

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

8(168)
2010

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ

Издательство "Новые технологии"

СОДЕРЖАНИЕ

ОГРАНИЗАЦИОННО-ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Карнаух Д. И. Разработка единой технологической платформы как ключевой инфраструктурный элемент построения электронного правительства 2

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

Норенков И. П., Уваров М. Ю. Поиск решений по прецедентам на основе кластеризации онтологий 8
Карпов А. А., Цирульник Л. И., Железны М. Разработка компьютерной системы "говорящая голова" для аудиовизуального синтеза русской речи по тексту . . 13
Попов С. И., Попов Ю. И., Шалыто А. А. Задача о муравьеде и муравьях 18

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

Жук Д. М., Маничев В. Б., Ильницкий А. О. Методы и алгоритмы решения дифференциально-алгебраических уравнений для моделирования систем и объектов во временной области. Часть 2 23
Черемисинова Л. Д. Оценка энергопотребления КМОП-схем на логическом уровне 27

КОДИРОВАНИЕ И ОБРАБОТКА СИГНАЛОВ

Кухаренко Б. Г. Байесовская фильтрация в технологии спектрального анализа на основе быстрого преобразования Фурье 36
Агиевич С. Н. Применение локальных сплайнов для оценивания параметров распределений 42

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

Кравченко Т. К. Разработка системы ключевых показателей для оценки эффективности процессов службы информационных технологий 48
Димитриенко Ю. И., Димитриенко О. Ю. Кластерно-континуальное моделирование в экономике на основе методов механики многомерных сплошных сред 54
Штрик А. А. Анализ воздействия кризиса 2009 года на состояние и перспективы развития ИТ-рынка России. 63

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В МЕДИЦИНЕ

Кузнецов А. А. Информационные технологии донозологической диагностики 68
Митрушкин Е. И., Робский С. Ю. Оценка вклада информационной системы в эффективность медицинского учреждения 74

ИНФОРМАЦИЯ

Государственный университет Высшая школа экономики подписал соглашение с IEEE CS 78
Contents 79
Приложение. Макаров И. М., Лохин В. М., Манько С. В., Романов М. П., Кадочников М. В. Технологии обработки знаний в задачах управления автономными мехатронно-модульными реконфигурируемыми роботами

Главный редактор
НОРЕНКОВ И. П.

Зам. гл. редактора
ФИЛИМОНОВ Н. Б.

Редакционная
коллегия:

АВДОШИН С. М.
АНТОНОВ Б. И.
БАТИЩЕВ Д. И.
БАРСКИЙ А. Б.
БОЖКО А. Н.
ВАСЕНИН В. А.
ГАЛУШКИН А. И.
ГЛОРИОЗОВ Е. Л.
ГОРБАТОВ В. А.
ДОМРАЧЕВ В. Г.
ЗАГИДУЛЛИН Р. Ш.
ЗАРУБИН В. С.
ИВАННИКОВ А. Д.
ИСАЕНКО Р. О.
КОЛИН К. К.
КУЛАГИН В. П.
КУРЕЙЧИК В. М.
ЛЬВОВИЧ Я. Е.
МАЛЬЦЕВ П. П.
МЕДВЕДЕВ Н. В.
МИХАЙЛОВ Б. М.
НЕЧАЕВ В. В.
ПАВЛОВ В. В.
ПУЗАНКОВ Д. В.
РЯБОВ Г. Г.
СОКОЛОВ Б. В.
СТЕМПКОВСКИЙ А. Л.
УСКОВ В. Л.
ФОМИЧЕВ В. А.
ЧЕРМОШЕНЦЕВ С. Ф.
ШИЛОВ В. В.

Редакция:

БЕЗМЕНОВА М. Ю.
ГРИГОРИН-РЯБОВА Е. В.
ЛЫСЕНКО А. В.
ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://www.informika.ru/text/magaz/it/> или <http://novtex.ru/IT>.

Журнал включен в систему Российского индекса научного цитирования.

Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

ОГРАНИЗАЦИОННО-ИНФОРМАЦИОННЫЕ СИСТЕМЫ

От редакции

Одной из важных и актуальных сфер применения информационных технологий становится создание и развитие электронного правительства. Под электронным правительством понимается такая форма организации государственного управления и деятельности органов государственной власти, при которой за счет широкого применения информационно-коммуникационных технологий обеспечивается высокий уровень эффективности принимаемых решений, оперативности и удобства получения организациями и гражданами государственных услуг и информации о результатах деятельности государственных органов. Обычно термином "электронное правительство" обозначают и собственно информационные системы, поддерживающие деятельность органов государственной власти.

Актуальность проблем электронного правительства обусловлена тем, что государственное управление без целенаправленного применения информационных технологий не может соответствовать современному состоянию производственных ресурсов и социальных отношений и, следовательно, не может быть эффективным.

Сегодня проблемы электронного правительства находятся на начальной стадии постановки задач. Для их решения необходимо развитие методов интеллектуальных систем поддержки принятия решений, средств семантического Веб, технологий интеграции информационных систем, адаптации программного обеспечения к изменяющимся требованиям и условиям, повышения квалификации пользователей и т. п. Сложность и важность решения возникающих задач требуют значительных усилий соответствующих структур и организаций, широкого обсуждения научной общественностью путей реализации электронного правительства.

Хотя часть высказанных в статье Д. И. Карнауха рекомендаций и предложений носит частный и дискуссионный характер, основная мысль автора о необходимости системного подхода к решению проблем электронного правительства заслуживает безусловной поддержки. Публикуя статью Д. И. Карнауха, редакция надеется на привлечение внимания специалистов в области информационных технологий к этой важной теме, в том числе на появление на страницах журнала новых публикаций по проблемам электронного правительства.

УДК 004.41 + 004.75

Д. И. Карнаух, канд. эконом. наук, нач. отдела информационно-аналитических систем,
ЗАО "Строительное управление № 155" (ЗАО СУ-155),
e-mail: qdaniel@mail.ru

Разработка единой технологической платформы как ключевой инфраструктурный элемент построения электронного правительства

Проведен обзор существующих проблем в области автоматизации государственного управления в Российской Федерации, проанализированы недостатки реализации концепции "электронного правительства". Предложен подход к построению электронного правительства в части, касающейся автоматизации процессов государственного управления на основе разработки единой технологической платформы.

Новизна предложенного подхода заключается в идее интеграции всех субъектов государственного управления и различных отраслевых программных модулей, баз данных и иных информационных ресурсов в едином информационном пространстве, что позволит сформировать инфраструктурную основу электронного правительства и обеспечит возможности по его устойчивому развитию.

Область применения научных результатов статьи — построение единой технологической платформы в качестве инфраструктурной основы электронного правительства; формирование системы требований к разработке и интеграции модулей различных экономических отраслей и функциональных типов для обеспечения условий устойчивого развития электронного правительства.

Ключевые слова: государственное управление, электронное правительство, информационное обеспечение, единая технологическая платформа, отрасли экономики, интеграция, единое информационное пространство

Введение

Одним из важных факторов конкурентоспособности любого государства на мировой экономической арене является уровень информационной обеспеченности функций государственного управления. Под информационной обеспеченно-

стью в данном случае следует понимать наличие информационных систем, обеспечивающих полный цикл процессов государственного управления в какой-либо экономической или административной сфере, связанных с информационным взаимодействием субъектов системы управления,

планированием, прогнозированием, мониторингом и анализом информации.

Уровень информационной обеспеченности государственного управления в России по признанию многих аналитиков, управленцев и государственных служащих, включая президента РФ, является недостаточным, многие положения Концепции формирования электронного правительства недостигнуты и уже очевидно, что не будут достигнуты в срок. В то же время и в США, и в Европе, при большей эффективности и информационной прозрачности работы государственного аппарата, в настоящее время активизируются дополнительные усилия по повышению набора и качества электронных услуг и интерактивных сервисов, предоставляемых государством населению. Имеющая место тенденция свидетельствует о том, что нашей стране необходимо активизировать усилия по достижению запланированных показателей развития информационного общества и государства, для того чтобы не допустить снижения конкурентоспособности нашей страны по отношению к США и развитым европейским странам.

Информационная обеспеченность государственного управления достаточно емкое понятие, которое охватывает не только процессы взаимодействия государства с населением (так называемые информационные и интерактивные сервисы), но и внутренние информационные процессы, связанные с управлением отдельными экономическими отраслями. Уровень качественной автоматизации этих процессов вызывает определенное беспокойство, поскольку число информационных систем в различных государственных ведомствах и учреждениях постоянно растет, а интегрируемость информационных потоков и способность персонала к эксплуатации всего парка систем снижается.

Анализ существующих недостатков в комплексной информатизации процессов государственного управления в Российской Федерации

Если проанализировать состояние информационной обеспеченности государственного управления в такой ключевой социальной сфере как здравоохранение, то становится ясно, что наличие разрозненных нестандартизированных информационных систем, охватывающих только определенные уровни управления (например федеральный уровень), либо определенный набор учреждений (например министерство на федеральном уровне или определенные учреждения на региональном уровне), либо ограниченный набор процессов управления, является недостаточным для полноценного автоматизированного управления отраслью. Для автоматизации государственного управления

в масштабах всей страны нужен принципиально новый, системный подход, связанный с полным охватом всех взаимодействующих субъектов, основанный на стандартизации и комплексной интеграции информационных ресурсов и потоков.

Приведенный пример можно транслировать на большинство других отраслей экономики, в которых ситуация с информационной обеспеченностью обстоит не лучшим образом: образование, строительство и ЖКХ, промышленность и др.

К основным проблемам, связанным с недостаточной автоматизацией процессов государственного управления, можно отнести:

- невозможность оперативного нисходящего планирования: от федеральных министерств и ведомств к региональному уровню и конкретным учреждениям;
- невозможность оперативного анализа состояния отрасли;
- невозможность централизованного мониторинга системы показателей на региональном уровне и федеральном уровне;
- недостаточная согласованность действий федеральных ведомств при планировании государственных расходов и управленческих мероприятий;
- недостаточная согласованность действий федерального и регионального уровня государственного управления;
- отсутствие автоматизированных средств комплексного моделирования социальной и экономической ситуации на стыке различных экономических отраслей.

Описанные проблемы зачастую проявляются в открытой межведомственной и межуровневой борьбе за сферы влияния и ресурсы и ведут к снижению прозрачности информационных процессов, повышению риска принятия ошибочных решений, снижению оперативности управления и существенному увеличению государственных издержек.

Несмотря на перечисленные проблемы, положительные моменты в информатизации процессов государственного управления все же есть.

Практически во всех государственных учреждениях установлено программное обеспечение для ведения кадрового и бухгалтерского учета, во многих министерствах реализованы системы электронного документооборота. Все эти решения можно отнести к классу типовых стандартизированных решений, с развитой документацией и средствами поддержки и обучения.

Но если говорить о специализированных программных решениях по автоматизации государственного управления конкретными экономическими отраслями, то очевидно, что функционал таких решений более требователен к квалификации пользователей: к знаниям методологии работы с

информацией и методологии коллективного, в том числе и межведомственного, взаимодействия.

Следует признать, что существующие специализированные решения, при всех их недостатках, все же позволяют решать определенный класс задач, но в значительной мере не имеют развитой системы поддержки и обучения пользователей, а в силу недостаточной стандартизации и слабой интегрируемости с информационными системами других учреждений и ведомств, не позволяют говорить о построении методологии коллективного информационного взаимодействия в терминах единого информационного пространства.

По мере происходящего увеличения числа специализированных отраслевых программных решений, а во многих министерствах программные решения внедряются регулярно, способность пользователей к эксплуатации этих решений (в силу их многообразия и периодических замещений) снижается, требования к общей квалификации пользователей растут, а расходы на сопровождение информационных систем увеличиваются.

В дополнение необходимо отметить, что ходом реализации концепции электронного правительства недоволен не только президент РФ, но и обычные граждане. Позитивные ожидания простых граждан от процесса создания электронного правительства, о чем свидетельствуют многочисленные опросы, снижаются.

Разработка подхода к формированию электронного правительства на основе единой технологической платформы

Электронное правительство можно условно разделить на две составляющие:

- государственные услуги населению и организациям (взаимодействие государственных и муниципальных учреждений, а также учреждений, работающих в системе государственных услуг, с населением и организациями), которые включают как информационные услуги, так и интерактивные;
- государственное управление; эффективность работы государственного аппарата связана в большей степени именно с этим направлением.

В дальнейшем мы сконцентрируемся именно на втором направлении.

Информационной основой электронного правительства должна стать единая технологическая платформа, позволяющая интегрировать различные программные модули. Под единой технологической платформой в данном случае понимается не программная архитектура (например, Microsoft .NET или 1С Предприятие), а каркасное приложение,

которое может служить основой для интеграции разнообразных программных модулей.

Модули можно разделить:

- **по сферам использования** (их можно соотнести с отраслями экономики: здравоохранение, образование, строительство и ЖКХ, промышленность и т. д.);
- **по функциональным типам** (модули прогнозирования и моделирования трендов на основе сценарных условий, модули планирования, модули мониторинга, модули план-фактного анализа и контроля и др.).

Поскольку речь идет о единой технологической платформе, без которой немыслима действенная интеграция различных информационных ресурсов, распределенных по различным министерствам и ведомствам, а также зачастую разделенных по функциональным типам, то встает вопрос о разработке унифицированных механизмов, которые бы позволили подключать в единую технологическую платформу различные отраслевые модули каждого функционального типа. В силу специфики функциональной типизации унифицированные механизмы платформы могут быть разделены на **логические слои, соответствующие каждому классу решаемых задач**. Такой подход позволит существенно улучшить возможности модернизации и распределенного масштабирования как самой технологической платформы, так и ее компонентов.

Приоритетом при разработке единой технологической платформы как инфраструктурной основы построения электронного правительства должно быть использование стандартизированных, и преимущественно открытых, технологий. Использование открытых технологий позволит значительно снизить зависимость от разработчиков коммерческих пакетов создания информационных систем и обеспечить соответствие принятым правительством и президентом РФ решениям по обеспечению национальной безопасности в сфере информационных технологий на основе внедрения открытого и свободного программного обеспечения в государственных и бюджетных организациях.

Ключевым техническим аспектом при разработке единой технологической платформы является выбор программной архитектуры. Такой архитектуры, которая бы не только поддерживала модульность, но и позволяла интегрировать распределенные программные решения и базы данных и обеспечить масштабы поставленной задачи по построению электронного правительства, в первую очередь, с точки зрения числа пользователей и их территориальной распределенности, а также объемов обрабатываемой информации. Идеальным вариантом такой архитектуры является

Концептуальная структура единой технологической платформы

Функции	Отрасль экономики			
	Здравоохранение	Образование	Строительство и ЖКХ	Промышленность
Прогнозирование				
Сценарное моделирование				
Планирование				
Мониторинг				
План-фактный анализ				
Контроль				

ся "архитектура тонкого клиента" — с использованием сети Интернет в качестве коммуникационной инфраструктуры и интернет-браузера в качестве основного инструмента пользовательского взаимодействия. "Архитектура тонкого клиента" во многом решает задачи масштабирования и централизованного обслуживания единой технологической платформы.

Важными техническими аспектами разработки единой технологической платформы являются выбор базового языка программирования и выбор системы управления базами данных. Выбор языка программирования определяет производительность платформы, ее независимость от операционной системы, гибкость ее сопровождения и развития, способность к интеграции новых модулей и сторонних информационных ресурсов. Ярким примером открытого стандарта среди языков программирования служит язык *Java*, основанный на принципах *Open Source* (программное обеспечение с открытым исходным кодом). Развитием языка *Java* и поддержкой сообщества независимых разработчиков, которое насчитывает по всему миру более 5 млн человек, занимается *Oracle Corporation* — правопреемник создателя языка *Java* корпорации *Sun Microsystems*. Система управления базами данных необходима для интеграции отраслевых информационных ресурсов, информационные блоки которых могут храниться как в централизованном банке данных, так и в распределенных региональных базах данных. Поскольку при выборе СУБД ставится вопрос о сохранности критичной для государства информации, то к СУБД предъявляют достаточно жесткие требования по обеспечению отказоустойчивости. В этом отношении наилучшим примером СУБД служит *Oracle*. Несмотря на то, что СУБД *Oracle* является коммерческой, она зарекомендовала себя как наилучшее решение для управления базами данных промышленного масштаба как с точки зрения надежности, так и с точки зрения производительности и масштабируемости.

Рассмотрев технические аспекты как рекомендации при выборе методов практического воплощения технологической платформы, перейдем к рассмотрению наиболее значимого концептуального момента — структуры единой технологической платформы. Для того чтобы лучше понять структуру платформы, построим ее наглядное отображение в виде таблицы 1 с определением логических слоев, отражающих укрупненный функциональный состав, и примерами сфер применения единой технологической платформы.

В приведенной таблице показан укрупненный пример структуры технологической платформы. В реальном воплощении эта структура будет намного сложнее и многообразнее, в том числе в си-

лу пересечения систем функционирования различных отраслей.

Например, решение задачи по моделированию уровня занятости населения необходимо проводить на стыке отраслей, выбрав в качестве базовой — образование (как отрасль, дающую стране специалистов определенного профиля и категории). Остальные же отрасли необходимо рассматривать в качестве потребителей рабочей силы: это и промышленность, и здравоохранение, и то же образование, и государственная служба и т. д.

Другой пример — анализ влияния заболеваемости на экономику промышленных предприятий. Здесь нужно учитывать и информацию системы здравоохранения, и статистическую информацию о болезнях сотрудников, собираемую на промышленных предприятиях. Для анализа этой информации должен быть разработан специальный алгоритм, который учитывал бы не только выбытие сотрудников по болезни, но и такие факторы, как временное замещение заболевших сотрудников здоровыми (с повышением нагрузки на последних), презентизм (продолжение нахождения заболевшего сотрудника на рабочем месте с естественным снижением его трудоспособности и повышением риска заражения других сотрудников).

Еще один пример касается реализации социальных функций государства по жилищному направлению. На данный момент правительство не располагает единой базой проектируемых, строящихся и построенных объектов жилой недвижимости, с использованием которой можно было бы оперативно формировать государственное задание на строительство и закупку квартир с заданными параметрами в разрезе регионов и конкретных поселений. Поэтому анализ объемов необходимого жилищного строительства по государственному заказу для выделения очередникам в части соотношения с возможностями компаний-застройщиков существенно затруднен. При заведомо более низких темпах строительства социаль-

Схема вертикального и горизонтального развития единой технологической платформы на примере отрасли здравоохранения

Уровень управления	Субъекты
Федеральный	Минздрав, Минфин, Минэкономразвития, Федеральный фонд ОМС. FN (число субъектов) = 4
Региональный N (число регионов)	Территориальные фонды ОМС, Департаменты здравоохранения. RN (число субъектов) = 2
Учреждений N $\sum_{i=1} Q_i$	Лечебно-профилактические учреждения различного уровня подчинения (федеральные, региональные, муниципальные, ведомственные, коммерческие — в системе ОМС), Страховые медицинские организации. Q_i (число субъектов)

ного жилья, чем это необходимо, правительство должно иметь возможность оптимизации планов за счет оперативного моделирования: в каких регионах, какие классы жилых помещений и в каких количествах нужно строить в следующем году, через год, через два года и т. д.

Из вышеизложенного не следует делать вывод, что правительство контролирует или должно контролировать все социально-экономические отрасли и все процессы, к ним относящиеся. Степень участия государства в различных отраслях экономики может варьироваться: в здравоохранении, например, на данном этапе без всестороннего государственного контроля обойтись просто невозможно (государство выступает в качестве гаранта и менеджера процессов в области оказания медицинской помощи населению в рамках Конституции РФ и Программы государственных гарантий оказания гражданам РФ бесплатной медицинской помощи), в строительстве же государство выступает в качестве заказчика социального жилья и надзорного органа (разрешение на строительство, разрешение на ввод в эксплуатацию и др.).

Вмешательство государства в управление отраслями экономики не должно быть однозначно связано с мониторингом состояния отраслей (в основном это касается регулярного сбора статистики), поскольку постоянный мониторинг ключевых показателей позволяет осуществлять оперативную координацию правительственных инициатив и реформ даже для тех отраслей, где участие государства как контролирующего органа незначительно.

Приведенная структура наглядно демонстрирует принцип модульности единой технологической платформы, что определяет необходимые условия для ее масштабирования.

Единая технологическая платформа должна иметь способность к горизонтальному и вертикальному развитию. Выполнение этого условия является необходимым для построения электронного правительства на базе единого информационного пространства, которое позволит объединить все субъекты государственного управления, распределенные как по уровням вертикали управления — от федерального уровня до уровня учреждений, так и в рамках одного уровня управления (например, в управлении системой здравоохранения на федеральном уровне участвует не только Министерство здравоохранения, но и Министерство финансов, Министерство экономического развития, Федеральный фонд обязательного медицинского страхования). Схема вертикальных и горизонтальных связей единой технологической платформы применительно к государственному управлению здравоохранением в укрупненном виде представлена в табл. 2.

Число субъектов управления регионального уровня кратно числу субъектов (регионов) РФ (N). Число субъектов управления уровня учреждений определяется как сумма числа учреждений в каждом из регионов (Q_i) (действительно для лечебно-профилактических учреждений; страховые медицинские организации могут работать в нескольких регионах одновременно).

При числе регионов, равном 83, и общем числе лечебно-профилактических учреждений (по данным Росстата на конец 2008 года), равном 6500, можно провести предварительную оценку числа субъектов управления всех уровней в здравоохранении: $4 + 83 \cdot 2 + 6500 = 6670$. В каждом из субъектов управления может быть от одного (для некоторых небольших ЛПУ) до нескольких десятков (для отдельных министерств) пользователей. Цифры, полученные только для одной отрасли — для здравоохранения — свидетельствуют об исключительной масштабности задачи автоматизации функций государственного управления.

В графе "Субъекты" представлены субъекты управления для каждого уровня управления. Горизонтальные связи, поддержка которых должна быть заложена в единую технологическую платформу, определяют возможности взаимодействия одноуровневых субъектов. Особенно актуальна проблема взаимодействия на федеральном уровне, поскольку между министерствами до сих пор в качестве основы информационного взаимодействия практикуется файловый обмен и практически не используются средства корпоративного информационного взаимодействия.

Необходимо отметить, что обязательным требованием к единой технологической платформе является ее универсальная способность к интеграции модулей и функциональных блоков, а следовательно, неотъемлемой частью платформы должна стать подсистема метаданных, которая предназначена для логического и структурного

описания интегрируемых отраслевых и технических модулей и функций, а также взаимосвязей между ними.

Процесс разработки единой технологической платформы может быть разбит на следующие этапы.

1. *Концептуальное проектирование и разработка методического обеспечения.* Оно включает определение смысловых блоков и функциональных модулей, выделение основных информационных сущностей, взаимосвязей между ними и их значимых характеристик, первичное определение категорий пользователей (детальная проработка прав доступа осуществляется при разработке и интеграции конкретных модулей технологической платформы). Концептуальное проектирование во многом выполняет роль детальной постановки задачи, а также позволяет сформировать систему требований к разработке и интеграции модулей и функциональных блоков. Поскольку от того, насколько детально проработаны эти требования, зависит универсальность единой технологической платформы, то требования к разработке и интеграции новых модулей и функциональных блоков должны быть изложены в виде подробного методического описания.

2. *Логическое проектирование.* Оно включает: построение таблиц на основе выделенных информационных сущностей; определение показателей — полей таблиц; описание характера взаимосвязи таблиц; определение требований к поддержке логической целостности базы данных, в том числе на уровне отдельных информационных сущностей; логическое описание функций на уровне блок-схем и иных логических схем, необходимых для последующего формирования программных алгоритмов.

3. *Формулировка требований к аппаратному обеспечению для построения программно-аппаратного комплекса.* На этом этапе важно предусмотреть гибкие возможности по масштабированию технологической платформы при прогнозируемом росте числа пользователей, модулей, функциональных блоков, хранимой и обрабатываемой информации.

4. *Физическая реализация* — это развертывание аппаратного обеспечения и вспомогательных систем, создание структуры и первичное наполнение базы данных, разработка интерфейса, развертывание единой технологической платформы на установленном оборудовании.

Поскольку на практике одномоментная разработка модулей различных функциональных типов применительно ко всем отраслям экономики не представляется реальной, предполагается, что процесс разработки модулей осуществляется поэтапно и, более того, по циклическому принципу:

постановка задачи — разработка — тестирование и оценка результатов — отладка — корректировка задачи и т. д. Каждая следующая итерация призвана максимально приблизить ожидаемый результат к заданными параметрами качества. В процессе развития и трансформации отраслей экономики очевидно возникнет потребность в доработке и совершенствовании отраслевых модулей.

В то же время необходимость доработки самой единой технологической платформы за счет использования принципа универсальности должна быть минимальна. Это позволит существенно сократить не только финансовые, но и временные издержки и обеспечит конкурентоспособные на международной арене условия развития электронного правительства в Российской Федерации.

Выводы

Для того чтобы электронное правительство представляло собой по-настоящему действующий инструмент коллективной работы государственных ведомств и учреждений, а не набор разрозненных приложений, необходимо строить его на едином системном подходе, основанном на комплексной автоматизации процессов государственного управления с применением централизованного стандартизированного программного обеспечения.

В статье рассмотрен подход к построению электронного правительства в части, касающейся, в первую очередь, автоматизации его внутренних процессов, т. е. процессов государственного управления, на основе разработки единой технологической платформы. Единая технологическая платформа позволит не только объединить в едином информационном пространстве все субъекты государственного управления и различные отраслевые программные модули, базы данных и иные информационные ресурсы, но и решить следующие концептуальные задачи, без выполнения которых построение электронного правительства не представляется реальным:

- обеспечить возможность нисходящего планирования: от федеральных министерств и ведомств к региональному уровню и конкретным учреждениям;
- обеспечить возможность оперативного анализа состояния отраслей экономики;
- предоставить механизмы централизованного мониторинга не отдельных показателей, а системы показателей как на региональном, так и на федеральном уровне;
- обеспечить согласованность действий федеральных ведомств при планировании государственных расходов и управленческих мероприятий;

- повысить согласованность действий федерального и регионального уровня государственного управления;
- предоставить механизмы полноценного моделирования социальной и экономической ситуации на стыке различных экономических отраслей.

Таким образом, электронное правительство, основанное на единой технологической платформе, позволит решать полный спектр задач государственного управления — от сквозного мониторинга до моделирования условий развития и реформирования отраслей экономики — с использованием автоматизированного механизма коллективной работы, основанного на принципах единого информационного пространства.

Предложенный инновационный подход позволит существенно улучшить параметры информационного взаимодействия субъектов государственного управления, повысить скорость обработки и анализа информации, скорость принятия стратегических и оперативных управленческих

решений, обеспечить рост качественных параметров государственного управления, сформировать надежные условия для роста конкурентоспособности государственного управления Российской Федерации на международной арене.

Список литературы

1. **Стенографический** отчет о совместном заседании Государственного совета и Совета по развитию информационного общества. 23 декабря 2009 года. Москва, Кремль.
2. **Концепция** формирования в Российской Федерации электронного правительства до 2010 года (в ред. Постановления Правительства РФ от 10.03.2009 № 219): одобрена распоряжением Правительства Российской Федерации от 6 мая 2008 г. № 632-р.
3. **Попова М., Рудычева Н., Миссинг А.** ИТ в медицине. Как информатизированы российские клиники? // CNews Analytics. 2009. № 3.
4. **Исследование** Агентства социальной информации Санкт-Петербурга 2006—2009 гг. о реакции населения на реализацию концепции электронного правительства. URL: <http://www.egov.prior.nw.ru>.
5. **Доктрина** информационной безопасности Российской Федерации. Утв. Президентом РФ 09.09.2000 № Пр-1895.
6. **Янг С.** Системное управление организацией / Пер. с англ. — М.: Советское радио, 1972.

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК 004.89

И. П. Норенков, д-р техн. наук, проф., зав. каф.,
М. Ю. Уваров, инж.,
 МГТУ им. Н. Э. Баумана,
 e-mail: norenkov@www.cdl.bmstu.ru

Поиск решений по прецедентам на основе кластеризации онтологий¹

Статья посвящена проблеме формализации поиска решений, содержащихся в документах корпоративных баз знаний. Предлагается подход к построению паттернов принятия решений, основанный на кластеризации предметных онтологий.

Ключевые слова: интеллектуальная система, принятие решений, рассуждение по прецедентам, кластеризация онтологий

Введение

Принятие решений (ПР) требуется во многих сферах человеческой деятельности. Как правило, задачи ПР плохо формализуемы и слабо структу-

рированы. Стремление повысить качество решений и сократить затраты на их выработку диктует необходимость использования для задач ПР, несмотря на их плохую формализуемость, возможностей современной компьютерной техники.

Системы поддержки принятия решений (СППР) и способы их классификации довольно разнообразны. В соответствии с одним из них можно выделить системы с типовыми решениями, с типовыми правилами синтеза решений и системы поиска прецедентов.

Примерами систем, относящихся к двум первым типам, могут служить многие из автоматизированных систем технологической подготовки производства. Так, в системе КОМПАС-Автопроект-Технология [1] имеется база типовых технологических процессов, разработанных для конкретных узкоспециализированных приложений, определяемых классами заготовок и типами производства. Решения получаются ручной доработкой выбранного техпроцесса-аналога или путем интерактивного синтеза маршрутной технологии из отдельных компонентов техпроцессов, хранящихся в библиотеке типовых операций и переходов.

К системам с типовыми правилами синтеза решений относятся экспертные системы [2] и системы, в которых используется технология рассуждения на основе правил RBR (*Rule-Based Reasoning*) [3].

¹ Работа выполнена при финансовой поддержке РФФИ (код проекта 10-07-00401-а).

В случае экспертных систем знания определены в виде продукций, семантических сетей или фреймов, а в случае RBR — в виде правил или аксиом, которые позволяют выводить новые следствия из начального набора данных. Сложность разработки баз экспертных знаний и правил ограничивает, как и в предыдущем случае, применение таких систем узкоспециализированными приложениями.

В общем случае для слабоструктурированных приложений не удастся предварительно разработать достаточно полную базу типовых решений и их элементов, обрабатываемых для конкретных требований по определенным алгоритмам. Поэтому база знаний представляет собой множество документов, в которых с той или иной степенью подробности описаны прецеденты — некоторые ситуации и соответствующие им частные решения. В отличие от предыдущих случаев такие документы являются внешними, т. е. пришедшими не от разработчиков СППР, в то время как обобщенные (типовые) решения составляют содержание специально разрабатываемых документов. В подобных СППР получаемое решение есть не результат исполнения алгоритмов структурного синтеза, а результат поиска наиболее подходящих документов в корпоративной базе знаний. Но, как отмечено в работе [4], систем автоматической обработки документов в целях извлечения знаний в настоящее время не создано, за исключением отдельных узкоспециализированных систем, примерами которых могут служить системы выделения из текстов ссылок на именованные объекты или обработки опубликованных в Интернете сведений о характеристиках и качестве продукции определенного типа.

