованием компьютеров; повышение интеллектуальности компьютерных систем; создание и использование коммуникационных и мультимедийных сред;

поиск и сбор релевантной для конкретных целей информации и пр. Академическая же дисциплина *computing* рассматривается как интегральная дисциплина, охватывающая широкий спектр более специализированных научно-прикладных дисциплин (поддисциплин), таких как компьютерные науки, искусственный интеллект, компьютерные сети, вычислительная математика, технологии баз данных, информационные системы, мультимедиа, биоинформатика и пр.

Именно широта области ИТ (или компьютеринга), а также ее приложений предопределяет необходимость построения многопрофильной системы подготовки ИТ-кадров, т. е. включающей модули профилированной подготовки по наиболее явно выраженным направлениям специализации (профилям и трекам профессиональной подготовки). Поэтому система стандартных курсов строится на принципах целостности (на основе единой архитектурной модели и единой методологии) и модульности. Архитектуре этой системы и методологическим основам ее построения и посвящено руководство СС2005 [13].

Значительное место в СС2005 уделяется определению характеристик каждого из определенных в этом документе профилей, определению общности между ними. Для этого используются следующие способы:

1. Краткое описание профессиональных характеристик профилей (*descriptions of the computing disciplines*).
2. Графическое описание проблемных областей, характерных для каждого определяемого профиля, на основе модели пространства задач (*graphical views of the computing disciplines with problem space model*).
3. Сравнительный анализ профилей по тематическому содержанию профессиональной подготовки с помощью шкалированной табличной формы (*a tabular comparison of computing degree programs*).
4. Описание итоговых или исходящих профессиональных способностей выпускников (*professional capabilities expected of the graduates*) в соответствии с профилем подготовки.

Рассмотрим основные решения, представленные в данном документе.

2.1 Архитектура компьютеринга

Система стандартов учебных программ бакалаврской подготовки по дисциплине компьютеринг (в качестве эквивалента часто будем использовать термин «направление ИТ»), т. е. система курсов, имеет древовидную структуру.

На верхнем уровне этого дерева располагается рассматриваемый методологический документ СС2005, а нижняя часть дерева включает руководства по составлению программ учебных курсов (*curricula guidelines* или *curriculum standards*) для следующих профилей:

- вычислительная техника (computer engineering — CE);
- компьютерные науки (computer science — CS);
- информационные системы (information systems — IS);
- программная инженерия (software engineering — SE);
- системы ИТ (information technology — IT).

Для последнего профиля дается пояснение, что существуют две трактовки понятия ИТ. В широком смысле под ИТ понимается весь объем понятия computing. В узком смысле под ИТ понимаются собственно системы ИТ (IT-systems или systems of IT), формирующие современную информационную инфраструктуру бизнеса. Таким образом, данный профиль ориентирован на подготовку интеграторов систем, разработчиков и эксплуатационщиков информационной инфраструктуры предприятий и ее компонентов, корпоративных сетей.

Далее указанные выше профили подготовки, вошедшие в состав СС2005, а именно профили CE, CS, IS, SE, IT, будем называть базовыми или классическими.

2.2 Краткое описание профессиональных характеристик базовых профилей

В СС2005 определены основные профессиональные характеристики каждого из базовых профилей. Рассмотрим эти характеристики, за исключением характеристик профиля computer engineering (вычислительная техника), который относится к подготовке кадров более специализированной области, требующей определенной инженерной подготовки, и который в соответствии с традициями отечественной высшей школы реализуется, как правило, в технических университетах.

1. Компьютерные науки (computer science)

Выпускников по профилю «компьютерные науки» ожидает широкий спектр деятельности — от исследований и разработок теоретических и программно-алгоритмических решений в области обработки информации до участия в разработках в таких наукоемких областях, как робототехника, компьютерное видение, системы искусственного интеллекта, биоинформатика и др.

Выделяются следующие виды деятельности:

- разработка и реализация программного обеспечения для исследовательских и проектных работ в области создания новых ИТ, а также руководство наукоемкими разработками в области ИТ;
- разработка новых методов использования компьютеров и обработки информации, в том числе в интересах прикладных областей;
- разработка эффективных алгоритмов и методов реализации функций систем ИТ.

2. Информационные системы (information systems)

Профессионалы по информационным системам должны быть компетентны в интеграции ИТ-решений с бизнес-процессами для достижения конечных целей предприятия (корпоративных целей). При этом акцент в деятельности такого рода делается на информации (а не на ИТ, которые рассматриваются в качестве инструмента, позволяющего производить, обрабатывать, распространять необходимую информацию), а также на процессах, которые предприятие может осуществлять, используя ИТ.

Выпускники данного профиля должны:

- понимать технические и организационные факторы в своей деятельности;
- быть способными определять каким образом информация и информатизированные бизнес-процессы должны обеспечить конкурентоспособность предприятия;
- играть ключевую роль в определении требований для корпоративных информационных систем (КИС);
- разрабатывать спецификации КИС;
- осуществлять проектирование и реализацию КИС;
- осуществлять тестирование и комплексные испытания КИС;
- отвечать за оптимизацию бизнес-процессов и т. п.

По существу такие специалисты должны служить мостом между техническими специалистами и управленцами.

3. Системы информационных технологий (information technology)

Профессионалы по профилю «системы ИТ», в отличие от специалистов по информационным системам, делают акцент в своей деятельности не на корпоративной информации, а на собственно самих ИТ, точнее системах ИТ, составляющих информационную инфраструктуру предприятий. Они должны обеспечивать необходимый уровень качества функционирования систем ИТ, их эксплуатацию, модернизацию, обеспечение информационной безопасности, повышение эффективности использования информационных и технических ресурсов и т. д.

Таким образом, специалисты данного профиля должны сочетать хорошую академическую подготовку, позволяющую им быстро адаптироваться к новым технологиям и стандартам, с практическими умениями решать текущие производственные задачи.

Примерами компетенций профессионалов по профилю системы ИТ могут служить:

- инсталляция сетей;
- сетевое администрирование;
- управление сетевой безопасностью;
- разработка и поддержание мультимедийных ресурсов предприятия;
- установка и настройка коммуникационного оборудования компьютерных сетей;
- администрирование почтовых серверов и систем;
- управление жизненным циклом ИТ-сервисов;
- модернизация ИТ-систем и их оборудования;
- администрирование операционных систем и т. п.

4. Программная инженерия (*software engineering*)

Профессионалы по профилю «программная инженерия» должны быть компетентными в области создания и сопровождения систем программного обеспечения, отвечающих требованиям надежности, эффективности, сопровождаемости, открытости и т. п.

Особенно важной сферой деятельности SE-профессионалов служат сложные критические и военные приложения. Специалисты данного профиля должны иметь высокий уровень подготовки в математике и компьютерной науке.

Примерами компетенций профессионалов по профилю «программная инженерия» могут служить:

- владение методами и средствами разработки программного обеспечения, удовлетворяющего требованиям надежности;
- управление процессами жизненного цикла программных систем;
- разработка комплектов тестов;
- разработка и реализация методов тестирования и испытания программных комплексов,
- интеграции и сопровождение программных систем,
- моделирование окружений функционирования программных систем и др.

2.3 Графическое описание проблемных областей базовых профилей на основе модели пространства задач

Для иллюстрации различий и общности рассмотренных выше базовых профилей в документе СС2005 предложен графический метод для характеристики областей деятельности выпускников данных профилей на основе модели пространства задач в области компьютеринга.

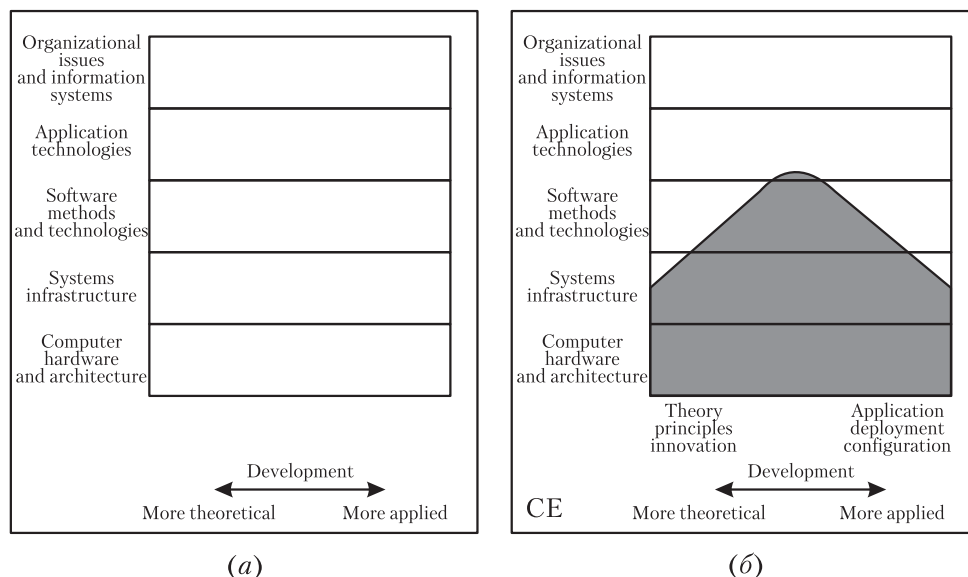


Рис. 1 Модель пространства задач (а) и пространство задач для профиля СЕ (б)

В данном методе пространство задач, или проблемная область (problem space), ИТ-деятельности моделируется посредством двумерной диаграммы, показанной на рис. 1, а. Ось X на диаграмме представляет собой непрерывную шкалу, характеризующую (вообще говоря, шкалированное) свойство «научность/практичность» некоторой профессиональной деятельности, т. е. предполагается, что в левой части плоскости концентрируются наукоемкие работы, отличающиеся теоретическим характером, научностью, новизной, инновационностью, а в правой — виды деятельности практического характера, например: использование приложений, установка программного обеспечения, конфигурирование систем, администрирование сетевых ресурсов и пр. А ось Y прошкалирована так, что отражает уровни абстракции видов деятельности. Нижний слой диаграммы (Computer hardware and architecture) включает виды деятельности, связанные с разработкой компьютерных архитектур, с созданием и эксплуатацией аппаратного обеспечения. Следующий слой (Systems infrastructure) — с созданием и эксплуатацией системной инфраструктуры предприятий (например, корпоративной сетевой инфраструктуры), последующий слой (Software methods and technologies) охватывает деятельности по разработке программного обеспечения и программных систем. Слой (Application technologies), предшествующий наивысшему слою в иерархии, включает работы по созданию прикладных технологий (например, систем автоматизации научных исследований, e-коммерции). Самый верхний слой (Organizational issues and information systems) соответ-

ствует созданию КИС и систем автоматизации организационной деятельности предприятий. В качестве примера использования данного метода на рис. 1, б представлено пространство задач, характерное для профиля «вычислительная техника» (computer engineering) и показывающее тяготение данной профессии к нижним уровням моделируемого пространства.

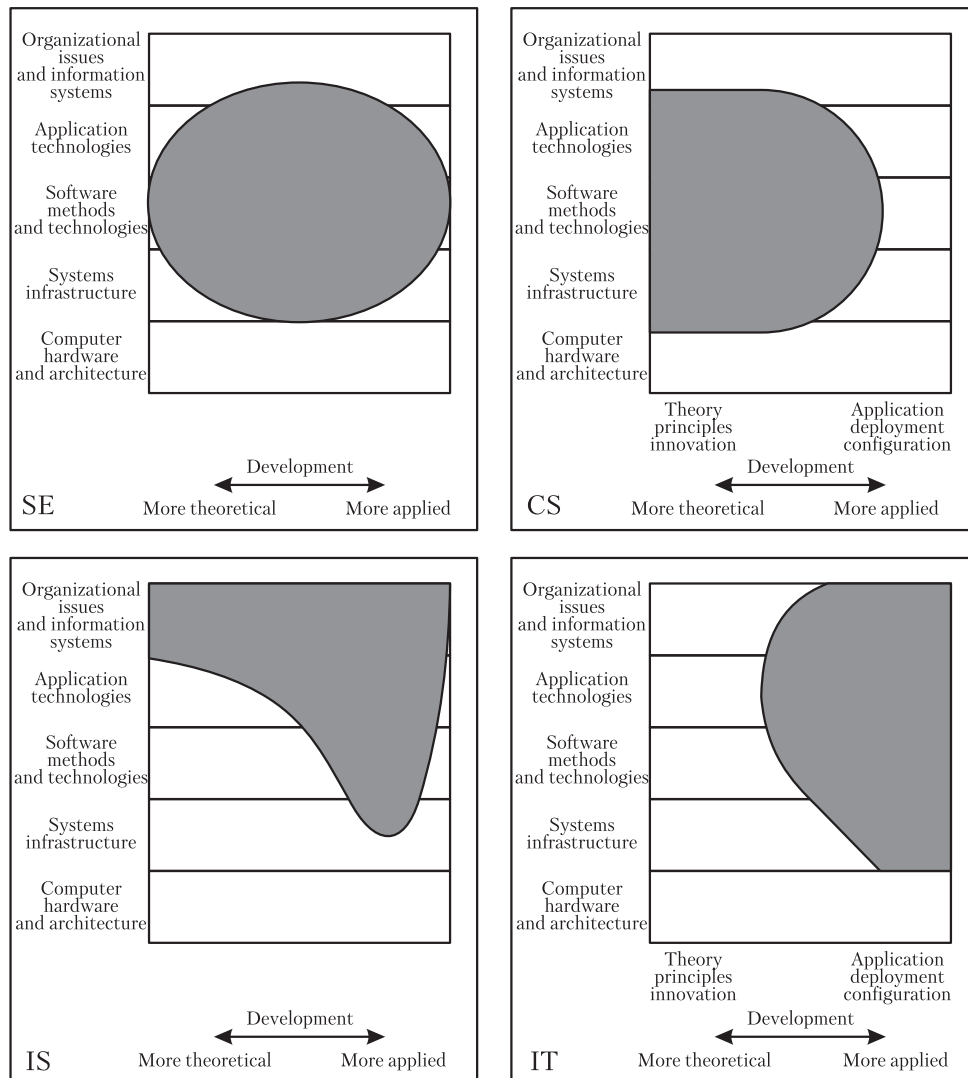


Рис. 2 Сравнение областей деятельности для базовых профилей SE, CS, IS, IT

На рис. 2 для сравнения представлены области деятельности, характерные для оставшихся, так сказать, «нехадверных» базовых профилей, анализ руководств для которых предполагается сделать ниже, а именно профилей CS, IS, IT, SE.

Рассмотренный выше метод описания областей деятельности посредством соответствующих диаграмм позволяет сравнивать различные профили ИТ-профессии на качественном уровне.

Для более детального описания различий и общности классических профилей применяется другой подход, рассмотренный ниже.

2.4 Сравнительный анализ базовых профилей по тематическому содержанию профессиональной подготовки с помощью шкалированной табличной формы

Анализ требований, предъявляемых к ИТ-профессионалам со стороны ИТ-отрасли, выполненный под руководством проф. П. Деннинга [16], а также рабочими группами в процессе разработки стандартов курсикулов для бакалаврских программ по базовым профилям, показал, что значительный акцент в подготовке бакалавров должен быть сделан на изучении (а также на развитии соответствующих умений и навыков) порядка 40 так называемых базовых или ключевых технологий, знание которых и умение пользоваться которыми определяют профессиональную состоятельность выпускника независимо от его профилизации, хотя уровень владения конкретной темой может существенно меняться в зависимости от профиля подготовки.

Например, администраторы сетей практически не сталкиваются с необходимостью разрабатывать большие программы на языках программирования, однако должны уметь пользоваться современными языками для решения собственных локальных задач, возникающих в процессе сетевого администрирования. В то же время разработчики программного обеспечения должны владеть языками и средами, поддерживающими разработку программ, а также процессами жизненного цикла программных средств на экспертном уровне.

Список таких ключевых технологий составлен разработчиками CC2005 на основе обобщения тем, описанных в упомянутых выше руководствах по созданию программ учебных курсов для базовых профилей. Этот список технологий, дополненный шкалированными весовыми характеристиками, отражающими уровень освоения каждой темы для учащихся соответствующих профилей, позволяет в компактной табличной форме сравнить содержание профессиональной подготовки по базовым профилям на тематическом уровне. При этом для определения уровня профессиональной подготовки по ключевым темам в зависимости от профиля используется некоторая модификация метода Блума [17], в котором оценки уровня подготовки выражаются в некоторых абстрактных значениях в диапазоне от 0 до 5. Высшая оценка соответствует критической важности данной темы в под-

Knowledge Area	CE		CS		IS		IT		SE	
	min	max	min	max	min	max	min	max	min	max
Programming Fundamentals	4	4	4	5	2	4	2	4	5	5
Integrative Programming	0	2	1	3	2	4	3	5	1	3
Algorithms and Complexity	2	4	4	5	1	2	1	2	3	4
Computer Architecture and Organization	5	5	2	4	1	2	1	2	2	4
Operating Systems Principles & Design	2	5	3	5	1	1	1	2	3	4
Operating Systems Configuration & Use	2	3	2	4	2	3	3	5	2	4
Net Centric Principles and Design	1	3	2	4	1	3	3	4	2	4
Net Centric Use and configuration	1	2	2	3	2	4	4	5	2	3
Platform technologies	0	1	0	2	1	3	2	4	0	3
Theory of Programming Languages	1	2	3	5	0	1	0	1	2	4
Human-Computer Interaction	2	5	2	4	2	5	4	5	3	5
Graphics and Visualization	1	3	1	5	1	1	0	1	1	3
Intelligent Systems (AI)	1	3	2	5	1	1	0	0	0	0
Information Management (DB) Theory	1	3	2	5	1	3	1	1	2	5
Information Management (DB) Practice	1	2	1	4	4	5	3	4	1	4
Scientific computing (Numerical mthds)	0	2	0	5	0	0	0	0	0	0
Legal / Professional / Ethics / Society	2	5	2	4	2	5	2	4	2	5
Information Systems Development	0	2	0	2	5	5	1	3	2	4
Analysis of Business Requirements	0	1	0	1	5	5	1	2	1	3
E-business	0	0	0	0	4	5	1	2	0	3
Analysis of Technical Requirements	2	5	2	4	2	4	3	5	3	5
Engineering Foundations for SW	1	2	1	2	1	1	0	0	2	5
Engineering Economics for SW	1	3	0	1	1	2	0	1	2	3
Software Modeling and Analysis	1	3	2	3	3	3	1	3	4	5
Software Design	2	4	3	5	1	3	1	2	5	5
Software Verification and Validation	1	3	1	2	1	2	1	2	4	5
Software Evolution (maintenance)	1	3	1	1	1	2	1	2	2	4
Software Process	1	1	1	2	1	2	1	1	2	5
Software Quality	1	2	1	2	1	2	1	2	2	4
Comp Systems Engineering	5	5	1	2	0	0	0	0	2	3
Digital logic	5	5	2	3	1	1	1	1	0	3
Embedded Systems	2	5	0	3	0	0	0	1	0	4
Distributed Systems	3	5	1	3	2	4	1	3	2	4
Security: issues and principles	2	3	1	4	2	3	1	3	1	3
Security: implementation and mgt	1	2	1	3	1	3	3	5	1	3
Systems administration	1	2	1	1	1	3	3	5	1	2
Management of Info Systems Org.	0	0	0	0	3	5	0	0	0	0
Systems integration	1	4	1	2	1	4	4	5	1	4
Digital media development	0	2	0	1	1	2	3	5	0	1
Technical support	0	1	0	1	1	3	5	5	0	1

Рис. 3 Табличная шкалированная модель уровня подготовки выпускников по ключевым темам для классических профилей подготовки (min — соответствует минимальному уровню компетенций, max — максимальному)

готовке бакалавра заданного профиля, определяя необходимость максимального акцента на данной теме. Оценка «0» соответствует ситуации, когда компетенция для конкретного профиля является несущественной. Остальные, ненулевые, значения шкалы характеризуют соответствующий их значению уровень значимости данной темы для конкретного профиля подготовки.