Системы поиска прецедентов, обычно называемые системами CBR (*Case-Based Reasoning*) [5], должны находить сведения о ранее успешно решенных задачах, выраженные в базах слабоструктурированных документов. Эти документы могут содержать оригинальные описания решений широкого круга проблем, касающихся деятельности предприятия. Попыткой создания подобной системы можно считать проект "Перспективные технологии и новые разработки" [6]. Примерами практического применения систем CBR являются решения задач в области контактной механики [7], в области технической диагностики [8] и др.

Эффективность поиска решений в базах прецедентов определяется не только содержимым собственно базы прецедентов, но и в значительной мере методами, используемыми для организации поиска.

Статья посвящена обсуждению вопросов формализации поиска релевантных документов на основе применения онтологий для формирования

паттернов принятия решений. Паттерны принятия решений — это метаданные документов, в слотах которых содержатся описание признаков ситуации и решаемой проблемы, а также ссылки на документы, содержащие решения или элементы решений.

Организация поиска решений

Рассматривается известная схема поиска и принятия решения:

Формирование запроса — сопоставление атрибутов запроса и поискового образа (метаданных) каждого документа из базы знаний — отбор группы документов, признанных релевантными, — упорядочение отобранных документов по степени релевантности и их анализ в целях выделения аналогов искомого решения.

Для современных поисковых систем характерно использование полнотекстового поиска. Полнотекстовый поиск — поиск по всему содержанию документа. Большинство поисковых машин работает по принципу определения наличия в документе терминов запроса, учета их встречаемости в документах и их средней языковой частотности [9, 10]. Как правило, Интернет-поисковики относятся к системам с полнотекстовым поиском.

Недостатками полнотекстового поиска являются отсутствие контроля за информативностью тех терминов, которые составляют поисковые образы документов, и повышенные затраты вычислительных ресурсов на сопоставление полей запроса и поискового образа.

Альтернативой полнотекстовому поиску является поиск по метаданным. Поиск по метаданным — обычно это атрибутивный поиск по параметрам документов, характерным для систем метаданных, подобных Дублинскому ядру [11] (преимущественно это регистрационные атрибуты — автор документа, заголовок, дата создания, тема и т. п.). Однако для ПР атрибутивный поиск по регистрационным атрибутам может играть лишь вспомогательную роль, так как в первую очередь поиск должен вестись по параметрам, характеризующим проектную ситуацию и признаки решений. Именно такие параметры должны составлять основу метаданных при поиске решений.

В последнее время для организации семантического поиска по метаданным в корпоративных системах управления знаниями начинают использовать онтологии тех предметных областей, с которыми связана деятельность предприятия. Именно с помощью онтологий в виде концептов описываются признаки проблемной ситуации и ссылки на решения. Однако, как отмечено в работе [9], связывание содержащихся в текстах знаний с онтологиями является непростой задачей.

Для ее решения требуется особая структура метаданных.

В большинстве приложений формирование метаданных выполняется вручную на основе имеющихся стандартов, таких как Dublin Core, стандарт IEEE Learning Object Metadata [12], относящийся к области информационных ресурсов сферы образования, или стандарт ISO 19115 "Географическая информация. Метаданные" (ГОСТ Р 52573—2006) [13]. Попытки полуавтоматического формирования метаданных основаны на поиске в тексте документа отношений типа "субъект/объект" [14] (причем дескрипторы субъектов и объектов берутся из онтологии), которые и составляют основу метаданных. Также применяют формирование аннотаций документов путем выделения из текста характерных фрагментов [15, 16], как, например, это делается в системе аналитической обработки текстовой информации "Аналитический курьер" компании Ай-Теко [17].

Использование онтологий и паттернов принятия решений в системах CBR должно быть направлено на повышение точности и полноты поиска решений. В данной статье предлагается подход к построению метаданных документов на основе кластеризации предметных онтологий. Кластеризация ведет к взвешенному учету разных аспектов проектной ситуации и заключается в разделении концептов на группы с учетом имеющихся в онтологии отношений "целое—часть", "род—вид", "объект—свойство", "объект—действие" и "действие—средство". В целях расширения пространства поиска решений учитываются не только факты совпадения концептов запроса и документа, но и свойство "близости" концептов в семантических сетях, выражающих онтологию.

Для использования онтологического подхода, кроме создания собственно прикладных онтологий и их кластеризации, нужно решить такие задачи, как выбор структуры поисковых образов документов (ПОД) и запросов (ПОЗ) и определение релевантности запросов и документов.

Структура поисковых образов документов и запроса

При выборе структуры ПОД и ПОЗ следует учитывать тот факт, что разные концепты могут соответствовать одному и тому же аспекту задачи. Такие концепты целесообразно объединять в группы и относить к одному и тому же слоту в ПОД, поисковые образы документов при этом становятся паттернами ПР. Другими словами, целесообразна кластеризация концептов используе-

мых онтологий по признакам проблемной ситуации и атрибутам решения. Таким образом:

$$\text{ПОЗ} = \{Z_1, Z_2, \dots, Z_n\};$$

$$\text{ПОД} = \{Y_1, Y_2, \dots, Y_n\},$$

где Z_i и Y_i — i -е слоты ПОЗ и ПОД соответственно; n — число слотов; $Z_i = \{z_{1i}, z_{2i}, \dots, z_{ki}\}$, k_i — число дескрипторов концептов в i -м слоте запроса; аналогично определяются слоты Y_i документов.

Кластеризация концептов должна выполняться на основе выделения ролей концептов. Для задач ПР, как минимум, нужно выделить следующие роли, рассматриваемые как ситуационные слоты паттернов принятия решений:

1. Предметная область.
2. Объект.
3. Свойство.
4. Действие (процесс).
5. Средство (инструмент).
6. Проблема (задача).

Кроме того, выделяются слот "ссылка на документ" и регистрационные слоты документа, подобные метаданным из Дублинского ядра, включая аннотацию — описание решаемой проблемы на естественном языке, служащее для неформализованного отбора подходящих документов.

Формирование ПОД основано на выделении в документе всех слов, совпадающих с дескрипторами, имеющимися в предметной онтологии, и в их включении в один из слотов "Объект", "Действие", "Свойство" или "Средство", соответствующих кластерам онтологии. Что касается формулировок задач ПР, их разнообразие весьма велико и, как правило, их дескрипторы в предварительно разработанной онтологии отсутствуют. Так, значительная часть концептов слота "Проблема" представлена словосочетаниями типа "диагностика неисправностей компьютера", "методы расчета производительности", "извлечение знаний" и т. д. Поэтому дескрипторы для слота "Проблема" должны формироваться из терминов слотов "Объект" и "Действие" и, возможно, дополнительно слотов "Свойство" и "Средство". Обнаружение в документе и включение в ПОД таких словосочетаний происходит, если составляющие их дескрипторы оказываются расположенными в тексте документа близко друг от друга, т. е. на основе контекстных ограничений на расстояние между словами.

Следует отметить, что разрешение проблемы омонимии (многозначности слов) возлагается на разработчика онтологии так же, как это делается в существующих системах анализа и обработки текстов, например, в продуктах компании RCO (*Russian Context Optimizer*) [18].

Кроме ролевой, возможна нишевая кластеризация документов, имеющая целью отразить це-

левую направленность документа. Возможные ниши — научная, образовательная (университетская, школьная), популярная, рекламная. О нишевой принадлежности документа можно судить по наличию или, наоборот, отсутствию в нем концептов определенных кластеров.

Отнесение концептов к тому или иному кластеру, т. е. задача классификации концептов, решается создателем онтологии. Как и любая слабоструктурированная проблема, эта задача не имеет строгих формальных методов решения. Так, очевидно, что часть концептов, отнесенных к кластеру "Свойство", могла бы быть включена и в слот "Объект". В неоднозначных ситуациях вопросы кластеризации концептов решает разработчик онтологии при соблюдении условий $C_i \cap C_j = \emptyset$, где C_i — i -й кластер, $i \neq j$. Другими словами, при формировании кластеров любой концепт может быть отнесен только к одному из кластеров "Объект", "Свойство", "Действие", "Средство". При этом внутри кластера отражаются отношения "целое/часть" и "род/вид", а отношения "объект/свойство", "объект/действие" и "действие/средство" служат

ориентиром для распределения концептов по разным кластерам.

При кластеризации онтологии представленные выше слоты составляют верхний уровень кластеризации. На следующем уровне целесообразно кластер "Объект" разделить на ряд специфичных слотов, характеризующих более узкую предметную (дисциплинарную) ориентацию. Например, для предметной области "Информационные технологии" можно распределить концепты по кластерам "Программная инженерия", "Компьютеры и вычислительные системы", "Вычислительные сети", "Автоматизированные системы", "Безопасность информации" и др.

После кластеризации онтология приложения может быть представлена семантическими сетями отдельных слотов (ССС), и становится возможным автоматическое формирование ПОД для документов коллекций (индексация документов). Формирование ПОД включает следующие операции.

1. Выделение в документе всех концептов онтологии с подсчетом числа повторений в документе каждого концепта; упорядочение концеп-

Объект	Свойство	Действие	Средство	Проблема
В.Ю. Осипов. Рекуррентная нейронная сеть с управляемыми синапсами				
Нейронная сеть, синапс, информация, заряд, сигнал, перцептрон	Структура, архитектура, характеристика	Анализ, моделирование, прогнозирование, распознавание, запоминание, управление, эксперимент		Синапс/управление, нейронная сеть/поиск, нейронная сеть/моделирование, нейронная сеть/анализ, информация/запоминание, объект/управление, нейронная сеть/преобразование, нейронная сеть/структура, нейронная сеть/разработка, структура/запоминание, сигнал/структура
Спирин В.Г. Математическая модель сопротивления тонкопленочного резистора				
Математическая модель, резистор	Сопротивление, топологический размер, допуск	Разработка, расчет	Метод, алгоритм	Математическая модель/сопротивление, математическая модель/сопротивление/разработка, метод/расчет/сопротивление, резистор/проектирование
Судаков В.А. Автоматизация процесса управления разработкой корпоративной информационной системы				
Корпоративная информационная система, методология, проект, ЛПР, информационная технология, текстовый документ, язык, система управления, объект, программное обеспечение, OLAP, жизненный цикл	Эффективность	Управление, разработка, автоматизация, анализ	Средство, методология	Корпоративная информационная система/разработка, программное обеспечение/разработка/методология, OLAP/разработка, жизненный цикл/управление/корпоративная информационная система, автоматизированная система управления/разработка, средство/управление/проект, эффективность/разработка/программное обеспечение
Серков Л.А. Синергетические аспекты моделирования инновационных процессов				
Инновация, модель, имитационная модель, логистическая кривая, синергетическая модель, система управления	Стохастичность, устойчивость, частота	Расознавание, моделирование, прогнозирование, управление, синтез, анализ, самоорганизация, эксперимент	Подход	Модель/социально-экономическая система/управление, синергетическая модель/анализ, экономика/синергетическое моделирование, подход/стохастичность/процесс, анализ/устойчивость

тов по числу повторений и помещению в соответствующие слоты паттерна ПР тех концептов, у которых повторяемость выше заданного порога.

2. Формирование словосочетаний концептов из разных кластеров ("Объект", "Свойство", "Действие" и "Средство") по критерию близости расположения соответствующих концептов в тексте документа (например, в пределах одного предложения).

3. Определение повторяемости отношений, определенных в п. 2, и формирование из них слота "Проблема" в паттерне ПР документа.

В качестве примеров в таблице приведены поисковые образы научных статей из журнала "Информационные технологии" (с некоторыми сокращениями списков концептов и без учета частотности концептов).

Содержимое слота "Проблема" указывает на описываемые в документе задачи, включая задачи ПР.

Оценка релевантности документа и запроса

Релевантность документа и запроса в информационно-поисковых системах обычно оценивается коэффициентом релевантности

$$r_k = \frac{\sum_{j=1}^m (a_j x_{kj})}{\sum_{j=1}^m a_j} \quad (1)$$

где $x_{kj} = 1$ при совпадении j -го дескриптора запроса с каким-либо дескриптором того же синтета в k -м документе, иначе $x_{kj} = 0$; a_j — вес (информативность) j -го дескриптора запроса; m — число дескрипторов в запросе. В простейшем случае $a_j = 1$, часто в качестве a_j используют величину $\ln(N/N_j)$, где N — число документов в базе, N_j — число документов, содержащих j -й дескриптор.

При использовании онтологий, разделенных на кластеры, формула (1) предназначена для оценки релевантности отдельных слотов в ПОД и ПОЗ. В формулу оценки R_k общей релевантности запроса и k -го документа могут входить коэффициенты релевантности r_{ki} всех основных слотов, определенные по формуле (1), с весовыми коэффициентами w_i , характеризующими важность слотов:

$$R_k = \sum_{i=1}^n w_i r_{ki},$$

где n — число слотов, или R_k определяется только по совпадению дескрипторов из слотов "Проблема" запроса и документа. При определении весов a_j дескрипторов слота "Проблема" целесообразно учитывать типичные для аннотирования документов факторы частотности, месторасположения дескриптора и т. п. [16].

Возможно введение промежуточных между 0 и 1 значений для x_{kj} , характеризующих, например, "близость" концептов, не совпадающих в ПОД и ПОЗ, или частотность появления дескрипторов в документе. Так, промежуточное значение присваивается x_{kj} , если j -й концепт запроса не описан в ПОД, но связан отношением "целое/часть" или "род/вид" с каким-либо из концептов, описанных в ПОД.

Заключение

Для реализации предложенного поиска решений по прецедентам нужна система, в которой дополнительно к обычным функциям системы онтологического инжиниринга поддерживаются следующие возможности:

- объединение концептов в кластеры с присвоением каждому концепту признаков, определяющих его принадлежность к тому или иному ролевому или нишевому кластеру;
- явное отражение отношений "целое/часть" и "род/вид" между концептами с определением окрестностей концептов в семантических сетях слотов;
- формирование поисковых образов документов путем извлечения из документов терминов, отображающих концепты, относящиеся к отдельным слотам;
- поиск терминов из кластеров "Свойство" и "Действие" в пределах малой окрестности (например, в пределах одного предложения) термина из кластера "Объекты" в тексте документа.

Кроме того, для повышения степени автоматизации поиска и анализа документов полезно использовать и ряд других приемов таких, как использование не только внешних, но и внутренних документов с типовыми решениями и типовыми правилами синтеза решений, выделение в документах фактографической информации об объектах и др.

Список литературы

1. Андрюченко А. КОМПАС-Автопроект: скорость и эффективность технологического проектирования // САПР и графика. 2002. № 9.
2. Лорьер Ж.-Л. Системы искусственного интеллекта. М.: Мир, 1991. 568 с.
3. Liebowitz J. The Handbook of Applied Expert Systems. URL: http://cylib.iit.nau.edu.ua/Books/Prog/Prolog/The_Handbook_of_APPLIED_EXPERT_SYSTM/ewtoc.html
4. Ермаков А. Е. Извлечение знаний из текста и их обработка: состояние и перспективы // Информационные технологии. 2009. № 7. С. 50—55.
5. Aamodt A., Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches // AI Communications. 1994. 7 (1). P. 39—59.
6. Перспективные технологии и новые разработки. URL: <http://www.sibpatent.ru/default.asp?khid = 55040&code = 550177&sort = 2>

7. Дворянкин А. М., Сипливая М. Б., Жукова И. Г., Капш А. С., Кульцов А. Е. Интеграция рассуждений по прецедентам и онтологии в интеллектуальной системе поддержки инженерного анализа в области контактной механики // Известия Волгоградского ГТУ, Сер. "Актуальные проблемы управления, вычислительной техники и информатики в технических системах". 2008. № 2. С. 90—94.
8. Варшавский П. Р. Реализация метода правдоподобных рассуждений на основе прецедентов для интеллектуальных систем поддержки принятия решений. URL: <http://www.raai.org/resurs/papers/kii-2006/doklad/Varshavsky.doc>
9. Добров Б. В., Иванов В. В., Лукашевич Н. В., Соловьев В. Д. Онтологии и тезаурусы: модели, инструменты, приложения. Казань: Казанский государственный университет, 2008. 220 с.
10. Толчеев В. О. Методы выявления информативных признаков в задаче классификации текстовых документов // Информационные технологии. 2005. № 8. С. 14—21.
11. The Dublin Core Metadata Initiative. URL: <http://dublin-core.org/>
12. IEEE 1484.12.1—2002. Learning Object Metadata standard. New York: IEEE, 2002.
13. Географическая информация. Метаданные. ГОСТ Р 52573—2006. URL: <http://tehstandart.com/gost/R-52573-2006>
14. Тузовский А. Ф. Формирование семантических метаданных для объектов системы управления знаниями // Изв. Томского политехнического университета. 2007. № 3. С. 108—112.
15. Барсегай А. А., Куприянов М. С., Степаненко В. В., Холлод И. И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. СПб.: БХВ-Петербург. 2007. 384 с.
16. Башмаков А. И., Башмаков И. А. Интеллектуальные информационные технологии. М.: Изд-во МГТУ им. Н. Э. Баумана, 2005. 304 с.
17. Киселев С. Л. "Аналитический курьер" — современная система извлечения знаний из хранилищ документов. URL: <http://www.it2b.ru/it2b2.view4.page43.html>
18. Морфологический словарь. URL: http://www.rco.ru/product.asp?ob_no = 18

УДК 004.522

А. А. Карпов, канд. техн. наук, ст. науч. сотр.,
Санкт-Петербургский институт информатики и автоматизации РАН, Россия,
e-mail: karpov_a@mail.ru

Л. И. Цирульник, канд. техн. наук, ст. науч. сотр.,
Объединенный институт проблем информатики НАН Беларуси, г. Минск, Беларусь,

М. Железны, канд. техн. наук, доц.
Западно-Чешский университет, г. Пльзень, Чехия

Разработка компьютерной системы "говорящая голова" для аудиовизуального синтеза русской речи по тексту

Представлен первый высококачественный синтезатор аудиовизуальной русской речи, разработанный в результате комплексных междисциплинарных исследований. Система является объединением подсистемы синтеза русской речи по произвольному тексту с компьютерной трехмерной моделью головы человека, имеющей управляемые губы и лицевые органы. Система преобразует произвольный текст на русском языке в синтезированную речь, синхронизируемую с движениями лицевых органов виртуальной модели. Новизна этой системы состоит в учете специфики русского языка и речи, а также в оригинальной модели асинхронности речевых модальностей, которая позволяет улучшить разборчивость синтезированной бимодальной речи и естественность "виртуальной говорящей головы".

Ключевые слова: говорящая голова, аудиовизуальная речь, моделирование асинхронности речи, речевой интерфейс, компьютерный синтез речи, многомодальные интерфейсы, искусственный интеллект

Введение

Как показывают междисциплинарные исследования, визуальная информация очень важна для лучшего понимания произносимой речи. Сигналы от визуальных и слуховых каналов дублируют и дополняют друг друга, что помогает правильно воспринимать речь во многих сложных ситуациях, например, при воздействии динамических акустических шумов, при одновременном разговоре нескольких человек (так называемый "cocktail-party effect"). Кроме того, некоторые фонемы слабо отличимы на слух (например, /м/ и /н/), но

сильно разнятся визуально (/м/ произносится с закрытым ртом, а /н/ — с открытым). Таким образом, добавление визуальной информации при синтезе речи должно способствовать увеличению разборчивости синтезируемой речи.

Еще одним преимуществом компьютерных систем синтеза речи, способных генерировать не только звуковые, но и визуальные речевые сигналы, является возможность передавать слушателю больше информации об интонации речи и выражаемых эмоциях.

"Говорящая голова" — это виртуальная трехмерная модель лица человека с двигающимися

глазами, ртом и лицевыми мускулами, которая может разговаривать, синхронизируя движения артикуляторных органов с синтезированной речью. Подвижность глаз, век и лицевых мускулов модели позволяет выразительно отображать широкий спектр интонаций и эмоций. Комбинация компьютерного синтеза речи с визуализацией органов артикуляции способствует значительному улучшению восприятия синтезированной речи человеком, а также делает компьютерные системы доступными для пользователей с ослабленным слухом или зрением и пожилых людей.

Архитектура системы аудиовизуального синтеза русской речи

Архитектура многомодальной системы аудиовизуального синтеза русской речи, так называемой "говорящей головы", представлена на рис. 1 (см. четвертую сторону обложки). На вход система принимает произвольный текст на русском языке; выходными данными являются речевой сигнал и визуализация движений видимых органов артикуляции, соответствующих синтезированному речевому сигналу.

В основу имитационного метода синтеза видеоизображений речи положена параметрическая трехмерная модель головы человека, разработанная учеными из Западно-Чешского университета для чешского языка [1] и адаптированная для русского. Для построения "говорящей головы" разработана общая трехмерная модель головы и создан набор визем, задаваемых значениями параметров модели.

Модель головы представляет собой набор точек — вершин виртуального пространства, которые соединены дугами, образуя треугольные поверхности, формирующие каркас трехмерной модели (рис. 2, см. четвертую сторону обложки). Создание модели головы осуществляется полуавтоматическим методом в ходе видеозаписи сканирования речи реального диктора с использованием проектора, видеокамеры и системы из четырех зеркал для создания эффекта стереоизображения на изображении от одной видеокамеры. Полученные данные обрабатываются и сохраняются в файле формата виртуальной реальности VRML в виде набора координат вершин каркаса, треугольных плоскостей и соответствующих текстур лица диктора. Полный каркас модели головы описывается несколькими десятками тысяч вершин (рис. 2, а), из которых лишь некоторые являются активными, т. е. ими может управлять программа, имитируя движения лицевых мускулов. К активным вершинам относится набор точек области губ, управление которыми позволяет отображать визему.

В данной модели синтеза видеоизображений речи используется не только общая модель голо-

вы, но и модели отдельных ее подвижных органов: глаз, языка, нижней и верхней челюстей, внутренних артикуляторных органов. Эти элементы описываются по тем же принципам, что и модель головы, однако созданы они не на основе обработки видеоизображений, а на базе знаний из антропологической физиологии. Каждой из моделей может управлять программа независимо от самой головы. Благодаря созданию отдельных моделей артикуляторных органов и функций управления разработанную систему можно использовать для синтеза видеоизображений речи не только для чешского и русского, но и для других языков, в которых такие элементы, как боковинки и тело языка, являются видимыми артикуляторными органами. Еще одним преимуществом создания отдельных моделей элементов "говорящей головы" является возможность управления морганием глаз, что создает иллюзию "живой головы".

В основу системы синтеза речевого сигнала положена модель компиляционного синтеза русской речи, разработанная в Объединенном институте проблем информатики НАН Беларуси [2]. В компьютерной системе синтеза речи по произвольному тексту входной орфографический текст последовательно подвергается преобразованиям, осуществляемым несколькими программными процессорами: текстовым, фонетическим, просодическим и акустическим (см. рис. 1).

Текстовый процессор предназначен для преобразования входного орфографического текста в просодически-размеченный текст. Процессор выполняет следующие задачи: разбиение текста на предложения; преобразование чисел, аббревиатур, сокращений и т. д.; разбиение предложений на просодические синтагмы; расстановка сильных и слабых словесных ударений; разбиение синтагм на акцентные единицы (АЕ); определение интонационного типа синтагм.

Просодически-размеченный текст поступает в фонетический процессор, который выполняет следующие задачи: преобразование орфографического текста в последовательность фонем; преобразование последовательности фонем в аллофонную последовательность. Сформированная аллофонная последовательность поступает на вход двух процессоров: просодического и визуального.

Просодический процессор выполняет следующие задачи: разбиение АЕ на элементы акцентных единиц (предядро, ядро и заядро); вычисление требуемых значений частоты основного тона F_0 , амплитуды A и длительности аллофонов T в соответствии с портретами акцентных единиц для каждой АЕ.

Акустический процессор использует информацию, поступающую от фонетического и просодического процессоров, для выполнения следующих

операций: модификации просодических параметров звуковых волн аллофонов и мультифонов; конкатенации звуковых волн аллофонов и мультифонов в соответствующую последовательность звуковых сигналов.

Отличительной особенностью предложенного аудиовизуального синтезатора является моделирование эффекта естественного временного рассогласования между потоками соответствующих аудио- и видеоединиц русской речи. "Говорящей головой" управляют акустические данные, а визуальная подсистема контролируется результатами работы подсистемы преобразования "текст—речь" с помощью модуля синхронизации (моделирования асинхронности) звуковой и визуальной модальностей.

Модель асинхронности аудио- и визуального потоков речи

Реализация правильного способа синхронизации речевых модальностей (подобного естественной человеческой речи) является одной из основных задач для высококачественного компьютерного синтеза аудиовизуальной речи на любом языке [3]. Суть проблемы состоит в естественном рассогласовании двух основных речевых модальностей, т. е. потоки соответствующих фонем и визем в естественной речи не являются полностью синхронными по времени, как это может показаться на первый взгляд, хотя значительно перекрываются. Такой феномен вызван динамикой самого процесса речеобразования и эффектом коартикуляции, который по-разному проявляется на акустическом и визуальном компонентах речи, что, в основном, и вызывает асинхронность между ними. Правильная синхронизация модальностей речи имеет ключевое значение при восприятии речи человеком, так как она напрямую влияет как на разборчивость, так и на естественность речевого высказывания.

Мировые исследования показывают, что степень синхронности между потоками фонем и визем различна для разных языков. Разными исследователями было обнаружено, что движения губ и звучащая речь практически одновременны для японской речи [4], в противоположность этому для американского диалекта английского языка временные расхождения значительны (до нескольких сотен миллисекунд), что обусловлено богатой артикуляцией и зачастую даже гиперартикуляцией многих американских дикторов.

Для оценки асинхронности аудиовизуальной русской речи авторами проведены корпусные исследования. Был записан бимодальный корпус русской слитной речи, который содержит аудиовидеозаписи произнесения специального фоне-

тически сбалансированного текста¹ четырьмя различными дикторами, как мужчинами, так и женщинами. Все дикторы — нормативные носители русского языка возрастом от 20 до 70 лет. Сессия записи речи для каждого диктора проходила 20—25 мин, чтобы обеспечить высокое качество сигналов, все сеансы записи проходили в шумоизолированных помещениях. Для захвата видеосигнала применяли высококачественную цифровую видеокамеру *Sony*, а для записи аудиосигнала использовали профессиональный стационарный микрофон. Устройства мультимедийной записи находились на расстоянии около 20 см ото рта диктора. Записи проводили с помощью специализированного программного обеспечения, основанного на технологии *DirectShow*, что гарантирует синхронизацию входных аудио- и видеопотоков.

Собранные бимодальные записи речи были обработаны и размечены экспертным путем. Разметку и сегментацию проводили на уровнях фраз, слов и элементарных речевых единиц для аудио- и видеоданных независимо. Для разметки корпуса использовали набор из 42 фонем русского языка, соответствующих фонетическому алфавиту SAMPA с добавлением ударных вариантов гласных звуков и паузы (акустическая тишина). Число классов визем было равно 18 [5]. Таким образом, нескольким различным фонемам могла соответствовать одна визема. Сегментация бимодального корпуса на фонемы и виземы выполнялась "один к одному", т. е. каждой фонеме в потоке соответствовала одна визема, несмотря на то, что некоторые последовательности фонем могут быть представлены одной виземой. В последнем случае ставилась виртуальная граница посередине общей визуальной единицы для того, чтобы сохранить соответствие виземной и фонемной разметок.

Основываясь на анализе бимодального корпуса речи, были сформулированы и программно реализованы следующие правила согласования потоков фонем и визем в слитной русской речи.

1. Начало виземы практически всегда предшествует началу соответствующей фонемы; таким образом, переход между парой визем всегда выполняется на первой фонеме из соответствующей пары фонем. В естественной речи наблюдаются единичные случаи, когда визема заканчивается после соответствующей фонемы (только для огубленных гласных, расположенных перед паузой в случае прижатого положения губ в паузе), но такие случаи можно не учитывать.

¹ Фонетически сбалансированный текст — текст, в котором распределение частот встречаемости фонем и других фонетических единиц близко к теоретическому, полученному на достаточно представительных и больших по объему выборках.

2. Наибольшее отставание фонем наблюдается для огубленных гласных /о/ и /у/ (в среднем около 80 мс), чуть меньше — для губных шумных согласных /в/, /ф/, /п/, /б/ (около 60 мс), затем для неогубленных гласных /а/, /и/, /э/, /ы/ (до 50 мс). Остальные виземы показывают незначительное опережение фонем.

3. Наиболее синхронны во времени пары визема—фонема для сонорных негубных согласных /р/, /л/, /н/.

4. Ударные огубленные фонемы (только /y₀/ и /y₁/) в большей степени отстают от фонем, чем безударные. Для остальных гласных это неверно.

5. В начале фразы виземы обычно больше опережают фонемы, чем в середине или конце фразы. Бывают случаи, когда виземы гласных могут начинаться раньше, чем короткая согласная фонема перед ней, в этом случае согласная визема может просто пропадать.

На основании результатов исследований асинхронности фонемного и виземного потоков в естественной речи определен набор контекстно-зависимых правил переходов между виземами в потоке аудиовизуального синтеза речи [5].

Оценивание синтезированной аудиовизуальной русской речи

Для тестирования и оценки качества аудиовизуальной речи проведен ряд когнитивных экспериментов, которые состояли из двух основных взаимосвязанных частей:

- оценка разборчивости речи (как одномодальной, так и многомодальной), синтезированной в условиях различных акустических шумов;
- оценка естественности речи, синтезированной с использованием различных моделей синхронизации аудио- и визуальной модальностей. При проведении когнитивных экспериментов использовались стимулы трех различных типов:
 - 1) синтезированная звучащая речь;
 - 2) синтезированная аудиовизуальная речь, полученная с использованием системы "говорящая голова";
 - 3) предварительно записанная естественная речь диктора (записи естественной речи того же диктора использовались для создания синтезированной речи).

Для проведения эксперимента из собранного корпуса русской речи были выбраны 20 фонетически-сбалансированных фраз, которые впоследствии предъявляли добровольцам—участникам эксперимента (аудиторам) в произвольном порядке. Каждая фраза состояла из 4—5 общеупотребительных слов, при этом вся фраза целиком была бессмысленной, что исключает использование ау-

диторами априорных семантических знаний при тестировании.

На первом этапе тестирования, целью которого являлось определение разборчивости аудио- и аудиовизуальной речи, аудиторы прослушивали каждую аудиофразу и записывали последовательность услышанных слов, затем прослушивали и просматривали аудиовизуальную синтезированную фразу и записывали последовательность слов.

На втором этапе, целью которого являлось тестирование естественности различных моделей синхронизации аудио- и видеопотоков, аудиторы оценивали следующие четыре модели:

1. Базовая система "говорящая голова" с полностью синхронными потоками фонем и визем.

2. Система "говорящая голова" с моделью асинхронности аудиовизуальной речи, предложенной авторами данной статьи.

3. Модель асинхронности, в которой осуществляется стационарная задержка звукового сигнала относительно видеосигнала в 150 мс (обозначим ее модель B150A).

4. Модель асинхронности, в которой осуществляется стационарная задержка визуального сигнала относительно аудиосигнала в 150 мс (обозначим ее модель A150B).

Естественность каждой из моделей оценивалась аудитором по пятибалльной шкале, где наилучшая оценка — 5 означала, что модальности отлично синхронизированы. Аудиторы должны были также записать последовательность слов, которую они распознали.

На последнем этапе тестирования аудиторам предлагалось прослушать и оценить ту же самую фразу, произнесенную естественным, предварительно записанным голосом.

Такой цикл с различными фразами повторялся 20 раз (по числу фраз) для каждого испытуемого. Нужно отметить также, что в чистый акустический сигнал добавляли аддитивный шум (белый шум, либо шум толпы, создающий "cocktail-party effect") с изменяющейся интенсивностью (отношение сигнал/шум варьировалось от 5 до 25 дБ).

Число аудиторов — 10 человек, в возрасте от 20 до 35 лет, носители русского языка, без выявленных дефектов слуха и зрения.

Распределение средних (по всем аудиторам) оценок четырех моделей синхронизации в зависимости от отношения сигнал/шум акустического сигнала показаны на рис. 3 (см. четвертую сторону обложки). Легко заметить, что с уменьшением отношения сигнал/шум аудиторские оценки также уменьшаются — это справедливо для всех моделей, кроме A150B, которая и без того получила очень низкие баллы. Когда пользователи понимают синтезированную речь хуже, они имеют общую тенденцию ставить более низкие оценки.

Также при низких соотношениях сигнал/шум расстояния между элементами приведенной гистограммы сокращаются, что означает: пользователи хуже чувствуют разницу между моделями синхронизации в зашумленных сигналах. Вероятно, это вызывается тем, что в условиях шумов тяжело идентифицировать моменты начала и конца фразы. Преимущество же предложенной модели асинхронности модальностей речи в условиях незашумленного сигнала очевидно.

Анализируя средние оценки каждого из аудиторов в отдельности, отметим следующее:

- все аудиторы идентифицировали рассинхронизацию аудио- и визуальной речи для модели A150B;
- два аудитора из 10 не ощутили разницу в синхронизации для базовой синхронной модели, предложенной асинхронной модели и модели B150A;
- два аудитора не отличили B150A и асинхронную модель.

Остальные аудиторы определили разницу во всех четырех системах.

Оказалось довольно неожиданно, что пользователи выставили низкие оценки базовой синхронной модели, которая заняла лишь 3-е место в "соревновании"; большинство аудиторов предпочли новую асинхронную модель и один пользователь — модель B150A. Примечательно, что для всех аудиторов модель B150A оказалась более приемлема, чем модель A150B. Этот факт еще раз подтверждает, что видеопоток в естественной речи опережает аудиопоток и задержка видеопотока по отношению к аудио ощущается аудиторами и оценивается как неестественная [6].

На рис. 4 представлены результаты той части эксперимента, которая относилась к количественной оценке разборчивости² синтезированной русской речи. Как и ожидалось, реальный голос оказался более разборчив, чем синтетический. Рис. 4 доказывает, что визуальная речевая модальность действительно помогает понимать речь лучше, особенно в зашумленных условиях. Нужно отметить, что расстояния между функциями разборчивости растут с увеличением уровня шума. В наших экспериментах система бимодального синтеза превзошла одномодальную систему синтеза в среднем на 6 % по показателю разборчивости.

При анализе результатов тестирования разборчивости мы неоднократно наблюдали так называемый эффект МакГурка [7], когда правильное распознавание звука в слове возникает лишь при объединении акустических и визуальных сигналов, например, для таких слов, как "сотка" и "соп-

² Эта величина определяется как отношение правильно распознанных слов к общему числу слов в высказывании.

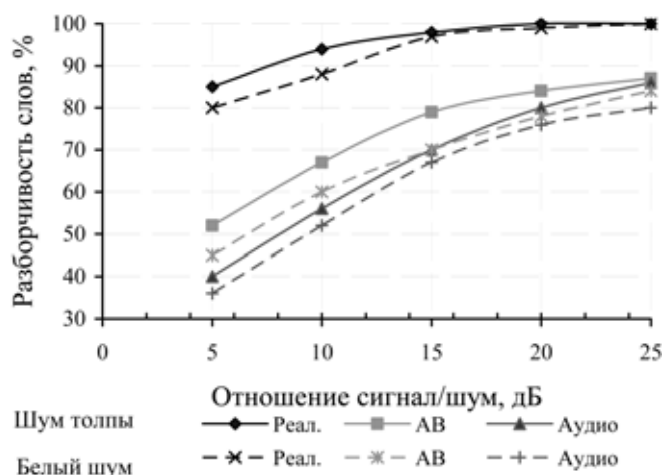


Рис. 4. Результаты экспериментов по оценке разборчивости речи в условиях аддитивных акустических шумов

ка"; "ода" и "оба". В обоих случаях только по движениям губ можно правильно отличить звуки в середине слов.

Полученные результаты разборчивости речи различны для аддитивного белого шума и шума толпы. В акустических условиях с присутствием белого шума разборчивость речи слегка ниже, чем в условиях шума толпы. Этот эффект можно, пожалуй, объяснить следующим утверждением: белый шум более напрягает слуховую систему человека, чем шум толпы, который более естественен и "приятен" для человека. В условиях аддитивного шума толпы сложнее правильно определить границы полезной речи, но сама речь более разборчива; белый же шум отрицательно влияет на разборчивость речи.

Заключение

В статье представлена система аудиовизуального синтеза русской речи, так называемая "говорящая голова". Когнитивные эксперименты позволили количественно и качественно оценить естественность и разборчивость синтезированной аудио- и аудиовизуальной речи.

Полученные результаты показали, что предложенная авторами модель синхронизации аудио- и визуального потоков воспринимается аудиторами как наиболее естественная и улучшает восприятие синтезированной речи.

Результаты по разборчивости речи продемонстрировали преимущество аудиовизуального синтеза русской речи перед звуковой синтезированной речью, более того, достоинства визуальной компоненты особенно очевидны в зашумленных условиях.

Бимодальные речевые системы являются востребованными в практических приложениях, поскольку повышают разборчивость, комфортность

и естественность синтезированной речи, а также делают их доступными для людей с ограниченными физическими возможностями, т. е. с дефектами слуха, зрения. Разрабатываемая бимодальная система синтеза речи может быть эффективно использована для образовательных задач, в развлекательных приложениях и системах виртуальной реальности, в информационно-справочных системах и автоматах самообслуживания, аудиовизуальных инсталляциях и на телевидении.

Данное исследование поддержано Грантом Президента РФ (проект МК-64898.2010.8); Министерством образования и науки РФ в рамках ФЦП "Научные и научно-педагогические кадры инновационной России" (госконтракты П2579 и П2360); фондами РФФИ и БРФФИ в рамках совместного проекта № 08-07-90002/№ Ф08Р-016а; КНВШ Администрации Санкт-Петербурга (проект 26-05/130); Министерством Образования, Молодежи и Спорта Чешской Республики (проект DIMAS-CZ № ME08106).

Список литературы

1. **Železný M., Krňoul Z., Čísař P., Matoušek J.** Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis // *Signal Processing*. 2006. Т. 86. № 12. С. 3657—3673.
2. **Lobanov B., Tsirulnik L.** Development of multi-voice and multi-language TTS synthesizer (languages: Belarussian, Polish, Russian) // *Труды 11-й Международной конференции "Речь и Компьютер" SPECOM'2006*, Санкт-Петербург, 2006. Р. 274—283.
3. **Govokhina O., Bailly G., Breton G.** Learning optimal audiovisual phasing for a HMM-based control model for facial animation // *Proc. of ISCA Speech Synthesis Workshop*, Bonn, Germany, 2007, CD.
4. **Sekiyama K.** Differences in auditory-visual speech perception between Japanese and America: McGurk effect as a function of incompatibility // *Journal of the Acoustical Society of Japan*. 1994. Т. 15. С. 143—158.
5. **Karpov A., Tsirulnik L., Zelezny M., Krnoul Z., Ronzhin A., Lobanov B.** Study of Audio-Visual Asynchrony of Russian Speech for Improvement of Talking Head Naturalness // *Труды 13-й Международной конференции SPECOM'2009*, Санкт-Петербург, 2009. С. 130—135.
6. **Conrey B., Pisoni D.** Audiovisual asynchrony detection for speech and nonspeech signals // *Audio-Visual Speech Processing AVSP'2003*, St. Jorioz, France, 2003. С. 25—30.
7. **McGurk H., MacDonald J.** Hearing Lips and Seeing Voices // *Nature*. 1976. Т. 264, № 5588. С. 746—748.

УДК 004.8

С. И. Попов, студент,

Ю. И. Попов, студент,

А. А. Шалыто, д-р техн. наук, проф., зав. каф.,

Санкт-Петербургский государственный университет информационных технологий, механики и оптики,
e-mail: yugropov@rambler.ru

Задача о муравьеде и муравьях

Рассматривается расширение "Задачи об умном муравье-3", названное авторами "Задача о муравьеде и муравьях". В задаче используется двумерный тор произвольного размера с расположенными на нем муравьедом, муравьями (пища для муравьеда) и яблоками (пища для муравьев). Муравьедом и муравьями управляют автоматы Мили.

Цель работы — построить с помощью генетического алгоритма автоматы, управляющие муравьедом и муравьями, которые позволят эффективно решить поставленную задачу. Для построения генетического алгоритма необходимо разработать способы представления автоматов, выбрать функции приспособленности, операторы скрещивания, мутации и отбора.

Авторами разработано инструментальное средство для решения "Задачи о муравьеде и муравьях" при различных значениях используемых параметров.

Ключевые слова: задача об умном муравье, искусственный интеллект, коэволюция, генетический алгоритм, автомат Мили

Известна классическая "Задача об умном муравье" [1, 2], в которой с помощью генетического алгоритма должен быть сгенерирован автомат, управляющий одним муравьем с простым полем обзора. Муравей за фиксированное число шагов съедает максимальное число яблок, расположенных определенным образом на двумерном торе. В работах [3, 4] рассмотрена более сложная "Задача об умном муравье-3", в которой муравей имеет большее поле обзора, а число яблок при их случайном расположении на торе не фиксировано.

Однако для практики представляют интерес задачи, в которых особи взаимодействуют. Такие задачи принадлежат классу "коэволюций" [2, 5, 6]. Однако среди задач этого класса неизвестны задачи, решением которых были бы автоматы, сгенерированные генетическими алгоритмами.

Для устранения этого пробела авторами была предложена "Задача о муравьеде и муравьях". Настоящая работа посвящена описанию и решению этой задачи.

1. Постановка задачи

Задано поле — двумерный тор произвольного размера с координатной сеткой, которая делит его поверхность на клетки. На поле расположены муравьед, муравьи (еда для муравьеда) и яблоки (еда для муравьев). Яблоки и муравьи располагаются случайным образом. На одной из клеток находится муравьед (рис. 1).

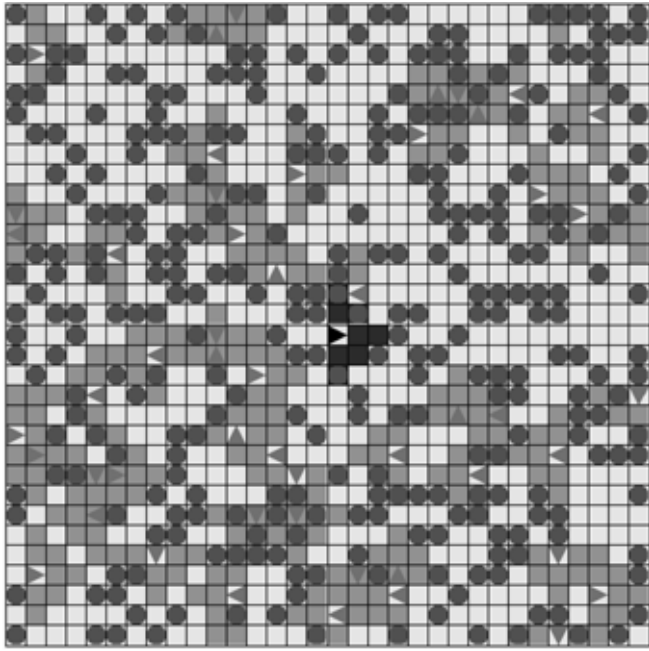


Рис. 1. Тор с муравьедом, муравьями и яблоками. Черный треугольник изображает муравьеда, черные клетки — это те клетки, которые он видит; серые треугольники — муравьи, серые клетки — та часть поверхности, которую видят муравьи; темно-серые кружочки обозначают места расположения яблок

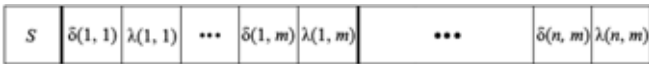


Рис. 2. Представление автомата Мили

Цель каждого муравья — за заданное число шагов съесть как можно больше яблок и "не попасться" муравьеде. Цель муравьеда — помешать им сделать это, съев за заданное число шагов максимальное число муравьев. Муравьи не должны "наступать" на других муравьев и муравьеда, так как они при этом погибают.

Будем предполагать, что муравьедом и муравьями управляют автоматы Мили [1].

Яблоки лежат на поле неподвижно, а муравьед и муравьи на каждом шаге могут совершить одно из трех действий (выходные воздействия соответствующих автоматов):

- повернуть налево;
- повернуть направо;
- пойти на одну клетку вперед, и, если там есть еда (муравей для муравьеда или яблоко для муравья), то съесть ее.

Муравьед и муравьи видят перед собой по восемь клеток: две слева, две справа, две спереди и по одной спереди по диагонали.

Задача, решаемая в настоящей работе: построить генетический алгоритм [2, 5–7], позволяющий сгенерировать автоматы Мили, управляющие муравьедом и муравьями, которые обеспечат достижение целей, указанных выше.

Для построения генетического алгоритма необходимо разработать способы представления автоматов, выбрать функции приспособленности, операторы скрещивания, мутации и отбора.

2. Представление автоматов

Для представления автомата Мили в работе используется массив байтов, в котором первый элемент содержит номер начального состояния автомата S . Затем следуют n секторов (n — число состояний автомата), каждый из которых содержит m пар чисел (m — число входных воздействий), каждая из которых состоит из следующего состояния δ и выходного воздействия λ (рис. 2).

В рассматриваемой задаче автоматы могут принимать большое число входных воздействий, так как каждая клетка поля может содержать муравьеда, муравья, яблоко или не содержать их.

Муравьед не различает яблоки и пустые клетки, и поэтому его автомат может принять $2^8 = 256$ входных воздействий (на клетке есть муравей либо его нет).

Муравей различает клетки с яблоками и пустые клетки. К тому же он может видеть (если они находятся в поле его зрения) других муравьев (для того, чтобы не наткнуться на них) и муравьеда (для того, чтобы вовремя убежать от него). Таким образом, автомат муравья может принять $3^8 + 8 \cdot 3^7 = 6561 + 8 \cdot 2187 = 24057$ входных воздействий.

3. Функции приспособленности

Функция приспособленности автомата, управляющего муравьем, равна сумме числа съеденных последним яблок и дроби, числителем которой является число этих яблок, а знаменателем — номер шага, на котором было съедено последнее яблоко. Функция приспособленности автомата, управляющего муравьедом, вычисляется аналогично, но при этом вместо яблок используются муравьи.

При решении задачи эти функции приспособленности вычисляются следующим образом: по очереди из текущего поколения выбирается один автомат, управляющий муравьедом, и для него случайным образом из текущего поколения автоматов, управляющих муравьями, выбирается заданное в эксперименте число различных автоматов k .

На поле размещают муравьеда и k муравьев, управляемых этими автоматами. При этом для яблок и муравьев генерируются случайные начальные координаты, а для каждого муравья также случайно указывается начальное направление движения. Все участвующие в эксперименте особи (муравьед и муравьи) совершают заданное число шагов.

Выбранные муравьед и муравьи имеют заданное число попыток для случайного размещения

на поле муравьев и яблок. При этом для каждой особи в отдельности значения функции приспособленности автомата на всех полях суммируются, а в конце эксперимента сумма делится на число совершенных попыток. Если число попыток у особей разное, то последние попытки не учитываются у тех особей, для которых было задано меньшее число попыток. В результате для каждой особи получается среднее (за несколько попыток) число съеденных ею единиц еды. Если несколько муравьев хотят пойти на одну и ту же клетку, то приоритет имеет тот муравей, чей автомат имеет наибольшее значение функции приспособленности (учитываются и предыдущие попытки). При этом автоматы, управляющие муравьями, которые не смогли сделать шаг, остаются в предыдущих состояниях.

После вычисления значений функций приспособленности всех автоматов муравьедов, если остаются автоматы муравьев, которые не участвовали в эксперименте, то для каждого из них выбирают автомат муравьеда и необходимое число автоматов муравьев и вычисляют значения их функций приспособленности. При этом функции приспособленности автоматов, вычисленные ранее, имеют после эксперимента предыдущие значения.

В заключение раздела отметим, что в ходе коэволюции в общем случае автоматы муравьев, даже при одинаковом числе состояний, отличаются друг от друга.

4. Оператор скрещивания

Авторами был выбран один и тот же оператор скрещивания для автоматов муравьедов и муравьев. Он обеспечивает компромисс между временем работы и скоростью улучшения автоматов. Оператор скрещивания сначала копирует автомат первого предка в одного потомка, а автомат второго предка — в другого потомка. Затем он случайно устанавливает потомкам различные начальные состояния от предков. После этого 30 раз для каждого из потомков выбирается фрагмент другого предка (чьей копией потомок до этого не был) длиной в 1 % от общей длины (напомним, что автоматы хранятся в виде байтовой строки), и этот фрагмент устанавливается на те же позиции потомка, на которых он был у предка.

5. Оператор мутации

В операторе мутации для муравьеда и муравья с заданной вероятностью изменяется начальное состояние автомата. Затем выполняется проход по всем состояниям (в каждом из них, в свою очередь, выполняется проход по всем входным воздействиям) и с той же вероятностью на случайные

изменяются номера состояний для переходов и выходные воздействия.

6. Оператор отбора автоматов в следующее поколение

Каждое поколение в генетическом алгоритме делится на элиту (автоматы с наибольшими значениями функций приспособленности) и остальные автоматы. Сначала в следующее поколение переходят автоматы из элиты. Далее для отбора автоматов используется метод рулетки [2, 5, 7]. К автоматам, которые были отобраны методом рулетки, применяется оператор скрещивания, а к их потомкам — оператор мутации. После этого они добавляются в следующее поколение к элите.

7. Инструментальное средство

Для решения поставленной задачи разработано инструментальное средство, поддерживающее описанные особенности рассмотренного генетического алгоритма. Это средство позволяет проследить процесс "выращивания" автоматов, а также визуализировать поведение особей, управляемых этими автоматами, на случайном поле с муравьями и яблоками.

Инструментальное средство позволяет задать значения параметров эксперимента (рис. 3): ширины поля, высоты поля, заполненности поля яблоками, заполненности поля муравьями, на-

Изменение конфигурации	
Укажите новую конфигурацию эксперимента:	
Ширина поля (от 10 до 100):	32
Высота поля (от 10 до 100):	32
Заполненность поля яблоками (от 0 до 1):	0.3
Заполненность поля муравьями (от 0 до 1):	0.05
Координаты начальной позиции муравьеда:	
X	16
Y	16
Начальное направление движения муравьеда:	Восток
Число шагов в эксперименте (от 1 до 999):	200
Параметры:	Муравьед
Размер поколения (от 20 до 500):	200
Число состояний в автомате (от 1 до 20):	5
Вероятность мутации (от 0 до 1):	0.1
Доля элиты в поколении (от 0 до 1):	0.1
Число попыток в эксперименте (от 1 до 99):	5
Сохранить Сбросить Отмена	

Рис. 3. Выбор параметров эксперимента и особей

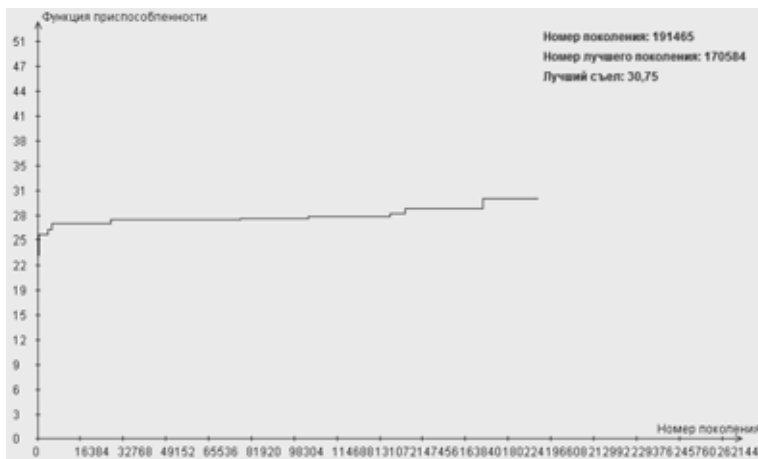


Рис. 4. График зависимости числа съеденных муравьев от номера поколения муравьеда

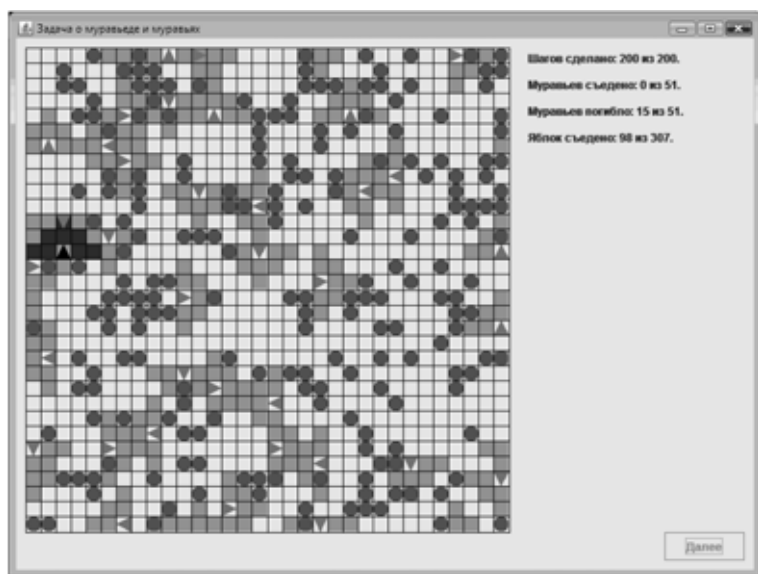


Рис. 5. Вид поля в начале генерации поколений после 200 шагов. Муравьед никого не съел, муравьи съели меньше половины яблок

чальных координат муравьеда, начального направления движения муравьеда, число шагов в эксперименте.

Для автоматов муравьеда и муравьев (все автоматы муравьев имеют одинаковые параметры) отдельно можно задать следующие параметры: размер поколения, число состояний в автомате, вероятность мутации, долю элиты в поколении, число попыток в эксперименте (число случайных полей, на которых особь пытается добиться своей цели и улучшить значение функции приспособленности своего автомата).

Инструментальное средство занимает около 200 Мбайт оперативной памяти. Для вычисления следующего поколения требуется примерно 2,5 с при использовании процессора *Intel Pentium 4* с частотой 3,2 ГГц.

8. Результаты экспериментов

Для быстрого обучения автоматов, управляющих муравьедом, необходимо иметь большое число муравьев на поле. Аналогично для автоматов муравьев: если на поле больше яблок, то муравьи учатся есть быстрее. Авторы решили рассмотреть задачу "Задачу о муравьеде и муравьях" при следующих параметрах:

Параметр	Значение
Ширина поля	32
Высота поля	32
Заполненность поля яблоками	0,3
Заполненность поля муравьями.	0,05
Координата X начальной позиции муравьеда	16
Координата Y начальной позиции муравьеда	16
Начальное направление движения муравьеда	Восток
Число шагов в эксперименте.	200

Для автоматов муравьедов и муравьев были выбраны одинаковые значения параметров, указанные ниже:

Параметр	Значение
Размер поколения	200
Число состояний в автомате	5
Вероятность мутации	0,1
Доля элиты в поколении.	0,1
Число попыток в эксперименте.	5

Из графика на рис. 4 видно, что особь постепенно учится съесть все большее число муравьев (исходно на поле был размещен 51 муравей). При этом, если вначале муравьед передвигается случайным образом и чаще всего вращается на месте (рис. 5), то, как показали эксперименты, уже к 50-му поколению прослеживается его стремление догнать жертву, хотя иногда в нужный момент он сворачивает не туда.

Чем больше номер поколения, тем меньше подобных ошибок. Ближе к 2000-му поколению муравьед уже уверенно преследует муравьев, несмотря на то, что они пытаются от него убежать (муравьи отличают муравьеда от себе подобных и яблок и постепенно учатся к нему не приближаться) (рис. 6).

При дальнейшем увеличении числа поколений муравьед поедает больше муравьев. Так, в 170584-м поколении значение функции приспособленности автомата муравьеда равно 30,75 (см. рис. 4).

На втором графике (рис. 7) виден процесс роста числа съеденных яблок у муравьев от поколения к поколению (сравните рис. 5 и 6). На результат (значение функции приспособленности равно 15,87 в 127473-м поколении) повлияло то, что му-

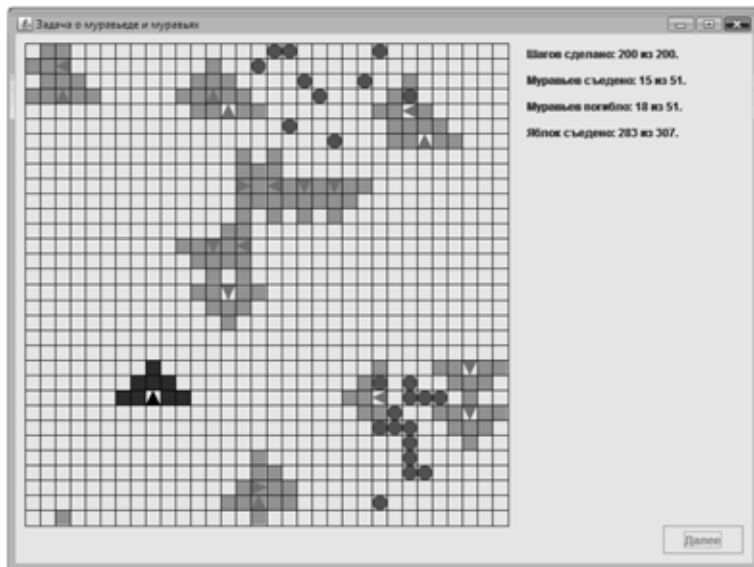


Рис. 6. Вид поля в процессе генерации поколений после 200 шагов. Муравьед съел 15 муравьев, муравьи съели большинство яблок

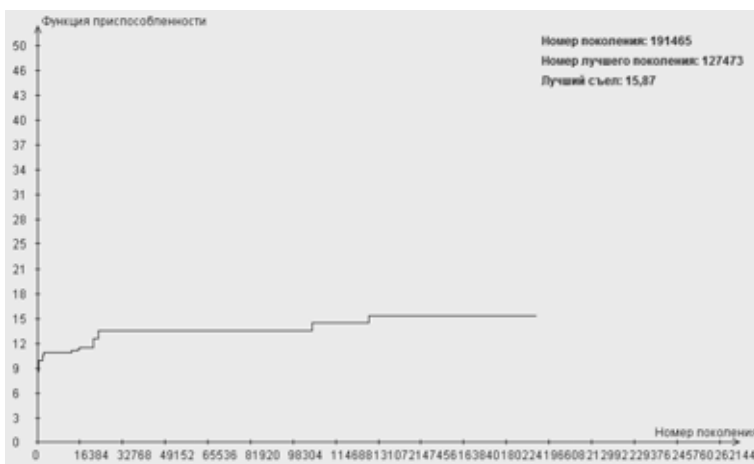


Рис. 7. График зависимости числа съеденных яблок от номера поколения муравьев

равьи фактически соревнуются между собой за яблоки, число которых на поле исходно равнялось 307.

Заключение

Авторы сформулировали новую коэволюционную задачу и для ее решения разработали инструментальное средство (все графики и рисунки к статье получены с помощью этого средства).

Результаты, приведенные выше, свидетельствуют о том, что генетические алгоритмы "справляются" с этой задачей не хуже, чем со значительно более простой "Задачей об умном муравье-3". При этом отметим, что в рассматриваемой задаче генерируется большее число поколений по сравнению с "Задачей об умном муравье-3" для того, чтобы муравьед съел более половины муравьев. Это связано с тем, что если в известной задаче расположение еды муравья фиксировано, то в рассматриваемой задаче еда от муравьеда может "убегать".

Список литературы

1. Полицарова Н. И., Шальто А. А. Автоматное программирование. СПб.: Питер, 2009. URL: http://is.ifmo.ru/books/_book.pdf
2. Koza J. Genetic programming. On the Programming of Computers by Means of Natural Selection. MA: The MIT Press, 1998.
3. Бедный Ю. Д., Шальто А. А. Применение генетических алгоритмов для построения автоматов в задаче "Умный муравей". СПбГУ: ИТМО, 2007. URL: <http://is.ifmo.ru/works/ant>.
4. Давыдов А. А., Соколов Д. О., Царев Ф. Н., Шальто А. А. Виртуальная лаборатория обучения генетическому программированию для генерации управляющих конечных автоматов // Сборник докладов III Международной научно-практической конференции "Современные информационные технологии и ИТ-образование". ВМК МГУ. М.: МАКС Пресс. 2008. С. 179—183. URL: http://is.ifmo.ru/works/_2_93_davidov_sokolov.pdf
5. Mitchell M. An Introduction to Genetic Algorithms. MA: The MIT Press, 1999.
6. Джонс М. Программирование искусственного интеллекта в приложениях. М.: ДМК Пресс, 2006.
7. Гладков Л. А., Курейчик В. В., Курейчик В. М. Генетические алгоритмы. М.: Физматлит, 2006.



МЕЖДУНАРОДНАЯ СПЕЦИАЛИЗИРОВАННАЯ ВЫСТАВКА
ИНФОРМАЦИЯ:
ТЕХНИКА И ТЕХНОЛОГИИ ЗАЩИТЫ
5-8 ОКТЯБРЯ 2010, Санкт-Петербург, Ленэкспо

По вопросам участия в выставке/конференции, пожалуйста, обращайтесь:

Семенова Анна, Зориков Константин

Тел.: +7 (812) 380 6009,

E-mail: sfitex@primexpo.ru

<http://iscs-expo.primexpo.ru/>

УДК 519.622.2

Д. М. Жук, канд. техн. наук, доц.,
 В. Б. Маничев, канд. техн. наук, доц.,
 А. О. Ильницкий, аспирант,
 МГТУ имени Н. Э. Баумана,
 e-mail: manichev@bmstu.ru

Методы и алгоритмы решения дифференциально-алгебраических уравнений для моделирования систем и объектов во временной области. Часть 2*

Рассматривается выбор методов и алгоритмов для решения систем линейных алгебраических уравнений (СЛАУ) при моделировании систем и объектов во временной области. Обосновывается использование метода LU-разложения с выбором главного элемента в качестве базового метода решения СЛАУ, а также повышенной разрядности арифметических операций, когда это необходимо, для обеспечения требуемой точности расчета всех неизвестных систем СЛАУ.