На основе экспертных оценок, анализа опыта реализации большого ряда наиболее успешных университетских программ в СС2005 предложены шкалированные оценки уровня подготовки выпускников для каждого базового профиля, сведенные вместе для удобства их сравнения в одну таблицу (рис. 3). В левом столбце этой таблицы перечислены «ключевые технологии/темы профессио-

нальной подготовки», а каждый из следующих столбцов содержит диапазоны числовых оценок важности этой темы для каждого профиля, представленного столбцом таблицы. Таким образом, на пересечении строк и столбцов стоят два весовых значения шкалы, левое значение соответствует минимальному уровню подготовки по данной теме для данного профиля, правое — максимальному уровню.

Минимальное значение отражает акцент на подготовке по данной теме, рекомендуемый для программ образовательного стандарта по конкретному профилю подготовки. Максимальное значение используется для ограничения объема внимания к конкретной теме в учебных планах с учетом выбранных студентом дополнительных курсов как некоторая ограничительная мера для предотвращения однобокости в подготовке выпускников.

Рассмотренная таблица, по существу, представляет собой систему целей обучения по соответствующим профилям подготовки.

Аналогичный подход применяется и при определении требований к подготовке ИТ-бакалавров по непрофильным темам.

2.5 Описание исходящих профессиональных характеристик выпускников базовых профилей

С точки зрения работодателя, ценность выпускника университета определяется его способностями (исходящими, или рабочими, характеристиками) выполнять конкретные виды работ.

По аналогии с примененным выше методом шкалирования для моделирования целей обучения по ключевым темам/технологиям для классических профилей, в СС2005 предложена система рабочих, или исходящих, характеристик для выпускников в соответствии с профилем подготовки, представленная в табличной форме (Table 3.3. Relative performance capabilities of computing graduates by discipline [11]).

Данная таблица включает более шестидесяти примерных видов работ, разбитых на одиннадцать классов по видам технологий. Она (или ее модификация) может служить опорным элементом при построении практико-ориентированной части учебных программ (спецкурсов, семинаров, мастер-классов, курсовых и дипломных работ, производственных практик). В частности, участие в реализации этой части программы представителей работодателей является чрезвычайно важным.

Рассмотрев общую методологию разработки стандартов curriculums для области ИТ, перейдем к анализу современных curriculums для бакалавриата и магистратуры.

3 Стандарты куррикулумов бакалавриата

Современный стек стандартов куррикулумов дисциплины компьютеринг, ориентированный на подготовку бакалавров, включает следующие основные документы:

- «Computing Science 2001» (CS2001) [8];
- «Computer Science 2008» (CS2008) [14];
- «Information Systems 2010» (IS2010) [15];
- «Software Engineering 2004» (SE2004) [11, 18];
- «Information Technology 2008» (IT2008) [12].

Ниже сделаем краткий анализ этих документов. Заметим, что в стек не включен документ «Computer Engineering 2004» (CE2004), тяготеющий к инженерному образованию, а также документ «Computing Curricula 2009: Guidelines for Associate-Degree Transfer Curriculum in Computer Science», ориентированный на доучивание на бакалаврскую степень выпускников колледжей с двухлетней профессиональной подготовкой.

3.1 Компьютерные науки (computer science — CS)

В настоящее время методический базис в сфере подготовки бакалавров по профилю computer science составляют два руководства:

- «Computing Science 2001» (CS2001) [8];
- «Computer Science 2008» (CS2008) [14].

Несмотря на то что руководство CS2001 было разработано более десятилетия назад, при его создании были заложены столь основательные решения, что многие из них оказались актуальными и по сей день. В частности, в отличие от других куррикулумов, в документ CS2001 вошел обширный методический материал по педагогическим стратегиям и моделям реализации описанного в документе объема знаний, а также были включены описания (в приложениях) большого числа курсов для разных педагогических стратегий реализации данного куррикулума.

Руководство CS2008 разрабатывалось с целью актуализации основных разделов CS2001, а именно спецификаций объема знаний и его ядра, целей обучения и итоговых профессиональных характеристик выпускника. Таким образом, документ CS2008 не стал полной ревизией своего предшественника, а только редакцией его отдельных разделов. Структура объема знаний осталась неизменной. Поэтому при составлении реальных учебных программ руководство CS2001 также может успешно использоваться с учетом того, что отдельные разделы этого документа актуализированы в CS2008.

Основными методическими положениями указанных куррикулов для профиля computer science являются следующие решения:

1. Модель архитектуры для представления объема знаний CS типовая — иерархическая. На верхнем уровне она содержит предметные области (areas), которые подразделяются на разделы или модули знаний (units), последние, в свою очередь, разбиваются на темы (topics). Структура знаний осталась неизменной и в документе CS2008. Она содержит 14 предметных областей, включая: «Дискретные структуры» (DS), «Основы программирования» (PF), «Алгоритмы и теория сложности» (AL), «Архитектура и организация ЭВМ» (AR), «Операционные системы» (OS), «Распределенные вычисления» (NC), «Языки программирования» (PL), «Взаимодействие человека и машины» (HC), «Графика и визуализация» (GV), «Интеллектуальные системы» (IS), «Управление информацией» (IM), «Социальные и профессиональные вопросы программирования» (SP), «Программная инженерия» (SE), «Методы вычислений» (CN).
2. Одной из центральных в куррикуле является концепция ядра, или обязательной части объема знаний, которая должна присутствовать во всех учебных программах по данному профилю. Объем почасовой лекционной нагрузки ядра составляет 280 лекционных часов (и в CS2001, и в CS2008). При этом понимается, что объем внеаудиторных занятий должен примерно в три раза превосходить объем аудиторных, т. е. раздел, требующий трех аудиторных часов, должен изучаться примерно 12 ч (3 ч в аудитории и 9 ч самостоятельно). А для пятнадцатинедельного семестрового цикла трехкредитный курс потребует около 40 лекционных часов при общей учебной нагрузке в 160 ч.
3. CS2008 содержит обзор пересмотренного объема знаний (приложение А), детальное описание обновленного объема знаний (приложение В), описание некоторых новых рекомендованных курсов (приложение С), дополняющих описания курсов документа CS2001.
4. Определена типовая модель учебной программы и классификация курсов на вводные (introductory), основные (intermediate) и углубленные, или специальные (advanced).
5. В качестве инструментов диверсификации учебных программ документ CS2001 предоставляет широкий спектр решений построения программ на основе различных образовательных и тематических парадигм. В частности, в данном руководстве определены шесть различных стратегий реализации вводных курсов, четыре подхода к компоновке основных курсов. Комбинация этих способов дает широкие возможности вузам для творческого подхода к построению конкретных учебных программ. В руководстве CS2008 поддерживается принцип многообразия учебных программ, в том

числе акцентируется внимание на целесообразности использования подхода к диверсификации программ на основе их ориентации по выбранному научно-прикладному направлению.

6. Определена классификация целей обучения (learning objectives) и система целей, связанная с элементами объема знаний и пронизывающая его на всех уровнях описания.
7. В CS2008 приведен обновленный весьма обширный список результатов обучения, т. е. список ожидаемых характеристик выпускника, состоящий из наборов общих характеристик, когнитивных способностей и навыков, дополнительных профессиональных навыков.

В заключение этого раздела дадим краткую качественную характеристику основных изменений в ядре объема знаний, что представляет наибольший интерес, так как это влечет необходимость внесения соответствующих изменений во вводные и базовые курсы соответствующих учебных программ российских вузов:

- в области «Дискретные структуры» (DS) существенно больший акцент делается на логичности рассуждения, способности студентов выполнять строго обоснованные и аргументированные доказательства, в то же время снижается внимание к формальным (символическим) доказательствам;
- в области «Архитектура и организация ЭВМ» (AR) введено рассмотрение многоядерных и мультитредовых процессоров;
- в области «Распределенные вычисления» (NC) введены новые темы, посвященные сервисно-ориентированной архитектуре и распределенным вычислениям, включая гриды, в то же время темы, посвященные коммутации каналов и пакетной коммутации, потокам и дейтаграммам, программному обеспечению шлюзов были удалены из ядра;
- в области «Интеллектуальные системы» (IS) включены темы, посвященные изучению концепции понимания (perception), онтологий, теории планирования, понятий игрового программного обеспечения;
- в области «Методы вычислений» (CN) акцент переместился с методов вычислений на высокопроизводительные вычислительные технологии и параллелизм, и теперь эти темы становятся претендентами на включение в ядро.

3.2 Информационные системы (information systems — IS)

Для профиля «информационные системы» актуальными являются два технически эквивалентных руководства:

- «Information Systems 2010» (IS2010), а также его эквивалент, реализованный в виде Wiki-ресурса — IS Curriculum Wiki [15];

– второй материал, выполненный в виде веб-ресурса, служил технологической площадкой для коллективной разработки стандарта куррикулума IS. После того как веб-версия объема знаний и других разделов куррикулума достигла высокой степени консенсуса, ресурс был закрыт для внесения изменений и на основе его материала был издан том куррикулума в традиционном исполнении. Такой подход позволил вовлечь в проект широкий круг профессиональной общественности.

Основной причиной, обусловившей актуальность разработки IS2010, явились значительные изменения в данной области, в том числе в стандартизации процессов проектирования IS, широком внедрении веб-технологий, создании новых архитектурных парадигм, широком использовании крупномасштабных систем ERP, повсеместном распространении мобильных компьютеров и гаджетов, широком использовании инфраструктурных фреймворков (таких как ITIL, COBIT, ISO 17799) и пр.

Данный куррикулум разрабатывался в сложное кризисное время, поэтому с некоторой осторожностью декларирует следующие ожидаемые результаты обучения, которые он может обеспечить:

- совершенствование организационных процессов;
- использование возможностей, созданных технологическими инновациями;
- понимание и решение информационных потребностей;
- проектирование и управление архитектурой предприятия;
- определение и оценка решений, поиска альтернатив;
- обеспечение безопасности данных и инфраструктуры;
- понимание, управление и контроль ИТ-рисков.

Рассмотрим основные решения IS2010.

1. Модель архитектуры для представления объема знаний IS типовая — иерархическая. На верхнем уровне она содержит курсы, которые разбиваются на модули знаний, в свою очередь, представляемые как наборы тем. Но понятие курса трактуется не жестко. Отмечается возможность комбинирования курсами и их содержанием.
2. Объем знаний профиля IS состоит из двух категорий курсов: основных курсов, или курсов ядра (core courses), и факультативных курсов, или курсов по выбору (electives). Первые содержат знания, необходимые для всех треков профессиональной подготовки, а вторые — модули знаний, из которых строится профилизация подготовки, или треки профессиональной подготовки.

Структура знаний включает в себя семь основных курсов: «Основы информационных систем», «Управление данными и информацией», «Архитектура предприятия», «Управление проектами», «ИТ-инфраструктуры»,

«Системный анализ и проектирование», «Стратегия, управление и приобретение информационных систем». Указанный в kurikulumе примерный список дополнительных курсов включает следующие курсы: «Разработка приложений», «Менеджмент бизнес-процессов», «Корпоративные системы», «Введение в человеко-машинное взаимодействие», «Аудит и управление в ИТ», «ИС-инновации и новые технологии», «ИТ-безопасность», «ИТ-безопасность и управление рисками».

3. Объем почасовой лекционной нагрузки обязательных курсов сопоставим с объемом ядра kurikulumа для рассмотренного ранее профиля CS — семь курсов ядра в среднем по 40 лекционных часов, т. е. примерно 280 лекционных часов.
4. IS2010 содержит пересмотренную модель ожиданий (результатов) подготовки IS-выпускников, а также определение хорошо систематизированной системы учебных целей в форме требований «знания–умения–навыки» (knowledge-skills).
5. Определен примерный набор факультативных курсов, являющихся строительными блоками для треков профессиональной подготовки (17 треков-специализаций).
6. Основным инструментом диверсификации программ служит модель представления объема знаний в весьма удобной для практики табличной форме, представленной на рис. 4, в которой строкам соответствуют основные или факультативные курсы, а столбцам — треки специализации. На пересечении строк и столбцов таблицы ставится черный или белый кружок. В случае черного кружка считается, что курс должен читаться в полном объеме, в случае белого возможно неполное покрытие тем курса. Всего разработано 17 треков-специализаций, список которых включает следующие профессиональные позиции:
 - (1) разработчик приложений;
 - (2) бизнес-аналитик;
 - (3) аналитик бизнес-процессов;
 - (4) аналитик технологий управления инфокоммуникациями;
 - (5) администратор баз данных;
 - (6) аналитик баз данных;
 - (7) менеджер e-бизнеса;
 - (8) erp-специалист (erp specialist);
 - (9) специалист по информационному аудиту и совместимости данных;
 - (10) разработчик ИТ;

Structure of the IS Model Curriculum: Information Systems specific courses

Career Track:	Information Systems specific courses																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Core IS Courses:																	
Foundations of IS	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Enterprise Architecture	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
IS Strategy, Management and Acquisition	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Data and Information Management	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Systems Analysis & Design	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
IT Infrastructure	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
IT Project Management	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Elective IS Courses:																	
Application Development	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Business Process Management	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Collaborative Computing																	
Data Mining / Business Intelligence																	
Enterprise Systems	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Human-Computer Interaction	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Information Search and Retrieval	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
IT Audit and Controls	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
IT Security and Risk Management	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Knowledge Management	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Social Informatics																	

Key:
 ● = Significant Coverage
 ○ = Some Coverage
 Blank Cell = Not Required

Рис. 4 Модель куррикулума IS

- (11) менеджер по обработке информационных ресурсов;
 - (12) консультант по ИТ;
 - (13) менеджер операций по ИТ;
 - (14) менеджер по рискам и безопасности ИТ;
 - (15) сетевой администратор;
 - (16) менеджер проекта;
 - (17) менеджер веб-контента.
7. Документ IS2010 содержит детальное описание изменений в ядре по сравнению с IS2002, а также рассматривает три модели реализации куррикулума — в Североамериканской школе бизнеса, имеющей AACSB-аккредитацию, в Школе информатики и в Европейской бизнес-школе с трехлетним бакалавриатом.
 8. Куррикулум содержит развернутую спецификацию основных и дополнительных курсов.
 9. Для оценки глубины знаний и умений используется пятиуровневая шкала, основанная на методе Блума.

3.3 Программная инженерия (software engineering — SE)

Для профиля «программная инженерия» продолжает оставаться актуальным куррикулум «Software Engineering 2004» (SE2004) [11, 13]. Учитывая, что имеется качественный перевод этого руководства на русский язык, будем предельно кратко при его описании. Рассмотрим основные характеристики этого документа. В нем:

1. Определены сферы деятельности профессионалов в области программной инженерии, уточняющие описание области программной инженерии, представленное в документе CC2005.
2. Определены результаты обучения в виде списка фундаментальных навыков и знаний, которыми необходимо обладать всем выпускникам программ данного профиля, сформулированных достаточно полно и в то же время общо, что позволяет легко адаптировать эти требования к конкретным учебным программам.
3. Основным содержанием SE2004 является спецификация объема знаний, называемого здесь SEEK (Software Engineering Education Knowledge) и определяющего содержание обучения по данному профилю, при этом модель SEEK аналогична модели для объема знаний документа CS2001 (Computing Curricula: Computer Science). Основой для проектирования структуры и содержания SEEK послужил общепризнанный документ SWEBOOK [19],

определяющий набор знаний, ожидаемых от профессионала с четырехлетним опытом работы в области программной инженерии.

4. SEEK, построенный структурно по типовому иерархическому принципу: предметные области (arears) → модули (units) → темы (topics), включает следующие десять предметных областей знаний: «Основы компьютинга» (CMP), «Основы математики и инженерии» (FND), «Профессиональная практика» (PRF), «Моделирование и анализ программного обеспечения» (MAA), «Проектирование программного обеспечения» (DES), «Верификация и аттестация программного обеспечения» (VAV), «Эволюция программного обеспечения» (EVL), «Процессы разработки программного обеспечения» (PRO), «Качество программного обеспечения» (QUA), «Управление программными проектами» (MGT).
5. Как и в других куррикулах, в SEEK центральной является концепция ядра, содержание которого рассматривается как обязательная составляющая всех учебных программ по профилю SE. Ядро определяется на уровне тем модулей совокупности знаний. Число основных (ядерных) тем — 235. Минимальный объем ядра весит более 400 аудиторных часов, что значительно больше объема ядра профиля CS. Однако если сделать анализ содержания основных тем для SE, то он покажет, что они включают большую часть содержания ядра для профиля CS. Таким образом, профиль SE легко реализуется как некоторая специализация профиля CS. Статистика подтверждает, что такой подход весьма распространен, а именно 50% учебных программ по программной инженерии реализуется на факультетах компьютерных наук (computer science).
6. SEEK заимствовал из SWEBOOK рейтинговую систему для классификации разделов знаний, основанную на классификации образовательных целей по Блуму [20], т. е. для каждой темы определено некоторое значение, соответствующее классификации Блума и показывающее, каким уровнем владения данной темой должен обладать выпускник конкретного трека подготовки. Заметим, что в описании SEEK используется модифицированный метод Блума, охватывающий только три из шести уровней целей обучения: знание, понимание, применение. Кроме того, для каждой темы определена степень ее значимости, показывающая, является ли данная тема необходимой (E), желаемой (D) либо факультативной (O) по отношению к основному набору знаний по программной инженерии.
7. Дипломный проект в программе обучения программной инженерии оценивается как критически важный. По учебной нагрузке он приравнивается к годовому курсу (т. е. 80 лекционных часов).
8. Значительное место в куррикуле отведено описанию возможных стратегий реализации учебной программы, приводятся рекомендации по разработке

учебных планов, типовые шаблоны построения учебного плана, рекомендации по компоновке отдельных курсов.

9. В качестве основного способа диверсификации учебных программ предлагается использовать специализацию в одной из предметных областей SEEK или в некоторой прикладной области из рекомендованного в SE 2004 (но не исчерпывающего) списка. Этот список включает следующие прикладные области, характеризующиеся спецификой создаваемого для них программного обеспечения:
 - (1) SAS.net — «Распределенные системы»;
 - (2) SAS.inf — «Информационные системы и обработка данных»;
 - (3) SAS.fin — «Финансовые системы и системы электронной коммерции»;
 - (4) SAS.sur — «Отказоустойчивые и живучие (survivable) системы»;
 - (5) SAS.sec — «Хорошо защищенные системы»;
 - (6) SAS.sfy — «Системы с повышенными требованиями к безопасности»;
 - (7) SAS.emb — «Встроенные системы и системы реального времени»;
 - (8) SAS.bio — «Биомедицинские системы»;
 - (9) SAS.sci — «Научно-исследовательские системы»;
 - (10) SAS.tel — «Телекоммуникационные системы»;
 - (11) SAS.av — «Авиационное электронное оборудование и транспортные системы»;
 - (12) SAS.ind — «Системы контроля промышленного процесса»;
 - (13) SAS.mm — «Мультимедийные, игровые и развлекательные системы»;
 - (14) SAS.mob — «Системы для малых и мобильных платформ»;
 - (15) SAS.ab — «Системы, основанные на агентах» (agent-based systems).
10. Приведено описание категорий курсов учебной программы, их способа кодирования, шаблонов построения циклов для вводных курсов и промежуточных курсов. Завершается руководство приложением, содержащим детальное описание учебных курсов.

3.4 Системы информационных технологий (information technology — IT)

Методическим руководством для подготовки бакалавров по профилю information technology, или сокращенно IT, является стандарт куррикулума под названием «Information Technology 2008» (IT2008) [12].

Как и для рассмотренных выше стандартов куррикулов, основная цель данного руководства состояла в определении объема профессиональных знаний

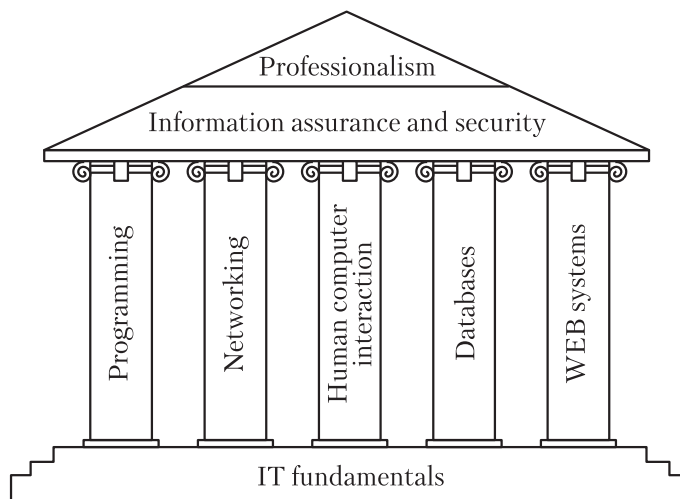


Рис. 5 Модель куррикулума ИТ

по данному профилю подготовки, а также подходов и методов по реализации процесса обучения.

Состав и взаимосвязь важнейших тем, определяющих содержание профиля ИТ как академической дисциплины, раскрывает архитектурная модель профиля в виде аркадной конструкции, изображенная на рис. 5. Фундаментом такой конструкции служит область знаний, названная «Основы информационных технологий», на которой и устанавливаются пилоны-столпы, представляющие базовые для профиля области знаний — «Программирование», «Компьютерные сети», «Взаимодействие человека с компьютером», «Базы данных» и «Веб-системы». На этих основах и формируется завершение арочной конструкции в виде таких областей знания, как «Информационная безопасность и защита информации» и «Профессионализм». Последняя аккумулирует широкий свод знаний, начиная от профессиональной этики, патентоведения, лицензионной практики и кончая принципами взаимоотношений в коллективе и с пользователями.

Основное содержание данного куррикулума включает:

1. Описание объема знаний для бакалаврских учебных программ профиля ИТ, разработанного на основе подхода руководства CS2001 в виде иерархической системы — предметные области, структурированные на модули знаний (units), которые, в свою очередь, детализируются до уровня тем и целей (результатов) обучения (learning outcomes). На верхнем уровне иерархии декомпозицию объема знаний составляют следующие тринадцать предметных областей: «Основы информационных технологий» (Information Technology

Fundamentals — ITF), «Взаимодействие человека с компьютером» (Human Computer Interaction — HCI), «Информационная безопасность и защита данных» (Information Assurance and Security — IAS), «Управление информацией (включая технологии баз данных)» (Information Management — IM), «Интегративное программирование и технологии» (Integrative Programming and Technologies — IPT), «Математика и статистика для ИТ» (Math and Statistics for IT — MS), «Сетевые технологии» (Networking — NET), «Основы программирования» (Programming Fundamentals — PF), «Платформенные технологии» (Platform Technologies — PT), «Администрирование и сопровождение систем» (Systems Administration and Maintenance — SA), «Архитектура и интеграция систем» (System Integration & Architecture — SIA), «Социальные и профессиональные вопросы» (Social and Professional Issues — SP), «Веб-технологии и системы» (Web Systems and Technologies — WS).

2. Описание результатов/целей обучения, определенных для каждого модуля объема знаний, которые используются при подробном описании объема знаний, представленного в приложении *A*. Для большинства модулей определяются обязательные цели обучения (core outcomes), т. е. связанные с изучением обязательных элементов объема знаний.
3. Ядро объема знаний определяется на уровне помодульной декомпозиции предметных областей объема знаний. Оно включает 81 модуль (unit) из 85 модулей, составляющих объем знаний данного профиля. Объем ядра оценивается в 314 аудиторных часов. Анализ содержания ядра показывает, что реализация профиля ИТ на основе ядра модуля CS в рамках бакалаврской программы не представляет сложности.
4. Итоговые результаты обучения (the it advanced outcomes), которые ожидается получить в результате обучения по программе, соответствующей настоящему руководству.
5. Описание моделей построения учебной программы для двух основных подходов подготовки бакалавров по профилю ИТ.
6. В качестве основного способа диверсификации учебных программ по профилю ИТ рассматривается два подхода к подготовке бакалавров, которые условно можно назвать «сначала интеграция» (integration first) и «сначала пилоны» (pillars first), т. е. базовые знания.

Первый подход призван дать студентам как можно раньше интегрированное представление о профессиональной сфере деятельности. В качестве основного инструмента для реализации данного подхода предлагаются два курса второго года обучения — «ИТ-системы» и «Веб-системы». Пример реализации такого подхода представлен в приложении *B*.

Второй подход предполагает раннее погружение в области знаний, изображенные на рис. 5 в виде пилонов, или столпов, а интеграционные аспекты вводятся по мере накопления «столповых» знаний. Реализация такого подхода детально описана в том же приложении В.

Дальнейшее варьирование учебных программ предполагается осуществлять введением специальных курсов на четвертом году обучения. При такой методике специализации достигается независимость построения специализированных учебных программ от выбранных образовательных парадигм, т. е. от того, каким из двух основных способов реализуется бакалаврская программа.

Для поддержки специализации учебных программ в руководстве предложен список из более чем 60 потенциальных спецкурсов, содержание и актуализация которых должны поддерживаться на сайте рабочей группы — разработчика данного куррикулума. Еще одним методическим решением, предлагаемым в ИТ2008 и способствующим интегративному характеру обучения столь многогранной дисциплине, является рекомендация использовать на финише обучения так называемый *capstone*-подход (*integrative capstone experience*) — по существу, актуальный коллективный проект, способствующий интеграции полученных учащимися в процессе обучения знаний.

Учитывая практическую направленность деятельности выпускников учебных программ по данному профилю, большое место в рассматриваемом куррикулуме уделено вопросам организации профессиональной практики.

4 Стандарты куррикулов магистратуры

Рассмотренные выше стандарты куррикулов были разработаны с целью формирования глобального образовательного пространства в сфере самого массового вида ИТ-образования — подготовки бакалавров по родственным направлениям дисциплины компьютеринг (называемыми нами профилями).

Как было показано, основное содержание этих руководств составляло описание актуальных объемов знаний для обучения соответствующим профессиям, а также описание методик построения совместимых по базовым знаниям образовательных программ и рекомендаций по реализации процессов обучения.

В конце 2009 г. появился куррикулум, регламентирующий подготовку магистров по профилю «программная инженерия» — «Graduate Software Engineering 2009» (GSWE2009) [20], который определил новые тенденции в магистерском обучении. А именно перенос в магистратуру технологий разработки учебных программ на базе куррикулов с их характерными чертами — четким описанием целей и результатов обучения, детальной спецификацией объемов знаний профессионального образовательного поля, выделением обязательного набора знаний (ядра) для всех учебных программ, определением примерного перечня актуальных направлений специализации.

Несколько слов о разработке руководства GSwE2009. Этот документ стал первым стандартом куррикулума магистерского уровня, созданным в рамках нового амбициозного iSSEc-проекта (Integrated Software & Systems Engineering Curriculum (iSSEc) Project) — проекта куррикулумов по интегрированной программной и системной инженерии.

Этот проект, стартовавший в 2007 г., реализуется коалицией из академических, промышленных, правительственных и профессиональных организаций. Его основным спонсором является Министерство обороны США. Активную роль в проекте играют профессиональные организации, включая Международный совет по системной инженерии (INCOSE), Промышленную ассоциацию национальной обороны США (NDIA), Компьютерное сообщество Института инженеров по электронике и электротехнике (IEEE-CS), Ассоциацию компьютерной техники (ACM) и др. Документ GSwE2009 нацелен прежде всего на подготовку магистров в области программной инженерии с акцентом на практическую деятельность. Разработчики данного руководства уверены, что оно окажет сильное влияние на развитие магистерского образования в целом.

GSwE2009 включает описание:

- набора исходных требований к выпускникам или результатов подготовки магистров по программам, соответствующим GSwE2009 (далее — GSwE2009-программы или программы GSwE2009);
- входных требований к подготовке студентов, желающих обучаться по GSwE2009-программам;
- архитектурной модели куррикулума;
- ядра объема знаний (Core Body of Knowledge — СВОК), определяющего обязательный свод знаний для GSwE2009-программ;
- модифицированного метода Блума, используемого для спецификации учебных целей при изучении объема знаний;
- учебных курсов, содержащих материал СВОК, дополняющий свод знаний SWEBOK [19], взятый за основу содержания СВОК и др.

Объем знаний GSwE2009 (и соответственно СВОК) построен в виде четырехуровневой иерархической системы структурных элементов (дидактических единиц), включающей предметные области на высшем уровне иерархии, структурированные далее на модули знаний (второй уровень), которые детализируются до уровня тем, а темы — до уровня подтем (третий и четвертый уровни соответственно). С каждой дидактической единицей связан некоторый индекс, определяющий необходимый уровень освоения этой единицы учащимся и шкалируемый с помощью модифицированного метода Блума.

Архитектура учебных программ приведена на рис. 6.

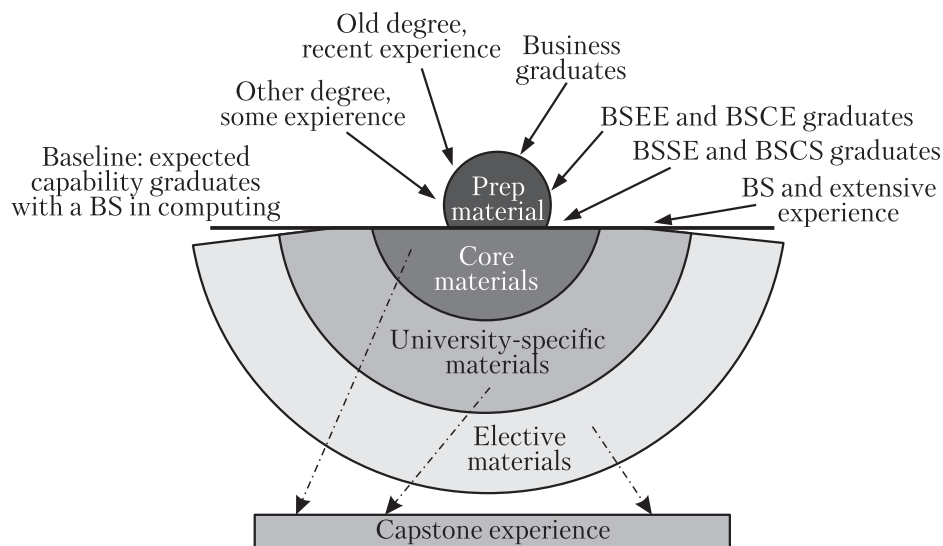


Рис. 6 Архитектура учебных программ GSwE2009

Изображенная на рис. 6 архитектура куррикулума включает:

- подготовительный материал (preparatory material), владение которым необходимо при поступлении на GSwE2009-программы;
- материалы ядра (core materials), т. е. СВОК;
- материалы университета (university-specific materials);
- материалы по выбору студента (elective materials);
- обязательный capstone-проект (mandatory capstone experience), ниже которого на рисунке простирается пространство профессиональной деятельности магистра, удовлетворяющего исходящим требованиям.

Важно отметить, что в перечне исходящих требований (или результатов подготовки) по программам GSwE2009 первым стоит требование к владению на магистерском, т. е. экспертном, уровне входящими в СВОК знаниями, формируемыми на базе свода знаний SWEВОК, дополненного рядом тем по системной инженерии, информационной безопасности, профессиональной подготовке, человеко-машинному интерфейсу, инженерной экономике, управлению рисками, качеству программного обеспечения.

Объем СВОК оценивается в 200 аудиторных или контактных часов, необходимых для его изучения (т. е. общих часов в четыре раза больше — 800), что эквивалентно 5 семестровым учебным курсам по 40 аудиторных часов за семестр (160 общих часов на каждый курс). Структура ядра показана на рис. 7 в виде

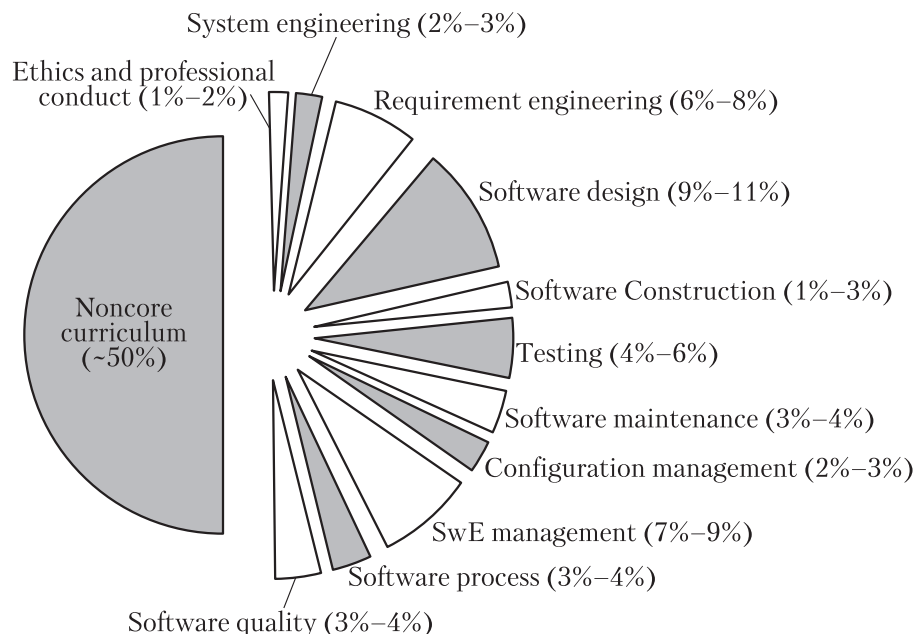


Рис. 7 Структура СВБК

правого полукруга, состоящего из секторов, соответствующих ядерной части некоторой предметной области знаний, при этом размер сектора соответствует доли этой части в процентах относительно самого ядра. Всего в ядро входят модули из 11 предметных областей, взятых в основном из SWEBOK: А. Ethics and Professional Conduct; В. System Engineering; С. Requirements Engineering; D. Software Design; E. Software Construction; F. Testing; G. Software Maintenance; H. Configuration Management (CM); I. Software Engineering Management; J. Software Engineering Process; K. Software Quality.

В заключение обзора СВБК отметим, что объем содержащегося в нем обязательного для изучения материала в 200 аудиторных часов представляется весьма значительным. Это, по существу, около 50% всей учебной программы. И это, безусловно, новое веяние в подготовке магистров. Далее анализ содержания СВБК показывает исключительно большое значение, которое отводится в GSwE2009 изучению современных международных стандартов, прежде всего в области системной и программной инженерии, включая SWEBOK, CMMI, ISO/IEC 12207, ISO/IEC 15288, пакет стандартов программной инженерии IEEE (порядка 40). Также от магистров требуется знание и рассмотренных образовательных стандартов компьютеринга.

5 Заключение

5.1 Основные принципы и особенности разработки куррикулов

Анализ рассмотренной выше методологической базы построения учебных программ по направлению «информационные технологии, или компьютеринг», позволяет увидеть общую картину процесса стандартизации учебных программ в наиболее динамично развивающейся научно-прикладной области, выявить основные тенденции, закономерности, характерные особенности этого процесса. В частности, к общим принципам его реализации можно отнести такие стороны, как целенаправленность, системность, концептуальная целостность, модульность, консорциумный подход при создании куррикулов, оценка их качества на основе консенсуса.

К важнейшим характерным чертам стандартов куррикулов относятся следующие аспекты:

- единая структура построения и единый понятийный контекст;
- знание-ориентированность — спецификация структуры и собственно объемов знаний (body of knowledge) по профилям подготовки (до уровня тем / подтем) является основным содержанием любого куррикула;
- единый способ структурирования и представления объемов знаний в виде трех- или четырехуровневой иерархической структуры — на верхнем уровне иерархии располагаются предметные области (areas) — самые крупные части объема знаний, которые подразделяются на разделы или модули знаний (units), последние, в свою очередь, разбиваются на темы (topics), в некоторых случаях темы делятся на подтемы (subtopics);
- концепция ядра — выделение в объемах знаний минимально необходимого образовательного содержания, реализация которого во всех учебных программах обеспечивает единство образовательного пространства, мобильность учащихся в рамках профиля или всего направления, гарантию качества базовой подготовки;
- четкая спецификация профессиональных характеристик профилей, системы целей обучения, итоговых профессиональных характеристик выпускников;
- рекомендации методического характера по диверсификации направлений подготовки, составлению учебных планов, компоновки курсов из модулей знаний в соответствии с выбранной педагогической стратегией реализации учебной программы, организации профессиональной практики, реализации процессов обучения;
- описание учебных курсов и пакетов курсов для различных педагогических стратегий реализации куррикулов.

5.2 Об использовании международных образовательных стандартов информационно-технологического образования в существующем методическом обеспечении российской высшей школы

Рассмотренная выше технология развития системы ИТ-образования на основе стандартизованных на международном уровне curriculums создавалась для достижения вполне очевидных целей, а именно: для глобализации процессов подготовки востребованных мировой экономикой ИТ-кадров, для обеспечения единства и однородности образовательного пространства, мобильности студентов и преподавателей, интеграции усилий академических, промышленных, коммерческих и правительственных организаций в создании и непрерывной актуализации современного методического и научного обеспечения системы ИТ-образования. Все это обуславливает необходимость развития национальной образовательной системы в области ИТ с учетом наработанного международного задела. В противном случае российское образование может быстро потерять даже надежды на конкурентоспособность на мировом образовательном рынке.

Литература

1. *Перекатов В. И.* Компьютерные дисциплины в представлении профессиональных обществ США: вехи академической легенды // Информационные технологии и вычислительные системы, 2002. № 1. С. 11–28.
2. *Перекатов В. И.* Компьютерные дисциплины в представлении профессиональных обществ США: последний curriculum? // Информационные технологии и вычислительные системы, 2002. № 4. С. 46–63.
3. Association for Computing Machinery, Curriculum Committee on Computer Science. An undergraduate program in computer science — preliminary recommendations // Comm. ACM, 1965. Vol. 8. No. 9. P. 543–552.
4. Curriculum 68. Recommendations for Academic Programs in Computer Science // Comm. ACM, 1968. Vol. 11. No. 3. P. 543–552.
5. Curriculum'78. Recommendations for the Undergraduate Program in Computer Science // Comm. ACM, 1979. Vol. 22. No. 3. P. 147–166.
6. Computing as a Discipline: Report // Comm. ACM, 1989. Vol. 32. No. 1. P. 9–23.
7. Computing Curricula 1991: Report of the ACM/IEEE-CS Joint Task Force. — IEEE Computer Society Press, 1991.
8. Computing Curricula 2001. Computer Science Volume. Association for Computing Machinery and Computer Society of IEEE. <http://www.acm.org/education/cc2001/final>.
9. Information Systems 2002 (IS2002). Association for Computing Machinery and Computer Society of IEEE.
10. Computer Engineering 2004 (CE2004). Association for Computing Machinery and Computer Society of IEEE.
11. Software Engineering 2004 (SE2004). Association for Computing Machinery and Computer Society of IEEE.