Ключевые слова: моделирование, обыкновенные дифференциальные уравнения (ОДУ), дифференциально-алгебраические уравнения (ДАУ), методы интегрирования, временная область, линейные алгебраические уравнения (ЛАУ)

Введение

В части 1 рассматривается выбор методов интегрирования для решения систем ДАУ, неразрешенных относительно производных и представленных в виде:

$$F(X, XP, Y, t) = 0, \quad (1)$$

где X — вектор дифференцируемых переменных размерностью m ; XP — вектор производных этих переменных по переменной t размерностью m , т. е. $XP = dX/dt$; Y — вектор алгебраических переменных размерностью k ; t — независимая переменная (обычно — время); F — вектор-функция размерностью l , где $l = m + k$. Заданы начальные условия $X_0 = X(0)$ и отрезок интегрирования.

Для выбранных методов интегрирования используется следующий алгоритм.

* Часть 1 см. в № 7, 2010 г.

На каждом шаге интегрирования h_n решается система нелинейных алгебраических уравнений (НАУ), размерностью $2m + k$ (система НАУ приведена для неявного метода Эйлера)

$$\begin{cases} XP_n - \frac{X_n - X_{n-1}}{h_n} = 0; \\ F(X_n, XP_n, Y_n, t_n) = 0 \end{cases}$$

относительно векторов X_n, XP_n, Y_n общей размерностью $2m + k$ методом Ньютона, который сводится к решению соответствующей системы линейных алгебраических уравнений (СЛАУ). Здесь $h_n = t_n - t_{n-1}$ — шаг интегрирования; n — номер шага интегрирования.

В связи с этим выбор и обоснование методов решения СЛАУ очень важны при разработке программно-технического комплекса FMS PA10.

Выбор методов решения СЛАУ при моделировании систем и объектов во временной области

Выбранные в части 1 статьи неявные методы интегрирования сводятся в общем случае к решению СЛАУ вида

$$Ax = b,$$

где $A = (a_{ij})$ — матрица коэффициентов размерности $n \times n$, $\det(A) \neq 0$; $b = (b_1, \dots, b_m)^T$ — вектор-столбец правых частей; x — вектор неизвестных.

Так как системы вида (2) приходится решать большое число раз, возникает необходимость в эффективных и точных численных методах решения такого рода систем.

Под *эффективностью* здесь и далее понимается как можно меньшие затраты машинного времени на получение решения, под *точностью* — гарантия того, что погрешность (в смысле числа верных знаков в решении для каждой неизвестной системы (2)) полученного решения не превосходит заранее заданной величины.

В настоящее время разработано очень большое число разнообразных методов решения СЛАУ, но все их можно разделить на два класса: точные и итерационные. Метод решения СЛАУ относится к классу точных, если в предположении отсутствия округлений он дает точное решение задачи после конечного числа арифметических и логических операций. Итерационные методы постепенно приближают точность решения к заданной.

Точные методы вследствие плохой устойчивости к накоплению арифметической погрешности применяются для систем небольшой и средней размерности до порядка 10^4 [1]. Итерационные методы применяются в основном для решения СЛАУ очень большой размерности, до порядка 10^7 , но их использование накладывает ограничения на матрицу системы [1].

Наиболее простым, производительным и точным методом решения СЛАУ общего вида является метод исключения Гаусса. Метод Гаусса требует $(2/3)n^3 + O(n^2)$ ($n \rightarrow \infty$) арифметических операций [2]. В связи с тем, что СЛАУ, которые возникают при моделировании динамики технических систем, являются сильно разреженными (порядка 3–5 ненулевых элементов в строке), несложно подсчитать, что при отсутствии заполнения матрицы коэффициентов новыми ненулевыми элементами при проведении исключений метод Гаусса требует примерно $O(n)m^2$ ($n \rightarrow \infty$) арифметических операций, где m — среднее число ненулевых элементов в строке. Главным недостатком метода Гаусса является плохая устойчивость к накоплению ошибки округления. Устойчивость метода можно повысить, выбирая на каждом шаге исключения неизвестных главный элемент (по строке, по столбцу или по всей матрице). Вместе с тем, можно использовать стратегию минимизации заполнения матрицы ненулевыми элементами, что минимизирует число арифметических операций и, следовательно, накопление погрешности.

Анализ результатов многочисленных тестов показал, что наиболее приемлемой стратегией с точки зрения минимизации числа арифметических операций и в то же время минимизации накопления ошибки округления, являются совмещение стратегии выбора наименее заполненного столбца и затем выбор главного элемента по данному столбцу. Для примера проведем результаты следующих тестов. Будем генерировать матрицу коэффициентов размерности $n = 100 \times 100$, $n = 600 \times 600$ и $n = 1100 \times 1100$. Для каждой размерности будем увеличивать число ненулевых элементов в строке с $m = 3$ до $m = 9$ с шагом 2 (3, 5, 7, 9). Для каждого m проведем по 100 испытаний. Будем анализировать среднее число ненулевых элементов до и после решения, а также время, затраченное на решение. Данные испытания проведем для метода Гаусса с выбором главного элемента по столбцу и для метода Гаусса с описанной выше стратегией минимизации заполнения. Результаты проведенных испытаний приведены в таблице.

Используя данные, приведенные в таблице, можно подсчитать, что при проведении испытаний для размерности 100×100 , эффективность стратегии минимизации составила примерно 25;

Стратегия выбора главного элемента	$n \times n$	m	N	N'	T , мс
Выбор главного элемента по столбцу	100×100	3	300	728,22	1,09
		5	500	2 004,70	2,80
		7	700	2 809,41	4,36
		9	900	3 312,40	5,92
	600×600	3	1800	16 557,76	46,78
		5	3000	61 260,26	392,92
		7	4900	90 561,11	736,90
		9	5400	108 694,60	1034,98
	1100×1100	3	3300	52 253,14	273,71
		5	5500	202 770,70	2409,45
		7	7700	299 608,60	5498,14
		9	9900	361 301,40	7769,99
Выбор наименее заполненного столбца вместе с выбором главного элемента по столбцу	100×100	3	300	542,48	0,94
		5	500	1 690,58	2,49
		7	700	2 600,71	5,01
		9	900	3 163,72	6,25
	600×600	3	1800	9 138,95	35,48
		5	3000	44 790,39	251,59
		7	4900	75 132,33	634,64
		9	5400	97 087,25	955,30
	1100×1100	3	3300	26 072,17	144,24
		5	5500	142 246,06	1649,39
		7	7700	241 786,67	3841,41
		9	9900	313 588,86	6395,12

Примечание: n — размерность решаемой СЛАУ; m — число ненулевых элементов в строках матрицы коэффициентов; N — общее среднее число ненулевых элементов матрицы коэффициентов до проведения испытаний; N' — общее среднее число ненулевых элементов матрицы коэффициентов после проведения испытаний; T — среднее время, затраченное на проведение одного испытания.

16; 8; 5 % при начальном заполнении 3; 5; 7; 9 ненулевых элементов в строке соответственно. Для матриц размерности 600×600 — соответственно 45; 27; 18; 11 %. Для матриц размерности 1100×1100 — соответственно 51; 30; 20; 14 %. Таким образом, можно сказать, что стратегия выбора столбца с минимальным числом ненулевых элементов в сочетании с выбором главного элемента по столбцу наиболее эффективна на матрицах большой размерности при небольшом начальном заполнении и наименее эффективна на матрицах небольшой размерности с максимальным начальным заполнением.

Сопоставив вышесказанное с результатами тестов, можно предложить стратегию выбора главного элемента при выполнении k -го шага прямого хода по методу Гаусса, состоящую из следующих этапов:

1. Выбор столбца с наименьшим числом ненулевых элементов из столбцов с номерами от k до $n - 1$.
2. Перестановка выбранного столбца на место k -го столбца.
3. Выбор строки с наибольшим по модулю элементом в найденном столбце.

4. Перестановка выбранной строки на место k -й строки.

Приведенная стратегия минимизирует заполнение матрицы коэффициентов ненулевыми элементами, о чем свидетельствуют результаты приведенных выше испытаний, а также накопление вычислительной погрешности, так как используется выбор главного элемента по столбцу.

В результате анализа численной устойчивости метода Гаусса без выбора главного элемента и с использованием описанной выше стратегии были получены графики зависимости погрешности результата от порядка СЛАУ и разброса порядков коэффициентов исходной матрицы (рис. 1 и 2). На рисунках por — это разброс порядков коэффициентов исходной матрицы, x' — точное решение, x — расчетное решение, n — размерность решаемой СЛАУ.

Проанализировав приведенные графики, можно заметить, что при разбросе порядков коэффициентов $\text{por} > 2$ и при размерностях матрицы $n > 1000$ погрешность получаемого результата становится недопустимой. В таких случаях необходимо использовать арифметику повышенной разрядности.

В любом случае необходимо сначала оценить погрешность полученного результата. Если помимо треугольного разложения матрицы A сохранены коэффициенты исходной системы, то можно оценить погрешность решения x с помощью *итерационного уточнения* по формулам

$$\begin{aligned} r_m &= b - Ax_m; \\ Ad_m &= r_m; \\ x_{m+1} &= x_m + d_m. \end{aligned}$$

В качестве начального приближения берется вычисленный методом Гаусса вектор x . Приращение d_m вычисляется как решение линейной системы $Ax = d_m$, что сводится к решению треугольных систем $Ly = r_m$, $Ux = y$. Итерационный процесс заканчивается при значении m , для которого

$$\begin{aligned} \|x_m - x_{m-1}\| &\leq \varepsilon \|x_m\|; \\ \|d_m\| &\geq d_{m-1} \wedge m > 2, \end{aligned}$$

где $m = \text{MAXIT}$ — заданное максимальное число итераций; ε — машинная точность, равная 10^{-16} . В ходе итерационного процесса можно сделать оценку для класса нормализованных систем линейных уравнений

$$\lim_{m \rightarrow \text{MAXIT}} \frac{r_{mi}}{b_i} < 10^{-p+1},$$

где p — длина мантиссы.

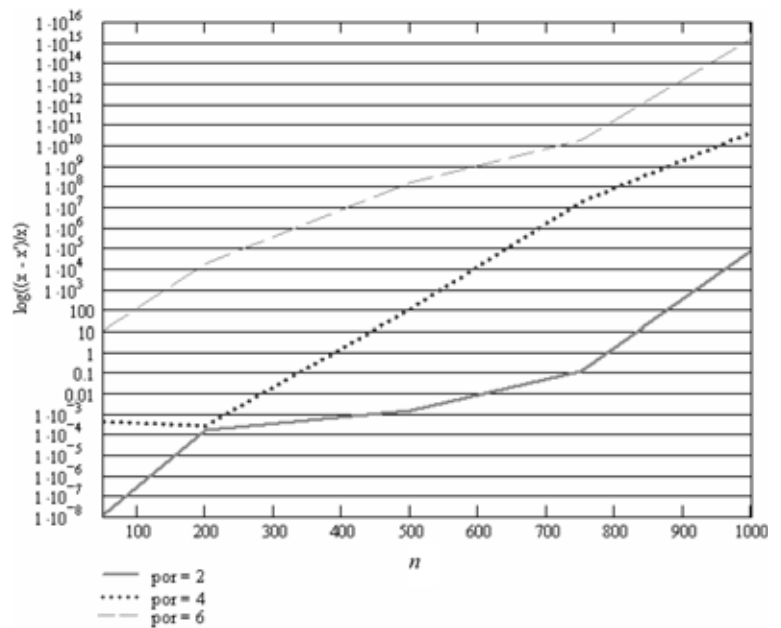


Рис. 1. Графики зависимости погрешности от порядка СЛАУ и разброса порядков коэффициентов для метода Гаусса без выбора главного элемента

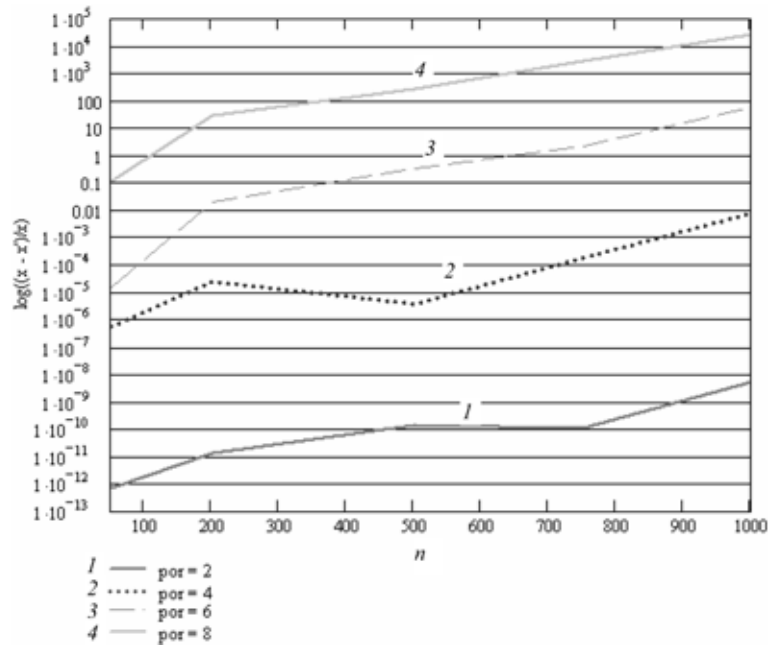


Рис. 2. Графики зависимости погрешности от порядка СЛАУ и разброса порядков коэффициентов для метода Гаусса с использованием описанной стратегии

В процессе итерационного уточнения мы можем судить только о векторах d_m и r_m . Как только перестает меняться r_m , то сразу перестает меняться и d_m . О погрешности решения в ходе итерационного уточнения можно сказать только одно: в большинстве случаев она уменьшается. Но до каких пределов? Об этом сказать ничего нельзя. Так, при тестировании (при известном эталонном решении) процесс итерационного уточнения мо-

жет нам улучшить решение от относительной погрешности 10^2 до 10^{-9} , а может и от 10^{-3} до 10^{-4} только. По векторам d_m и r_m нельзя сказать, как хорошо улучшит нам решение процесс итерационного уточнения, хотя нам нужно только знать порядок погрешности.

О точности решения системы линейных уравнений методом Гаусса мы можем судить по d_1 . При тестировании 5000 тестов d_1 дало правильную относительную погрешность в 4705 (94 %) случаях. Обозначим x^* — точное решение СЛАУ. В 244 (5 %) случаях $\|d_1\|/\|x^*\|$ было не больше, чем на два порядка больше погрешности полученного решения и в 51 (1 %) случае $\|d_1\|/\|x^*\|$ было не более, чем на два порядка меньше. Таким образом, если $\|d_1\|/\|x^*\| = 10^{-9}$, то можно сказать, что решение имеет как минимум семь верных значащих цифр. В целях повышения эффективности анализа погрешности удобно вместо метода Гаусса использовать метод LU-разложения, который по устойчивости к вычислительной погрешности и по трудоемкости равнозначен методу Гаусса. Таким образом, можно описать алгоритм решения СЛАУ с заданной точностью, который состоит из следующих шагов:

1. Решение СЛАУ методом LU-разложения с применением описанной выше стратегии выбора главного элемента.

2. Оценка ошибки полученного на предыдущем шаге решения методом итерационного уточнения.

3. Если порядок ошибки решения, вычисленной на предыдущем шаге, не превосходит заданного значения, то переход к п. 5; в противном случае — переход к п. 4.

4. Выдача предупреждения о невозможности решения системы с требуемой точностью и необходимости решения СЛАУ с использованием арифметики повышенной разрядности.

5. Полученное решение удовлетворяет заданной точности. Выход.

Заключение

В результате проведенных исследований, рассмотренных в частях 1 и 2 данной работы, выполнен анализ и обоснование выбора базовых неявных методов для интегрирования систем ДАУ и разработан новый комбинированный корректирующий алгоритм реализации неявного метода трапеций, который будет положен в основу при разработке нового универсального, высоконадежного программно-технического комплекса FMS-PA10 для моделирования систем и объектов во временной области. В результате проведенных исследований выполнены также анализ и обоснование выбора базовых методов для решения СЛАУ, возникающих при реализации неявных методов интегрирования систем ДАУ, а также предложен

алгоритм решения СЛАУ, гарантирующий точность получаемого решения.

Предполагается, что программно-технический комплекс FMS-PA10 будет базовым математическим ядром систем моделирования динамических процессов нового поколения для моделирования и анализа разнообразных и разнородных систем, технических устройств и объектов. Программно-технический комплекс FMS-PA10 будет иметь следующие принципиальные отличия от ранее разработанных программных комплексов этого класса:

- 100 %-ная гарантированная правильность интегрирования систем ДАУ при моделировании динамических процессов (в случае возможного неправильного результата будет выдано соответствующее предупреждение о необходимости перехода на вычисления с повышенной разрядностью чисел);
- возможность интегрирования функций с разрывами производных без потери точности интегрирования;
- возможность моделирования любых разнородных (электрических, электронных, механических, гидравлических, пневматических, термодинамических и др.) систем и подсистем, модели которых описываются системами ДАУ общего вида (1).
Дальнейшие исследования будут включать:
- решение проблемы вычисления элементов матрицы Якоби при реализации неявных методов интегрирования систем ДАУ;
- разработку новых методов решения СЛАУ с учетом перехода на вычисления с повышенной разрядностью чисел при необходимости;
- разработку новой методики тестирования программно-технического комплекса FMS-PA10, основанной на всестороннем тестировании модуля решения систем ДАУ, и сравнения результатов тестирования с лучшими математическими программами — MATLAB, Maple, Mathematica, MathCAD, Mathematica и т. п. (с добавлением в классические математические тесты задач для неустойчивых систем ОДУ и ДАУ с известным аналитическим решением, а также включением наборов тестовых задач моделирования реальных систем и объектов с известными значениями правильных результатов моделирования).

Результаты работы могут быть использованы при разработке модулей моделирования и анализа разнообразных систем и объектов.

Список литературы

1. Бахвалов Н. С., Жидков Н. П., Кобельников Г. М. Численные методы. М.: Наука, 1987. 487 с.
2. Богачев К. Ю. Практикум на ЭВМ. Методы решения линейных систем и нахождения собственных значений. М.: Кафедра вычислительной математики, МГУ им. М. В. Ломоносова, 1998. 367 с.

Л. Д. Черемисинова, д-р техн. наук,
гл. науч. сотр.,
Объединенный институт проблем информатики
НАН Беларуси,
e-mail: cld@newman.bas-net.by

Оценка энергопотребления КМОП-схем на логическом уровне

Рассматриваются и анализируются критерии оценки энергопотребления синхронных схем, реализуемых по КМОП-технологии, на этапе логического проектирования. Эти оценки позволяют давать сравнительные прогнозные оценки вариантов синтеза по критерию энергопотребления. При минимизации энергопотребления на логическом уровне предполагается, что схемотехнические решения проектируемой схемы, такие как частота синхронизации и напряжение питания, фиксированы.

Ключевые слова: логическое проектирование СБИС, энергопотребление, КМОП-технология

Введение

Прогресс в области микроэлектронных технологий обуславливает постоянное повышение степени интеграции, что позволяет, в свою очередь, создавать на одном кристалле все более быстродействующие и функционально сложные устройства. Одним из существенных препятствий на пути повышения плотности монтажа больших интегральных схем (БИС) является чрезмерное рассеивание ими энергии, оно приводит к перегреву устройства, нарушая его функциональность и сокращая время безотказной работы [1, 4, 11]. Это связано с тем, что быстрый рост плотности и размеров кристаллов БИС приводит к росту рассеивания теплоты и, соответственно, росту затрат на охлаждение и снижение шума. В последние годы фактор минимизации энергопотребления при проектировании интегральных схем стал играть такую же важную роль, как площадь и быстродействие. В большинстве случаев при проектировании электронных устройств приходится добиваться высокой эффективности при ограниченном потреблении электроэнергии.

Одной из основных причин повышенного внимания к вопросу энергопотребления является значительный рост числа устройств, которые по условиям применения должны быть портативными (например, портативные вычислительные и мультимедийные устройства). Для последних очень важно увеличить время их автономной работы без подзарядки. Однако и для непортативных систем

фактор энергопотребления в настоящее время выходит на первый план. Другими словами, достижение высокой энергоэффективности является в настоящее время одной из главных задач при проектировании всех электронных устройств.

В работе рассматривается проблема оценки энергопотребления микросхем, выполненных на основе КМОП-технологии [1, 2]. Известно, что компоненты КМОП-схем потребляют подавляющую часть всей необходимой для функционирования энергии во время их переключения. Таким образом, энергопотребление существенно зависит от последовательности входных воздействий на КМОП-схемы. САПР микроэлектронных схем должны иметь средства, позволяющие оценить энергопотребление схемы на стадии ее проектирования и выбрать наименее энергоемкий вариант. Приведем основные из упоминаемых в литературе [1, 2] возможных путей сокращения рассеиваемой мощности:

- уменьшение напряжения питания (технология должна допускать функционирование схемы при малых питающих напряжениях, но уменьшение значения напряжения ограничено снизу допустимым отношением сигнала к шуму);
- уменьшение емкостей микросхем (внутренней диффузионной емкости самой микросхемы, емкости межсоединений и разветвлений по выходу);
- управление синхронизацией и подачей питания (определяются интервалы времени, в течение которых некоторые части схемы не выполняют полезную работу, затем такие блоки выводятся из работы на это время путем отключения питания или блокировки синхронизации);
- минимизация интенсивности переключений сигналов на входах микросхем (ведет к оптимизации динамического рассеивания энергии на логическом уровне).

В настоящей работе анализируются основные из известных критериев оценки энергопотребления синхронных комбинационных схем, реализуемых по КМОП-технологии, на этапе логического проектирования. Эти оценки тесно увязаны с энергопотреблением реализованной на кристалле СБИС схемы, позволяя в процессе логического синтеза давать сравнительные оценки вариантов схемы по критерию энергопотребления. При минимизации энергопотребления на логическом уровне предполагается, что схемотехнические решения проектируемой схемы, такие как частота синхронизации и напряжение питания, фиксированы.

Под оценкой энергопотребления далее понимается оценка среднего значения энергии, рассеиваемой микросхемой (эта оценка существенно

отличается от оценки максимальной мощности, потребляемой мгновенно). При минимизации энергопотребления на логическом уровне предполагается, что схемотехнические решения проектируемой схемы, такие как частота синхронизации и напряжение питания и земли, фиксированы.

1. Составляющие потребляемой микросхемой мощности

При обсуждении вопроса, какая энергия (мощность) потребляется при работе микросхемы, понятия потребляемая и рассеиваемая мощность, как правило, считаются синонимами. В принципе, следует различать понятия энергии и мощности. Потребляемая микросхемой мощность зависит от тактовой частоты и может быть сокращена за счет ее уменьшения. Однако в этом случае на выполнение одних и тех же вычислений потребуется большее время, и соответственно, расход энергии питающей батареи не уменьшится. Отсюда более правильно было бы говорить о важности снижения энергопотребления, а не потребляемой мощности. Учитывая, что далее считается, что тактовая частота проектируемой схемы неизменна, снижение потребляемой мощности означает и снижение энергопотребления.

В КМОП-технологии всю рассеиваемую микросхемой мощность можно разделить на статическую и динамическую составляющие: $P = P_{\text{stat}} + P_{\text{dyn}}$. Статическая составляющая обусловлена наличием статических проводящих путей между шинами питания или токов утечки. Она присутствует даже тогда, когда схема бездействует — находится в режиме ожидания. Статическое или стационарное рассеяние энергии микросхемы зависит от ее топологии и определяется током утечки, протекающим между шинами питания при нулевой активности переключений [1, 3]. У большинства хорошо спроектированных КМОП-схем статическая рассеиваемая мощность очень мала.

Около 80 % всей рассеиваемой энергии приходится на ее динамическую составляющую [5]. Динамическая мощность рассеивается только во время переходных процессов, когда сигналы на выходах узлов микросхемы переключаются. Эта мощность обусловлена зарядом емкостей и возникновением временных путей протекания тока между шинами питания. В КМОП-технологии основной вклад в суммарную динамическую мощность вносит [2—5] мощность P_{dyn} , обусловленная зарядом и разрядом емкостной нагрузки узла, и мощность P_{sc} рассеивания за счет протекания сквозных токов через микросхему. Составляющая P_{sc} вызывается токами, проходящими между источником питания и землей через пары p -МОП/ n -МОП транзисторов, когда во время из-

менения значения сигнала на входе узла они оба частично открыты (так как в реальных схемах времена нарастания и спада входного сигнала не равны нулю) и проводят ток одновременно [5]. В типичных КМОП-цепях доминирующим фактором является емкостное рассеяние P_{dyn} , а потребление мощности P_{sc} удерживается в приемлемых пределах на технологическом уровне, поэтому этот фактор при оптимизации энергопотребления схемы на логическом уровне, как правило, не учитывается [1, 9, 11].

Наиболее существенной причиной потребления мощности КМОП-схемой в динамике является динамическая рассеиваемая мощность P_{dyn} , обусловленная токами заряда/разряда паразитных емкостей транзисторов и линий связи. Эти емкости учитываются в виде емкостной нагрузки C_L выхода узла, значение которой практически равномерно распределено между внутренней емкостью, образованной диффузионной и перекрывающейся емкостями, и внешней емкостью, обусловленной внешней нагрузкой проводника и затвора [1]. При смене входного сигнала с 0 на 1 ток, протекающий через транзистор с p -каналом, заряжает емкость C_L , при смене входного сигнала с 1 на 0 емкость C_L разряжается током, текущим через транзистор с n -каналом. В каждом из этих случаев на сопротивлении открытого транзистора рассеивается мощность, выражаемая известным соотношением [6, 7, 8] $P_{\text{dyn}} = \frac{1}{2} V_{dd}^2 C_L A_s$, где V_{dd} — напряжение питания; A_s — интенсивность переключений узла микросхемы, равная математическому ожиданию числа переключений узла в единицу времени, при которых потребляется мощность.

Можно выделить две важные составляющие интенсивности A_s переключений схемы: частоту синхронизации f_{clk} , определяющую среднюю периодичность изменения состояния входа схемы, и величину E_s , значение которой равно среднему числу переключений сигнала на выходе узла в течение одного такта работы схемы: $A_s = f_{clk} E_s$. Составляющая E_s определяется как математическое ожидание числа логических переходов (из 1 в 0 или из 0 в 1) за один период синхронизации и называется в литературе *переключательной активностью*. Исходя из определения, значение переключательной активности равно 1 или 0. Отсюда получается известное соотношение [6—8], выражающее рассеиваемую узлом мощность P_{dyn} через его переключательную активность, $P_{\text{dyn}} = \frac{1}{2} V_{dd}^2 f_{clk} E_s C_L$. Произведение $E_s C_L$ часто называют переключае-

мой емкостью [10, 11], задающей среднее значение перезаряжаемой емкости в течение одного такта (периода). Рассеивание энергии микросхемы в целом вычисляется [5, 11] как сумма:

$$P_s = \frac{1}{2} V_{dd}^2 f_{clk} \sum_{i=1}^n E_i C_i, \quad (1)$$

где n — число узлов в схеме (суммирование проводится по всем узлам схемы); C_i — емкостная нагрузка i -го узла; E_i — переключательная активность i -го узла схемы.

2. Подходы к оценке энергопотребления

Известны два подхода [4, 11–14] к оценке энергопотребления схем:

- методы, основанные на динамическом анализе;
- методы, основанные на статическом анализе.

Первый подход предполагает моделирование исследуемой схемы на множестве возможных сценариев ее функционирования (упорядоченных последовательностей наборов значений сигналов на ее входах) [15, 17]. Моделирование широко используется в практике проектирования для установления разных параметров (в том числе и энергопотребления) уже спроектированных схем. Наиболее известным и достаточно эффективным средством моделирования электрических схем является система SPICE (*Simulation Program with Integrated Circuit Emphasis*) — программа моделирования с ориентацией на интегральные схемы (разработка Калифорнийского университета Беркли). В процессе моделирования схемы на заданной последовательности наборов значений входных сигналов в форме диаграмм входных напряжений вычисляются усредненные формы диаграмм выходных напряжений и токов узлов схемы, по которым вычисляется среднее энергопотребление схемы с учетом технологии ее изготовления. Однако результаты моделирования заданной электрической схемы существенно зависят от выбора тестируемой последовательности, от того, какие и сколько наборов значений входных сигналов используется, в какой последовательности они подаются, какова форма сигналов и т. д. Эти данные могут быть известны для уже спроектированной схемы и ее окружения на кристалле.

Методы оценки энергопотребления схем, основанные на моделировании, к сожалению, не годятся для использования в процессе проектирования схемы, когда ее окружение также еще не спроектировано или даже не определено. В этом случае практически ничего неизвестно о возможных ее входных воздействиях. Кроме того, для получения достоверной оценки энергопотребления необходимо рассмотреть большое число входных воздействий, что делает использование методов

оценки энергопотребления на основе моделирования практически невозможным для сложных схем [18]. Основная проблема при моделировании таких схем заключается в определении априори числа стимулирующих воздействий и порядка их следования.

В основе статических методов оценки энергопотребления схем лежит вероятностный подход к определению переключательной активности сигналов, он основан на вероятностных характеристиках входных сигналов и функционально-структурных свойствах исследуемой схемы. Эти методы обеспечивают компактное описание последовательности наборов входных сигналов. Подход предполагает задание вероятности переключения каждого из сигналов на входе схемы, которая отражает частоту смены его значения на входе. Методы основаны на распространении вероятностной информации о смене значений сигналов через всю схему, от входов к выходам [17]. Вероятностный подход позволяет: 1) компактно определить возможные последовательности входных воздействий на схему и оценить энергопотребление схемы, исходя из взаимного влияния входных сигналов в последовательные моменты времени; 2) избежать повторения циклов моделирования схемы на задаваемом большом множестве наборов значений входных сигналов.

Вероятностные методы хотя и уступают методу моделирования по степени соответствия реальному поведению схемы, однако позволяют просто оценить, как часто сигналы на выходах узлов схемы меняют свои значения, и использовать эту информацию для оценки динамической составляющей (1) рассеиваемой мощности. Именно этот подход к оценке энергопотребления позволяет сравнивать варианты реализации схемы в процессе ее проектирования, что дает возможность уже на логическом уровне проектировать схемы с низким энергопотреблением.

Наряду с упомянутыми подходами развиваются гибридные методы, например методы вероятностного моделирования [19], в которых исходное задание последовательности наборов диаграмм входных сигналов заменяется вероятностными диаграммами входных сигналов. Вероятностная диаграмма представляет собой последовательность значений, задающих вероятности, с которыми сигнал принимает значение 1 для определенного интервала времени, а также вероятности смены его значения с 0 на 1 для определенных моментов времени. Вероятностная диаграмма входных сигналов позволяет вычислить усредненные диаграммы тока через узлы схемы за один шаг моделирования, по которым и определяется среднее энергопотребление элементов и схемы в целом.

3. Вероятностные оценки переключательной активности сигналов

Вычисление переключательных активностей узлов для реальных схем весьма трудоемкая процедура, требующая учета множества факторов, возможных при функционировании схемы. Используемые в практике проектирования оценки переключательной активности схемы выведены, как правило, в предположении нулевой и ненулевой задержки сигнала узлами схемы. В первых моделях оценки предполагается, что изменения на входах схемы распространяются через все ее элементы мгновенно и значит одновременно. Во вторых моделях оценок предполагается, что элементы схемы имеют конечные (но не нулевые) задержки, и принимаются во внимание дополнительные переключения, не предусмотренные функцией, реализуемой узлом, и обусловленные переходными процессами (глюками и гонками), происходящими во время смены значений сигналов на входе до тех пор, пока узел не установится в конечное устойчивое состояние. Паразитные переходы часто возникают в тех точках схемы, где сходятся ветки схемы с разбалансированными по отношению друг к другу задержками относительно входа схемы.

Переключения, обусловленные переходными процессами, также рассеивают энергию, не выполняя при этом предусмотренные схемой вычисления. Этот компонент энергопотребления называют иногда паразитной активностью. Переключения, вызванные сменой значения на входе схемы и предписанные функцией, которую реализует узел, являются полезным компонентом энергопотребления и называются *функциональной активностью*.

Использование моделей реальных задержек существенно повышает время вычисления оценок переключательной активности схемы, одновременно повышая их точность. В работе [16] показано, что отношение паразитного рассеивания мощности к суммарной мощности рассеивания меняется в достаточно широких пределах для разных схем (от 9 до 38 %) и особенно велико в некоторых схемах обработки данных (например в умножителях). Учитывая трудоемкость вычисления реальной переключательной активности схемы, упомянутую погрешность вычислений можно считать допустимой, особенно в случае вычисления этих оценок в процессе синтеза логической схемы.

Большая часть энергии статических КМОП-схем потребляется во время зарядки и разрядки емкостной нагрузки. Для того чтобы оценить потребление энергии логической схемой, необходимо вычислить переключательные активности ее внутренних полюсов. Оценка переключательной

активности некоторого полюса i схемы основана на вычислении вероятности p_i появления на нем сигнала 1. При этом, как правило, предполагается, что для полюсов схемы имеет место *пространственная* и *временная* независимость. Пространственная независимость предполагает отсутствие корреляции сигналов на разных полюсах. Она может возникать, когда пространственно связанные сигналы зависят друг от друга, это вызывается, например, наличием разветвлений на выходах элементов или обратных связей, но на практике между сигналами может существовать устойчивая зависимость, вызываемая и другими причинами. Временная независимость предполагает, что значение сигнала в любом такте синхронизации не зависит от его значений в предшествующих тактах.

В литературе предложено множество вероятностных методов оценки энергопотребления логических схем [4, 5, 11, 17], основная масса из них применима только для комбинационных схем. Методы оценки энергопотребления сформулированы относительно разных предположений о задержке сигналов узлами схемы, возможности разных видов зависимостей между сигналами, учете переходных процессов при смене сигнала и т. д.; с использованием разных статистических характеристик (вероятности появления сигнала 1, вероятности смены сигнала на полюсе, интенсивностей переключений сигналов — плотностей переключений сигналов, равновесных вероятностей, вероятностных форм сигналов и т. д.); с использованием разных моделей: на основе бинарных диаграмм решений (БДР, от англ. *Binary Decision Diagram* — BDD); коэффициентов корреляции. Основные из этих оценок рассмотрены ниже.

4. Оценки переключательной активности в предположении нулевой задержки сигналов

Различают [17] вероятность p_i^1 появления сигнала 1 (0) на некотором i -м полюсе и вероятность смены сигнала на полюсе. Вероятность p_i^1 называется *сигнальной вероятностью* (вероятностью сигнала 1) и определяется средней долей тактов, на которых сигнал на i -м полюсе имеет единичное значение. Вероятность p_i^t ($p_i^{1 \rightarrow 0}$ или $p_i^{0 \rightarrow 1}$) называется *переключательной активностью* и определяется средней долей тактов, на которых сигнал на i -м полюсе меняет свое значение по сравнению со значением в предшествующем такте.

В предположении нулевых задержек элементов (что исключает переключения за счет переходных процессов) и временной независимости сигналов вероятность $p_i^{1 \rightarrow 0}$ ($p_i^{0 \rightarrow 1}$) смены сигнала на i -м полюсе с 1 на 0 (с 0 на 1) равна произведению p_i^1

вероятности появления на нем сигнала 1 (0) в одном такте на вероятность того, что в следующем такте на нем появится 0 (1). Соответственно переключательная активность i -го полюса схемы равна $E_i = p_i^1 \rightarrow 0 \ p_i^0 \rightarrow 1 = 2p_i^1 p_i^0$, и в предположении, что $p_i^1, p_i^0 < 1$:

$$E_i = 2p_i^1 (1 - p_i^1). \quad (2)$$

Например, если сигнальная вероятность полюса $p_y^1 = 0,2$, то $E_y = 0,32$, если $p_y^1 = 0,5$, то переключательная активность $E_y = 0,5$ имеет максимальное значение.

Вероятность p_y появления сигнала 1 на выходе элемента e зависит от функции, реализуемой этим элементом. Если предположить, что значения сигналов на входах статического КМОП-элемента независимы и распределены равномерно (т. е. сигнальные вероятности всех входных полюсов равны 0,5), то вероятности появления сигналов 1 и 0 на его выходе легко найти по таблице истинности, реализуемой им функции (они равны отношению числа соответственно единиц k^1 и нулей k^0 в столбце значений функции к числу всех наборов): $p_e^1 = \frac{k^1}{2^n}, p_e^0 = \frac{k^0}{2^n}$. Соответственно, выход элемента с n входами имеет переключательную активность $E_e = 2 \frac{k^1 k^0}{2^n 2^n} = 2 \frac{k^1(2^n - k^1)}{2^{2n}}$.

Например, сигнальные вероятности и переключательные активности выходов простых элементов И, ИЛИ, И-НЕ, ИЛИ-НЕ, исключаящее ИЛИ (см. таблицу) выражаются следующим образом: $p_e^{\wedge 1} = p_e^{\vee 1} = p_e^{\vee 0} = p_e^{\bar{\wedge} 0} = \frac{1}{4}$; $p_e^{\vee 1} = p_e^{\bar{\wedge} 1} = p_e^{\wedge 0} = p_e^{\bar{\vee} 0} = \frac{3}{4}$; $p_e^{\oplus 1} = p_e^{\oplus 0} = \frac{1}{2}$; $E_e^{\wedge} = E_e^{\vee} = E_e^{\bar{\wedge}} = E_e^{\bar{\vee}} = \frac{3}{8}$; $E_e^{\oplus} = \frac{1}{2}$.

В реальных случаях использование равномерного распределения значений сигналов на входах элемента неправомерно. Переключательная активность выхода логического элемента существенно зависит от вероятностных характеристик сигналов на его входах. В случае, когда сигналы на

Функции, реализуемые элементами НЕ, И, ИЛИ, исключаящее ИЛИ, И-НЕ, ИЛИ-НЕ

X	Y	\bar{x}	$x \wedge y$	$x \vee y$	$x \oplus y$	$\overline{x \wedge y}$	$\overline{x \vee y}$
0	0	1	0	0	0	1	1
0	1	1	0	1	1	1	0
1	0	0	0	1	1	1	0
1	1	0	1	1	0	0	0

входах элемента не коррелируют в пространстве и во времени, сигнальные вероятности простых элементов, типа инвертора, И, ИЛИ, И-НЕ, ИЛИ-НЕ с $n(e)$ входными полюсами и двухвходового элемента "исключающее ИЛИ", легко могут быть подсчитаны, исходя из таблиц истинности реализуемых ими функций (см. таблицу, столбец 6 для случая $n(e) = 2$):

$$p_e^{\bar{\vee}} = 1 - p_1; p_e^{\wedge} = \prod_{i=1}^{n(e)} p_i; p_e^{\vee} = 1 - \prod_{i=1}^{n(e)} (1 - p_i); \quad (3)$$

$$p_e^{\bar{\wedge}} = 1 - \prod_{i=1}^{n(e)} p_i; p_e^{\bar{\vee}} = \prod_{i=1}^{n(e)} (1 - p_i);$$

$$p_e^{\oplus} = p_1 + p_2 - 2p_1 p_2; \quad (4)$$

где p_i — сигнальная вероятность i -го входа элемента.

В случае элементов, реализующих более сложные функции, для которых затруднительно построение таблиц истинности, эти оценки могут быть вычислены методом, основанным на построении упорядоченных бинарных диаграмм решений (OBDD от англ. Ordered Binary Decision Diagram) [20].

5. Распространение вероятностной информации через схему

Если заданы сигнальные вероятности входных сигналов схемы, то они могут быть распространены на выходы элементов схемы и через всю схему на ее выходные полюсы. Таким образом могут быть подсчитаны переключательные активности всех полюсов схемы и, соответственно, переключательная активность схемы в целом. Следует принимать во внимание, что сигнальная вероятность для узлов схемы может быть подсчитана по формулам (3), (4) только в том случае, если для всех его входов выполняется условие пространственной (и временной) независимости, а сам элемент имеет нулевую задержку. Но даже если требование независимости выполняется для входных сигналов схемы, оно может не иметь места для входных сигналов внутренних элементов схемы. Нарушение условия независимости может быть результатом наличия разветвлений на выходах элементов и линий обратной связи.

Частным случаем схем, удовлетворяющих условиям независимости входов всех их элементов, являются древообразные схемы. Для этих схем вычисление сигнальных вероятностей для элементов по формулам (3), (4) дает в итоге точное значение (в рамках принятых ограничений) этой вероятности для выходного полюса схемы. Для других схем (где существует хотя бы один полюс, питающий более чем один элемент) вычисление вероятностей по этим формулам порождает приближенные значения сигнальных вероятностей.

Точные методы подсчета сигнальной вероятности для некоторого узла схемы, сигналы на входе которого независимы, основаны на нахождении глобальной функции $f(X)$, реализуемой этим узлом, т. е. функции, аргументами которой являются входные переменные x_i , соответствующие входам схемы. Один из таких методов основан на представлении функции $f(X)$ в виде совершенной дизъюнктивно-нормальной формы (ДНФ) D . Отличительная особенность этой ДНФ заключается в том, что все ее конъюнкции являются полными и соответственно взаимоисключающими: для всех $i \neq j$ вероятность того, что $p_{ki}p_{kj} = 1$, равна 0, так как найдется хотя бы одна переменная, входящая в эти конъюнкции во взаимно инверсном виде, а $p_{xi}p_{x\bar{i}} = 0$. Вероятность того, что ДНФ D примет

значение 1, равна $p_D = 1 - \prod_{i=1}^{n(D)} (1 - p_{ki})$ где $n(D)$

число конъюнкций ДНФ. Эта формула для $n(D) = 2$ трансформируется в $p_D = p_1 + p_2 - p_1p_2$, а в общем случае, когда $n(D) > 2$, она содержит

сумму $\sum_{i=1}^{n(D)} p_{ki}$ значений сигнальных вероятностей

p_{ki} для конъюнкций ДНФ минус всевозможные произведения вероятностей p_{ki} для разных сочетаний конъюнкций, для случая полных конъюнкций все эти произведения равны 0. Соответственно,

но, сигнальная вероятность ДНФ D : $p_D = \sum_{i=1}^{n(D)} p_{ki}$

Другой точный метод [20] подсчета сигнальных вероятностей для внутренних полюсов схемы основан на использовании упорядоченных бинарных диаграмм решений (УБДР) (рис. 1), которые строят для глобальных функций, реализуемых на полюсах анализируемой схемы. Этот метод имеет линейную сложность относительно числа вершин диаграммы, но экспоненциальную сложность относительно числа входных полюсов схемы (что связано с построением самой диаграммы двоичных решений). При подсчете сигнальной вероятности для рассматриваемого узла УБДР выполняется просмотр вершин УБДР, начиная от терминальных (листьевых) вершин, помеченных символами 0 и 1, к корню. Для каждой нетерминальной вершины b , помеченной символом аргумента функции x_i , подсчитывается вероятность достижения этой вершины следующим образом:

$$p_b = p_{xi}p_{b1} + p_{x\bar{i}}p_{b0}, \quad (5)$$

где b_1 и b_0 — переменные, помечающие вершины-предшественники, связанные с вершиной b дугами, помеченными символами 1 и 0, и соответст-

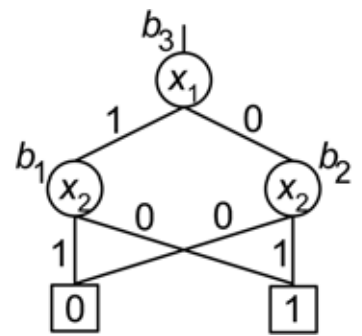


Рис. 1. Упорядоченная бинарная диаграмма решений функции "исключающее ИЛИ"

вующие коэффициентам разложения функции по переменной x_i (b_1 и b_0 соответствуют коэффициентам f_{xi} и $f_{x\bar{i}}$ в дизъюнктивном разложении функции f по переменной x_i : $f = x_i f_{xi} \vee \bar{x}_i f_{x\bar{i}}$; p_{xi}

и $p_{x\bar{i}}$ — сигнальные вероятности для 1 i -го аргумента функции; p_{b1} и p_{b0} — вероятности, рассчитанные для вершин b_1 и b_0 диаграммы. Справедливость формулы (5) следует из тех же соображений,

что и справедливость формулы $p_D = \prod_{i=1}^{n(D)} p_{ki}$

для совершенной ДНФ.

Оба этих алгоритма, к сожалению, достаточно трудоемки, так как требуют построения совершенной ДНФ или УБДР глобальной функции, реализуемой каждым полюсом схемы. Более простым является алгоритм [5], основанный на распространении вероятностей от входов (сигнальные вероятности которых заданы) к выходам схемы. Для каждого элемента e , для входов p_{ei} которого уже найдены алгебраические выражения, задающие сигнальные вероятности, записываются алгебраические выражения, задающие сигнальные вероятности, в соответствии с функциями, реализуемыми этими элементами, и формулами (3), (4), выведенными в предположении пространственной независимости входных сигналов. Во всех найденных выражениях для вероятностей опускаются знаки степени, т. е. все p_{ei}^k ($k > 1$) заменяются на p_{ei} .

В связи с тем, что в процессе синтеза схемы с малым энергопотреблением подсчет (и пересчет) вероятностей проводится многократно, на практике, как правило, приходится иногда довольствоваться не трудоемким методом, дающим точную (в рамках сделанных допущений) оценку, а более простым, не гарантирующим (в общем случае) точность полученной оценки, методом подсчета вероятностей. Этот метод основан на распространении значений сигнальных вероятностей к выходам схемы: начиная от входов схемы с заданными сигнальными вероятностями p_{xi} , вычисляются по формулам (3), (4) вероятности для всех

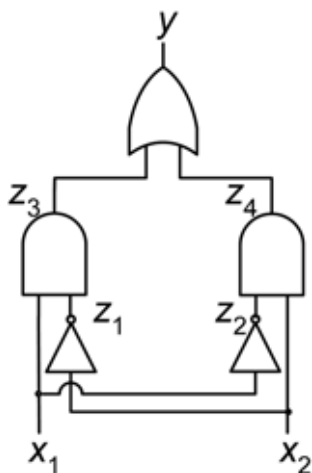


Рис. 2. Схема, реализующая функцию "исключающее ИЛИ"

полюсов схемы. Разница значений вероятностей, вычисленных таким образом (по сравнению с точными методами), для схем, отличных от древовидных, получается за счет упомянутых выше экспонент.

Например, пусть сигнальные вероятности и соответствующие им переключательные активности для входных полюсов схемы имеют значения

$$p_{x1} = 0,2; p_{x2} = 0,7; E_{x1} = 0,32; E_{x2} = 0,42. \quad (6)$$

Подсчитаем разными способами сигнальные вероятности и переключательные активности полюсов простенькой схемы (рис. 2):

$$\begin{aligned} p_{z1} = p_{\bar{x}2} = 0,3; p_{z2} = p_{\bar{x}1} = 0,8; \\ p_{z3} = 0,06; p_{z4} = 0,56; p_y^1 = 0,5864; \\ E_{z3} = 0,1128; E_{z4} = 0,4928; E_y^1 = 0,4851. \end{aligned} \quad (7)$$

Подсчет сигнальных вероятностей для выходного полюса анализируемой схемы (рис. 2) по УБДР (см. рис. 1) функции, реализуемой этой схемой, приводит в соответствии с (5) к следующим результатам:

$$\begin{aligned} p_{b1} = p_{\bar{x}2} = 0,3; p_{b2} = p_{x2} = 0,7; \\ p_{b3} = p_{x1}p_{\bar{x}2} + p_{\bar{x}1}p_{x2} = 0,62; E_y^2 = 0,4712. \end{aligned}$$

Как и должно получиться, результат подсчета сигнальной вероятности для функции "исключающее ИЛИ" в соответствии с формулой (4) совпадает с результатом, полученным при использовании УБДР:

$$\begin{aligned} p_y^{\oplus} = p_{x1} + p_{x2} - 2p_{x1}p_{x2} = 0,9 - 0,28 = 0,62, \\ E_y^2 = 0,4712. \end{aligned} \quad (8)$$

Разница между подсчетами E_y^1 (7) и E_y^2 (8) переключательной активности, в основе которых лежат вероятности, вычисляемые по формулам (3), (4) и по УБДР, равна $E_y^1 - E_y^2 = 0,0139$, т. е.

разница в сотых долях, или погрешность составляет 2,9 %, для более сложных схем эта разница будет нарастать. Однако во втором случае приходится строить УБДР, что для сложных схем (а соответственно сложных функций) весьма затруднительно. Разница в подсчетах, как уже говорилось, получается вследствие того, что значения сигналов на входах внутренних элементов схемы коррелируют, из-за чего нарушается условие пространственной независимости полюсов схемы.

6. Оценки переключательной активности в предположении ненулевой задержки сигналов

Как уже упоминалось, ненулевая задержка сигнала реальными элементами является причиной того, что на входы внутреннего узла схемы сигналы приходят с разной задержкой. Это приводит к наличию паразитных переключений выходного сигнала на выходе узла до его установки в устойчивое состояние, и соответственно, к росту динамической составляющей энергопотребления узла, который никак не отражается в оценках, подсчитываемых в предположении нулевой задержки сигналов. Методы оценки энергопотребления внутренних узлов схемы в предположении ненулевой задержки сигналов основаны на более общих, по сравнению с сигнальной вероятностью и переключательной активностью, понятиях — *равновесной вероятности и плотности переключений* [17]. Эти характеристики определяют поведение сигнала во времени, а не только поведение в устойчивом состоянии в течение такта. Предполагается, что логический сигнал $x(t)$ представляет собой стохастический процесс смены значений с 0 на 1 и с 1 на 0 в случайные моменты времени в интервале времени $t \in (-\infty, +\infty)$.

Равновесная вероятность P_x логического сигнала $x(t)$ определяется как средняя доля времени, в течение которого сигнал имеет значение 1:

$$P_x = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} x(t) dt.$$

В отличие от сигнальной вероятности равновесная вероятность зависит от внутренних задержек схемы и учитывает их. В предположении нулевой задержки сигналов равновесная вероятность сводится к сигнальной вероятности. Плотность переключений логического сигнала $x(t)$ определяется как $A_x = \lim_{T \rightarrow \infty} \left(\frac{n(T)}{T} \right)$, где $n(T)$ — число переключений сигнала $x(t)$ в интервале времени длины T .

Так как плотность переключений зависит от внутренних задержек схемы, то переключательная

активность является нижней оценкой значений переходной плотности. В синхронной схеме с длительностью такта T_c имеет место следующее соотношение между плотностью переключений и переключательной активностью: $A_x \geq \frac{E_x}{T_c}$. Равенство

имеет место для случая нулевой задержки сигналов.

При вычислении плотности переключений внутреннего узла схемы предполагается, что плотности переключений сигналов на его входе заданы. Показано [10], что если входы x_i элемента y пространственно независимы, то плотность переключений сигнала на его выходе вычисляется как

$$A_y = \sum_{i=1}^n P\left(\frac{dy}{dx_i}\right) A_{xi}. \quad (9)$$

Если функция y зависит от булевой переменной x , то булева разность y относительно переменной x определяется как $\frac{dy}{dx} = y|_{x=1} \oplus y|_{x=0}$.

Булева разность задает в логическом виде зависимость значения функции y от значения ее аргумента x или степень влияния входного сигнала x на значение сигнала на выходе элемента y . Если

$\frac{dy}{dx} = 1$, то всякое изменение значения переменной x влечет за собой изменение значения функции y ; $P\left(\frac{dy}{dx}\right)$ определяет вероятность проявления

изменения значения входного сигнала x на выходе элемента y , а $P\left(\frac{dy}{dx}\right)A_x$ — вклад сигнала x в переключательную активность выходного полюса элемента y .

При вычислении интенсивности переключений на основе формулы (9) не принимается во внимание пространственная корреляция сигналов, но зато, хоть и приближенно, но учитываются глюки, сопровождающие переходные процессы.

Следует отметить, что формула (9) выведена в предположении одновременного переключения сигналов на входе элемента. Ее применение для случая одновременного переключения сигналов на входе приводит к некорректно подсчитанной интенсивности переключений. В работе [5] предлагается обобщение модели (9) на случай наличия одновременно переключающихся входов.

7. Распространение вероятностной информации через схему

Если даны значения сигнальной вероятности и плотности переключений на входах схемы, то за один проход схемы могут быть подсчитаны плот-

ности переключений на всех узлах схемы, а стало быть, и оценка значения рассеиваемой схемой мощности в соответствии с формулой

$$P_s = \frac{1}{2} V_{dd}^2 f_{clk} \sum_{i=1}^n A_i C_i, \text{ аналогичной (1).}$$

Для того чтобы вычислить вероятности $P\left(\frac{dy}{dx_i}\right)$, нужно распространить по схеме от входов

также и равновесные вероятности P_x . Показано [10], что распространение этих вероятностей может осуществляться аналогично тому, как это делается при распространении сигнальных вероятностей при допущении нулевых задержек сигналов в схеме, если принять, что все задержки вынесены из схемы в виде одной задержки, привязанной к выходу блока, реализующего логику схемы. Соответственно, при подсчете равновесной вероятности P_y узла y схемы, можно использовать один из точных методов, лежащих в основе подсчета сигнальных вероятностей и упомянутых в разд. 5, например, предлагается использовать УБДР, с помощью которой легко найти и булевы разности.

Например, для элементов, реализующих простые функции $y = \bar{x}$, $y = x_1 x_2$, $y = \bar{x}_1 \bar{x}_2$, $y = x_1 \vee x_2$, $y = x_1 \vee x_2$, $y = x_1 \oplus x_2$, плотности переключений сигналов на выходах вычисляют следующим образом:

$$\frac{dy}{dx} = 1, A_y^- = A_x; \quad (10)$$

$$\begin{aligned} \frac{dy}{dx_1} = x_2, \frac{dy}{dx_2} = x_1 \text{ и } A_y^{\wedge} = A_y^{\bar{\vee}} = \\ = P_{x_2} A_{x_1} + P_{x_1} A_{x_2}; \end{aligned} \quad (11)$$

$$\frac{dy}{dx_1} = x_2, \frac{dy}{dx_2} = \bar{x}_1$$

$$\text{и } A_y^{\vee} = A_y^{\bar{\wedge}} = (1 - P_{x_2}) A_{x_1} + (1 - P_{x_1}) A_{x_2}; \quad (12)$$

$$\frac{dy}{dx_1} = \frac{dy}{dx_2} = 1 \text{ и } A_y^{\oplus} = A_{x_1} + A_{x_2}. \quad (13)$$

Для примера подсчитаем переключательную активность на выходе уже рассмотренной выше схемы (см. рис. 2). Пусть равновесные вероятности для входных полюсов схемы и соответствующие им плотности переключений имеют те же значения, что и сигнальные вероятности и переключательные активности (6) для рассмотренного выше примера: $P_{x_1} = 0,2$; $P_{x_2} = 0,7$; $A_{x_1} = 0,32$; $A_{x_2} = 0,42$.

Равновесные вероятности и плотности переключений для полюсов схемы согласно формулам (3), (10)—(12) имеют следующие значения: $p_{z_1} = 0,3$;

$p_{z2} = 0,8$; $p_{z3} = 0,06$; $p_{z4} = 0,56$; $A_{z1} = 0,42$;
 $A_{z2} = 0,32$; $A_{z3} = 0,18$; $A_{z4} = 0,56$; $A_y^1 = 0,6056$.

Подсчет плотности переключения для функции "исключающее ИЛИ" в соответствии с формулой (13) дает $A_y^2 = 0,74$. Разница между результатами подсчетов $A_y^1 = 0,6056$ и $A_y^2 = 0,74$ плотности переключений путем распространения вероятностной информации через схему, реализующую функцию "исключающее ИЛИ", и по формуле (13) равна $A_y^2 - A_y^1 = 0,1344$, т. е. разница составила 18,16 %. Основной вклад в получившуюся разницу, помимо корреляции значений сигналов на входах внутренних элементов, вносит разбалансированность во времени поступления сигналов на входы этих элементов вследствие ненулевых задержек сигналов. Разница между результатами подсчетов $A_y^2 = 0,74$ плотности переключений и $E_y^2 = 0,4712$ (8) переключательной активности составляет 0,2688, или 36 % от всей потребляемой динамической мощности A_y^2 (что не противоречит оценкам этой доли, приведенным в работе [16]). Эта доля мощности приходится на не учитываемую (оценкой переключательной активности) интенсивность переключений за счет паразитных переходов.

Заключение

Проанализированы основные оценки энергопотребления статических КМОП-схем, лежащие в основе других предложенных в литературе оценок (см. работы [4], [5], [8], [10], [12], [21]), введенных в предположениях, в разной степени ослабляющих такие требования к сигналам (входным для схемы и внутренним), как временная и пространственная независимость.

Список литературы

1. Рабаи Ж. М. Цифровые интегральные схемы. Методология проектирования. М.: Вильямс, 2007. 912 с.
2. Уэйкерли Дж. Проектирование цифровых устройств. Т. 1. М.: Постмаркет, 2002. 544 с.

3. Bellaouar A. B., Elmasry M. I., Allen J. Low-Power Digital VLSI Design Circuits and Systems. USA, Norwell: Kluwer Academic Publishers, 1995. 530 с.
4. Benini L., Micheli G. De. Logic Synthesis for Low Power // Logic Synthesis and Verification / Eds. S. Hassoun, T. Sasao, R. K. Brayton. — Boston, Dordrecht, London: Kluwer Academic Publishers, 2002.
5. Roy K., Prasad S. C. Low Power CMOS VLSI Circuit Design. New York: John Wiley and Sons Inc., 2000. 376 p.
6. Chandrakasan A. P. Low Power CMOS Digital Design // IEEE Journal on Solid-State Circuits. 1992. Vol. 27, N 4. P. 473—482.
7. Holt G., Tyagi A. Minimizing Interconnect Energy through Integrated Low-Power Placement and Combinational Logic Synthesis // In ISPD, California USA. 1997. P. 48—53.
8. Yeap G. P. Practical Low Power Digital VLSI Design. — Kluwer Academic Publisher, 1998.
9. Turgis S., Azemard N., Auvergne D. Explicit evaluation of short-circuit power dissipation for CMOS logic structures // Proc. of the 1995 Intern. Symposium on Low Power Design. April 1995. P. 129—134.
10. Najm F. N. Transition density: A new measure of activity of digital circuits // IEEE Trans. on Computer-Aided Design of Integr. Circuit and Systems. 1993. Vol. 12, N 2. P. 310—323.
11. Pedram M. Power Minimization in IC Design: Principles and Applications // ACM Transactions Design Automation Electronic Systems. 1996. Vol. 1. P. 3—56.
12. Buch P., Lin S., Nagasamy V., Kuh E. S. Techniques for fast circuit simulation applied to power estimation of CMOS circuits // Proc. of the Intern. Symposium on Low Power Design. — Dana Point, CA. 1995. April 23—26. P. 135—138.
13. Najm F., Hajj I., Yang P. An extension of probabilistic simulation for reliability analysis of CMOS VLSI circuits // 28th ACM / IEEE Design Automation Conference, Tech. Dig. June 1991. P. 644—649.
14. Ghosh A., Devadas S., Keutzer K., White J. Estimation of average switching activity in combinational and sequential circuits // 29th ACM / IEEE Design Automation Conference, Tech. Dig. June 1992. P. 253—259.
15. Xakellis M., Najm F. Statistical estimation of the switching activity in digital circuits // 31st ACM/IEEE Design Automation Conf. — San Diego, CA, 1994. P. 728—733.
16. Benini L., Favalli M., Risco B. Analysis of hazard contribution to power dissipation in CMOS IC's // Proc. of the 1994 Intern. Workshop on Low Power Design. — April 1994. P. 27—32.
17. Najm F. N. A survey of Power Estimation Techniques in VLSI Circuits // IEEE Trans. on VLSI. 1994. N 12. P. 446—455.
18. Rajgopal S., Mehta G. Experiences with simulation-based schematic level current estimation // Proc. of the Intern. Workshop on Low Power Design. 1994. April. P. 9—14.
19. Najm F., Burch R., Yang P., Hajj I. Probabilistic simulation for reliability analysis of CMOS VLSI circuits // IEEE Trans. on Computer-Aided Design. 1990. Vol. 9. N 4. P. 439—450.
20. Akers A. Binary decision diagrams // IEEE Trans. on Computers. 1978. N 27. P. 509—516.



2010
Разработка ПО

Шестая восточно-европейская научно-практическая конференция по программной инженерии "Разработка ПО 2010" (CEE-SECR 2010).

11—15 октября 2010 г. в Москве состоится Шестая восточно-европейская научно-практическая конференция "Разработка ПО 2010" CEE-SECR 2010. Подробная информация на сайте <http://cee-secr.org>

Б. Г. Кухаренко, канд. физ.-мат. наук,
ст. науч. сотр., вед. науч. сотр.,
Институт машиноведения РАН, г. Москва,
e-mail: kukharenko@imash.ru

Байесовская фильтрация в технологии спектрального анализа на основе быстрого преобразования Прони

Описан предварительный анализ временных рядов посредством расширения их размерности за счет добавления разности ряда, сглаживания многомерного временного ряда по методу Рауха на основе адаптивного восстановления линейной модели временного ряда посредством алгоритма ожидания и максимизации правдоподобия и анализа независимых компонент сглаженного временного ряда. Предварительный анализ повышает точность и увеличивает частотное разрешение спектрального метода Прони.