12. Information Technology 2008 (IT2008). Association for Computing Machinery and Computer Society of IEEE.
13. Computing Curricula 2005 (CC2005). Association for Computing Machinery and Computer Society of IEEE.
14. Computer Science 2008 (CS2008). Association for Computing Machinery and Computer Society of IEEE.
15. Information Systems 2010 (IS2010). Association for Computing Machinery and Computer Society of IEEE. (IS Curriculum Wiki http://blogsandwikis.bentley.edu/iscurriculum/index.php/Main_Page).
16. *Сухомлин В. А.* Реформа высшей школы — анализ итогов // Современные информационные технологии и ИТ-технологии: Сб. тр. V Междунар. науч.-практич. конф. / Под ред. В. А. Сухомлина. — М.: ИНТУИТ, 2010. С. 3–22.
17. Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain / Ed. B. S. Bloom. — Longmans, 1956.
18. Санкт-Петербургский государственный университет. Рекомендации по преподаванию информатики в университетах. — СПб., 2002.
19. SWEBOOK. <http://www.computer.org/portal/web/swebok>.
20. Graduate Software Engineering 2009(GSwE2009). Association for Computing Machinery and Computer Society of IEEE.

ИНФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ОРГАНИЧЕСКИХ МОЛЕКУЛ

*И. М. Гуревич*¹, *М. П. Евстигнеев*²

Аннотация: Излагается и обосновывается методика оценки объема информации в молекулах, в том числе тех, описание которых хранится в Protein Data Bank. Рассматривается соединение разнообразных пар оснований и приводятся оценки объемов информации в соединениях цепей ДНК для специфических спариваний и нетиповых соединений. Показано, что природа сформировала типовые соединения, связывающие основания в единое целое, структура которых содержит максимальный объем информации. Подтверждена эффективность информационного подхода к исследованию естественных (физических, химических и биологических) систем.

Ключевые слова: информация; физические системы; молекулы; ДНК; соединения

1 Введение

Для произвольных физических систем и объектов наряду с физическими характеристиками (массой, энергией, зарядом, . . .) определяются и информационные характеристики — информационная энтропия, информационная дивергенция, совместная информационная энтропия, информация связи, дифференциальная информационная емкость [1–5]. Информационная энтропия и информационная дивергенция определены на наблюдаемых (наблюдаемыми в квантовой механике называют физические величины, которые можно измерить, причем результатами измерения должны быть действительные числа) и состояниях физической системы, совместная информационная энтропия — на унитарных преобразованиях, связывающих эти состояния, а информация связи — на наблюдаемых и состояниях взаимодействующих подсистем физической системы и связывающих их унитарных преобразованиях.

Информационная энтропия (по Шеннону) N случайной величины q , с плотностью распределения (вероятности) $p(q)$ ($\int p(q) dq = 1$) равна $-\int_0^a p(q) \log_2 p(q) dq$ бит и характеризует неопределенность в q (до измерения)

¹Институт проблем информатики Российской академии наук; ООО «ГЕТНЕТ Консалтинг», iggurevich@gmail.com

²Севастопольский национальный технический университет, max_evstigneev@mail.ru

и/или информацию в q (после измерения). Пусть H — гильбертово пространство, $\psi(q) \subset H$ — вектор H , а q — обобщенная координата. Физическая система, отождествляемая с вектором (волновой функцией) $\psi(q)$, наделена вероятностной структурой: множество Ω элементарных событий (исходов) составляют значения q , а вероятностная мера задается плотностью распределения (вероятности) $p(q) = |\psi(q)|^2$, $p(x_0) dx = P\{x_0 < x < x_0 + dx\}$ — вероятность реализации состояния q при измерении физической системы; $p(q) = |\psi(q)|^2$, $\int p(q) dq = \int |\psi(q)|^2 dq = 1$. Таким образом, физическая система, отождествляемая с вектором (волновой функцией) $\psi(q)$, несет (содержит) $N = - \int_0^a p(q) \log_2 p(q) dq$ бит неопределенности (информации).

В дискретном случае $|e_i\rangle$, $i = 1, \dots, n$, — система базисных ортонормированных векторов пространства H ;

$$\langle e_j | e_j \rangle = \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$$

Физическая система отождествляется с вектором (волновой функцией) $\psi \subset H$: $\psi = \sum c_i |e_j\rangle$, $\sum_i |c_i|^2 = 1$. На ψ определено вероятностное пространство:

множество Ω элементарных событий (исходов) составляют базисные вектора $|e_i\rangle$; вероятностная мера задается квадратами модулей коэффициентов c_i : $p_i = |c_i|^2$ — вероятность реализации состояния e_i при измерении физической системы в базисе (e_i) . $\sum p_i = \sum |c_i|^2 = 1$. Информационная энтропия (по Шеннону) N случайной величины x , принимающей значения x_i с вероятностями p_i ($\sum p_i = 1$), равна $-\sum p_i \log_2 p_i$ бит и характеризует неопределенность в x (до измерения) и/или информацию в x (после измерения). Таким образом, физическая система, отождествляемая с вектором (волновой функцией) $\psi = \sum c_i |e_i\rangle$, содержит $N = -\sum p_i \log_2 p_i = -\sum |c_i|^2 \log_2 |c_i|^2$ бит неопределенности (информации).

Единицы измерения информации. Объем информации I измеряется в битах (двоичных единицах) [1, 3, 4]. Для системы с n равновероятными состояниями объем информации, получаемой при реализации одной из альтернатив, равен $\log_2 n$, а один бит — объем информации, получаемой при реализации одной из двух равновероятных альтернатив: 1 бит = $\log_2 2$. Информационные модели и законы являются содержательной интерпретацией математических выражений, формул, уравнений в информационной системе понятий, дающей объяснение результатам наблюдений. Информационные характеристики физических систем (например: объем информации в расширяющейся Вселенной, зависимость объема информации в черной дыре от ее массы) исследуются с привлечением информационных методов, с использованием общепризнанных физических моделей. Так как информационные характеристики связаны с физическими характеристика-

ми, то они исследуются одновременно (например: структура, излучение черной дыры, масса начальных неоднородностей Вселенной).

Связь между энтропией и информацией. В общем случае энтропия S системы, содержащей I бит информации, равна $kI \ln_2$, где k — постоянная Больцмана.

Связь между энергией и энтропией. Дифференциальное уравнение $dE = TdS - PdV$ — основное термодинамическое тождество, где E — энергия, T — температура, P — давление, V — объем, объединяет первый и второй законы термодинамики и справедливо для бесконечно малых обратимых процессов. Оно определяет связь между энергией и энтропией [3, 4]. Из него следует, что при постоянном объеме $dE = TdS$, а интегрирование при постоянной температуре T дает связь между энергией и энтропией: $E = TS$.

2 Методика оценки объема информации в молекулах

Вначале оценивается объем информации в объектах нижнего уровня (лептонах и кварках). Согласно основному принципу квантовой механики Цайлинге-ра [1] считаем, что в этих объектах — фундаментальных частицах — содержится 1 бит информации. Далее оценивается объем информации в объектах второго уровня, который равен сумме объемов информации в объектах нижнего уровня и информации, заключенной в структуре объекта второго уровня (мезоны, барионы). Объем информации в структуре объекта второго уровня оценивается по его волновой функции и/или по графу его структуры. Затем оценивается объем информации в объектах третьего уровня (атомах), который равен сумме объемов информации входящих в его состав объектов предыдущих уровней и информации, заключенной в структуре объекта последующего уровня (атомы). Объем информации в структуре объекта третьего уровня оценивается по его волновой функции. Затем оценивается объем информации в молекулах (объектах четвертого уровня), который равен сумме объемов информации входящих в его состав объектов предыдущих уровней и информации, заключенной в структуре молекулы. Объем информации в структуре молекулы оценивается по ее волновой функции или по соответствующему графу [3–7]. Объем информации в молекуле I_{ml} равен сумме объемов информации в атомах $N_i I_{at i}$ и информации в структуре молекулы $I_{str ml}$: $I_{ml} = \sum N_i I_{at i} + I_{str ml}$, где N_i — число атомов типа i в молекуле, $I_{at i}$ — объем информации в атоме типа i .

2.1 Оценка объема информации в структуре молекул

Предлагается использовать оценку объема информации в структуре графа, соответствующей структуре молекулы, $I_{gr} = \sum_{i=1}^m n_i \log_2 v_i$, приведенную в рабо-

тах Гуревича [2–5], где m — количество классов топологически эквивалентных вершин графа (вершин заданной степени v_i), а n_i — число вершин графа степени v_i . Данная оценка является оценкой сверху. Рассматриваемое количество направлений выхода из каждой вершины обеспечивает обход графа с использованием всех возможных путей (направлений перехода) от атома к атому. Оценка объема информации в структуре графа, соответствующего структуре молекулы, использующая степени вершин, уменьшенные на единицу, $I'_{gr} \sum n_j \log_2(v_j - 1)$, является оценкой снизу. Оценки снизу объема информации в соединениях примерно в два раза меньше оценок сверху.

Представим оценки N. Rashevsky [7] и E. Труссо [6].

Оценка Rashevsky: $I_{gr} = - \sum_{i=1}^m w_i \log_2 w_i$, где m — количество классов топологически эквивалентных вершин графа; $w_i = n_i/n$ — вероятность принадлежности вершины графа к классу i ; а n — число вершин графа. Объем информации в структуре однородного графа ($w_1 = 1$) равен нулю.

Оценка Труссо: $I_{gr} = - \sum_{i=1}^m w_i \log_2 w_i - \sum_{i=1}^m \gamma_i \log_2 \gamma_i$, где $\gamma_i = 1 - w_i$. Объем информации в структуре однородного графа равен нулю.

Рассмотрим несколько примеров.

1. Оценка объема информации в структуре линейного графа, представленного на рис. 1.

Две вершины имеют степень 1 и две вершины — степень 2. Объем информации в структуре графа равен 2 битам.

2. Оценка объема информации в структуре молекулы метана CH_4 , представленной на рис. 2.

Степени вершин графа, описывающего структуру молекулы метана, равны единице для вершин, соответствующих атомам водорода, и 4 для вершины, соответствующей атому углерода.

Объем информации в структуре молекулы метана равен 2 битам.

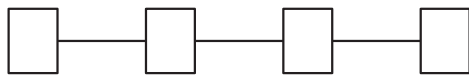


Рис. 1 Линейный граф

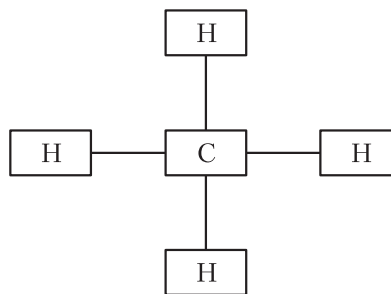


Рис. 2 Молекула метана

3. Оценка объема информации в структуре молекулы бутана C_4H_{10} ($CH_3-CH_2-CH_2-CH_3$), представленной на рис. 3. Для бутана десять вершин графа имеют степень 1 и четыре — степень 4.

Объем информации в структуре молекулы бутана составляет 8 бит.

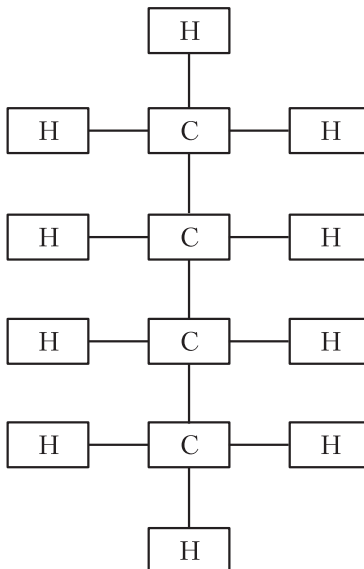


Рис. 3 Молекула бутана

2.2 Сравнение оценок объемов информации в структурах молекул по Rashevsky, Trusso и Гуревичу

Оценки объемов информации по Rashevsky, Trusso и Гуревичу в приведенных структурах молекул представлены в табл. 1.

Оценки объемов информации по N. Rashevsky и E. Trusso в графах, описывающих структуру молекул типа бутана, не зависят от числа атомов углерода:

– по Rashevsky

$$I_{gr} = -\frac{n/(3n+2) \log_2 n/(3n+2) + (2n+2)/(3n+2) \log_2(2n+2)}{3n+2};$$

– по Trusso

$$I_{gr} = -\frac{2(n/(3n+2) \log_2 n/(3n+2) + (2n+2)/(3n+2) \log_2(2n+2))}{3n+2}.$$

Таблица 1 Сравнение оценок

Наименование молекулы	Оценка объема информации в структуре молекулы, бит		
	По Rashevsky	По Trusso	По Гуревичу
Линейная молекула	1,0	2,0	2,0
Молекула метана	0,722	1,444	2,0
Молекула бутана	0,863	1,726241	8,0
Молекула типа бутана при большем числе атомов углерода	0,918	1,836	2n

Оценки объемов информации по Rashevsky и Trusso в структурах молекул, описываемых однородными графами, равны нулю, а в графах, описывающих структуры молекул типа бутана, практически не зависят от числа атомов углерода в молекуле. Это не адекватно описанию объединения молекул, так как при формировании единой молекулы из двух объем информации в структуре должен возрастать (не может уменьшаться). Следовательно, объем информации в структурах молекул и в самих молекулах целесообразно оценивать по методике Гуревича.

2.3 Оценки объема информации в соединениях цепей ДНК

Применим изложенную методику для оценки объема информации в соединениях цепей ДНК. Рассмотрим соединения цепей ДНК, обозначенные как AA_1 , AA_2 , AC_1 , AC_2 , AC_3 , AC_4 , AT_{canon} , CC_1 , CC_2 , GA_1 , GA_2 , GA_3 , GA_4 , GA_5 , GC_{canon} , $GC_{\text{noncanon1}}$, $GC_{\text{noncanon2}}$, $GC_{\text{noncanon3}}$, GG_1 , GG_2 , GG_3 , GT , TT_1 , TT_2 . Символ А обозначает основание аденин, символ С — основание цитозин, символ G — основание гуанин и символ Т — основание тимин. Пара символов обозначает спаривание соответствующих оснований [8]. Индекс canon обозначает специфическое (типовое, реализованное природой) спаривание оснований, при котором образуются две водородные связи между основаниями А (аденин) и Т (тимин) и три водородные связи между основаниями G (гуанин) и С (цитозин). Индекс noncanon обозначает нетиповое спаривание оснований. В статье рас-

Таблица 2 Характеристики атомов

Элемент	Масса, а. е.	Объем информации, бит
Водород	1	10,422
Азот	14	138,908
Углерод	12	113,064
Кислород	16	149,33

смачивается соединение разнообразных пар оснований. Индексы указывают на конкретную реализацию спаривания (соединения). Масса и объем информации в атомах, входящих в основания, указаны в табл. 2 [2–5].

2.4 Описание молекул в Protein Data Bank¹

Наименование молекулы: наименование pdb-файла.

Описание молекулы в формате pdb: формат описания приведен в табл. 3.

Оценки объема информации в атомах H, C, N, O, P и S представлены в табл. 4 [2–5].

Таблица 3 Формат описания молекул

Columns	Data type	Contents
1–6	Record name	«АТОМ»
7–11	Integer	Atom serial number
13–16	Atom	Atom name
17	Character	Alternate location indicator
18–20	Residue name	Residue name
22	Character	Chain identifier
23–26	Integer	Residue sequence number
27	AChar	Space group
31–38	Real (8.3)	Orthogonal coordinates for <i>X</i> in Angstroms
39–46	Real (8.3)	Orthogonal coordinates for <i>Y</i> in Angstroms
47–54	Real (8.3)	Orthogonal coordinates for <i>Z</i> in Angstroms
55–60	Real (6.2)	Occupancy
61–66	Real (6.2)	Temperature factor (Default = 0.0)
73–76	LString (4)	Segment identifier, left-justified
77–78	LString (2)	Element symbol, right-justified
79–80	LString (2)	Charge on the atom

2.5 Исходные данные для расчетов

Имя записи: столбцы 1–6 pdb-файла.

Нумерация атомов ($i = 1, \dots, N$), составляющих молекулу, — столбцы 7–11.

N — количество атомов в молекуле — последнее число в столбцах 7–11.

Наименования атомов, составляющих молекулу: столбцы 13–16.

Наименование основания: столбцы 18–20.

¹Protein Data Bank (PDB, <http://www.rcsb.org/pdb/home/home.do>, an Information Portal to Biological Macromolecular Structures) — хранилище для больших биологических молекул, таких как белки и нуклеиновые кислоты; является ключевым ресурсом в области структурной биологии, и крупные научные журналы в настоящее время требуют от ученых представлять свои структуры данных в PDB.

Таблица 4 Оценки объема информации

Обозначение элемента	Название элемента	Заряд (атомный номер) — число протонов, электронов	Масса	Число нейтронов	Объем информации в атомах (в неориентированных протонах, нейтронах и электронах с учетом цвета кварков), бит
1	2	3	4	5	6
H	Водород	1	1,00794	0	10,422
C	Углерод	6	12,011	5	113,064
N	Азот	7	14,006	7	138,908
O	Кислород	8	15,999	7	149,33
P	Фосфор	15	30,973	15	297,66
S	Сера	16	32,066	16	317,504

Координаты атомов:

координата x_i ($i = 1, \dots, N$) — столбцы 31–38;

координата y_i ($i = 1, \dots, N$) — столбцы 39–46;

координата z_i ($i = 1, \dots, N$) — столбцы 47–54.

Объем информации V_i в атоме i ($i = 1, \dots, N$): тип атома определяется по 13–16 столбцам pdb-файла, а объем информации для данного типа атома берется из колонки 6 табл. 4.

Постоянная Больцмана $k = 1,380662 \cdot 10^{-23}$ Дж/К [9].

2.6 Вычисление расстояний между атомами i и j , составляющими молекулу

$$R_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2.$$

2.7 Определение множества вершин, смежных данной

Смежной с вершиной i является вершина, отстоящая от нее на расстояние, не превышающее $1,5 \text{ \AA}$. Множество M_i смежных для вершины i вершин равно $\{j : R_{ij} \leq 1,5\}$ [9].

2.8 Вычисление степени вершины

S_i — степень вершины i , равная мощности $|M_i|$ множества смежных вершин [9].

2.9 Определение множеств вершин, формирующих водородные связи

Если D–H–A — это водородная связь (D — атом донора, A — атом акцептора), то расстояние между D и A не должно превышать 3,2 Å, а между H и A — не должно превышать 2,4 Å. В качестве D и A могут выступать атомы N, O, C. Рассматриваются атомы водорода H, которые ковалентно связаны с N, или O, или C. Это и есть D–H. Далее в окрестности D–H снова ищутся атомы N, O, C по указанному критерию.

2.10 Вычисление степени вершины с учетом водородных связей

Ранее полученные оценки степеней вершин уточняются с учетом водородных связей. Уточненная степень вершины для атома, входящего в водородную связь, равна полученной оценке плюс 1. Степени вершин атомов, не входящих в состав водородных связей, не изменяются [9].

2.11 Оценка объема информации в структуре молекулы V_{Str}

1. Оценка объема информации в структуре молекулы по методике Гуревича.
2. оценка объема информации в структуре молекулы по методике Rashevsky.
3. оценка объема информации в структуре молекулы по методике Труссо.