Ключевые слова: многомерные временные ряды, фильтр Калмана, сглаживатель Рауха, линейные динамические системы, алгоритм ожидания и максимизации правдоподобия, анализ независимых компонент, спектральный анализ, метод Прони

Предварительный анализ временных рядов

Статья посвящена описанию предварительного анализа временных рядов в технологии спектрального анализа на основе метода Прони (Prony) [1]. В работе [2] алгоритм быстрого преобразования Прони используется для спектрального анализа неустойчивых коллективных колебаний (флаттера) лопаток ротора турбокомпрессора. Для этого используется сегментирование записи нестационарного колебания (нестационарного временного ряда). Пусть $y[\overline{1, M}]$, $M \ll N$, представляет собой некоторый сегмент полного временного ряда (записи колебания) $y[\overline{1, N}]$. Для краткости выбран шаг дискретизации времени $\Delta t = 1$. Спектральная декомпозиция Прони сегмента $y[\overline{1, M}]$ имеет вид

$$y[k] = \sum_{l=1}^p r[l]z[l]^{k-1} + n[k], \quad k = \overline{1, M}. \quad (1)$$

Здесь индекс k представляет дискретное время; p — число полюсов сегмента; $z[l] = \exp(\delta[l] + j2\pi f[l])$, $l = \overline{1, p}$, — полюса сегмента, где $\delta[l]$ и $f[l]$ — соответственно фактор демпфирования (логарифмический декремент) и частота; $r[l] = A[l]\exp(j\phi[l])$, $l = \overline{1, p}$, — вычеты в полюсах, где $A[l]$ и $\phi[l]$ — соответственно амплитуда и фаза; $n[k]$ — аддитивный шум.

После определения полюсов $z[l]$, $l = \overline{1, p}$, вычеты $r[l]$, $l = \overline{1, p}$, в полюсах определяются из (1) по методу наименьших квадратов. Оценка зависящих от времени спектров факторов демпфирования и частот и соответствующих им спектров амплитуд и фаз для полного временного ряда (записи колебания) $y[\overline{1, N}]$ является результатом последовательных сдвигов временного окна фиксированной длины $M\Delta t$.

Частота нелинейных колебаний распределенной системы зависит от времени. Однако локально во времени нелинейный характер колебаний распределенной системы проявляется появлением в спектре частот гармонических серий от одной или нескольких фундаментальных частот [3]. На практике гармонические серии от различных фундаментальных частот перекрываются, поэтому гармоники конкретной фундаментальной частоты невозможно выделить линейной фильтрацией. В результате трудно оценить степень нелинейности колебаний на конкретной фундаментальной частоте распределенной системы.

Анализ независимых компонент (Independent Component Analysis — ICA) способен разделить сигнал, представленный временными рядами, на смеси заданного числа компонент с перекрывающимися спектрами частот при условии различных статистических характеристик этих компонент, но требует для этого такого же или большего числа наблюдаемых временных рядов. В настоящей статье необходимый набор наблюдаемых временных рядов получается из единственного временного ряда в три шага. Сначала дополнительный компонент получается как разность исходного временного ряда

$$\Delta y = y[\overline{2, N}] - y[\overline{1, (N-1)}]. \quad (2)$$

Численное дифференцирование временного ряда $y[\overline{1, N}]$ по формуле (2) делает его амплитудно-частотную характеристику относительно более высокочастотной. После этого используется бай-

есовская фильтрация двухмерного наблюдаемого временного ряда

$$\mathbf{Y} = \{\mathbf{y}; \Delta\mathbf{y}\}^T, \quad (3)$$

где $\Delta\mathbf{y}$ определяется формулой (2); τ — символ транспонирования.

В принципе, для описания колебаний с возбуждением конечного числа мод распределенной системы могут использоваться модели с сосредоточенными параметрами. Например, для исследованных в работе [2] колебаний лопадки ротора компрессора такой моделью является сосредоточенная изгибно-крутильная модель [4]. В канонической форме она представляет собой систему четырех линейных дифференциальных уравнений первого порядка для четырехмерного вектора состояний. Такая четырехмерная сосредоточенная модель после ее дискретизации во времени может быть использована в качестве линейной динамической модели для байесовской фильтрации на основе линейного фильтра Калмана или сглаживателя Рауха, которые оценивают временные ряды для четырех скрытых переменных на основе двух временных рядов наблюдаемых переменных [5, 6]. Однако если не определены спектральные параметры временных рядов наблюдаемых переменных, параметры этой линейной динамической модели неизвестны. Поэтому в настоящей работе на основе двухмерного наблюдаемого временного ряда \mathbf{Y} (3) посредством алгоритма ожидания и максимизации правдоподобия адаптивно проводится оценка параметров линейной динамической модели и скрытого временного ряда состояний \mathbf{X} размерности $p \times N$, $p = 4$. Статистически независимые компоненты для скрытого многомерного временного ряда состояний \mathbf{X} анализируются на основе алгоритма анализа независимых компонент. Независимые компоненты представляют различные нелинейные моды распределенной системы, и их спектры частот являются гармоническими сериями от различных доминантных фундаментальных частот.

Байесовская фильтрация и оценка параметров линейной динамической системы

Пусть $\mathbf{X} = \mathbf{X}[\bar{1}, q; \bar{1}, N]$ — многомерный временной ряд для вектора состояния размерности $p \times N$, а $\mathbf{Y} = \mathbf{Y}[\bar{1}, p; \bar{1}, N]$ — многомерный наблюдаемый временной ряд измерений этого вектора состояний размерности $q \times N$, причем в общем случае $p \neq q$ (например, некоторые компоненты вектора состояния \mathbf{X} не измеряются). Как шум состояния (процесса) $\mathbf{W} = \mathbf{W}[\bar{1}, p; \bar{1}, N]$, так и шум выхода $\mathbf{V} = \mathbf{V}[\bar{1}, q; \bar{1}, N]$ — многомерные времен-

ные ряды для случайных переменных с нулевым средним, распределенных по нормальному закону с ковариационными матрицами \mathbf{Q} и \mathbf{R} соответственно. Линейные инвариантные относительно времени динамические системы, называемые также линейными Гауссовыми моделями в пространстве состояний, описываются двумя уравнениями для вектор-столбцов $\mathbf{x}[k] = \mathbf{X}[\bar{1}, p; k]$, $\mathbf{y}[k] = \mathbf{Y}[\bar{1}, q; k]$, $\mathbf{W}[k] = \mathbf{W}[\bar{1}, p; k]$ и $\mathbf{v}[k] = \mathbf{V}[\bar{1}, q; k]$ (индекс k представляет дискретное время):

$$\mathbf{x}[k + 1] = \mathbf{F} \cdot \mathbf{x}[k] + \mathbf{w}[k]; \quad (4)$$

$$\mathbf{y}[k] = \mathbf{G} \cdot \mathbf{x}[k] + \mathbf{v}[k], \quad (5)$$

где \mathbf{F} — матрица переходов; \mathbf{G} — матрица измерений.

В теории линейной фильтрации временной ряд \mathbf{Y} для вектора измерений рассматривается как зашумленный детерминированный временной ряд \mathbf{X} для вектора состояния. В теории байесовской фильтрации отличающиеся одним временным шагом вектор-столбцы (значения переменных состояния в последовательные дискретные моменты времени) на основе (4) объединяются в Гауссову случайную переменную со статистическими характеристиками шума состояния [5]. Аналогичная комбинация формируется на основе (5) для одно-временных вектор-столбцов переменной состояния системы и переменной измерения [5]. В результате условные плотности вероятностей для векторов наблюдения и состояния линейной динамической системы имеют вид

$$P(\mathbf{y}[k]|\mathbf{x}[k]) = (2p)^{-q/2} \|\mathbf{R}\|^{-1/2} \exp\left(-\frac{1}{2} (\mathbf{y}[k] - \mathbf{G} \cdot \mathbf{x}[k])^T \mathbf{R}^{-1} (\mathbf{y}[k] - \mathbf{G} \cdot \mathbf{x}[k])\right); \quad (6)$$

$$P(\mathbf{x}[k]|\mathbf{x}[k-1]) = (2p)^{-p/2} \|\mathbf{Q}\|^{-1/2} \exp\left(-\frac{1}{2} (\mathbf{x}[k] - \mathbf{F} \cdot \mathbf{x}[k-1])^T \mathbf{Q}^{-1} (\mathbf{x}[k] - \mathbf{F} \cdot \mathbf{x}[k-1])\right). \quad (7)$$

Исходная плотность состояний в момент времени $t = 1$ также предполагается Гауссовой со средним $\pi[1]$ и вариацией $\mathbf{V}[1]$:

$$P(\mathbf{x}[1]) = (2p)^{-p/2} \|\mathbf{V}[1]\|^{-1/2} \exp\left(-\frac{1}{2} (\mathbf{x}[1] - \pi[1])^T \mathbf{V}[1]^{-1} (\mathbf{x}[1] - \pi[1])\right). \quad (8)$$

В (4) состояние системы (линейно) зависит только от предыдущего состояния, отстоящего на

один временной шаг. Отсюда следует свойство марковости условных вероятностей:

$$P(\mathbf{X}, \mathbf{Y}) = P(\mathbf{x}[1]) \prod_{k=2}^N P(\mathbf{x}[k]|\mathbf{x}[k-1]) \prod_{k=1}^N P(\mathbf{y}[k]|\mathbf{x}[k]). \quad (9)$$

Из (6)—(8) следует, что логарифм $P(\mathbf{X}, \mathbf{Y})$ (9) является квадратичной формой (постоянный член опущен):

$$\begin{aligned} \log(P(\mathbf{X}, \mathbf{Y})) = & -\frac{1}{2} \sum_{k=1}^N (\mathbf{y}[k] - \mathbf{G} \cdot \mathbf{x}[k])^T \mathbf{R}^{-1} \cdot (\mathbf{y}[k] - \\ & - \mathbf{G} \cdot \mathbf{x}[k]) - \frac{N}{2} \log(\|\mathbf{R}\|) - \frac{1}{2} \sum_{k=2}^N (\mathbf{x}[k] - \\ & - \mathbf{F} \cdot \mathbf{x}[k-1])^T \cdot \mathbf{Q}^{-1} \cdot (\mathbf{x}[k] - \mathbf{F} \cdot \mathbf{x}[k-1]) - \\ & - \frac{N-1}{2} \log(\|\mathbf{Q}\|) - \frac{1}{2} (\mathbf{x}[1] - \boldsymbol{\pi}[1])^T \cdot \mathbf{V}[1]^{-1} \cdot (\mathbf{x}[1] - \\ & - \boldsymbol{\pi}[1]) - \frac{1}{2} \log(\|\mathbf{V}[1]\|). \quad (10) \end{aligned}$$

EM-алгоритм. В работах [7—9] описано использование алгоритма ожидания и максимизации правдоподобия (EM-алгоритма) для оценки параметров линейной динамической системы (4)—(5) на основе многомерного наблюдаемого временного ряда \mathbf{Y} . E-шаг описываемого EM-алгоритма состоит в вычислении условного среднего (логарифма ожидаемого правдоподобия):

$$LL = E[\log(P(\mathbf{X}, \mathbf{Y}))]. \quad (11)$$

Из формулы (10) следует, что LL (11) зависит от трех условных средних (ожиданий), для которых используются обозначения

$$\langle \mathbf{x}[k] \rangle = E[\mathbf{x}[k]|\mathbf{y}]; \quad (12)$$

$$\mathbf{P}[k] = E[\mathbf{x}[k] \cdot \mathbf{x}[k]^T|\mathbf{y}]; \quad (13)$$

$$\mathbf{P}[k; k-1] = E[\mathbf{x}[k] \cdot \mathbf{x}[k-1]^T|\mathbf{y}]. \quad (14)$$

Оценка состояния $\langle \mathbf{x}[k] \rangle$ в момент времени k зависит от прошлых $\mathbf{y}[1, k]$ и будущих $\mathbf{y}[k+1, N]$ наблюдений [6]. Поэтому она отличается от оценки, вычисленной посредством фильтра Калмана, который оценивает условное среднее только по отношению к прошлым наблюдениям [5]. Прежде чем показать вычисление условных средних (ожиданий) (12)—(14) на E-шаге EM-алгоритма, опишем M-шаг, т. е. оценку параметров линейной динамической системы (4)—(5).

M-шаг. Параметры линейной динамической системы (4) и (5) — это матрицы \mathbf{F} и \mathbf{G} , а также статистические характеристики \mathbf{R} (6), \mathbf{Q} (7), $\boldsymbol{\pi}[1]$ и $\mathbf{V}[1]$ (8) Гауссовых переменных. На основе $\langle \mathbf{x}[1, N] \rangle$ и $\mathbf{y}[1, N]$ каждый из параметров линейной динамической системы оценивается из равенства нулю соответствующей частной произ-

водной логарифма ожидаемого правдоподобия LL (11). В результате получаем:

$$\mathbf{G} = \left(\sum_{k=1}^N \mathbf{y}[k] \cdot \langle \mathbf{x}[k] \rangle^T \right) \left(\sum_{k=1}^N \mathbf{P}[k] \right)^{-1}; \quad (15)$$

$$\mathbf{R} = \frac{1}{N} \sum_{k=1}^N \left(\mathbf{y}[k] \cdot \mathbf{y}[k]^T - \mathbf{G} \cdot \langle \mathbf{x}[k] \rangle \cdot \mathbf{y}[k]^T \right); \quad (16)$$

$$\mathbf{F} = \left(\sum_{k=1}^N \mathbf{P}[k; k-1] \right) \left(\sum_{k=2}^N \mathbf{P}[k-1] \right)^{-1}; \quad (17)$$

$$\mathbf{Q} = \frac{1}{N-1} \left(\sum_{k=2}^N \mathbf{P}[k] - \mathbf{F} \cdot \sum_{k=2}^N \mathbf{P}[k; k-1] \right); \quad (18)$$

$$\boldsymbol{\pi}[1] = \langle \mathbf{x}[1] \rangle; \quad (19)$$

$$\begin{aligned} \mathbf{V}[1] = E[(\mathbf{x}[1] - \langle \mathbf{x}[1] \rangle)(\mathbf{x}[1] - \langle \mathbf{x}[1] \rangle)^T|\mathbf{y}] = \\ = \mathbf{P}[1] - \langle \mathbf{x}[1] \rangle \langle \mathbf{x}[1] \rangle^T. \quad (20) \end{aligned}$$

E-шаг. Используются следующие обозначения: $\langle \mathbf{x}[k; l] \rangle = E[\mathbf{x}[k]|\mathbf{y}[1, l]]$ (таким образом, для условного среднего (12) имеем $\langle \mathbf{x}[k] \rangle \equiv \langle \mathbf{x}[k; N] \rangle$) и $\mathbf{V}[k; l] = E[(\mathbf{x}[k] - \langle \mathbf{x}[k] \rangle)(\mathbf{x}[k] - \langle \mathbf{x}[k] \rangle)^T|\mathbf{y}[1, l]]$. Сначала приводятся рекурсии вперед для линейного фильтра Калмана:

$$\langle \mathbf{x}[k; k-1] \rangle = \mathbf{F} \cdot \langle \mathbf{x}[k-1; k-1] \rangle; \quad (21)$$

$$\mathbf{V}[k; k-1] = \mathbf{F} \cdot \mathbf{V}[k-1; k-1] \cdot \mathbf{F}^T + \mathbf{Q}, \quad (22)$$

вычисление матрицы усиления Калмана

$$\mathbf{K}[k] =$$

$$= \mathbf{V}[k; k-1] \cdot \mathbf{G}^T (\mathbf{G} \cdot \mathbf{V}[k; k-1] \cdot \mathbf{G}^T + \mathbf{R})^{-1}; \quad (23)$$

$$\begin{aligned} \langle \mathbf{x}[k; k] \rangle = \langle \mathbf{x}[k; k-1] \rangle + \\ + \mathbf{K}[k] \cdot (\langle \mathbf{y}[k] \rangle - \mathbf{G} \cdot \langle \mathbf{x}[k; k-1] \rangle); \quad (24) \end{aligned}$$

$$\mathbf{V}[k; k] = \mathbf{V}[k; k-1] - \mathbf{K}[k] \cdot \mathbf{G} \cdot \mathbf{V}[k; k-1], \quad (25)$$

где $\langle \mathbf{x}[1] \rangle = \boldsymbol{\pi}[1]$ из (19) и $\mathbf{V}[1]$ (20). Вывод рекурсий (21)—(25) основан на байесовском подходе [10].

Следуя [7—9], для вычисления $\langle \mathbf{x}[k] \rangle = \langle \mathbf{x}[k; N] \rangle$ и условного среднего (13) $\mathbf{P}[k] \equiv \mathbf{V}[k; N] + \langle \mathbf{x}[k; N] \rangle \langle \mathbf{x}[k; N] \rangle^T$ выполняется рекурсия назад ($\mathbf{J}[k]$ — вспомогательная матрица):

$$\mathbf{J}[k-1] = \mathbf{V}[k-1; k-1] \mathbf{F}^T (\mathbf{V}[k; k-1])^{-1}; \quad (26)$$

$$\begin{aligned} \langle \mathbf{x}[k-1; N] \rangle = \langle \mathbf{x}[k-1; k-1] \rangle + \\ + \mathbf{J}[k-1] (\langle \mathbf{x}[k; N] \rangle - \mathbf{F} \cdot \langle \mathbf{x}[k-1; k-1] \rangle); \quad (27) \end{aligned}$$

$$\begin{aligned} \mathbf{V}[k-1; N] = \mathbf{V}[k-1; k-1] + \mathbf{J}[k-1] (\mathbf{V}[k; N] - \\ - \mathbf{V}[k; k-1]) \mathbf{J}[k-1]^T. \quad (28) \end{aligned}$$

Условное среднее (14) имеет вид

$$\begin{aligned} \mathbf{P}[k; k-1] \equiv \mathbf{V}[k; k-1; N] + \\ + \langle \mathbf{x}[k; N] \rangle \langle \mathbf{x}[k-1; N] \rangle^T, \end{aligned}$$

где $\mathbf{V}[k; k-1; N] = E[(\mathbf{x}[k] - \langle \mathbf{x}[k] \rangle)(\mathbf{x}[k-1] - \langle \mathbf{x}[k-1] \rangle)^T|\mathbf{y}[1, N]]$.

Значения $\mathbf{V}[k; k-1; N]$ также вычисляются рекурсивно назад:

$$\mathbf{V}[k-1; k-2; N] = \mathbf{V}[k-1; k-1] \cdot \mathbf{J}[k-2]^T + \mathbf{J}[k-1] \cdot (\mathbf{V}[k; k-1; N] - \mathbf{F} \cdot \mathbf{V}[k-1; k-1]) \cdot \mathbf{J}[k-2], \quad (29)$$

которая инициализируется как

$$\mathbf{V}[N; N-1; N] = (\mathbf{I} - \mathbf{K}[N] \cdot \mathbf{G}) \cdot \mathbf{F} \cdot \mathbf{V}[N-1; N-1].$$

Последовательные итерации из М- и Е-шагов с оценкой изменения логарифма ожидаемого правдоподобия LL (11) используются для получения оценки $\langle \mathbf{x}[\overline{1}, N] \rangle$ многомерного временного ряда состояния \mathbf{X} размерности $p \times N$.

Анализ независимых компонент многомерного временного ряда состояния

В анализе независимых компонент многомерный временной ряд состояния $\mathbf{X} = \mathbf{X}[\overline{1}, q; \overline{1}, N]$ рассматривается как реализация случайного вектора размерности $q \times N$ с нулевым средним (для краткости выбран шаг дискретизации времени $\Delta t = 1$). Задача анализа независимых компонент состоит в оценке (глобальной во времени) линейной модели:

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S}, \quad (30)$$

где $\mathbf{S} = \mathbf{S}[\overline{1}, r; \overline{1}, N]$ — реализация многомерного (случайного) вектора размерности $r \times N$, вектор-строки которого $\mathbf{s}[i] = \mathbf{S}[i; \overline{1}, N]$, $i = \overline{1}, r$, предполагаются статистически взаимно независимыми; \mathbf{A} — постоянная матрица размерности $q \times r$ [11].

В (30) нельзя определить энергию независимых компонент $\mathbf{s}[i]$, $i = \overline{1}, r$, поскольку как независимые компоненты $\mathbf{s}[i]$, $i = \overline{1}, r$, так и матрица смешивания \mathbf{A} неизвестны (любой множитель при \mathbf{S} может быть компенсирован делением матрицы \mathbf{A} на этот множитель). Для независимых компонент с простыми и непересекающимися спектрами частот следует использовать алгоритм SOBI (Second-Order Blind Identification) [12]. Этот алгоритм основан на диагонализации ковариационных матриц с запаздыванием, т. е. на статистике второго порядка, и отражает различие в амплитудно-частотных характеристиках независимых компонент. Общая формулировка анализа независимых компонент основана на статистике четвертого порядка (экссесс — *kurtosis*) или понятии взаимной информации и отражает различие в фазово-частотных характеристиках независимых компонент [13].

Байесовская фильтрация для спектрального анализа записи колебаний лопатки ротора

Спектр колебаний лопатки ротора, исследованных в [2], содержит частоту оборотов ротора f_R , которая меняется от 56 до 60 Гц (рис. 1), и ее гармоники порядков 2 и 3, а также частоту крутиль-

ной моды ≈ 198 Гц и ее гармоники порядков 2 и 3. Это затрудняет выделение гармоник частоты изгибной моды ≈ 81 Гц посредством линейной фильтрации записи колебаний лопатки (рис. 2, а) для оценки степени нелинейности этой изгибной моды при изменении частоты оборотов.

При $t = 15 \dots 23,5$ с запись колебаний лопатки (рис. 2, а) демонстрирует пульсации на частоте 1 Гц, связанные с десинхронизацией крутильных частот лопаток ротора. Эти пульсации на частоте 1 Гц имеют значительную амплитуду и препятствуют точному определению фактора демпфирования (логарифмического декремента) на частоте изгибной моды и ее гармониках как на основе исходной записи колебаний лопатки (рис. 2, а), так и на основе результатов ее линейной фильтрации. Раз-

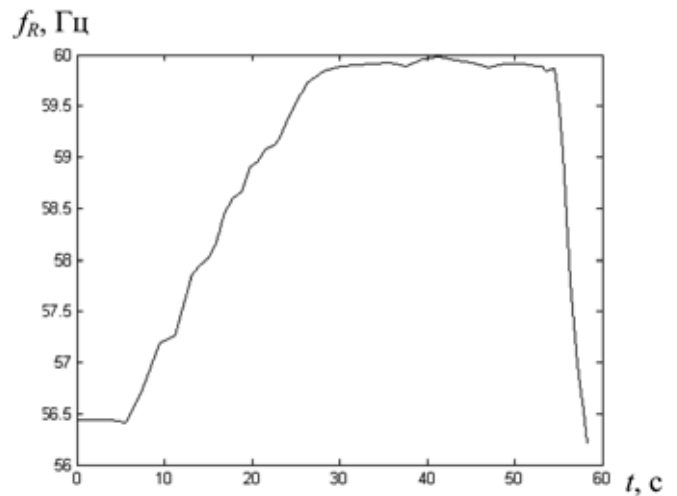


Рис. 1. Изменение частоты оборотов ротора компрессора

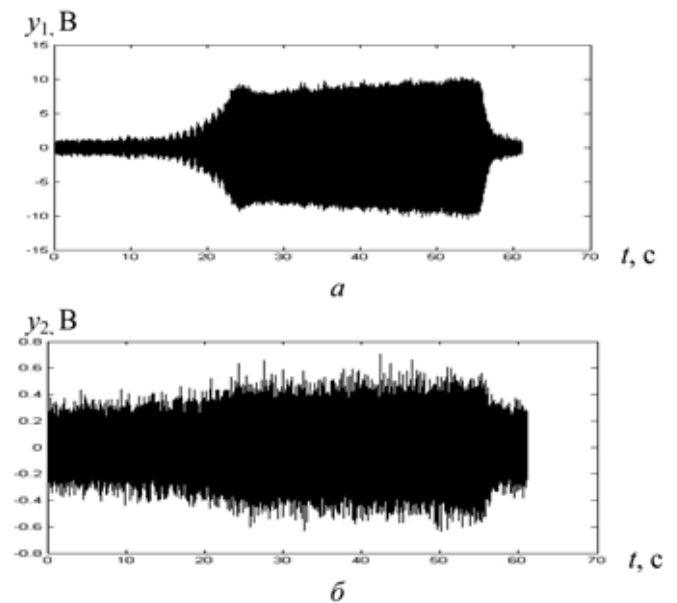


Рис. 2. Наблюдаемые временные ряды Y : а — запись колебания; б — ее разность

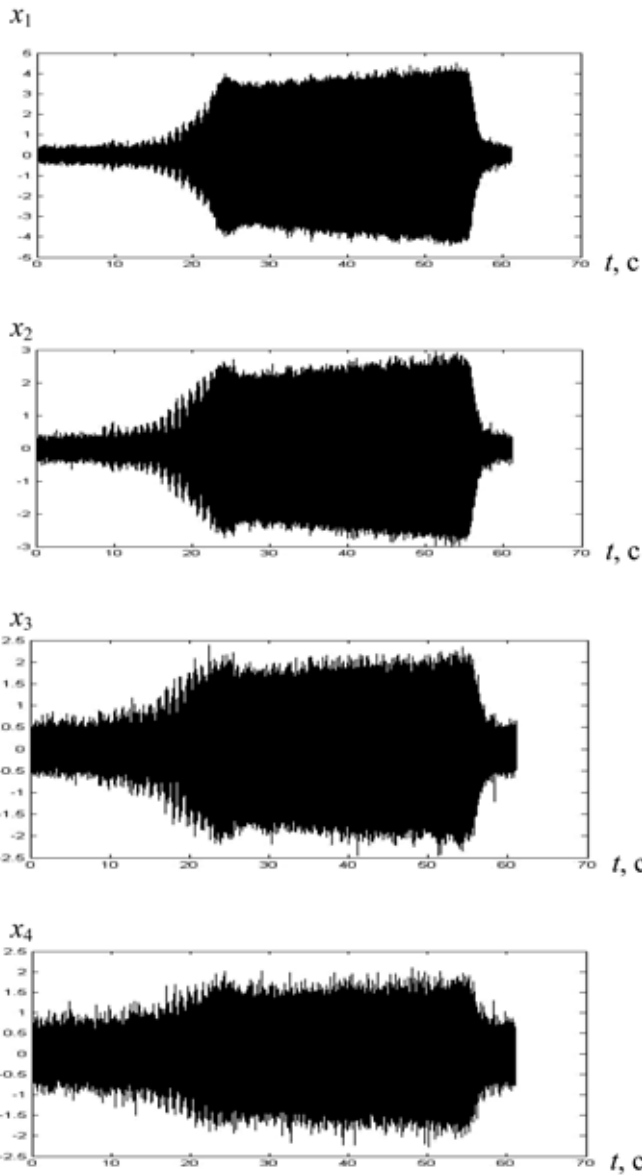


Рис. 3. Временные ряды состояния X для наблюдаемых временных рядов Y (рис. 2)

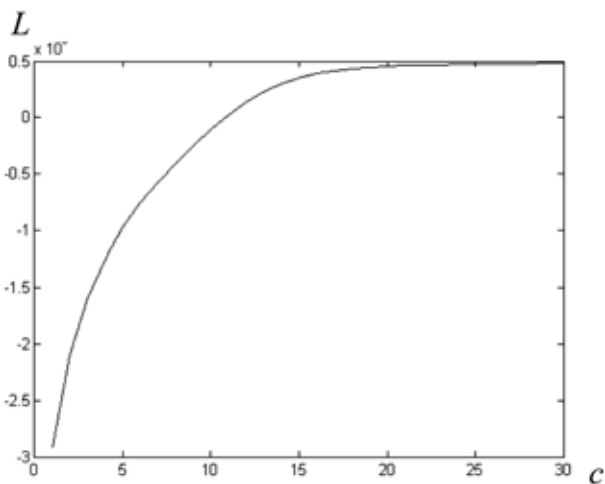


Рис. 4. Зависимость логарифма правдоподобия LL (11) от числа итераций c

ность (рис. 2, б) "обогащена" более высокими частотами, в том числе высокими гармониками изгибной частоты, высокими гармониками частоты оборотов и собственной частотой и гармониками крутильной моды. На основе двухмерного наблюдаемого временного ряда Y (рис. 2) по алгоритму ожидания и максимизации правдоподобия адаптивно проводится оценка параметров четырехмерной линейной динамической модели и четырехмерного временного ряда состояния X (рис. 3).