2.12 Оценка объема информации V_{At} в атомах, составляющих молекулу

$$V_{At} = \sum_{i=1}^N V_i$$

2.13 Оценка объема информации в молекуле

Объем информации в молекуле V_{Mol} равен объему информации в структуре молекулы V_{Str} (в подразд. 2.9 и 2.10 и далее в настоящем подразделе использованы оценки $V_{Str Gur} = \sum_{i=1}^m n_i \log_2 v_i$ и $V'_{Str Gur} = \sum n_j \log_2 (v_j - 1)$) плюс объем информации в атомах, составляющих молекулу, V_{At} : $V_{Mol} = V_{Str Gur} + V_{At}$ — оценка сверху, $V'_{Mol} = V'_{Str Gur} + V_{At}$ — оценка снизу.

2.14 Оценка энтропии S_{Str} структуры молекулы

Энтропия структуры молекулы равна постоянной Больцмана, умноженной на объем информации в структуре молекулы: $S = kV_{Str Gur}$ — оценка сверху; $S' = kV'_{Str Gur}$ — оценка снизу.

2.15 Оценка энтропии S_{Mol} молекулы

Энтропия молекулы равна постоянной Больцмана, умноженной на объем информации в молекуле: $S = kV_{\text{Mol}}$ — оценка сверху; $S' = kV'_{\text{Mol}}$ — оценка снизу.

3 Процесс вычислений

Оператор должен определить информационные характеристики оцениваемой молекулы, выбрать соответствующий файл в pdb-формате в Protein Data Bank, открыть его с помощью текстового редактора WordPad (MFC — Microsoft Foundation Classes) и работать с ним в формате описания молекулы в табл. 1. Если в pdb-файле содержится описание сложной системы молекул (ввиду недостаточно строгой формализации описания систем молекул), оператору необходимо предварительно разделить pdb-файл на систему pdb-файлов, в каждом из которых приводится описание отдельной молекулы.

Результаты вычислений хранятся в виде файла со следующим именем: наименование молекулы из имени исходного файла без расширения или имя молекулы из скорректированного файла плюс расширение log. Например: AA1.LOG, AC3.LOG.

Оценим объем информации в структуре соединения и в соединениях в целом [9–11] (табл. 5). Для оценок используются данные PDB. В колонке 2 табл. 5 описана структура соединения:



а в колонке 6 табл. 5 приведены степени вершин графа соединения:

Н	Н	С	Н	Н	С	Н
2	3	3	3	2	3	3

Приведенные оценки объема информации в соединениях цепей ДНК показывают, что реализованные в стандартных ДНК наиболее часто встречающиеся соединения имеют максимальный объем информации в структуре — 11,924 и 18,679 бит, что обеспечивает максимально возможное количество путей, соединяющих цепи ДНК — максимальную надежность связей между цепями ДНК.

Суммарный объем информации в атомах и в соединении в данном случае не характеризует взаимодействие между цепями ДНК.

Оценим объем информации в структуре оснований и в основаниях, формирующих соединения (табл. 6).

В табл. 6–8 пары чисел 1–4; 2–3; 3–6; 4–1 означают, что число вершин степени 1 равно 4, число вершин степени 2 равно 3, число вершин степени 3 равно 6, число вершин степени 4 равно 1.

Таблица 5 Оценки характеристик соединений цепей ДНК

Наименование	Состав	Формула	N_{At}	V_{At}	Степени вершин	$V_{str\ cв}$	$V_{соед}$
1	2	3	4	5	6	7	8
AA ₁	-HNCN- -HCN-	H ₂ N ₃ C ₂	7	663,696	HNCNHCN 2 3 3 3 2 3 3	9,924	673,620
AA ₂	-HNCN- -HNCN-	H ₂ N ₄ C ₂	8	802,604	HNCNHHCN 2 3 3 3 2 3 3 3	11,509	814,113
AC ₁	-NHN-	HN ₂	3	288,238	NHN 3 2 3	4,169	292,407
AC ₂	-NHN-	HN ₂	3	288,238	NHN 3 2 3	4,169	292,407
AC ₃	-NCNH- -HCNH-	H ₂ N ₄ C ₂	8	802,604	NCNHHCNH 2 3 3 2 3 3 3 2	11,509	814,113
AC ₄	-NHN-	HN ₂	3	288,238	NH=N 3 2 4	4,584	292,822
AT _{canon}	-HNCN- -HNCO-	H ₂ N ₃ C ₂ O	8	813,026	HNCNHNC=O 2 3 3 3 2 3 4 3	11,924	824,950
CC ₁	-NHN-	HN ₂	3	288,238	NHN 3 2 3	4,169	292,407
CC ₂	-NHO-	HNO	3	298,66	NHO 3 2 2	3,584	302,244
GA ₁	-OCNH- -NCNH-	H ₂ N ₃ C ₂ O	8	813,026	O=CNHHCNH 3 4 3 2 3 3 3 2	11,924	824,950
GA ₂	-OCNH- -NCCNH-	H ₂ N ₃ C ₃ O	9	926,09	O=CNHNCNH 3 4 3 2 4 3 3 3 2	13,924	940,014
GA ₃	-HNCN- -HNCN-	H ₂ N ₄ C ₂	8	802,604	HNCNHHCN 2 3 3 3 2 3 3 3	11,509	814,113
GA ₄	-NCCNH- -NCNH-	H ₂ N ₄ C ₃ O	9	1064,998	NCCNHHCNH 4 3 3 3 2 3 3 3 2	13,509	1078,507
GA ₅	-NHO-	HNO	3	298,66	NHO= 3 2 3	4,169	302,829
GC _{canon}	-OCNH- -NCNH- -CNHOC-	H ₃ N ₄ C ₄ O ₂	13	1327,392	O=CNHHCNHHCNHOC 3 4 3 2 3 3 3 3 3 3 2 2 3	18,679	1346,071
GC _{noncanon1}	-HNCNH- -NCO-	H ₂ N ₃ C ₂ O	8	926,09	HNCNHNCO 2 3 3 3 2 3 3 2	10,924	937,014
GC _{noncanon2}	-HNCNH- -OCN-	H ₂ N ₃ C ₂ O	8	926,09	HNCNHOCN 2 3 3 3 2 2 3 3	10,924	937,014
GC _{noncanon3}	-NHN-	HN ₂	3	288,238	NHN 3 2 3	4,169	292,407
GG ₁	-OCNH- -OCNH-	H ₂ N ₂ C ₂ O ₂	8	823,448	O=CNHO=CNH 3 4 3 2 3 3 3 2	13,924	837,372
GG ₂	-NCCO- -HNCNH-	H ₂ N ₃ C ₃ O	9	926,09	NCC=OHHCNH 3 3 4 3 2 3 3 3 2	13,509	939,599
GG ₃	-NCCO- -HNCNH-	H ₂ N ₃ C ₃ O	9	926,09	NCC=OHHCNH 3 3 4 3 2 3 3 3 2	13,509	939,599
GT	-HNCO- -HNCO-	H ₂ N ₃ C ₂ O ₂	8	823,448	HNC=OHNC=O 2 3 4 3 2 3 4 3	12,339	835,787
TT ₁	-OCNH- -OCNH-	H ₂ N ₂ C ₂ O ₂	8	823,448	O=CNHO=CNH 3 4 3 2 3 4 3 2	12,339	835,787
TT ₂	-HNCO- -HNCO-	H ₂ N ₂ C ₂ O ₂	8	823,448	HNC=OHNC=O 2 3 4 3 2 3 4 3	12,339	835,787

Замечание: N_{At} — количество атомов; V_{At} — объем информации в атомах; $V_{str\ cв}$ — объем информации в структуре связей; $V_{соед}$ — объем информации в соединении.

Таблица 6 Оценки характеристик оснований

Основание	Формулы молекул, формирующих основание	Количество атомов, формирующих основание	Масса основания, а. е.	Степени вершин атомов в основании	Объем информации в структуре основания, бит	Объем информации в основании, бит
A	$H_4N_5C_5$	14	134	1-4; 2-3; 3-6; 4-1	14,51	1301,55
C	$H_4N_3C_4O_1$	12	110	1-5; 2-2; 3-5; 4-0	9,92	1060,00
G	$H_4N_5C_5O_1$	15	150	1-4; 2-4; 3-6; 4-1	15,51	1450,88
T	$H_5N_2C_5O_2$	14	166	1-5; 2-3; 3-3; 4-3	13,75	1193,91

Таблица 7 Характеристики соединений

Обозначение соединения	Формулы молекул, формирующих соединение (левая, правая, полная)	N_{At}^I	N_{At}^{II}	N_{At}	Степени вершин атомов в первой молекуле	Степени вершин атомов во второй молекуле	Степени вершин атомов в соединении	$V_{соед}$	$V_{стр\ соед}$
1	2	3	4	5	6	7	8	9	10
AA ₁	$H_4N_5C_5$ $H_4N_5C_5$ $H_8N_{10}C_{10}$	14	14	28	1-4; 2-3; 3-6; 4-1	1-4; 2-3; 3-6; 4-1	1-6; 2-6; 3-14; 4-2	2603,09	32,19
AA ₂	$H_4N_5C_5$ $H_4N_5C_5$ $H_8N_{10}C_{10}$	14	14	28	1-4; 2-3; 3-6; 4-1	1-4; 2-3; 3-6; 4-1	1-6; 2-6; 3-14; 4-2	2603,09	32,19
AC ₁	$H_4N_3C_4O_1$ $H_4N_5C_5$ $H_8N_8C_9O_1$	12	14	26	1-5; 2-2; 3-5; 4-0	1-4; 2-3; 3-6; 4-1	1-8; 2-5; 3-12; 4-1	2361,54	26,02
AC ₂	$H_4N_3C_4O_1$ $H_4N_5C_5$ $H_8N_8C_9O_1$	12	14	26	1-5; 2-2; 3-5; 4-0	1-4; 2-3; 3-6; 4-1	1-8; 2-5; 3-12; 4-1	2361,54	26,02
AC ₃	$H_4N_3C_4O_1$ $H_4N_5C_5$ $H_8N_8C_9O_1$	12	14	26	1-5; 2-2; 3-5; 4-0	1-4; 2-3; 3-6; 4-1	1-7; 2-5; 3-13; 4-1	2361,54	27,60
AC ₄	$H_4N_3C_4O_1$ $H_4N_5C_5$ $H_8N_8C_9O_1$	12	14	26	1-5; 2-2; 3-5; 4-0	1-4; 2-3; 3-6; 4-1	1-8; 2-6; 3-10; 4-2	2361,54	25,85
AT _{canon}	$H_5N_2C_5O_2$ $H_3N_3C_5$ $H_8N_5C_{10}O_2$	14	14	28	1-5; 2-3; 3-3; 4-3	1-4; 2-3; 3-6; 4-1	1-7; 2-6; 3-11; 4-4	2207,21	31,43
CC ₁	$H_5N_3C_4O_1$ $H_4N_3C_4O_1$ $H_9N_6C_8O_2$	12	12	24	1-5; 2-2; 3-5; 4-0	1-5; 2-2; 3-5; 4-0	1-9; 2-4; 3-11; 4-0	2130,41	21,43
CC ₂	$H_5N_3C_4O_1$ $H_4N_3C_4O_1$ $H_9N_6C_8O_2$	12	12	24	1-5; 2-2; 3-5; 4-0	1-5; 2-2; 3-5; 4-0	1-8; 2-6; 3-10; 4-0	2130,41	21,85

Окончание табл. 7 на с. 320

Таблица 7 (окончание) Характеристики соединений

Обозначение соединения	Формулы молекул, формирующих соединение (левая, правая, полная)	N_{At}^I	N_{At}^{II}	N_{At}	Степени вершин атомов в первой молекуле	Степени вершин атомов во второй молекуле	Степени вершин атомов в соединении	$V_{соед}$	$V_{str\ соед}$
1	2	3	4	5	6	7	8	9	10
GA ₁	H ₄ N ₅ C ₅ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₁	14	15	29	1-4; 2-3; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-7; 3-14; 4-2	2752,42	33,19
GA ₂	H ₄ N ₅ C ₅ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₁	14	15	29	1-4; 2-3; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-12; 4-3	2752,42	33,02
GA ₃	H ₄ N ₅ C ₅ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₁	14	15	29	1-4; 2-3; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-7; 3-14; 4-2	2752,42	33,19
GA ₄	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ H ₈ N ₁₀ C ₁₀ O ₁	15	14	29	1-4; 2-4; 3-6; 4-1	1-4; 2-3; 3-6; 4-1	1-6; 2-8; 3-12; 4-3	2752,42	33,02
GA ₅	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ H ₈ N ₁₀ C ₁₀ O ₁	15	14	29	1-4; 2-4; 3-6; 4-1	1-4; 2-3; 3-6; 4-1	1-7; 2-7; 3-13; 4-2	2752,42	31,60
GC _{канон}	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₈ C ₉ O ₂	12	15	27	1-5; 2-2; 3-5; 4-0	1-4; 2-4; 3-6; 4-1	1-5; 2-8; 3-13; 4-1	2510,87	30,60
GC _{noncanon1}	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₈ C ₉ O ₂	12	15	27	1-5; 2-2; 3-5; 4-0	1-4; 2-4; 3-6; 4-1	1-8; 2-6; 3-12; 4-1	2510,87	29,02
GC _{noncanon2}	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₈ C ₉ O ₂	12	15	27	1-5; 2-2; 3-5; 4-0	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-12; 4-1	2510,87	29,02
GC _{noncanon3}	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₈ C ₉ O ₂	12	15	27	1-5; 2-2; 3-5; 4-0	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-12; 4-1	2510,87	29,02
GG ₁	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₂	15	15	30	1-4; 2-4; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-14; 4-2	2901,75	34,19
GG ₂	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₂	15	15	30	1-4; 2-4; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-14; 4-2	2901,75	34,19
GG ₃	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₂	15	15	30	1-4; 2-4; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-14; 4-2	2901,75	34,19
GT	H ₄ N ₅ C ₅ O ₁ H ₅ N ₂ C ₅ O ₃ H ₉ N ₇ C ₁₀ O ₃	15	14	29	1-4; 2-4; 3-6; 4-1	1-5; 2-3; 3-3; 4-3	1-7; 2-7; 3-11; 4-4	2644,78	32,43
TT ₁	H ₅ N ₂ C ₅ O ₂ H ₅ N ₂ C ₅ O ₂ H ₁₀ N ₄ C ₁₀ O ₄	14	14	28	1-5; 2-3; 3-3; 4-3	1-5; 2-3; 3-3; 4-3	1-8; 2-6; 3-8; 4-6	2387,81	30,68
TT ₂	H ₅ N ₂ C ₅ O ₂ H ₅ N ₂ C ₅ O ₂ H ₁₀ N ₄ C ₁₀ O ₄	14	14	28	1-5; 2-3; 3-3; 4-3	1-5; 2-3; 3-3; 4-3	1-8; 2-6; 3-8; 4-6	2387,81	30,68

Замечания: N_{At}^I — число атомов в первой молекуле; N_{At}^{II} — число атомов во второй молекуле; N_{At} — число атомов в молекулах, формирующих соединения; $V_{соед}$ — объем информации в соединении; $V_{str\ соед}$ — объем информации в структуре соединения.

Таблица 8 Объем информации, связывающей основания

Обозначение соединения	Формулы молекул, формирующих соединение (левая, правая, полная)	N_{At}^I	N_{At}^{II}	N_{At}	Степени вершин атомов в первой молекуле	Степени вершин атомов во второй молекуле	Степени вершин атомов в соединении	V_{str}^I	V_{str}^{II}	V_{str} соед	$V_{св}$
1	2	3	4	5	6	7	8	9	10	11	12
AA ₁	H ₄ N ₅ C ₅ H ₄ N ₅ C ₅ H ₈ N ₁₀ C ₁₀	14	14	28	1-4; 2-3; 3-6; 4-1	1-4; 2-3; 3-6; 4-1	1-6; 2-6; 3-14; 4-2	14,51	14,51	32,19	3,17
AA ₂	H ₄ N ₅ C ₅ H ₄ N ₅ C ₅ H ₈ N ₁₀ C ₁₀	14	14	28	1-4; 2-3; 3-6; 4-1	1-4; 2-3; 3-6; 4-1	1-6; 2-6; 3-14; 4-2	14,51	14,51	32,19	3,17
AC ₁	H ₄ N ₃ CO ₁ H ₄ N ₅ C ₅ H ₈ N ₈ C ₈ O ₁	12	14	26	1-5; 2-2; 3-5; 4-0	1-4; 2-3; 3-6; 4-1	1-8; 2-5; 3-12; 4-1	9,92	14,51	26,02	1,59
AC ₂	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ H ₈ N ₈ C ₈ O ₁	12	14	26	1-5; 2-2; 3-5; 4-0	1-4; 2-3; 3-6; 4-1	1-8; 2-5; 3-12; 4-1	9,92	14,51	26,02	1,59
AC ₃	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ H ₈ N ₈ C ₈ O ₁	12	14	26	1-5; 2-2; 3-5; 4-0	1-4; 2-3; 3-6; 4-1	1-7; 2-5; 3-13; 4-1	9,92	14,51	27,60	3,17
AC ₄	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ H ₈ N ₈ C ₈ O ₁	12	14	26	1-5; 2-2; 3-5; 4-0	1-4; 2-3; 3-6; 4-1	1-8; 2-6; 3-10; 4-2	9,92	14,51	25,85	1,42
AT _{canon}	H ₅ N ₂ C ₅ O ₂ H ₃ N ₃ C ₅ H ₈ N ₅ C ₁₀ O ₂	14	14	28	1-5; 2-3; 3-3; 4-3	1-4; 2-3; 3-6; 4-1	1-7; 2-6; 3-11; 4-4	13,75	14,511	31,43	3,17
CC ₁	H ₅ N ₃ C ₄ O ₁ H ₄ N ₃ C ₄ O ₁ H ₈ N ₆ C ₈ O ₂	12	12	24	1-5; 2-2; 3-5; 4-0	1-5; 2-2; 3-5; 4-0	1-9; 2-4; 3-11; 4-0	9,92	9,92	21,43	1,59
CC ₂	H ₅ N ₃ C ₄ O ₁ H ₄ N ₃ C ₄ O ₁ H ₈ N ₆ C ₈ O ₂	12	12	24	1-5; 2-2; 3-5; 4-0	1-5; 2-2; 3-5; 4-0	1-8; 2-6; 3-10; 4-0	9,92	9,92	21,85	2,01
GA ₁	H ₄ N ₅ C ₅ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₁	14	15	29	1-4; 2-3; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-7; 3-14; 4-2	14,51	11,924	33,19	3,17
GA ₂	H ₄ N ₅ C ₅ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₁	14	15	29	1-4; 2-3; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-12; 4-3	14,51	15,51	33,02	3
GA ₃	H ₄ N ₅ C ₅ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₁	14	15	29	1-4; 2-3; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-7; 3-14; 4-2	14,51	15,51	33,19	3,17
GA ₄	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ H ₈ N ₁₀ C ₁₀ O ₁	15	14	29	1-4; 2-4; 3-6; 4-1	1-4; 2-3; 3-6; 4-1	1-6; 2-8; 3-12; 4-3	15,51	14,51	33,02	3
GA ₅	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ H ₈ N ₁₀ C ₁₀ O ₁	15	14	29	1-4; 2-4; 3-6; 4-1	1-4; 2-3; 3-6; 4-1	1-7; 2-7; 3-13; 4-2	15,51	14,51	31,60	1,58
GC _{canon}	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₈ C ₉ O ₂	12	15	27	1-5; 2-2; 3-5; 4-0	1-4; 2-4; 3-6; 4-1	1-5; 2-8; 3-13; 4-1	15,51	9,92	30,60	5,17

Окончание табл. 8 на с. 322

Таблица 8 (окончание) Объем информации, связывающей основания

Обозначение соединения	Формулы молекул, формирующих соединение (левая, правая, полная)	N_{At}^I	N_{At}^{II}	N_{At}	Степени вершин атомов в первой молекуле	Степени вершин атомов во второй молекуле	Степени вершин атомов в соединении	V_{str}^I	V_{str}^{II}	$V_{str\ соед}$	$V_{св}$
1	2	3	4	5	6	7	8	9	10	11	12
GC _{нoнcaнoн1}	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₈ C ₉ O ₂	12	15	27	1-5; 2-2; 3-5; 4-0	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-12; 4-1	15,51	9,92	29,02	3,59
GC _{нoнcaнoн2}	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₈ C ₉ O ₂	12	15	27	1-5; 2-2; 3-5; 4-0	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-12; 4-1	15,51	9,92	29,02	3,59
GC _{нoнcaнoн3}	H ₄ N ₃ C ₄ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₂	12	15	27	1-5; 2-2; 3-5; 4-0	1-4; 2-4; 3-6; 4-1	1-8; 2-6; 3-12; 4-1	15,51	9,92	27,02	1,59
GG ₁	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₂	15	15	30	1-4; 2-4; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-14; 4-2	15,51	15,51	34,19	3,17
GG ₂	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₂	15	15	30	1-4; 2-4; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-14; 4-2	15,51	15,51	34,19	3,17
GG ₃	H ₄ N ₅ C ₅ O ₁ H ₄ N ₅ C ₅ O ₁ H ₈ N ₁₀ C ₁₀ O ₂	15	15	30	1-4; 2-4; 3-6; 4-1	1-4; 2-4; 3-6; 4-1	1-6; 2-8; 3-14; 4-2	15,51	15,51	34,19	3,17
GT	H ₄ N ₅ C ₅ O ₁ H ₅ N ₂ C ₅ O ₃ H ₈ N ₇ C ₁₀ O ₃	15	14	29	1-4; 2-4; 3-6; 4-1	1-5; 2-3; 3-3; 4-3	1-7; 2-7; 3-11; 4-4	15,51	13,75	32,43	3,17
TT ₁	H ₅ N ₂ C ₅ O ₂ H ₅ N ₂ C ₅ O ₂ H ₁₀ N ₁₀ C ₁₀ O ₄	14	14	28	1-5; 2-3; 3-3; 4-3	1-5; 2-3; 3-3; 4-3	1-8; 2-6; 3-8; 4-6	13,75	13,75	30,68	3,18
TT ₂	H ₅ N ₂ C ₅ O ₂ H ₅ N ₂ C ₅ O ₂ H ₁₀ N ₁₀ C ₁₀ O ₄	14	14	28	1-5; 2-3; 3-3; 4-3	1-5; 2-3; 3-3; 4-3	1-8; 2-6; 3-8; 4-6	13,75	13,75	30,68	3,18

Замечания: N_{At} — число атомов в обеих молекулах; V_{str}^I — объем информации в структуре 1-й молекулы; V_{str}^{II} — объем информации в структуре 2-й молекулы; $V_{св}$ — объем информации, связывающей молекулы оснований.

Оценим объем информации в структуре молекул и молекулах, формирующих соединение (табл. 7). Суммарный объем информации в соединении в данном случае не характеризует взаимодействие между цепями ДНК.

Оценим в структуре соединения объем информации, связывающей основания в единое целое. Объем информации, связывающей молекулы оснований, равен объему информации в структуре соединения минус объем информации в структуре первой молекулы, минус объем информации в структуре второй молекулы (табл. 8).

В табл. 9 представлены оценки объема информации в структурах азотистых оснований, полученные по методикам Гуревича, Rashevsky, Trucco.

Таблица 9 Объем информации в структурах оснований

Осно- вание	Степени вершин атомов в основании	Количество атомов, формирующих основание	Гуревич	Rashevsky		Trusso
A	1-4; 2-3; 3-6; 4-1	14	14,50978	1,78845	1,180726	2,969176
C	1-5; 2-2; 3-5; 4-0	12	9,924813	1,483356	1,126404	2,60976
G	1-4; 2-4; 3-6; 4-1	15	15,50978	1,806239	1,191352	2,997591
T	1-5; 2-3; 3-3; 4-3	14	13,75489	1,95919	1,229881	3,189072

В табл. 10 представлены данные, позволяющие сравнить оценки объема информации, связывающей молекулы оснований, по методикам Гуревича, Rashevsky, Trusso. Очевидно, что методики Rashevsky и Trusso не дают представления о связи цепей ДНК (более того, они дают отрицательные оценки — характеризующие отталкивание) и тем самым не адекватны задачам исследования взаимосвязей органических молекул.

В колонках 2; 6; 11 указаны объемы информации в структурах пар азотистых соединений, а в колонках 3, 4; 7, 8; 12, 13 указаны объемы информации в структурах азотистых соединений, полученные по методикам Гуревича, Rashevsky, Trusso. В колонках 5; 9; 14 указаны объемы информации в структурах соединений, связывающей основания в единое целое. Объем информации, связывающей молекулы оснований, равен объему информации в структуре соединения минус сумма объемов информации в структуре первой и второй молекул пар азотистых соединений, полученных по методикам Гуревича, Rashevsky, Trusso. В колонке 10 указана разность оценок Trusso и Rashevsky.

4 Заключение

1. Оценки объемов информации по Rashevsky и Trusso в структурах молекул, описываемых однородными графами, равны нулю, а оценки в графах, описывающих структуры молекул типа бутана, практически не зависят от числа атомов углерода в молекуле, что представляется не адекватным реальности. Приведенное сравнение на примере пар азотистых соединений показывает, что связь молекул по N. Rashevsky и E. Trusso характеризуется отрицательным объемом информации, что не адекватно описанию объединения молекул. При формировании единой молекулы из двух объем информации в структуре должен возрасти (не может уменьшаться) и объем информации в структурах молекул и в самих молекулах целесообразно оценивать по методике Гуревича.

Таблица 10 Сравнение оценок объемов информации, связывающей основания

Наименование	Гуревич				Rashevsky				+ Trucco	Trucco			
	1	2	3	4	5	6	7	8	9	10	11	12	13
AA ₁	32,19	14,51	14,51	3,17	1,724408	1,78845	1,78845	-1,85249	1,146015	2,870423	2,969176	2,969176	-3,06793
AA ₂	32,19	14,51	14,51	3,17	1,724408	1,78845	1,78845	-1,85249	1,146015	2,870423	2,969176	2,969176	-3,06793
AC ₁	26,02	9,92	14,51	1,59	1,67624	1,78845	1,483356	-1,59557	1,151446	2,827687	2,969176	2,60976	-2,75125
AC ₂	26,02	9,92	14,51	1,59	1,67624	1,78845	1,483356	-1,59557	1,151446	2,827687	2,969176	2,60976	-2,75125
AC ₃	27,60	9,92	14,51	3,17	1,647869	1,78845	1,483356	-1,62394	1,133957	2,781826	2,969176	2,60976	-2,79711
AC ₄	25,85	9,92	14,51	1,42	1,826245	1,78845	1,483356	-1,44556	1,196076	3,022322	2,969176	2,60976	-2,55661
AT _{canon}	31,43	14,51	13,75	3,17	1,845366	1,78845	1,95919	-1,90228	1,191973	3,037339	2,969176	3,189072	-3,12091
CC ₁	21,43	9,92	9,92	1,59	1,477334	1,483356	1,483356	-1,48938	1,122107	2,599441	2,60976	2,60976	-2,62008
CC ₂	21,85	9,92	9,92	2,01	1,477334	1,483356	1,483356	-1,48938	1,122107	2,599441	2,60976	2,60976	-2,62008
GA ₁	33,19	14,51	11,924	3,17	1,738525	1,78845	1,806239	-1,85616	1,155504	2,894028	2,997591	2,969176	-3,07274
GA ₂	33,02	14,51	15,51	3	1,848181	1,78845	1,806239	-1,74651	1,195361	3,043541	2,997591	2,969176	-2,92323
GA ₃	33,19	14,51	15,51	3,17	1,738525	1,78845	1,806239	-1,85616	1,155504	2,894028	2,997591	2,969176	-3,07274
GA ₄	33,02	15,51	14,51	3	1,848181	1,78845	1,806239	-1,74651	1,195361	3,043541	2,997591	2,969176	-2,92323
GA ₅	31,60	15,51	14,51	1,58	1,738525	1,78845	1,806239	-1,85616	1,192623	2,931147	2,997591	2,969176	-3,03562
GC _{canon}	30,60	15,51	9,92	5,17	1,654319	1,806239	1,483356	-1,63528	1,141236	2,795555	2,997591	2,60976	-2,8118
GC _{noncanon1}	29,02	15,51	9,92	3,59	1,698246	1,806239	1,483356	-1,59135	1,162289	2,860535	2,997591	2,60976	-2,74682
GC _{noncanon2}	29,02	15,51	9,92	3,59	1,698246	1,806239	1,483356	-1,59135	1,162289	2,860535	2,997591	2,60976	-2,74682
GC _{noncanon3}	27,02	15,51	9,92	1,59	1,698246	1,806239	1,483356	-1,59135	1,162289	2,860535	2,997591	2,60976	-2,74682
GG ₁	34,19	15,51	15,51	3,17	1,746466	1,806239	1,806239	-1,86601	1,162254	2,90872	2,997591	2,997591	-3,08646
GG ₂	34,19	15,51	15,51	3,17	1,746466	1,806239	1,806239	-1,86601	1,162254	2,90872	2,997591	2,997591	-3,08646
GG ₃	34,19	15,51	15,51	3,17	1,746466	1,806239	1,806239	-1,86601	1,162254	2,90872	2,997591	2,997591	-3,08646
GT	32,43	15,51	13,75	3,17	1,746466	1,95919	1,95919	-2,17192	1,172531	2,918997	2,997591	3,189072	-3,26767
TT ₁	30,68	13,75	13,75	3,18	1,588983	1,95919	1,95919	-2,3294	1,012747	2,60173	3,189072	3,189072	-3,77641
TT ₂	30,68	13,75	13,75	3,18	1,588983	1,95919	1,95919	-2,3294	1,012747	2,60173	3,189072	3,189072	-3,77641

2. Приведенные оценки объема информации в соединениях цепей ДНК показывают, что реализованные в стандартных ДНК соединения имеют максимальный объем информации в типовых соединениях, связывающих основания в единое целое, что обеспечивает максимально возможное число путей, соединяющих цепи ДНК, — максимальную надежность связей между цепями ДНК.
3. Суммарный объем информации в атомах и общий объем информации в данном случае не характеризует взаимодействие между цепями ДНК.
4. Предложенный авторами информационный подход к оценке взаимосвязи азотистых оснований носит универсальный характер и может быть использован для анализа характеристик и взаимосвязи произвольных молекул.

Литература

1. *Zeilinger A.* A foundational principle for quantum mechanics // *Foundations Phys.*, 1999. Vol. 29. No. 4. P. 631–643.
2. *Гуревич И. М.* Оценка объема неопределенности (информации) в элементарных частицах, атомах и молекулах // *Вестник СевНТУ. Сер. Физика и математика.* — Севастополь: СевНТУ, 2009. Вып. 99. С. 121–129.
3. *Гуревич И. М.* Информационные характеристики физических систем. — М.: 11-й формат, 2009. 170 с.
4. *Гуревич И. М.* Информационные характеристики физических систем. — Севастополь: Кипарис, 2010. 260 с.
5. *Гуревич И. М.* Оценка объема информации в структурах молекул и молекулах // *Актуальные вопросы биологической физики и химии (БФФХ-2011): Мат-лы междунар. науч.-техн. конф.* — Севастополь: СевНТУ, 2011. С. 185–187.
6. *Trusco E.* On the information content of graphs: Compound symbols; different states for each point // *Bull. Math. Biophys.*, 1956. Vol. 16. P. 237–258.
7. *Rashevsky N.* Same theorems in topology and a possible biological implication // *Bull. Math. Biophys.*, 1956. Vol. 17. P. 111–126.
8. *Льюин Б.* Гены. — М.: Мир, 1987. 544 с.
9. *Гуревич И. М., Евстигнеев М. П., Пучков М. А.* Автоматизация оценки объема информации в структурах молекул и молекулах // *Актуальные вопросы биологической физики и химии (БФФХ-2012): Мат-лы междунар. науч.-техн. конф.* — Севастополь: СевНТУ, 2012. С. 93–95.
10. *Гуревич И. М., Урсул А. Д.* Информация — всеобщее свойство материи. Характеристики. Оценки. Ограничения. Следствия. — М.: Либроком, 2012. 312 с.
11. *Гуревич И. М.* Оценка объема информации в соединениях цепей ДНК // *Актуальные вопросы биологической физики и химии (БФФХ-2012): Мат-лы междунар. науч.-техн. конф.* — Севастополь: СевНТУ, 2012. С. 90–92.

ANALYSIS AND MODELING METHODS AND TOOLS FOR INTEGRATED LOGISTIC SUPPORT OF STOCHASTIC SYSTEMS

*I. N. Sinitsyn¹, A. S. Shalamov², I. V. Sergeev³, V. I. Sinitsyn⁴, E. R. Korepanov⁵,
V. V. Belousov⁶, E. S. Agafonov⁷, and V. S. Shorgin⁸*

¹IPI RAN, sinitsin@dol.ru

²IPI RAN, a-shal5@yandex.ru

³IPI RAN, isergeev@ipiran.ru

⁴IPI RAN, vsinitsin@ipiran.ru

⁵IPI RAN, ekorepanov@ipiran.ru

⁶IPI RAN, vbelousov@ipiran.ru

⁷IPI RAN, eagafonov@ipiran.ru

⁸IPI RAN, vshorgin@ipiran.ru

Analysis and modeling methods and tools for one of the modern directions in the field of aftersale manufactured product service-integrated logistic support stochastic systems are considered. Special attention is paid to the problems of analytical modeling and operating costs minimization in hybrid and hereditary systems.

Keywords: software tools; stochastic systems; hereditary systems; stochastic shock impulse; analytical and statistical modeling; filtration; MATLAB

AUTOMATIZATION OF PROGRAMS DEBUGGING FOR RECURRENT DATAFLOW PROCESSOR

R. Zelenov¹, A. Prokofyev², and V. Volchek³

¹IPI RAN, graf.developer@gmail.com

²IPI RAN, a.a.prokofyev@mail.ru

³IPI RAN, v_volchek@inbox.ru

The paper deals with the problems of low-level programs development and the ways of their decisions for the development environment of programs for recurrent dataflow processor. Realization of errors reduction and tools of automatic localization of errors leading to exceptions creation in hardware is described.

Keywords: integrated development environment; debugging automatization; dataflow architecture; recurrence

**BUILT-IN DEBUGGING TOOLS
OF RECURRENT DATAFLOW PROCESSOR***A. Prokofyev¹, R. Zelenov², and V. Volchek³*¹IPI RAN, a.a.prokofyev@mail.ru²IPI RAN, graf.developer@gmail.com³IPI RAN, v_volchek@inbox.ru

The paper deals with the issues of observability and controllability of internal states of the recurrent dataflow processor. The implementation of tools that enable the developers to maximally quickly detect, isolate, and correct the error issues is considered. The most effective decisions considering architectural novelty and specificity of the recurrent dataflow processor organization are suggested.

Keywords: debugging; hardware; recurrence; dataflow architecture

**DATAFLOW ARCHITECTURE MODEL AND ITS USAGE
WITH A WORD RECOGNIZER PROGRAM AS AN EXAMPLE***D. Khilko¹ and Yu. Stepchenkov²*¹IPI RAN, dhilko@yandex.ru²IPI RAN, YStepchenkov@ipiran.ru

The imitational model of the dataflow recurrent architecture that is being implemented as application-specific integrated circuit at field-programmable gate array platform designed by Altera, is presented. This model lets the user to perform the debugging of the architecture's functional blocks before the hardware implementation and serves as an instrument of software development and debugging. The architecture key aspects, specialized programming language, and their projection at the model are also presented.

Keywords: imitational model; dataflow architecture; recurrence; programming

**AUTOMATION OF INFORMATION PROCESSING FOR DETECTION
OF RAIL FORMS ANOMALIES***S. K. Dulin¹, I. N. Rozenberg², and V. I. Umansky³*¹IPI RAN, s.dulin@ccas.ru²Research & Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport (JSC NIIAS), I.Rozenberg@gismps.ru³Closed Joint-Stock Company "IntechGeoTrans," umanvi@yandex.ru

Based on the analysis of experimental data, the analytical models of the information processing from the meters of inertial-satellite system, fixing the anomalies of the form of the rails are suggested. The models are described by the nonstationary differential equations with unknown parameters. The obtained equations for estimation of the state vector of an integrated navigation system allowed to develop a method for identifying the parameters of the rails anomaly, free from the need to expand the dimension of the filter and the assumption of stationarity of the identifiable parameters and providing the improvement of accuracy of identification by the use of probabilistic criteria nonlinearly dependent on the *a posteriori* probability density of the state vector.

Keywords: inertial-satellite system; rails form anomaly; *a posteriori* probability density of a state vector

METHODOLOGICAL APPROACHES TO THE SELECTION OF SENSOR TECHNOLOGIES FOR THE SITUATION CENTERS BASED ON USERS ACTIVITIES CLASSIFICATION

S. Denisov¹, A. Zatsarinny², V. Kondrashev³, and K. Chuprakov⁴

¹IPI RAN, sdn@ipi.ac.ru

²IPI RAN, azatsarinny@ipiran.ru

³IPI RAN, vd@ipi.ac.ru

⁴IPI RAN, chkos@rambler.ru

The questions of application of modern touch technologies for the systems of the situation centers to increase the efficiency of their functioning on the basis of improvement of the interface of interaction of the person making decision with information and analytical system of decision-making support are considered. The criteria of touch technologies choice for a situation center on the basis of the carried-out classification of kinds of their users activity are suggested.

Keywords: situation center; touch panel; interface of interaction; information and analytical system

MATHEMATICAL TOOLS AND SOFTWARE FOR LIGHT DEVICES MODELING

I. I. Bayneva¹ and V. V. Baynev²

¹N. P. Ogarev's Mordovian State University, baynevaii@rambler.ru

²N. P. Ogarev's Mordovian State University, bw14@mail.ru

The features of the light and light-distributing devices are described. The problems and possible solutions for their calculation and design are analyzed. The mathematical models used in the light and luminaire design with the mirror reflectors are described. The principles of work with the created software for the automated modeling and design of the light devices are considered.

Keywords: modeling; program; system; light devices; reflector; design

CITING DOCUMENTS IN PATENTS AS AN INDICATOR OF SCIENCE AND TECHNOLOGY FIELDS INTERRELATIONS

O. S. Kozhunova

IPI RAN, okozhunova@ipiran.ru

The paper is devoted to science and technology fields interrelations analysis involving patent data of the Netherlands, Belgium, Chile, and other countries, as well as to the evaluation of level of significance of scientific papers cited in the patents from the corresponding fields, including the technological ones.

Keywords: citing in patents; indicator of science and technology fields interrelations; information monitoring of science

INTELLECTUAL MECHANISMS FOR SEMANTIC SEARCHING IN INTERNET

I. P. Kuznetsov¹, M. M. Sharnin², and A. G. Matskevich³

¹IPI RAN, igor-kuz@mtu-net.ru

²IPI RAN, 1@keywen.com

³IPI RAN, xmag@mail.ru

The methods of searching in Internet based on knowledge extraction from texts are considered. During the first stage, the significant key words are extracted from the question. They are used by the program-spider to select adequate texts from Internet. In the second stage, semantic-oriented linguistic processor is used. It analyzes the text of question and selected texts to form their knowledge structures. Semantic search consists in matching these structures and forming the answer for users. Various kinds of semantic search are considered.