Отметим, что при $t = 15...23,5$ с записи временных рядов состояния X на рис. 3 демонстрируют пульсации на частоте 1 Гц, связанные (как и в записи исходного колебания на рис. 2, а) с десин-

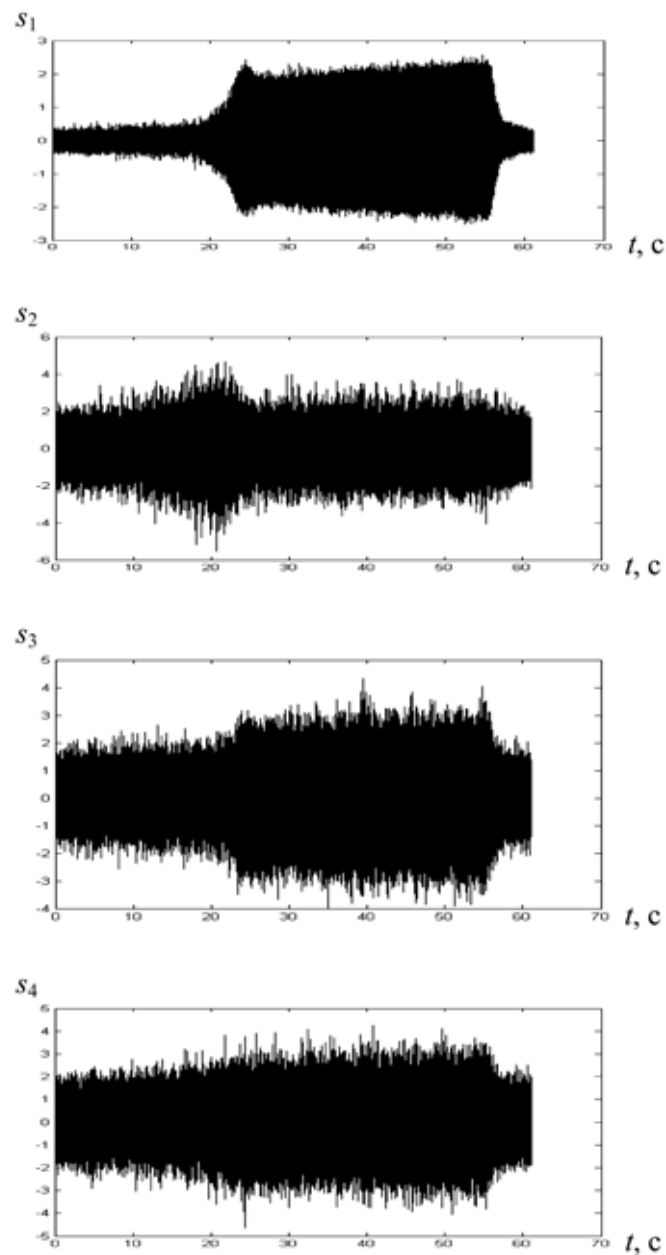


Рис. 5. Независимые компоненты S для временных рядов состояния X (рис. 3)

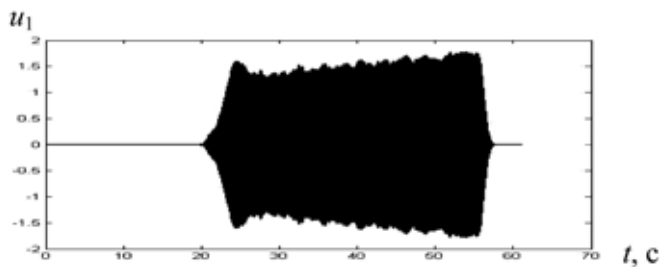


Рис. 6. Результат фильтрации независимой компоненты s_1 (рис. 5) на частоте 81 Гц, относительная полоса 0,05

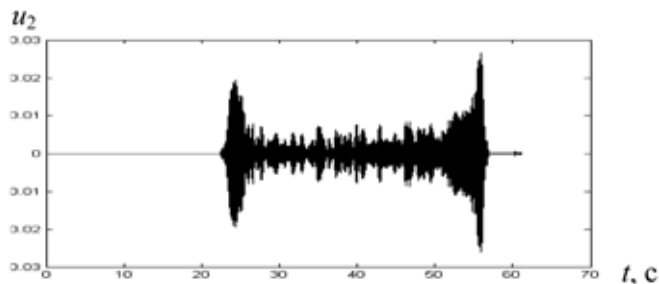


Рис. 7. Результат фильтрации независимой компоненты s_1 (рис. 5) на частоте 162 Гц, относительная полоса 0,01

хронизацией крутильных частот лопаток. Зависимость логарифма правдоподобия LL (11) от числа итераций s алгоритма ожидания и максимизации правдоподобия приведена на рис. 4.

Ниже приведены адаптивно определенные матрицы линейной динамической системы (4) и (5): динамическая матрица (матрица переходов)

$$\mathbf{F} = \begin{bmatrix} 0,8697 & -0,3147 & 0,1177 & -0,0480 \\ 0,1481 & 0,8947 & -0,1799 & 0,0653 \\ -0,1475 & -0,6101 & 0,1103 & 0,0090 \\ 0,1611 & 0,3572 & -0,1192 & 0,0209 \end{bmatrix}$$

и матрица измерений

$$\mathbf{G} = \begin{bmatrix} -2,1235 & 0,5094 & -0,1969 & 0,1080 \\ 0,0283 & 0,0556 & -0,0301 & 0,0418 \end{bmatrix}$$

Независимые компоненты \mathbf{S} для временных рядов состояния \mathbf{X} приведены на рис. 5.

Матрица смешивания (30) имеет вид

$$\mathbf{A} = \begin{bmatrix} 0,5843 & 0 & 0,6314 & 0,7608 \\ 0,6000 & 0,8392 & 1,0000 & 0,4118 \\ 0,8235 & 0,6157 & 0,5059 & 0,8863 \\ 0,6000 & 0,6824 & 0,7608 & 0,4902 \end{bmatrix}$$

Спектральный анализ по методу Прони независимых компонентов s_2 и s_3 (рис. 5) показывает, что они являются относительно высокочастотными (интервал времени $t = 10...23,5$ с возбуждения независимой компоненты s_2 является интервалом

интенсивного возбуждения крутильной моды [2]). Это позволяет квалифицировать временные ряды состояния \mathbf{X} (см. рис. 3) с точки зрения их высокочастотности. Важно, что при $t = 15...23,5$ с временной ряд независимой компоненты s_1 (рис. 5) не демонстрирует пульсации на частоте 1 Гц (которые характерны для временных рядов на рис. 2 и 3). Это гарантирует точное определение фактора демпфирования на частоте изгибной моды и ее гармониках во временном интервале $t = 15...23,5$ с их первоначального возбуждения (рис. 6, 7).

С учетом зависимости частоты оборотов f_R от времени (см. рис. 1) на рис. 8 и 9 приведены зависимости факторов демпфирования (логарифмического декремента) на изгибной частоте и ее второй гармонике от f_R соответственно.

При убывании частоты оборотов ротора f_R фактор демпфирования δ_1 на изгибной частоте 81 Гц отрицательный (рис. 8), а фактор демпфирования δ_2

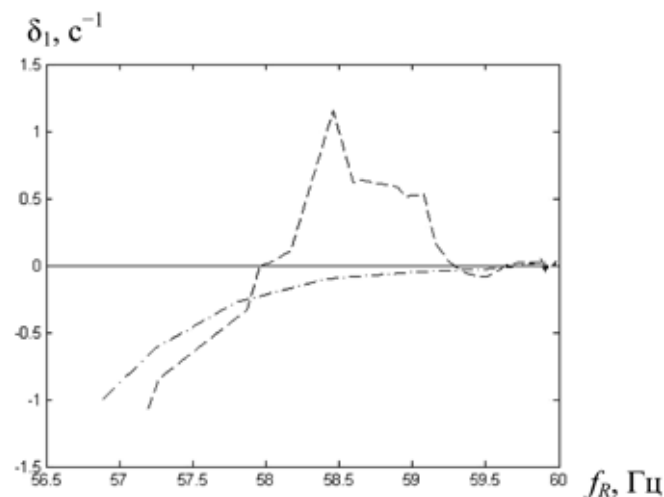


Рис. 8. Зависимость фактора демпфирования на частоте 81 Гц от частоты оборотов ротора

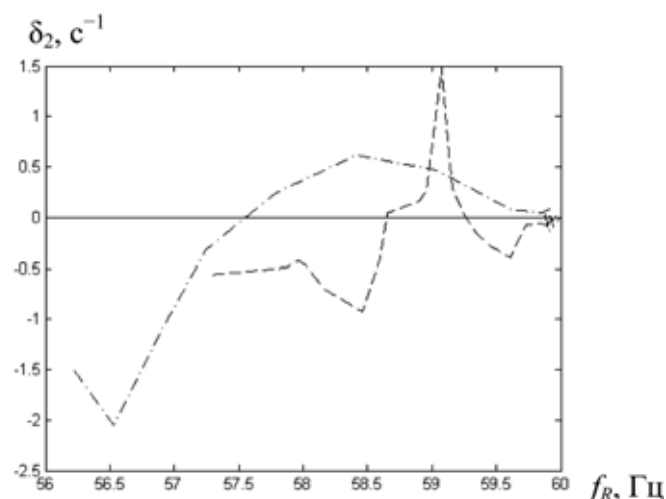


Рис. 9. Зависимость фактора демпфирования на частоте 162 Гц от частоты оборотов ротора

на второй гармонике изгибной частоты (162 Гц) — положительный (рис. 9). Причем зависимость фактора демпфирования δ_2 от частоты оборотов f_R при ее убывании аппроксимирует в среднем зависимость δ_2 при росте f_R . Поэтому фактор демпфирования δ_2 определяется в основном его аэроупругой составляющей [5].

Список литературы

1. Weiss L., McDonogh R. N. Prony's method, Z-transform, and Pade approximation // SIAM Review. 1963. V. 9. N 2. P. 145—149.
2. Кухаренко Б. Г. Технология спектрального анализа на основе быстрого преобразования Фурье // Информационные технологии. 2008. № 4. С. 38—42.
3. Фролов К. В., Балакшин О. Б., Кухаренко Б. Г., Минаев А. Я. Спектральный критерий и оценка нелинейности колебаний систем // Проблемы машиностроения и надежности машин. 2001. № 6. С. 3—7.
4. Bendiksen O. O., Friedmann P. P. The effect of bending-torsion coupling on fan and compressor blade flutter // Transactions of the ASME. 1982. V. 104. N 3. P. 617—623.

5. Simon D. Optimal State Estimation: Kalman, H_∞ and Non-linear Approaches. Hoboken, New Jersey: John Wiley & Sons, Inc. 2006.
6. Rauch H. E. Solutions to the linear smoothing problem // IEEE Transactions on Automatic Control. 1963. V. 8. P. 371—372.
7. Shumway R. H., Stoffer D. S. An approach to time series smoothing and forecasting using the EM algorithm // Journal of Time Series Analysis. 1982. V. 3. N 4. P. 253—264.
8. Shumway R., Stoffer D. Dynamic linear models with switching // Journal of the American Statistical Association. 1992. V. 86. P. 763—769.
9. Roweis S., Ghahramani Z. A unifying review of linear Gaussian models // Neural Computation. 1999. V. 11. N 2. P. 305—345.
10. Soderstrom T. Discrete-time Stochastic Systems. Prentice Hall International. 1994.
11. Cichocki A., Amari S. Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. John Wiley & Sons. 2002.
12. Belouchrani A., Abed-Meraim K., Cardoso J. F., Molines E. A blind source separation technique using second-order statistics // IEEE Transactions on Signal Processing. 1997. V. 45. P. 434—444.
13. Hyvarinen A. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis // IEEE Transactions on Neural Networks. 1999. V. 10. N 3. P. 626—634.

УДК 621.361.246

С. Н. Агиевич, канд. техн. наук, вед. науч. сотр.,
ФГНУ "ГНТЦ Наука"
e-mail: practicsdv@yandex.ru

Применение локальных сплайнов для оценивания параметров распределений

Предлагается подход к оцениванию частотных параметров сигналов на основе обработки их частотно-временных распределений плотности энергии с применением методов сплайн-гармонического анализа. Разработаны аналитические алгоритмы оценивания, базирующиеся на применении локальных сплайнов. Приводятся практические выражения для вычисления сплайн-функций и их производных, а также результаты практических экспериментов.

Ключевые слова: сплайн-анализ, вычисление сплайн-функций, частотные параметры сигналов, алгоритм оценивания параметров, обработка частотно-временных распределений

Введение

Применение совместных функций плотности распределения энергии (РЭ) сигналов в пространстве частота — время открывает новые возможности по оцениванию их параметров в условиях высокого уровня шумов и ограниченной информации о статистических характеристиках канала. Основу такого подхода составляет свойство одно-

компонентного сигнала (описываемого гладкими функциями огибающей и полной фазы) по локализации энергии в континууме точек его мгновенной частоты [1]. Использование данного свойства позволяет оценивать частотные параметры сигналов путем непосредственной идентификации соответствующих спектральных компонент в заданный момент времени в выбранном срезе формируемого РЭ.

К сожалению, предлагаемые методы по реализации указанного подхода [2—5] в большинстве случаев сложно реализуемы, что затрудняет их широкое практическое применение. Поэтому целью данной статьи является разработка теоретических основ оценивания частотных параметров сигналов с использованием элементов теории сплайн-гармонического анализа в интересах синтеза быстродействующих алгоритмов.

Анализ проблемы и определение путей ее решения

В [2] обосновано, что среди билинейных распределений наилучшую локализацию спектральной энергии сигнала обеспечивает функция Вигнера

$$\rho_w(k, n) = \sum_{m=-N+1}^{N-1} \exp(-j2\pi km/M) z_a(n+m) z_a^*(n-m), \quad (1)$$

где N — число дискретных отсчетов сигнала на интервале его наблюдения; $M = 2N - 1$ — значение, определяемое необходимостью реализации

операций удвоенной передискретизации [1]; $\{\exp(j2\pi kn/N)\}$ — базис дискретных экспоненциальных функций Фурье; j — знак мнимой части комплексного числа; z_a^* , z_a^* — аналитическая форма представления сигнала и его комплексно сопряженная по Гильберту копия соответственно; k — текущая переменная частоты; а $m, n = 0, 1, \dots, N-1$ — текущие переменные времени.

Форма представления (1) удобна с точки зрения ее реализации процедурами цифровой обработки сигналов (ЦОС), поскольку для ее синтеза могут быть использованы классические процедуры быстрого преобразования Фурье (БПФ). Между тем непосредственное применение распределения Вигнера связано с рядом ограничений, обусловленных возникновением так называемого интерференционного фона при обработке на его основе многокомпонентных сигналов [3].

В целях борьбы с указанным явлением в [3–5] предлагается ограничивать область непосредственной обработки сигналов применением дополнительной функции-окна. В частности, в [5] используется обобщенное РЭ "экспоненциального" типа:

$$\rho(k, n) = \sum_{l=-N+1}^{N-1} \exp(-j2\pi kl/M) \sum_{m=-N+1}^{N-1} W_G^\alpha(m-n, l) z_a(m+n+l) z_a^*(m+n-l), \quad (2)$$

где $G > 0$ — параметр подавления интерференционного фона; $0 \leq \alpha \leq 1/2$ — показатель, определяющий тип РЭ (от функции Вигнера при $\alpha = 0$ и $G \rightarrow \infty$ до функции Рихачека при $\alpha = 1/2$ и $G \rightarrow \infty$); l — переменная времени при обработке в пределах функции-окна $W_G^\alpha(*)$, которая в [5] определена как

$$W_G^\alpha(m, n) = \frac{\sqrt{G}}{2\sqrt{\pi}|l|} \left\{ \exp\left[-\frac{(m+\alpha l)^2}{4l^2/G}\right] + \exp\left[-\frac{(m-\alpha l)^2}{4l^2/G}\right] \right\}. \quad (3)$$

Экспериментальные исследования свойств (2), проведенные в работах [6–8], показали целесообразность выбора значения параметра G в диапазоне от 0,1 до 10 в зависимости от вида сигнала.

Таким образом, в рамках рассматриваемой проблемы задача оценивания параметров сигналов сводится к уточнению местоположения истинных спектральных компонент вдоль частотной оси на временном срезе обрабатываемого РЭ, т. е. поиску экстремума путем подбора дробных коэффициентов k' вместо их целого значения [6].

Как правило, на практике подобный процесс реализуется путем итеративных процедур градиентного поиска, алгоритм которых для функции

одной переменной можно представить следующим образом:

$$k'_{\langle r+1 \rangle} = k'_{\langle r \rangle} - \gamma \nabla_{\langle r \rangle},$$

где r — номер шага (итерации); γ — коэффициент, определяющий шаг коррекции значения текущего параметра k' ; $\nabla_{\langle r \rangle}$ — градиент при $k' = k'_{\langle r \rangle}$, причем

$$\nabla_{\langle r \rangle} = \left. \frac{\partial \rho}{\partial k'} \right|_{k' = k'_{\langle r \rangle}}. \quad (4)$$

В формуле (4) $\rho(k', n)$ — исходная функция, описывающая частотный срез РЭ для фиксированного момента времени n . Тогда при переходе к дробному значению k' центральный срез для распределения Вигнера (1) в терминах выражения (2) при $n = 0$ будет иметь вид

$$\rho_W(k', 0) = 2 \sum_{l=-N+1}^{N-1} \exp(-j2\pi k'l/M) z_a(l) z_a^*(-l). \quad (5)$$

Здесь $k' = k + \Delta k$; Δk — дискретный интервал.

Рассмотрим задачу введения дискретной переменной k в формулу (5) с позиций реализации процедур преобразования Фурье. Для этого интерпретируем указанное выражение следующим образом:

$$\rho_W(k', 0) = 2 \sum_{l=-N+1}^{N-1} \exp(-j2\pi kl/M) \times \{ \exp(-j2\pi \Delta k l/M) z_a(l) z_a^*(-l) \}. \quad (6)$$

Если теперь сигнал представить в форме

$$z_a(l) = A(l) \exp(j\Theta(l)),$$

где $A(l)$ и $\Theta(l)$ — соответственно амплитуда и полная фаза, то выражение в фигурных скобках в формуле (6) можно представить как

$$\bar{z}_a(l) \bar{z}_a^*(-l) = A(l)A(-l) \exp[j(\Theta(l) - \pi \Delta k l/M)] \exp[-j(\Theta(-l) + \pi \Delta k l/M)]. \quad (7)$$

В формуле (7) $\bar{z}_a(l)$ получается из $z_a(l)$ сдвигом спектра последнего на величину $(-\pi \Delta k l/M)$. Следовательно, по конечному результату метод дробных коэффициентов совпадает с методом цифрового гетеродинирования.

Необходимо отметить, что этап нахождения экстремумов функций с требуемой точностью при использовании метода цифрового гетеродинирования требует многократных вычислений функции РЭ [5]. Так, для достижения точности оценивания частотных параметров до десятых долей герца (при анализе сигналов ВЧ диапазона) потребуется несколько десятков итераций. Очевидно, что в такой ситуации вычисление всех спектральных коэффициентов процедурами БПФ нецелесообразно, поскольку достаточно рассчиты-

вать на каждом шаге требуемое значение дробного коэффициента в функции РЭ. Более того, во всех проводимых процедурах итераций будет использоваться один и тот же набор исходных данных $z_a(l)z_a^*(-l)$.

Для оценивания $\nabla_{(r)}$ в формуле (4) вычисляется разность между предыдущим и последующим спектральными отсчетами, поэтому простота получаемого алгоритма компенсирует сравнительно медленную скорость сходимости итерационной процедуры. Однако алгоритм резко усложняется при переходе к РЭ, обеспечивающим низкий уровень интерференции на в формируемом частотно-временном пространстве. Так, при $n = 0$ и $k' = k + \Delta k$ выражение (2) принимает вид

$$\begin{aligned} \rho(k', 0) &= \sum_{l=-N+1}^{N-1} \exp[-j2\pi kl/M] \times \\ &\times \sum_{m=-N+1}^{N-1} W_G^\alpha(m, l) [z_a(m+l) \exp(-j\pi \Delta k l/M)] \times \\ &\times [z_a^*(m-l) \exp(-j\pi \Delta k l/M)]. \end{aligned} \quad (8)$$

Для того чтобы метод дробных коэффициентов трансформировался в метод гетеродинирования, необходимо выражения в квадратных скобках преобразовать следующим образом:

$$\begin{aligned} \bar{z}_a(m+l) &= A(m+l) \exp(j\Theta(m+l)) \exp(-j\pi \Delta k(m+l)/M); \\ \bar{z}_a(m-l) &= A(m-l) \exp(-j\Theta(m-l)) \exp(-j\pi \Delta k(m-l)/M). \end{aligned} \quad (9)$$

Отметим, что смещение спектра сигнала и его инвертированной во времени копии осуществляется в одну сторону, хотя в действительности внутренняя сумма в формуле (8) преобразуется к виду

$$\begin{aligned} \sum_{m=-N+1}^{N-1} W_G^\alpha(m, l) z_a(m+l) \times \\ \times z_a^*(m-l) \exp[j2\pi \Delta k l/M]. \end{aligned} \quad (10)$$

Следовательно, в отличие от функции Вигнера (1) функция (2) не может быть вычислена применением дискретного спектрального преобразования с дробными коэффициентами. Такую задачу можно решить только повторным вычислением внутренней суммы в формуле (8) для процесса, смещенного по спектру на величину Δk . Однако вычислительная емкость такой процедуры соизмерима с вычислительной емкостью процедуры N -точечного дискретного преобразования Фурье, для которой отсутствует быстродействующая вычислительная форма. В таких условиях применение метода цифрового гетеродинирования сильно затрудняет обработку функций РЭ, так как в несколько десятков раз увеличивает объемы вычис-

лений. Кардинальным решением этой проблемы является отказ от многократного вычисления функции (8) при выполнении итеративной процедуры (4) и переход к другим методам повышения точности оценивания, например, с использованием процедур сплайн-аппроксимации. Следовательно, необходим поиск сплайн-функций, применение которых обеспечило бы синтез быстрых алгоритмов оценивания сигналов.

Выбор сплайна для алгоритма итеративного поиска экстремумов распределений

Учитывая, что в качестве основного результата рассматривается повышение скорости ЦОС, то трудно рассчитывать на достижение требуемого эффекта, ориентируясь, например, на глобальные интерполяционные и сглаживающие сплайны [10]. Так, при построении интерполяционного сплайна необходимо решать систему линейных алгебраических уравнений, размерность которой определяется числом интерполяционных условий и может быть достаточно велика. Однако для локальных сплайнов эта система распадается на совокупность подсистем, размерность которых значительно меньше. Тогда при вычислении каждого из параметров сплайна используется лишь несколько интерполяционных условий, что упрощает общую задачу.

Согласно [10], локальные сплайны характеризуются легкостью и быстротой построения, возможностью учитывать при вычислении лишь находящиеся на этом интервале и ближайшие к нему сеточные значения. Кроме того, локальный сплайн позволяет достичь максимально возможного порядка аппроксимации, хотя с константами в оценках несколько большими, чем у интерполяционных. Однако главное достоинство локальных сплайнов — в возможности их представления линейной комбинацией B -сплайнов [10], коэффициентами которых являются заданные явным образом линейные комбинации сеточных значений аппроксимируемой функции.

Простейшей локальной схемой построения сплайна на равномерной сетке с заданным шагом непосредственно по сеточным значениям аппроксимируемой функции является схема Шёнберга [11]. Существует несколько подходов к синтезу локальных сплайнов, но применительно к решению рассматриваемой проблемы наибольший интерес представляет направление, обеспечивающее интерполяцию как функции, так и ее производных, и при этом требующее при их построении лишь сеточных значений самой функции [12].

Для получения аналитических алгоритмов обработки функций РЭ введем следующие обозначения: C^s — пространство функций P , для кото-

рых непрерывны s -кратные производные; L_∞^s — пространство функций P , для которых $P^{(s-1)}$ абсолютно непрерывны, а $P^{(s)}$ локально ограничены; $x_+ = (x + |x|)/2$, причем $x_k^m = \begin{cases} hk, \\ h(k-1/2) \end{cases}$ в зависимости от того, m четно или нечетно, здесь h — шаг равномерной сетки. Тогда, если $x \in [x_k^m, x_{k+1}^m]$, где m — степень сплайна, то $x = x_k^m + ht$, $t \in [0, 1]$. Обозначим символом δ_h^s центральную разность:

$$\delta_h^s(g(x)) = h^{-s} \sum_{k=0}^s (-1)^k \binom{s}{k} g(x - h(k - s/2)). \quad (11)$$

Через выражение (11) в [12] определяется центральный B -сплайн 1-го дефекта степени $(m-1)$ с узлами в точках x_k^m :

$$b_h^m(x) = \delta_h^m(x_+^{m-1}/(m-1)!). \quad (12)$$

К основным свойствам B -сплайнов следует отнести:

$$b_h^m(x) \geq 0; \quad b_h^m(-x) = b_h^m(x)$$

$$\text{и } \text{supp } b_h^m(x) = (-mh/2, mh/2).$$

Пусть $P(x)$ — некоторая непрерывная функция и $P_k^m = P(x_k^m)$. Тогда локальный сплайн 1-го дефекта степени $m-1$ на сетке с шагом h , построенный по данным $\{P_k^m\}$, записывается следующим образом [12]:

$$S^m(P, x) = h \sum_k^m \varphi_k^m b_h^m(x - hk), \quad (13)$$

где φ_k^m — конечная линейная комбинация сеточных значений $\{P_k^m\}$.

Свойства сплайна определяются функционалом φ_k^m . Простейший сплайн такого рода был получен в [11] для $\varphi_k^m = P_k^m$ Шёнбергом, который, используя определение центральной разности (11), ввел более общее выражение для сплайна произвольной степени $m+s-1$:

$$S^{m+s}(P, x)^{(s)} = h \sum_k^{m+s} \varphi_k^{m+s} \delta_h^s(b_h^m(x - hk)). \quad (14)$$

Выражение (14) является сплайном степени $m-1$ и при $s=0$ вырождается в формулу (13).

Простейшие сплайны, аппроксимирующие $P^{(s)}$, имеют вид

$$S_0^{m+s}(P, x)^{(s)} = h \sum_k^{m+s} P_k^{m+s} \delta_h^s(b_h^m(x - hk)). \quad (15)$$

Исследование сплайнов вида (14), проведенное в [12], показало, что порядок аппроксимации зависит от степени сплайна и для простейших сплайнов (15) равен $O(h^2)$, если степень аппроксимирующего многочлена больше или равна двум. Для того чтобы получить более высокий порядок аппроксимации, необходимо в формуле (14) заменить P_k^{m+s} на $\varphi_k^{m+s}(P)$, где φ — функционал над сеточными значениями сплайна более сложной формы.

С точки зрения показателей сложности обработки дискретизированных процессов такой подход не продуктивен. Однако применение простейших сплайнов в форме (15) представляет практический интерес, так как позволяет получить не только сплайн любой степени заданной функции, но и ее производных. Кроме того, на базе выражения (15) в [12] удалось по-новому подойти к проблеме синтеза сплайн-функций и их производных, что дает возможность повысить точность приближения без сложных преобразований сеточных значений.

Суть идеи заключается в том, что для сплайнов (15) выводится асимптотическая формула (через операцию представления сеточных значений рядом Тейлора) и составляется линейная комбинация, которая уничтожает постоянные слагаемые в асимптотике:

$$S_r^{m+s}(P, x)^{(s)} = \sum_{l=0}^{r/2} \beta_{sl}^{s+m} h^{2l} S_0^{m+2l+s}(P, x)^{(2l+s)}, \quad \beta_0^{s+m} = 1. \quad (16)$$

В частном случае при $r=0$ выражение (16) сводится к формуле (15), при этом получаем следующий сплайн степени $m-1$:

$$S_r^{m+s}(P, x)^{(s)} = h \sum_k^{m+s} P_{kr}^{m+s} \delta_h^s(b_h^m(x - hk)), \quad (17)$$

где $P_{kr}^{m+s} = P_k^{m+s} + \sum_{l=1}^{r/2} h^{2l} \beta_{2l}^{m+s} \delta_h^{2l}(P_{k+l}^m)$, $2 \leq r \leq p$, $r < m$, m — четное значение, p — степень аппроксимируемого полинома. Коэффициенты β получаются из следующей рекуррентной формулы [12]:

$$\beta_0^{m+s} = 1, \quad \beta_{2n}^{m+s} = - \sum_{l=0}^{n-1} \beta_{2l}^{m+s} M_{2(n-l)}^{m+2l+s}, \quad \beta_{2n+1}^{m+s} = 0. \quad (18)$$

Здесь $M_s^m = \frac{h^{-s}}{s!} \int_0^{hm} \left(t - \frac{hm}{2}\right)^s b_h^m(t) dt$, $M_0^0 = 1$, $M_s^0 = 0$, $s \neq 0$.

Исследование асимптотических формул (17) и (18) показало, что если задана степень сплайна m , то наивысшего порядка аппроксимации можно достичь, если определить $r = 2[(m - 1)/2]$. Тогда порядок аппроксимации будет $O(h^r)$, а сплайн $S_r^{m+s}(P, x)^{(s)}$ точен на многочленах степени $m + s - 1$. Кстати, это наивысшая степень многочлена, на которой может быть точной производная порядка s от сплайна степени $m + s - 1$.

Шаблон сплайна (т. е. множество сеточных значений функции, задействованных для вычисления значения сплайна в данной точке) содержит $m + s + r$ таких значений:

$$\{P_{N-m-s+1-r/2}^{m+s}, \dots, P_{N+r/2}^{m+s}\}, \quad N = [x/h],$$

которое является минимальным для сплайнов, точно воспроизводящих $P^{(s)}$, если P — многочлен степени $m + s - 1$, поэтому такие сплайны можно называть сплайнами минимального шаблона (СМШ). В асимптотическом выражении СМШ могут быть выделены дополнительные постоянные составляющие, подавление которых возможно образованием новой комбинации с (17). Так, для $m = 2n$ имеем локальный сплайн степени $2n - 1$ дефекта 1:

$$S_q^{2n,s}(P, x) = h \sum_{k=N-2n+1}^N P_{qk}^{2n,s} \delta_h^s(b_n^m(x - hk)), \quad (19)$$

где $P_{qk}^{2n,s} = P_{kr}^{2n+s} + h^{2n} \eta_{2n}^s \delta_h^{2n}(P_k^{4+s})$;
 $\eta_{2n}^s = \frac{B_{2n}}{(2n)!} + \beta_{2n}^{2n+s}$.

Шаблон этого сплайна содержит $4n + s$ сеточных значений — на два больше, чем шаблон соответствующего СМШ. Во внутренних точках промежутка $[hN, h(N + 1)]$ сплайн $S_q^{2n,s}(P, x)$ аппроксимирует $P^{(s)}$ с точностью $O(h^{2n})$, а в узлах — с точностью $O(h^{2n+1})$. Если P — многочлен степени $2n + s + 1$, то $S_q^{2n,s}$ интерполирует $P^{(s)}$ в узлах $\{hN\}$.

В частности, для многочлена степени $2n + 1$ сплайн $S_q^{2n,0}$ является интерполяционным, поэтому (19) целесообразно определить квазиинтерполяционным сплайном (КИС).

Выражение для КИС четной степени ($m = 2n + 1$) имеет вид

$$S_q^{2n+1,s}(P, x) = h \sum_{k=N-2n}^N P_{qk}^{2n+1,s} \delta_h^s(b_n^{2n+1}(x - hk)). \quad (20)$$

Здесь $P_{qk}^{2n+1,s} = P_{kr}^{2n+1+s} + h^{2n} + 2\eta_{2n+1}^s \times \delta_h^{2n+2}(P_k^{4n+s+3})$; $r = 2n$, $\eta_{2n+1}^s = -\frac{2n+1}{(2n+2)!} \times B_{2n+2}(1/2) + \beta_{2n+2}^{2n+1+s}$; а $B_k(\tau)$ — многочлены Бернулли. Заметим, что сплайн $S_q^{2n+1,s}$ "квазиинтерполирует" функцию $P^{(s)}$ в точках $h(N + 0,5)$.

Сплайн $S_q^{ms}(P, x)$, будучи локальным, при использовании лишь ближайших к x сеточным значениям функции P приближает $P^{(s)}$ практически как сплайн, интерполирующий значения $P^{(s)}$ на сетке, тем самым обеспечивая повышение точности аппроксимации производных. Но главным достоинством приведенных сплайнов является то, что они, оставаясь локальными, практически совпадают со сплайнами, интерполирующими $P^{(s)}$, хотя при их построении используются лишь сеточные значения самой функции P .

Теперь рассмотрим возможность приложения представленных сплайнов к обработке функций совместных распределений.