Keywords: semantic-oriented linguistic processor; knowledge extraction from texts; semantic search; knowledge base

**METHOD OF INFORMATION CLASSIFICATION
BASED ON HIERARCHICAL TAGS AND ITS IMPLEMENTATION
ON THE EXAMPLE OF A FAMILY ARCHIVE**

I. M. Adamovich¹, O. I. Volkov², and N. A. Markova³

¹IPI RAN, Adam@amsd.com

²IPI RAN, Volkov@amsd.com

³IPI RAN, nMarkova@ipiran.ru

The contemporary methods of classifying information are examined and their benefits and limitations are analyzed. Classification of heterogeneous information based on hierarchical tags is suggested, and the set of the problems for which this kind of classification is possible and suitable is analyzed. The implementation of the suggested approach in the program "Family archive" is examined.

Keywords: classification; hierarchical tags; family archive

**SIGNAL PROCESSING METHODS FOR LOCALIZATION
OF NONRENEWABLE REGIONS**

*T. V. Zakharova¹, S. Yu. Nikiforov², M. B. Goncharenko³, M. A. Dranitsyna⁴,
G. A. Klimov⁵, M. Sh. Khaziakhmetov⁶, and N. V. Chayanov⁷*

¹M. V. Lomonosov Moscow State University, isa@cs.msu.ru

²M. V. Lomonosov Moscow State University, nikisimonmsu@gmail.com

³M. V. Lomonosov Moscow State University, goncharenko.mir@yandex.ru

⁴M. V. Lomonosov Moscow State University, margarita13april@mail.ru

⁵M. V. Lomonosov Moscow State University, gregklimov@yandex.ru

⁶M. V. Lomonosov Moscow State University, khaziakhmetov@yandex.ru

⁷Research and Education Center of Neuro-Cognitive Studies (MEG-Center),
oia.mgppu@gmail.com

This paper is devoted to the problem of myogram and magnetoencephalogram signal synchronizing. The evoked potentials have been used to localize the brain activity areas. The method is based on magnetoencephalogram data averaging with respect to the reference points. Voluntary movements start points were indicated as reference points. The main difficulty of magnetoencephalogram averaging was related to uncertainty in identification of voluntary movements start points. This paper presents a new accurate procedure for the reference points detection based on the stochastic characteristics of myogram and magnetoencephalogram. Two-stage

procedure was suggested for voluntary movement start points localization. First, these points are identified according to myogram and/or actogram signal records and then revised according to magnetoencephalogram signal records. The obtained results can be applied in many scientific fields to solve various localization problems.

Keywords: hypothesis testing; sample window variance; sample quantile; magnetoencephalography; magnetoencephalogram; actogram; myogram

THE ARCHITECTURE OF HYBRID RECOMMENDER SYSTEM GEFEST (GENERATION–EXPANSION–FILTERING–SORTING–TRUNCATION)

Yu. S. Nefedova

M. V. Lomonosov Moscow State University; IPI RAN, y.nefedova@gmail.com

Recommender system is a relatively new class of software with the task of prediction the items (movies, music, books, news, Web sites) that would be interesting to the user, through an analysis of his actions and estimates. This paper describes the construction of such system. Designed recommendation system generates a suitable user list of recommended items. The special emphasis is made on the ability to use all the available information, including the product data (description and characteristics), user data (user profile), and data about user behavior (information about the use or purchase of goods and/or explicit estimate of goods).

Keywords: recommender systems; collaborative filtering; method of nearest neighbor; classification algorithms

NONPARAMETRIC ESTIMATION OF THE DENSITY OF A MIXTURE OF PROBABILITY LAWS

V. Yu. Korolev¹, A. Yu. Korchagin², and O. A. Moreva³

¹M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@cs.msu.su

²M. V. Lomonosov Moscow State University, sasha.korchagin@gmail.com

³Ulm University, Germany, o.moreva@gmail.com

Kernel estimates of probability densities are empirically compared with the estimates obtained by some versions of the EM (expectation-maximization) algorithm on artificial samples generated from various classes of probability distributions. A grid modification of the EM-algorithm is also considered.

Keywords: kernel density estimator; EM-algorithm; grid method of separation of mixtures

**ON THE CONSISTENCY OF THE GRID METHOD ESTIMATORS
FOR PARAMETERS OF THE SCALE MIXTURES OF NORMAL
DISTRIBUTIONS**

A. L. Nazarov

M. V. Lomonosov Moscow State University, nazarov.vmik@gmail.com

It is shown that in some cases, the grid maximum likelihood method estimators in the Gaussian mixture decomposition problems are consistent. Sufficient consistency conditions describing the algorithm of the concordant change of the structure of parametric set (grid mesh) and the sample size are presented.

Keywords: mixture of normal distributions; grid method for separation of mixtures; consistency; Kullback–Leibler distance; Skorokhod space

**OPTIMIZING THE RESOURCES OF COMPUTER COMPLEX
WITH WEAKLY CONNECTED ARCHITECTURE**

M. Ya. Agalarov

IPI RAN, murad-agalarov@yandex.ru

The problem of calculating the optimum amount of computing resources in computer complex with weakly connected architecture and poisson incoming flows of a special class of tasks is considered. Heuristic algorithm is proposed for constrained optimization of computer system resources. The results of computational experiments and evaluation of algorithms complexity are presented.

Keywords: computer system; computing resources; flow of tasks; optimizing the amount of resources

**VIRTUAL MACHINES AS AN OPTIMIZING TECHNIQUE
FOR TEACHING PROCESS**

V. Petrukhin¹, A. Shneider², and S. Luchkin³

¹IPI RAN, cokrat2@rambler.ru

²IPI RAN, alexshnd@rambler.ru

³Secondary School No. 4, Kolchugino, Vladimir Region, Russia,
info32kab@gmail.com

The feasible ways for using virtual machines in a teaching process are discussed. The specific examples of a logical structure of a computer class built on the base of virtual machines are adduced.

Keywords: optimization; virtualization

**ELECTRONIC EDUCATIONAL RESOURCES:
EFFICIENCY OF APPLICATION**

A. A. Fedoseev

IPI RAN, A.Fedoseev@ipiran.ru

Based on the analysis of the processes of studying, teaching, and learning, the information content (core) of educational process has been identified and its information model has been developed. A desirable position of the efforts on creation of electronic educational resources application has been identified in the educational process. The resources features required to ensure their positive influence on the efficiency of educational process have been defined.

Keywords: teaching; learning; motivation; perception; drill and practice; educational process; information model; electronic educational resource

**ANALYSIS OF INTERNATIONAL EDUCATIONAL STANDARDS
IN THE FIELD OF INFORMATION TECHNOLOGY**

V. A. Sukhomlin

M. V. Lomonosov Moscow State University, sukhomlin@mail.ru

The aim of this work is to analyze the current state of international curricular standards for undergraduate and graduate degree programs in information technology (or computing), as well as seeking opportunities for using their potential in the methodical basis of the national high school.

Keywords: IT education; educational standards; curriculum; information technologies; computing

INFORMATION CHARACTERISTICS OF ORGANIC MOLECULES

I. M. Gurevich¹ and M. P. Evstigneev²

¹IPI RAN; LLC “HETNET Consulting,” iggurevich@gmail.com

²Sevastopol National Technical University, Ukraine, max_evstigneev@mail.ru

The method of estimating the amount of information in the molecules, including those whose description is stored in An Information Portal to Biological Macromolecular Structures (Protein Data Bank), is described and substantiated. The connection of different base pairs is considered and the estimates of the volume of information in the compounds of the DNA strands are provided for specific pairings and atypical compounds. It is shown that nature has formed standard connections, linking the bases into a coherent whole, the structure of which contains the maximum amount of information. The effectiveness of the information approach to explore the natural (physical, chemical, and biological) systems is confirmed.

Keywords: information; physical systems; molecules; DNA; connections

ОБ АВТОРАХ

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник ИПИ РАН

Агафонов Егор Сергеевич (р. 1981) — младший научный сотрудник ИПИ РАН

Адамович Игорь Михайлович (р. 1934) — кандидат технических наук, заведующий отделом ИПИ РАН

Байнев Виталий Валерьевич (р. 1993) — студент факультета электронной техники Мордовского государственного университета им. Н. П. Огарева, г. Саранск

Байнева Ирина Ивановна (р. 1971) — кандидат технических наук, доцент светотехнического факультета Мордовского государственного университета им. Н. П. Огарева, г. Саранск

Белуосов Василий Владимирович (р. 1977) — кандидат технических наук, заведующий сектором ИПИ

Волков Олег Игоревич (р. 1964) — ведущий программист ИПИ РАН

Волчек Виктор Николаевич (р. 1985) — аспирант, инженер-исследователь ИПИ РАН

Гончаренко Мирослав Богданович (р. 1991) — студент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Гуревич Игорь Михайлович (р. 1945) — кандидат технических наук, старший научный сотрудник ИПИ РАН, главный конструктор по системам ООО «ГЕТНЕТ Консалтинг»

Денисов Сергей Анатольевич (р. 1985) — программист 1-й категории ИПИ РАН

Драницына Маргарита Александровна (р. 1983) — аспирант кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Дулин Сергей Константинович (р. 1950) — доктор технических наук, профессор, старший научный сотрудник ИПИ РАН

Евстигнеев Максим Павлович (р. 1975) — доктор физико-математических наук, профессор Севастопольского национального технического университета, Украина

Захарова Татьяна Валерьевна (р. 1967) — кандидат физико-математических наук, научный сотрудник кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Зацаринный Александр Алексеевич (р. 1951) — доктор технических наук, профессор, заместитель директора ИПИ РАН

Зеленов Роман Альбертович (р. 1986) — аспирант, инженер-исследователь ИПИ РАН

Климов Григорий Анатольевич (р. 1991) — студент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Кожунова Ольга Сергеевна (р. 1982) — кандидат технических наук, заведующая сектором ИПИ РАН

Кондрашев Вадим Адольфович (р. 1963) — старший научный сотрудник ИПИ РАН

Корепанов Эдуард Рудольфович (р. 1966) — кандидат технических наук, заведующий сектором ИПИ РАН

Корчагин Александр Юрьевич (р. 1989) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Кузнецов Игорь Петрович (р. 1938) — доктор технических наук, профессор, главный научный сотрудник ИПИ РАН

Лучкин Сергей Алексеевич (р. 1961) — учитель информатики средней школы № 4 г. Кольчугино, Владимирская область РФ

Маркова Наталья Александровна (р. 1950) — кандидат физико-математических наук, ведущий научный сотрудник ИПИ РАН

Мацкевич Андрей Георгиевич (р. 1953) — старший научный сотрудник ИПИ РАН

Морева Ольга Андреевна (р. 1989) — студентка Университета г. Ульм, Германия

Назаров Алексей Леонидович (р. 1989) — аспирант кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Нефедова Юлия Сергеевна (р. 1985) — кандидат физико-математических наук, ассистент факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

Никифоров Семен Юрьевич (р. 1992) — студент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Петрухин Владимир Сергеевич (р. 1949) — старший научный сотрудник, ИПИ РАН

Прокофьев Александр Александрович (р. 1986) — аспирант, инженер-исследователь ИПИ РАН

Розенберг Игорь Наумович (р. 1965) — доктор технических наук, первый заместитель директора ОАО «Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте» (ОАО «НИИАС»)

Сергеев Игорь Викторович (р. 1965) — кандидат технических наук, заместитель директора ИПИ РАН

Синицын Владимир Игоревич (р. 1968) — доктор физико-математических наук, доцент, заведующий отделом ИПИ РАН

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом ИПИ РАН

Степченко Юрий Афанасьевич (р. 1951) — кандидат технических наук, заведующий отделом ИПИ РАН

Сухомлин Владимир Александрович (р. 1945) — доктор технических наук, профессор кафедры автоматизации систем вычислительных комплексов факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Уманский Владимир Ильич (р. 1954) — доктор технических наук, генеральный директор Закрытого акционерного общества «ИнтехГеоТранс»

Федосеев Андрей Алексеевич (р. 1946) — кандидат технических наук, ведущий научный сотрудник ИПИ РАН

Хазиахметов Максим Шамилович (р. 1989) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Хилько Дмитрий Владимирович (р. 1987) — аспирант, инженер-исследователь ИПИ РАН

Чаянов Никита Васильевич (р. 1954) — кандидат биологических наук, ведущий научный сотрудник Научно-образовательного центра нейрокогнитивных исследований (МЭГ-Центр)

Чупраков Константин Григорьевич (р. 1985) — кандидат технических наук, ведущий математик ИПИ РАН

Шаламов Анатолий Степанович (р. 1947) — доктор технических наук, профессор, консультант ИПИ РАН

Шарнин Михаил Михайлович (р. 1959) — кандидат технических наук, старший научный сотрудник ИПИ РАН

Шнейдер Александр Юльевич (р. 1955) — старший научный сотрудник, ИПИ РАН

Шоргин Всеволод Сергеевич (р. 1978) — кандидат технических наук, старший научный сотрудник ИПИ РАН

ABOUT AUTHORS

Adamovich Igor M. (b. 1934) — Candidate of Science (PhD) in informatics, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences

Agafonov Egor S. (b. 1981) — junior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Agalarov Yaver M. (b. 1952) — Candidate of Science (PhD) in technology, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences

Baynev Vitaly V. (b. 1993) — student, Faculty of Electronic Engineering, N. P. Ogarev Mordovian State University, Saransk

Bayneva Irina I. (b. 1971) — Candidate of Science (PhD) in technology, associate professor, Faculty of Light Engineering, N. P. Ogarev Mordovian State University, Saransk

Belousov Vasiliy V. (b. 1977) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences

Chayanov Nikita V. (b. 1954) — Candidate of Science (PhD) in biology, leading scientist, Research and Education Center of Neuro-Cognitive Studies (MEG-Center)

Chuprakov Konstantin G. (b. 1985) — Candidate of Science (PhD) in technology, leading mathematician, Institute of Informatics Problems, Russian Academy of Sciences

Denisov Sergej A. (b. 1985) — programmer I, Institute of Informatics Problems, Russian Academy of Sciences

Dranitsyna Margarita A. (b. 1983) — PhD student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Dulin Sergey K. (b. 1950) — Doctor of Science in technology, professor, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Evstigneev Maxim P. (b. 1975) — Doctor of Science in physics and mathematics, professor, Sevastopol National Technical University, Ukraine

Fedoseev Andrey A. (b. 1946) — Candidate of Science (PhD) in technology, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences

Goncharenko Miroslav B. (b. 1991) — student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Gurevich Igor M. (b. 1945) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences; Chief Systems Designer, LLC “HETNET Consulting”

- Khaziakhmetov Maxim Sh.** (b. 1989) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University
- Khilko Dmitri V.** (b. 1987) — PhD student, engineer-researcher, Institute of Informatics Problems, Russian Academy of Sciences
- Klimov Gregory A.** (b. 1991) — student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University
- Kondrashev Vadim A.** (b. 1963) — senior scientist, Institute of Informatics Problems, Russian Academy of Science
- Korchagin Alexander Yu.** (b. 1989) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University
- Korepanov Eduard R.** (b. 1966) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences
- Kozhunova Olga S.** (b. 1982) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute of Informatics Problems, Russian Academy of Sciences
- Kuznetsov Igor P.** (b. 1938) — Doctor of Science in technology, professor, principal scientist, Institute of Informatics Problems, Russian Academy of Sciences
- Luchkin Sergey A.** (b. 1961) — teacher of informatics, secondary school No. 4, Kolchugino, Vladimir Region, Russia
- Markova Natalia A.** (b. 1950) — Candidate of Science (PhD) in physics and mathematics, leading scientist, Institute of Informatics Problems, Russian Academy of Sciences
- Matskevich Andrey G.** (b. 1953) — senior scientist, Institute of Informatics Problems, Russian Academy of Science
- Moreva Olga A.** (b. 1989) — student, Ulm University, Germany
- Nazarov Alexey L.** (b. 1989) — PhD student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov State University
- Nefedova Yulia S.** (b. 1985) — Candidate of Science (PhD) in physics and mathematics, assistant professor, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; senior researcher, Institute of Informatics Problems, Russian Academy of Sciences
- Nikiforov Semen Yu.** (b. 1992) — student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University
- Petrukhin Vladimir S.** (b. 1949) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences
- Prokofyev Alexander A.** (b. 1986) — PhD student, engineer-researcher, Institute of Informatics Problems, Russian Academy of Sciences

Rozenberg Igor N. (b. 1965) — Doctor of Science in technology, First Deputy Director General, Research & Design Institute for Information Technology, Signalling and Telecommunications on Railway Transport (JSC NIIAS)

Sergeev Igor V. (b. 1965) — Candidate of Science (PhD) in technology, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences

Shalamov Anatoly S. (b. 1947) — Doctor of Science in technology, professor, consultant, Institute of Informatics Problems, Russian Academy of Sciences

Sharnin Mikhail M. (b. 1959) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Shneider Alexander Yu. (b. 1955) — senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Shorgin Vsevolod S. (b. 1978) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences

Sinitsyn Vladimir I. (b. 1968) — Doctor of Science in physics and mathematics, associate professor, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences

Stepchenkov Yuri A. (b. 1951) — Candidate of Science (PhD) in technology, Head of Department, Institute of Informatics Problems, Russian Academy of Sciences

Sukhomlin Vladimir A. (b. 1945) — Doctor of Science in technology, professor, Department of Automation of Computational Systems, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Umansky Vladimir I. (b. 1954) — Doctor of Science in technology, Director General, Closed Joint-Stock Company “IntechGeoTrans”

Volchek Victor N. (b. 1985) — PhD student, engineer-researcher, Institute of Informatics Problems, Russian Academy of Sciences

Volkov Oleg I. (b. 1964) — leading programmer, Institute of Informatics Problems, Russian Academy of Sciences

Zakharova Tatyana V. (b. 1967) — Candidate of Science (PhD) in physics and mathematics, scientist, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University

Zatsarinny Alexander A. (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences

Zelenov Roman A. (b. 1986) — PhD student, engineer-researcher, Institute of Informatics Problems, Russian Academy of Sciences

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2012 г.

	№	Стр.
Агаларов М. Я. Оптимизация ресурсов вычислительного комплекса со слабосвязанной архитектурой	2	244
Агафонов Е. С. см. Сеницын И. Н.		
Адамович И. М., Волков О. И., Маркова Н. А. Метод классификации информации на основе иерархических тегов и его реализация на примере Семейного архивного фонда	2	147
Архипов О. П., Зыкова З. П. Персонафицированное преобразование представлений цветных изображений на мониторе ПЭВМ	1	22
Байнев В. В. см. Байнева И. И.		
Байнева И. И., Байнев В. В. Математические и программные средства моделирования световых приборов	2	96
Белоусов В. В. см. Сеницын И. Н.		
Белоусов В. В. см. Сеницын И. Н.		
Бенинг В. Е., Закс Л. М., Королев В. Ю. Оценки скорости сходимости распределений случайных сумм к несимметричному распределению Стьюдента	1	132
Волков О. И. см. Адамович И. М.		
Волчек В. Н. см. Зеленов Р. А.		
Волчек В. Н. см. Зеленов Р. А.		
Волчек В. Н. см. Прокофьев А. А.		
Гончаренко М. Б. см. Захарова Т. В.		
Горшенин А. К. Устойчивость масштабных смесей нормальных законов относительно изменений смешивающего распределения	1	167
Григорьева М. Е., Попов С. В. О неравномерных оценках скорости сходимости в центральной предельной теореме	1	180
Гуревич И. М., Евстигнеев М. П. Информационные характеристики органических молекул	2	308
Денисов С. А., Зацаринный А. А., Кондрашев В. А., Чупраков К. Г. Методические подходы к выбору сенсорных технологий для ситуационных центров на основе классификации видов деятельности пользователей	2	79
Драницына М. А. см. Захарова Т. В.		
Дулин С. К., Розенберг И. Н., Уманский В. И. Автоматизация обработки информации для обнаружения аномалий формы рельсовых нитей	2	58
Дурново А. А. см. Зацман И. М.		

	№	Стр.
Дьяченко Ю. Г., Морозов Н. В., Степченков Д. Ю., Степченков Ю. А. Система характеристики самосинхронных элементов	1	38
Евстигнеев М. П. см. Гуревич И. М.		
Закс Л. М. см. Бенинг В. Е.		
Захарова Т. В., Никифоров С. Ю., Гончаренко М. Б., Драницына М. А., Климов Г. А., Хазиахметов М. Ш., Чайнов Н. В. Методы обработки сигналов для локализации невосполнимых областей головного мозга	2	157
Захарова Т. В. см. Хазиахметов М. Ш.		
Зацаринный А. А. см. Денисов С. А.		
Зацман И. М., Дурново А. А. Разработка и применение программно-ориентированных индикаторов в сфере науки	1	110
Зеленов Р. А., Прокофьев А. А., Волчек В. Н. Автоматизация отладки программ для рекуррентного операционного устройства	2	29
Зеленов Р. А., Прокофьев А. А., Степченков Ю. А., Волчек В. Н. Фиксация исключительных ситуаций в рекуррентном операционном устройстве	1	49
Зеленов Р. А. см. Прокофьев А. А.		
Зыкова З. П. см. Архипов О. П.		
Климов Г. А. см. Захарова Т. В.		
Кожунова О. С. Моделирование лексической семантики в задачах компьютерной лингвистики	1	86
Кожунова О. С. Цитирование документов в патентах как индикатор взаимосвязи областей науки и технологий	2	106
Конашенкова Т. Д. см. Сеницын И. Н.		
Кондрашев В. А. см. Денисов С. А.		
Корепанов Э. Р. см. Сеницын И. Н.		
Корепанов Э. Р. см. Сеницын И. Н.		
Королев В. Ю., Корчагин А. Ю., Морева О. А. Непараметрическое оценивание функции плотности смесей вероятностных законов с помощью EM-алгоритма	2	197
Королев В. Ю. см. Бенинг В. Е.		
Корчагин А. Ю. см. Королев В. Ю.		
Кузнецов И. П., Шарнин М. М., Мацкевич А. Г. Интеллектуальные механизмы семантического поиска в сети Интернет	2	129
Кузнецов И. П. см. Сомин Н. В.		
Лучкин С. А. см. Петрухин В. С.		
Маркова Н. А. см. Адамович И. М.		
Мацкевич А. Г. см. Кузнецов И. П.		
Морева О. А. см. Королев В. Ю.		

	№	Стр.
Морозов Н. В. см. Дьяченко Ю. Г.		
Назаров А. Л. О состоятельности оценок параметров масштабных смесей нормальных распределений, получаемых с помощью сеточных методов	2	227
Нефедова Ю. С. Архитектура гибридной рекомендательной системы GEFEST (Generation–Expansion–Filtering–Sorting–Truncation)	2	176
Никифоров С. Ю. см. Захарова Т. В.		
Николаев В. Г. см. Сомин Н. В.		
Петрухин В. С., Шнейдер А. Ю., Лучкин С. А. Виртуальные машины как средство оптимизации учебного процесса	2	256
Плеханов Л. П. Иерархический метод анализа самосинхронных электронных схем	1	62
Попов С. В. см. Григорьева М. Е.		
Прокофьев А. А., Зеленов Р. А., Волчек В. Н. Встроенные средства отладки рекуррентного операционного устройства	2	38
Прокофьев А. А. см. Зеленов Р. А.		
Прокофьев А. А. см. Зеленов Р. А.		
Розенберг И. Н. см. Дулин С. К.		
Сергеев И. В. см. Сеницын И. Н.		
Сеницын В. И. см. Сеницын И. Н.		
Сеницын И. Н., Корепанов Э. Р., Белоусов В. В., Конашенкова Т. Д. Развитие математического обеспечения для анализа нелинейных многоканальных круговых стохастических систем	1	3
Сеницын И. Н., Шаламов А. С., Сергеев И. В., Сеницын В. И., Корепанов Э. Р., Белоусов В. В., Агафонов Е. С., Шоргин В. С. Методы и средства анализа и моделирования стохастических систем интегрированной логистической поддержки	2	3
Сомин Н. В., Кузнецов И. П., Шарнин М. М., Николаев В. Г. Предметные словари: назначение, особенности и перспективы	1	121
Степченков Д. Ю. см. Дьяченко Ю. Г.		
Степченков Ю. А. см. Дьяченко Ю. Г.		
Степченков Ю. А. см. Зеленов Р. А.		
Степченков Ю. А. см. Хилько Д. В.		
Сухомлин В. А. Анализ международных образовательных стандартов в области информационных технологий	2	278
Уманский В. И. см. Дулин С. К.		
Федосеев А. А. Электронные образовательные ресурсы: эффективность применения	2	263

	№	Стр.
Хазиахметов М. Ш., Захарова Т. В. Использование вейвлет-анализа в климатических исследованиях	1	153
Хазиахметов М. Ш. см. Захарова Т. В.		
Хилько Д. В., Степченко Ю. А. Модель потоковой архитектуры на примере распознавателя слов	2	48
Чаянов Н. В. см. Захарова Т. В.		
Чупраков К. Г. см. Денисов С. А.		
Шаламов А. С. см. Синицын И. Н.		
Шарнин М. М. см. Кузнецов И. П.		
Шарнин М. М. см. Сомин Н. В.		
Шестаков О. В. О точности приближения распределения оценки риска пороговой обработки вейвлет-коэффициентов сигнала нормальным законом при неизвестном уровне шума	1	142
Шмейлин Б. З. Повышение отказоустойчивости данных в кэш-памяти путем их обновления	1	74
Шнейдер А. Ю. см. Петрухин В. С.		
Шоргин В. С. см. Синицын И. Н.		

2012 AUTHOR INDEX

	No.	Page
Adamovich I. M., Volkov O. I., and Markova N. A. Method of Information Classification Based on Hierarchical Tags and Its Implementation on the Example of a Family Archive	2	147
Agafonov E. S. see Sinitsyn I. N.		
Agalarov M. Ya. Optimizing the Resources of Computer Complex with Weakly Connected Architecture	2	244
Arkhipov O. P. and Zykova Z. P. Personified Transformation of Color Images Presentations on a PC Monitor	1	22
Baynev V. V. see Bayneva I. I.		
Bayneva I. I. and Baynev V. V. Mathematical Tools and Software for Light Devices Modeling	2	96
Belousov V. V. see Sinitsyn I. N.		
Belousov V. V. see Sinitsyn I. N.		

	No.	Page
Bening V. E., Zaks L. M., and Korolev V. Yu. Estimates of the Rate of Convergence of the Distributions of Random Sums to the Skew Student Distribution	1	132
Chayanov N. V. see Zakharova T. V.		
Chuprakov K. see Denisov S.		
Denisov S., Zatsarinnyi A., Kondrashev V., and Chuprakov K. Methodological Approaches to the Selection of Sensor Technologies for the Situation Centers Based on Users Activities Classification	2	79
Diachenko Yu., Morozov N., Stepchenkov D., and Stepchenkov Yu. Tools for Self-Timed Cells Characterization	1	38
Dranitsyna M. A. see Zakharova T. V.		
Dulin S. K., Rozenberg I. N., and Umansky V. I. Automation of Information Processing for Detection of Rail Forms Anomalies	2	58
Durnovo A. see Zatsman I.		
Evstigneev M. P. see Gurevich I. M.		
Fedoseev A. A. Electronic Educational Resources: Efficiency of Application	2	263
Goncharenko M. B. see Zakharova T. V.		
Gorshenin A. K. Stability of Normal Scale Mixtures with Respect to Variations in Mixing Distribution	1	167
Grigorieva M. E. and Popov S. V. On Nonuniform Estimates of the Rate of Convergence in the Central Limit Theorem	1	180
Gurevich I. M. and Evstigneev M. P. Information Characteristics of Organic Molecules	2	308
Khaziakhmetov M. Sh. and Zakharova T. V. Application of Wavelet-Analysis to Climatic Research	1	153
Khaziakhmetov M. Sh. see Zakharova T. V.		
Khilko D. and Stepchenkov Yu. Dataflow Architecture Model and Its Usage with a Word Recognizer Program as an Example	2	48
Klimov G. A. see Zakharova T. V.		
Konashenkova T. D. see Sinitsyn I. N.		
Kondrashev V. see Denisov S.		
Korchagin A. Yu. see Korolev V. Yu.		
Korepanov E. R. see Sinitsyn I. N.		
Korepanov E. R. see Sinitsyn I. N.		
Korolev V. Yu., Korchagin A. Yu., and Moreva O. A. Nonparametric Estimation of the Density of a Mixture of Probability Laws	2	197

	No.	Page
Korolev V. Yu. see Bening V. E.		
Kozhunova O. S. Citing Documents in Patents as an Indicator of Science and Technology Fields Interrelations.....	2	106
Kozhunova O. S. Lexical Semantics Modeling in the Tasks of Computational Linguistics	1	86
Kuznetsov I. P., Sharnin M. M., and Matskevich A. G. Intellectual Mechanisms for Semantic Searching in Internet.....	2	129
Kuznetsov I. P. see Somin N. V.		
Luchkin S. see Petrukhin V.		
Markova N. A. see Adamovich I. M.		
Matskevich A. G. see Kuznetsov I. P.		
Moreva O. A. see Korolev V. Yu.		
Morozov N. see Diachenko Yu.		
Nazarov A. L. On the Consistency of the Grid Method Estimators for Parameters of the Scale Mixtures of Normal Distributions	2	227
Nefedova Yu. S. The Architecture of Hybrid Recommender System GEFEST (Generation–Expansion–Filtering–Sorting–Truncation).....	2	176
Nikiforov S. Yu. see Zakharova T. V.		
Nikolaev V. G. see Somin N. V.		
Petrukhin V., Shneider A., and Luchkin S. Virtual Machines as an Optimizing Technique for Teaching Process.....	2	256
Plekhonov L. Hierarchical Method of the Analysis of Speed-Independent Electronic Circuits	1	62
Popov S. V. see Grigorieva M. E.		
Prokofyev A., Zelenov R., and Volchek V. Built-In Debugging Tools of Recurrent Dataflow Processor	2	38
Prokofyev A. see Zelenov R.		
Prokofyev A. see Zelenov R.		
Rozenberg I. N. see Dulin S. K.		
Sergeev I. V. see Sinitsyn I. N.		
Shalamov A. S. see Sinitsyn I. N.		
Sharnin M. M. see Kuznetsov I. P.		
Sharnin M. M. see Somin N. V.		
Shestakov O. V. On the Accuracy of Normal Approximation for Risk Estimate Distribution When Thresholding Signal Wavelet Coefficients in Case of Unknown Noise Level.....	1	142
Shmeilin B. The Fault Tolerance Increasing of Cache-Memory Data by Their Refreshing.....	1	74

	No.	Page
Shneider A. see Petrukhin V.		
Shorgin V. S. see Sinitsyn I. N.		
Sinitsyn I. N., Korepanov E. R., Belousov V. V., and Konashenkova T. D. Development of Software Tools for Analysis of Nonlinear Multichannel Stochastic Systems.....	1	3
Sinitsyn I. N., Shalamov A. S., Sergeev I. V., Sinitsyn V. I., Korepanov E. R., Belousov V. V., Agafonov E. S., and Shorgin V. S. Analysis and Modeling Methods and Tools for Integrated Logistic Support of Stochastic Systems.....	2	3
Sinitsyn V. I. see Sinitsyn I. N.		
Somin N. V., Kuznetsov I. P., Sharnin M. M., and Nikolaev V. G. Subject Vocabularies: Peculiarities, Applications, and Prospects	1	121
Stepchenkov D. see Diachenko Yu.		
Stepchenkov Yu. see Diachenko Yu.		
Stepchenkov Yu. see Khilko D.		
Stepchenkov Yu. see Zelenov R.		
Sukhomlin V. A. Analysis of International Educational Standards in the Field of Information Technology	2	278
Umansky V. I. see Dulin S. K.		
Volchek V. see Prokofyev A.		
Volchek V. see Zelenov R.		
Volchek V. see Zelenov R.		
Volkov O. I. see Adamovich I. M.		
Zakharova T. V., Nikiforov S. Yu., Goncharenko M. B., Dranitsyna M. A., Klimov G. A., Khaziakhmetov, M. Sh., and Chayanov N. V. Signal Processing Methods for Localization of Nonrenewable Regions.....	2	157
Zakharova T. V. see Khaziakhmetov M. Sh.		
Zaks L. M. see Bening V. E.		
Zatsarinnyi A. see Denisov S.		
Zatsman I. and Durnovo A. Program-Oriented Indicators: Production and Application in Science	1	110
Zelenov R. see Prokofyev A.		
Zelenov R., Prokofyev A., and Volchek V. Automatization of Programs Debugging for Recurrent Dataflow Processor	2	29
Zelenov R., Prokofyev A., Stepchenkov Yu., and Volchek V. Exceptions Fixation in Recurrent Dataflow Processor	1	49
Zykova Z. P. see Arkhipov O. P.		

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий. Журнал издается на русском языке. Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

1. В журнале печатаются результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях. Публикация не должна нарушать закон об авторских правах. Направляя свою рукопись в редакцию, авторы автоматически передают учредителям и редколлегии неисключительные права на издание данной статьи на русском языке и на ее распространение в России и за рубежом. При этом за авторами сохраняются все права как собственников данной рукописи. В связи с этим авторами должно быть представлено в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи “. . .”, передаем учредителям и редколлегии журнала «Системы и средства информатики» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторских прав других лиц или организаций. Подписи авторов: (ф. и. о., дата, адрес)».

Указанное соглашение может быть представлено как в бумажном виде, так и в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности опубликования представленной статьи в открытой печати.

2. Статья подписывается всеми авторами. На отдельном листе представляются данные автора (или всех авторов): фамилия, полное имя и отчество, телефон, факс, e-mail, почтовый адрес. Если работа выполнена несколькими авторами, указывается фамилия одного из них, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет самостоятельную экспертизу присланных статей. Возвращение рукописи на доработку не означает, что статья уже принята к печати. Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.
4. Решение редакционной коллегии о принятии статьи к печати или ее отклонении сообщается авторам.
Редколлегия не обязуется направлять рецензию авторам отклоненной статьи; дискуссия с авторами по поводу отклоненных статей не ведется.

5. Корректурa статей высылается авторам для просмотра. Редакция просит авторов присылать свои замечания в кратчайшие сроки.
6. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.
Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3. Основной текст: стиль — «Обычный»: шрифт Times New Roman, размер 14 пунктов, абзацный отступ 0,5 см, 1,5 интервала, выравнивание по ширине. Рекомендуемый объем рукописи — не свыше 25 страниц указанного формата. Ознакомиться с шаблонами, содержащими примеры оформления, можно по адресу в Интернете: <http://www.ipiran.ru/journal/template.doc>.
7. К рукописи, предоставляемой в 2-х экземплярах, обязательно прилагается электронная версия статьи, как правило, в форматах MS WORD (.doc) или LaTeX (.tex), а также — дополнительно — в формате .pdf на дискете, лазерном диске или по электронной почте. Сокращения слов, кроме стандартных, не применяются. Все страницы рукописи должны быть пронумерованы.
8. Статья должна содержать следующую информацию на русском и английском языках: название, Ф.И.О. авторов, места работы авторов и их электронные адреса, подробные сведения об авторах, оформленные в соответствии с форматом, определяемым файлами http://www.ipiran.ru/journal/issues/2011_05_01/authors.asp и http://www.ipiran.ru/journal/issues/2011_01_eng/authors.asp, аннотация (не более 100 слов), ключевые слова, литература. Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в порядке их первого упоминания. В списке литературы не должно быть позиций, на которые нет ссылки в тексте статьи. Все фамилии авторов, заглавия статей, названия книг, конференций и т. п. даются на языке оригинала, если этот язык использует кириллический или латинский алфавит.
9. Присланные в редакцию материалы авторам не возвращаются.
10. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в случае больших объемов информации возможно использование общеизвестных архиваторов (ZIP, RAR);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
11. Журнал является некоммерческим изданием. Плата за публикацию с авторов не взимается, гонорар авторам не выплачивается.

Адрес редакции: Москва 119333, ул. Вавилова, д. 44, корп. 2, ИПИ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05 e-mail: rust@ipiran.ru

RUSSIAN ACADEMY OF SCIENCES
SYSTEMS AND MEANS OF INFORMATICS

SCIENTIFIC JOURNAL

Volume 22 No.2 Year 2012

Editor-in-Chief and Chair of Editorial Council
Academician I. A. Sokolov

I N T H I S I S S U E :

ANALYSIS AND MODELING METHODS AND TOOLS FOR INTEGRATED LOGISTIC
SUPPORT OF STOCHASTIC SYSTEMS

***I. N. Sinitsyn, A. S. Shalamov, I. V. Sergeev, V. I. Sinitsyn, E. R. Korepanov,
V. V. Belousov, E. S. Agafonov, and V. S. Shorgin***

AUTOMATIZATION OF PROGRAMS DEBUGGING FOR RECURRENT DATAFLOW PROCESSOR

R. Zelenov, A. Prokofyev, and V. Volchek

BUILT-IN DEBUGGING TOOLS OF RECURRENT DATAFLOW PROCESSOR

A. Prokofyev, R. Zelenov, and V. Volchek

DATAFLOW ARCHITECTURE MODEL AND ITS USAGE WITH A WORD
RECOGNIZER PROGRAM AS AN EXAMPLE

D. Khilko and Yu. Stepchenkov

AUTOMATION OF INFORMATION PROCESSING FOR DETECTION OF RAIL FORMS ANOMALIES

S. K. Dulin, I. N. Rozenberg, and V. I. Umansky

METHODOLOGICAL APPROACHES TO THE SELECTION OF SENSOR
TECHNOLOGIES FOR THE SITUATION CENTERS BASED ON USERS
ACTIVITIES CLASSIFICATION

S. Denisov, A. Zatsarinnyi, V. Kondrashev, and K. Chuprakov

MATHEMATICAL TOOLS AND SOFTWARE FOR LIGHT DEVICES MODELING

I. I. Bayneva and V. V. Baynev

CITING DOCUMENTS IN PATENTS AS AN INDICATOR OF SCIENCE
AND TECHNOLOGY FIELDS INTERRELATIONS

O. S. Kozhunova

INTELLECTUAL MECHANISMS FOR SEMANTIC SEARCHING IN INTERNET

I. P. Kuznetsov, M. M. Sharnin, and A. G. Matskevich

METHOD OF INFORMATION CLASSIFICATION BASED ON HIERARCHICAL TAGS
AND ITS IMPLEMENTATION ON THE EXAMPLE OF A FAMILY ARCHIVE

I. M. Adamovich, O. I. Volkov, and N. A. Markova

SIGNAL PROCESSING METHODS FOR LOCALIZATION OF NONRENEWABLE REGIONS

***T. V. Zakharova, S. Yu. Nikiforov, M. B. Goncharenko, M. A. Dranitsyna,
G. A. Klimov, M. Sh. Khaziakhmetov, and N. V. Chayanov***

THE ARCHITECTURE OF HYBRID RECOMMENDER SYSTEM GEFEST
(GENERATION-EXPANSION-FILTERING-SORTING-TRUNCATION)

Yu. S. Nefedova