Простейший сплайн, аппроксимирующий производную по дискретным значениям одного среза функции распределения "экспоненциально" типа (2), имеет вид

$$S_0^{m+1}[P(k, n), x]^{(1)} = h \sum_{k=N-m-1}^N P(k, n) \delta_h^1(b_h^m(x - hk)). \quad (21)$$

Алгоритм итеративного процесса градиентного поиска экстремумов с использованием этого сплайна может быть записан следующим образом:

$$x_{(i+1)} = x_{(i)} - \gamma \{S_0^{m+1}[P(k, n), x]^{(1)}\}, \quad (22)$$

где i — шаг итерации.

В случае применения квадратичного сплайна сумма в формуле (21) будет содержать три слагаемых, но шаблон при этом должен включать четыре значения $P(k, n)$. Для кубических сплайнов число слагаемых в (21) и шаблон увеличатся на единицу и т. д. В целях повышения точности определения $S_0^{m+1}[P(k, n), x]^{(1)}$ используется СМШ, обеспечивающий максимальную точность восстановления производной при минимальном числе сеточных значений, или же КИС.

Заключение

Полученное выражение (22) можно рассматривать как аналитическую основу для построения быстросействующих алгоритмов оценивания параметров сигналов на основе обработки их РЭ. Практические формулы для вычисления сплайнов функций и их производных могут быть получены из

Таблица 1

Зависимость точности определения экстремумов распределения от значения ОСШ

ОСШ, дБ	0	4	8	12	16	18	20
σ_f для распределения, Гц	>3000	6	0,5	0,2	0,1	0,08	0,07
σ_f для частотного детектора, Гц	—	—	>3000	10	2	0,5	0,3

Таблица 2

Зависимость точности оценивания от числа реализаций при ОСШ = 4 дБ

$N_{\text{изм}}$	1	20	40	60	80	100	120
σ_f , Гц	6	2	1,2	1,1	0,9	0,85	0,84

приведенных выше общих соотношений (15), (17), (20) после соответствующих преобразований.

Дальнейшее направление работы видится в реализации процедур поиска частной производной двумерного сплайна, который строится по информации о частоте и времени. В этом случае алгоритм итеративного градиентного поиска можно представить в следующем виде:

$$f_{\langle i+1 \rangle} = f_{\langle i \rangle} - \gamma \frac{\partial S^{m+1}[P((k, n), f, t)]}{\partial f}.$$

За счет информации о соседних срезах функций РЭ сплайн позволяет учитывать динамику изменения мгновенной частоты. В конечном итоге это ведет к увеличению скорости сходимости итеративной процедуры и повышению точности оценивания экстремумов.

Так, в табл. 1 приведены результаты оценивания частотных параметров однокомпонентных узкополосных сигналов передач дискретных сообщений в зависимости от отношения сигнал/шум (ОСШ).

Оценивание проводилось на основе обработки функций РЭ при использовании алгоритма частотного гетеродинамирования, в котором применялись локальные сплайны, и с использованием метода частотного детектирования. По условиям эксперимента обработка осуществлялась в полосе 4 кГц, длительность обрабатываемой реализации составляла 16 мс ($M = 128$). Согласно данным табл. 1 при ОСШ ≥ 4 дБ точность оценивания

частоты σ_f составила 6...7 Гц, что приближается к границе потенциальной помехоустойчивости. Состоятельность и эффективность получаемых оценок сохраняется при ОСШ > 0 дБ. Более того, результаты слабо зависят от статистических свойств сигналов и помех. Последнее обстоятельство особенно важно при обработке сигналов в условиях высокого уровня априорной неопределенности параметров канала.

Дальнейшее повышение точности оценивания возможно за счет увеличения числа обрабатываемых реализаций. Так, в табл. 2 приведена зависимость требуемого числа реализаций $N_{\text{изм}}$, обеспечивающих необходимую точность оценивания для важного с практической точки зрения значения ОСШ = 4 дБ.

Анализ полученных результатов показал, что предлагаемые процедуры более чем на порядок позволяют повысить точность оценивания частотных параметров сигналов без увеличения длительности обрабатываемых фрагментов и объемов проводимых вычислений.

Список литературы

1. Коэн Л. Время-частотные распределения: Обзор // ТИИЭР. 1989. Т. 77. № 10. С. 72—121.
2. Jassen A. J. E. M. On locus and spread of pseudodensity function in time-frequency plane // Philips J. Res. 1982. V. 37. P. 79—110.
3. Claaseen T. A. C. M., Meclenbrauker W. F. G. The Wigner distribution a tool for time-frequency signal analysis. Part 1, 2, 3 // Philips J. Res. 1980. V. 35.
4. Choi H. I., Williams W. J. Improved time-frequency representation of multicomponent signals using exponential kernels // IEEE Trans. on ASSP. 1989. V. ASSP-37. № 6. P. 862—871.
5. Алексеев А. А., Чеченев С. Ю., Кириллов А. Б. Анализ сигналов на основе функций распределения мощности в условиях многосигнального воздействия // Радиотехника. 1993. № 10—12.
6. Алексеев А. А., Железняк В. К., Дворников С. В. и др. Применение методов частотно-временной обработки акустических сигналов для анализа параметров реверберации // Научное приборостроение. 2001. Т. 11. № 1. С. 65—76.
7. Дворников С. В., Алексеева Т. Е. Распределение Алексеева и его применение в задачах частотно-временной обработки сигналов // Информация и Космос. 2006. № 3. С. 9—21.
8. Дворников С. В. Теоретические основы синтеза билинейных распределений. СПб.: Изд-во Политехн. ун-та, 2007. 268 с.
9. Ярославский Л. П. Введение в цифровую обработку изображений. М.: Сов. радио, 1979. 312 с.
10. Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. Методы сплайн-функций. М.: Наука, 1980. 352 с.
11. Schoenberg I. J. On spline functions. Inequalities. N. Y.: Acad. Press, 1967.
12. Желудев В. А. Локальная сплайн-аппроксимация на равномерной сетке // Журнал вычислительной математики и математической физики. 1987. Т. 27. № 9. С. 1296—1309.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

УДК 004.051

Т. К. Кравченко, д-р эконом. наук, проф.,
Государственный университет —
Высшая школа экономики,
e-mail: tkravchenko@hse.ru

Разработка системы ключевых показателей для оценки эффективности процессов службы информационных технологий

Проведена сравнительная оценка применимости существующих подходов для разработки ключевых показателей эффективности процессов (KPI) службы информационных технологий компании. Предложен новый подход для разработки KPI, учитывающий специфику процессов ИТ-службы и использующий преимущества двух подходов: методика GQM и процессного подхода. Суть подхода: наложить методику GQM на процессный подход: выделять ключевые показатели в соответствии с методикой GQM для каждого элемента процесса.

Ключевые слова: служба информационных технологий (ИТ-служба), ключевые показатели эффективности процессов (KPI), набор лучших международных практик в области управления информационными технологиями — ITIL (Information Technology Infrastructure Library), стандарт CobiT (Control Objectives for Information and related Technology), система сбалансированных показателей BSC (Balanced Scorecard), процессный подход, методика GQM (Goal Question Metric — Цель Вопрос Метрика)

Вопрос оценки эффективности процессов службы информационных технологий (далее ИТ-службы) является актуальным для любой компании, поскольку ИТ-служба предоставляет и поддерживает множество критичных для бизнеса сервисов.

ИТ-служба является вспомогательным незарабатывающим подразделением компании. Поэтому стандартные ключевые показатели эффективности процессов (KPI), используемые для компании в целом, для нее не подходят. Одновременно ИТ-службу нельзя ставить в один ряд с другими вспомогательными подразделениями.

Известные в настоящее время подходы к разработке ключевых показателей для оценки эффективности процессов не в достаточной степени учитывают специфику ИТ-службы. Поэтому в

данной статье предпринята попытка решить следующие задачи:

- выявить признаки, по которым будут сравниваться различные подходы для оценки эффективности процессов ИТ-службы;
- оценить применимость существующих подходов для разработки KPI ИТ-службы компании с позиции выявленных признаков;
- предложить новый подход для разработки KPI, учитывающий специфику процессов ИТ-службы.

Для оценки эффективности процессов ИТ-службы можно выделить следующие признаки, которые должна учитывать система ключевых показателей: охват всех элементов процессов; отсутствие избыточности; учет специфики процессов ИТ-службы; учет целей процессов ИТ-службы; учет целевых аудиторий (адресность).

Охват всех элементов процессов: входы, выходы, виды деятельности, ресурсы и управляющие воздействия.

Отсутствие избыточности: система показателей должна быть необходимой и достаточной, которая не содержит при этом лишних показателей, не несущих ценности для управления этими процессами. Она должна ориентироваться на специфику деятельности ИТ-службы, опираться на типовые именно для ИТ-службы процессы.

Система показателей должна *учитывать цели этих процессов*, т. е. должна быть системой, создаваемой "сверху-вниз". Каждый показатель в этом случае будет измерять достижение той или иной цели, что позволит лицам, принимающим решения, видеть четкую взаимосвязь между детальными показателями и ощутимыми для деятельности ИТ-службы целями.

Система показателей должна *учитывать целевые аудитории конкретных отчетов*. Предполагается, что для каждого уровня управления и для каждой группы лиц, принимающих решения, должны быть разработаны свои специфические отчеты, состоящие из конкретных специально для этих аудиторий сгруппированных ключевых показателей эффективности. Помимо набора ключевых показателей эффективности, от целевой аудитории может зависеть периодичность отчетов, форма их отображения и детализация.

Выделим подходы, которые могут быть использованы при формировании системы ключе-

вых показателей для оценки эффективности процессов ИТ-службы.

Важнейшую роль при анализе эффективности процессов ИТ-службы играет *набор лучших международных практик в области управления информационными технологиями — ITIL (Information Technology Infrastructure Library)*.

В основе ITIL [1, 3, 9, 11] лежит сервисный подход, который говорит о том, что ИТ-служба должна оказывать бизнесу услуги, необходимые для реализации целей бизнеса.

Практический опыт ITIL в настоящее время включает в себя инструментальные средства (специализированное программное обеспечение), тренинговые и консалтинговые услуги, соответствующие методические рекомендации для организации всех видов деятельности ИТ-подразделений.

Второй подход, который может быть использован при оценке эффективности процессов ИТ-службы, это *CobiT (Control Objectives for Information and related Technology)*, продвигаемый ассоциацией ISACA (Ассоциация аудита и контроля информационных систем), которая была основана в 1969 г. для финансовых аудиторов в контроле ИТ [2, 10].

CobiT обеспечивает единый подход к сбору, анализу информации, подготовке выводов и заключений на всех этапах управления, контроля и аудита ИТ, возможность сравнения существующих ИТ-процессов с "лучшими" практиками, в том числе отраслевыми.

Основные инструменты, которые предлагает CobiT для анализа эффективности процессов ИТ-службы с точки зрения бизнес-целей, — это таблица соответствия бизнес- и ИТ-целей компании и подход к каскадированию целей ИТ до конкретных ключевых показателей и метрик.

Следующий подход — *формирование системы сбалансированных показателей BSC (Balanced Scorecard)*.

В рамках этого подхода [5, 6] показатели финансовых отчетов объединяются с операционными параметрами, что позволяет оценить нематериальные активы: уровень корпоративных инноваций, степень удовлетворенности сотрудников, эффективность приложений и т. д. Причем параметры рассматриваются с четырех точек зрения: финансовой; удовлетворения потребностей клиентов; внутренних процессов; дальнейшего роста и обучения. Менеджеры должны сопоставить перспективы каждого из этих четырех направлений с общей стратегией развития бизнеса.

По мнению ряда специалистов, причинно-следственные связи системы сбалансированных показателей (BSC) в чистом виде для оценки эффективности ИТ-службы не годны.

В качестве альтернативы выработан подход [4], ориентированный на информационные технологии и направленный на привлечение ИТ-ресурсов к решению стратегических задач — *ITSC (IT Scorecard)*.

Аналогично с BSC можно выделить четыре перспективы влияния ИТ на бизнес — повышение качества продукции; повышение производительности труда; повышение качества принятия решений; помощь в развитии бизнеса.

По каждой перспективе определяются цели, которые, в свою очередь, на операционном уровне каскадируются на инициативы по их достижению.

Деятельность ИТ-службы считается эффективной, если она соответствует выделенным инициативам и способствует достижению поставленных целей.

Существенную роль для оценки эффективности деятельности ИТ-службы играет *процессный подход в управлении*, определяющий рассмотрение любой компании как сети бизнес-процессов, увязанных с целями и миссией организации.

Процесс представляет собой систему последовательных, целенаправленных и регламентированных видов деятельности, в которой посредством управляющего воздействия и с помощью ресурсов входы процесса преобразуются в выходы, результаты процесса, представляющие ценность для потребителей.

Процессный подход основывается на следующих принципах: рассмотрение бизнеса как системы; восприятие деятельности как процесса; стандартизация и прозрачность ответственности.

Выстраивая систему ключевых показателей с использованием процессного подхода, необходимо помнить о пяти основных составляющих эффективности процесса (вход, выход, ресурсы, управляющее воздействие, виды деятельности) и измерять каждую из них.

Ключевые показатели эффективности "входа" отражают нагрузку на процесс. Например, для процесса управления инцидентами число инцидентов — ключевой показатель эффективности "входа".

Ключевые показатели эффективности "выхода", или показатели результативности, характеризуют, насколько процесс достигает своей цели.

Ключевые показатели эффективности "ресурсов" отражают загрузку и достаточность ресурсов, используемых процессом.

Ключевые показатели эффективности "управляющего воздействия" показывают, насколько процесс управляем и эффективны управляющие воздействия.

Ключевые показатели эффективности "видов деятельности" иллюстрируют, насколько правильно выполняется процесс.

В 1992 г. Виктор Базили (*University of Maryland*) предложил *методику GQM (Goal Question Metric —*

Цель—Вопрос—Метрика), обеспечивающую разработку метрик информационных систем, отвечающих целям [7, 8, 9].

При этом выделяются следующие уровни: концептуальный уровень *Goal — Цель*; операционный уровень *Question — Вопрос*; количественный уровень *Metric — Метрика*.

В соответствии с уровнями сформулированы следующие этапы разработки метрик:

- определить бизнес-цели;
- сформулировать вопросы (достигнуты ли цели?);
- разработать набор метрик, помогающих ответить на эти вопросы.

На первом этапе для определения бизнес-цели выделяются: направление; ключевая характеристика; объект процесса; точка зрения. Например: оптимизировать (направление), временные характеристики (ключевая характеристика), процесс обработки запроса на изменение (объект), с точки зрения менеджера процесса (точка зрения).

В качестве объектов измерения могут выступать процессы, продукты или ресурсы.

В качестве источников информации для формулирования целей — политики и стратегии (для определения направления), модели и описания (для определения объектов) и организационная структура (для определения точки зрения).

На втором этапе формулируются вопросы, помогающие ответить, каким путем идти к поставленным целям и достигнуты ли они. При этом выделяют три группы вопросов.

Первая группа вопросов касается характеристики объекта в отношении цели. Например, какова текущая скорость обработки запроса на изменение? Выполняется ли процесс обработки запроса на изменение в настоящее время?

Вторая группа вопросов касается характеристики атрибутов объекта в отношении к ключевой характеристике. Например: каковы временные отклонения процесса обработки запроса на изменение от допустимых? Улучшается ли выполнение процесса?

Третья группа вопросов касается характеристики объекта в отношении к сути. Например: удовлетворительно ли с точки зрения менеджера текущее выполнение процесса? Улучшается ли явно выполнение процесса?

После формирования вопросов на третьем этапе методики необходимо разработать соответствующие им метрики, позволяющие ответить на поставленные вопросы. Метрики могут быть как объективными, так и субъективными. Приведем пример набора метрик, разработанных по методике GQM в соответствии со сформулированными ранее целью и вопросами (табл. 1).

Оценим применимость существующих подходов, описанных выше, для оценки эффективно-

сти процессов ИТ-службы с точки зрения выделенных признаков (табл. 2).

В строках табл. 2 представлены существующие подходы к разработке системы ключевых показателей эффективности. В столбцах таблицы перечислены признаки, по которым сравниваются отдельные подходы. На пересечении строк и столбцов указано, учитывает ли система ключевых показателей эффективности, разработанная в соответствии с данным подходом, этот признак (соответственно, "+" и "-").

Как видно из табл. 2 ни один из рассмотренных подходов не позволяет разработать систему ключевых показателей, учитывающую одновременно все выявленные признаки, которые, с нашей точки зрения, необходимо учитывать при оценке эффективности процессов ИТ-службы.

В связи с этим в статье для оценки эффективности процессов ИТ-службы предлагается использовать преимущества двух подходов: методики GQM и процессного подхода.

Это вызвано тем, что методика GQM позволяет создать систему ключевых показателей для оценки ИТ-службы, отвечающую большинству выделенных признаков. Для того чтобы система ключевых пока-

Таблица 1

Пример метрик GQM

Цель	Оптимизировать временные характеристики процесса обработки запроса на изменение с точки зрения менеджера процесса
Вопрос	Какова текущая скорость обработки запроса на изменение?
Метрики	Среднее время обработки запросов на изменение, % случаев нарушения сроков обработки запроса на изменение
Вопрос	Каковы временные отклонения процесса обработки запроса на изменение от допустимых?
Метрики	% отклонений времени обработки запросов на изменение от допустимого Субъективная оценка отклонений менеджером процесса
...	...

Таблица 2

Оценка применимости существующих подходов

Подход, методика	Охват всех элементов процесса	Отсутствие избыточности	Учет специфики ИТ-службы	Учет целей процессов	Адресность системы ключевых показателей
ITIL	—	+	+	—	—
CobiT	—	+	+	+	—
BSC	—	—	—	+	—
ITSC	—	—	+	+	—
Процессный подход	+	—	—	—	—
GQM	—	+	+	+	+

зателей учитывала и недостающий признак, а именно "охват всех элементов процесса", необходимо интегрировать методику GQM с единственным из существующих подходов, удовлетворяющим этому условию — процессным подходом.

Процессный подход предусматривает структуризацию деятельности ИТ-службы как совокупности процессов с определенными входами, выходами, управляющим воздействием, ресурсами и видами деятельности.

Методика GQM предполагает построение метрик по определенному алгоритму в зависимости от целей.

Для того чтобы создать систему ключевых показателей для оценки эффективности процессов ИТ-службы, целесообразно наложить методику GQM на процессный подход — выделять ключевые показатели в соответствии с методикой GQM для каждого элемента процесса. Это позволит создать систему показателей, охватывающую все элементы процесса. При этом важным преимуществом методики GQM является ее многоуровневость. Методика предлагает смотреть на объекты измерения с различных точек зрения, что позволит создать систему показателей, ориентированную на поддержку принятия управленческих решений.

В соответствии с лучшими практиками ITIL процессы ИТ-службы выделяют согласно жизненному циклу услуг.

На этапе определения стратегии сервисов выделяют следующие процессы: разработка стратегии; управление ИТ-финансами; управление спросом; управление портфелем сервисов.

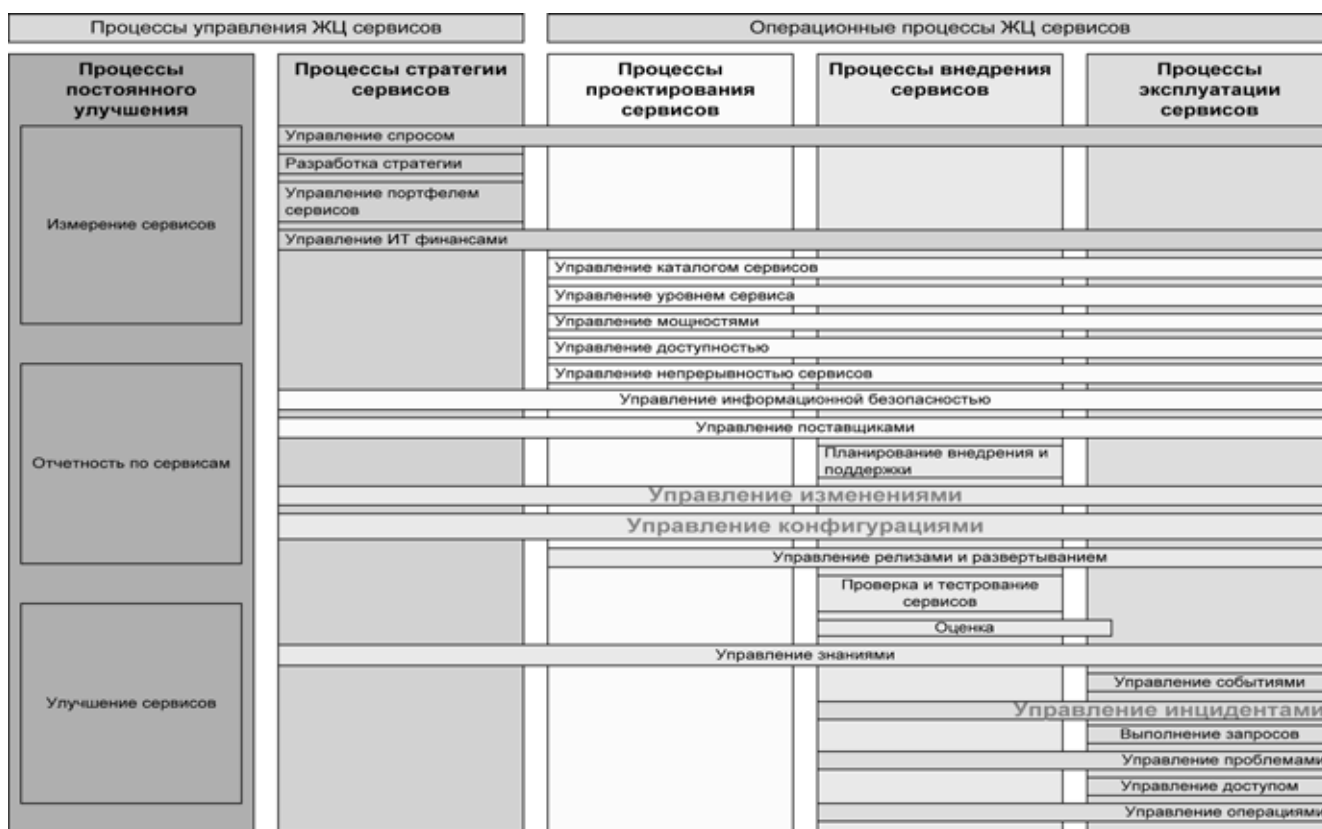
На этапе дизайна сервисов выделяют такие процессы управления: каталогом сервисов; уровнем сервиса; мощностями; доступностью; непрерывностью сервисов; информационной безопасностью; поставщиками.

На этапе внедрения сервисов выделяют следующие процессы: планирование внедрения и поддержки; управление изменениями, управление конфигурациями, управление релизами, проверка и тестирование сервисов, оценка, управление знаниями.

На этапе эксплуатации сервисов выделяют такие процессы управления: инцидентами, событиями, сервисными запросами, проблемами, доступом, операциями.

На этапе постоянного улучшения сервисов выделяют такие процессы, как измерение сервисов, отчетность по сервисам, улучшение сервисов.

Базовыми процессами целесообразно считать процессы управления инцидентами, конфигурациями и изменениями, так как это процессы, которые обязательно являются частью деятельности ИТ-службы. Процессы ИТ-службы в соответствии с ITIL v3. представлены на рисунке.



Процессы ИТ-службы

Цель процесса управления инцидентами — оперативное устранение инцидентов, связанных с предоставлением ИТ-услуг, а также ситуаций, которые могут привести к сбоям в предоставлении ИТ-услуг. Процесс позволяет минимизировать негативное влияние сбоев в ИТ-инфраструктуре на основную деятельность компании путем быстрого устранения этих сбоев.

Цель процесса управления изменениями — обеспечить стабильность в предоставлении ИТ-услуг путем тщательного планирования, согласования, тестирования и внедрения изменений в ИТ-инфраструктуре. Этот процесс является процессом более высокого уровня по сравнению с управлением инцидентами, но в любом случае присутствует в каждой ИТ-службе, хотя бы для наиболее крупных изменений и не в документированном виде.

Цель процесса управления конфигурациями — обеспечить качественные сбор, хранение, модификацию и верификацию информации об ИТ-ресурсах и активах, предоставление информации заинтересованным подразделениям, лицам и другим процессам. При отсутствии налаженного процесса управления конфигурациями сложно организовать эффективный процесс управления

инцидентами и практически невозможно организовать процесс управления изменениями достаточно высокого уровня зрелости.

Согласно рекомендациям ИТIL в выделенных базовых процессах основными видами деятельности являются:

- в области управления инцидентами — регистрация инцидентов (сбоев и нарушений в предоставлении ИТ-услуг), их классификация, назначение исполнителей, разрешение инцидентов и возобновление предоставления услуг;
- в области управления изменениями — регистрация запросов на изменение и их фильтрация; оценка влияния запрашиваемых изменений; согласование; планирование; контроль реализации изменений (в отношении объектов инфраструктуры, за которые отвечает ИТ-служба); оценка результатов внедрения изменений;
- в области управления конфигурациями — планирование процесса; идентификация конфигурационных единиц; организация конфигурационной базы данных; контроль конфигурационных единиц; верификация и аудит; предоставление информации о конфигурационных единицах.

Таблица 3

Таблица целей процесса

Объекты процесса		Точка зрения			
		1	...	N	
Продукты процесса	Продукт 1				
	...				
	Продукт N				
Виды деятельности	Вид деятельности 1		Направление Ключевая характеристика Объект Точка зрения		
	...				
	Вид деятельности N				
Ресурсы	Персонал	1			
		...			
		N			
	Система автоматизации				
	Документы	Документ 1			
		...			
Документ N					
Управляющее воздействие	Планирование				
	Выполнение				
	Проверка				
	Действие				

Основой создания системы ключевых показателей эффективности с помощью интегрированного подхода является таблица целей процесса (табл. 3).

В этой таблице строки — объекты процесса, выделенные в соответствии с процессным подходом, а именно: продукты процесса, виды деятельности, ресурсы и управляющее воздействие. Выделим три вида ресурсов: человеческие (персонал); система автоматизации; документационное обеспечение. Управляющее воздействие делим на четыре вида (в соответствии с циклом Деминга): планирование; выполнение; проверка; действие (корректировка).

Столбцы в табл. 3 — точки зрения, с позиций которых оценивается эффективность процесса. Точки зрения выделяются в соответствии с методикой GQM.

На пересечении точки зрения и объекта процесса формулируется цель процесса. Цель выделяется также в соответствии с методикой GQM: направление, ключевая характеристика, объект, точка зрения.

Следующий шаг после того, как все цели совершенствования будут сформулированы, — выделение показателей, определяющих достижение целей, и методик расчета этих показателей.

Сформулированные ключевые показатели эффективности описываются следующим образом: индекс ключевого показателя эффективности, название показателя, описание показателя и методика его расчета, цель, достижение которой измеряет показатель, единица измерения ключевого показателя эффективности.

После того как для каждой цели определены показатели, формируется перечень отчетов. Перечень должен включать номер отчета, краткое описание отчета, периодичность его формирования, получателей отчетов (его целевую аудиторию), а также параметры, которые можно задавать. Смысл использования изменяемых параметров заключается в том, что отчеты могут быть сформированы за различные периоды, по отдельным группам сотрудников и т. д.

Далее разрабатывается детальное описание отчетов. Описание отчетов включает: номер отчета;

название отчета; описание отчета (в том числе описание вида отображения данных); формат имени файла отчета; содержание отчета (набор используемых ключевых показателей эффективности); источник информации, используемый для формирования отчета; список получателей отчета, периодичность формирования отчета.

Последним шагом непосредственно перед реализацией отчетов в системе автоматизации является создание макетов отчетов.

Формирование и распространение отчетов, принятие на их основе решений и определение необходимых корректирующих действий происходит в соответствии с семиступенчатым процессом постоянного улучшения сервисов ИТ-службы, предлагаемым лучшими практиками ИТЛ.

На основе полученных отчетов руководство ИТ-службы может принимать различные решения, начиная от штрафных санкций либо поощрения сотрудников ИТ-службы и заканчивая оптимизацией и реинжинирингом процессов. Помимо этого, полученная отчетность может быть использована для обоснования ИТ-бюджета перед руководством компании.

Список литературы

1. **Basili V. R., Caldiera G., Rombach H. D.** The Goal Question Metric Paradigm. Encyclopedia of Software Engineering. Vol. 1. — John Wiley & Sons, 1994. P. 578—583.
2. **CobiT 4.1.** — М.: Аудит и контроль информационных систем. 2008. С. 1—14.
3. **ИТЛ® V3 Glossary Russian Translation v0.92.** — ITIL® R V3 Translation Project. 2009. 147 p.
4. **Гузик С. В.** Стандарт CobiT // Jet Info online: информационный бюллетень. 2003. № 1. С. 1—24.
5. **Легезо Д.** Третий пошел. ITIL v.3 // Intelligent enterprise. 2008. № 2.
6. **Мэйор Т.** Методологии оценки ИТ // Директор информационной службы. 2002. № 9. С. 56—59.
7. **Оценка работы процессов. Показатели и метрики.** URL: <http://www.itiportal.ru/rus/ITEMS/77-28/index.php>
8. **Пашанин И. В.** Сбалансированная система показателей как основа стратегического управления // Финансовая газета. 2007. № 11 (795) от 15 марта 2007.
9. **Каплан Р., Нортон Д.** Сбалансированная система показателей. От стратегии к действию. — М.: Олимп-Бизнес, 2006.
10. **Рудаков А. В.** Технология разработки программных продуктов. 4-е изд. М.: Академия, 2008.
11. **Ван Бон Я., Кеммерлинг Г., Пондман Д.** Введение в ИТ. Сервис-менеджмент / Пер. Потоцкий М. Ю. — М.: IT Expert, 2003. 225 с.

Ю. И. Димитриенко, д-р физ.-мат. наук, проф.,
О. Ю. Димитриенко, аспирант,
 МГТУ им. Н. Э. Баумана,
 e-mail: dimit@serv.bmstu.ru

Кластерно-континуальное моделирование в экономике на основе методов механики многомерных сплошных сред

Предложена модель рынка массовых продаж, основанная на кластеризации объектов в многомерном пространстве, а также на построении континуальной модели движения кластеров. Предложена система аксиом, описывающая движение кластеров в многомерном пространстве, с помощью которой выведены основные уравнения движения в рамках модели жестких кластеров. Представлены примеры численного расчета параметров меняющегося рынка продаж, проведено сравнение с экспериментальными данными.

Ключевые слова: кластеры, моделирование в экономике, многомерная механика сплошной среды, жесткий кластер, законы сохранения, численное моделирование

В настоящее время для моделирования экономических процессов активно развиваются методы нелинейного динамического анализа [1], которые в отличие от широко распространенных статистических методов основаны на более содержательных моделях экономических систем, учитывающих "физические" закономерности их развития и внутренние взаимосвязи составляющих элементов. Целью настоящей работы является дальнейшее развитие методов нелинейного динамического анализа для обработки информации в динамических экономических системах. Предложен новый подход к построению моделей экономических процессов, основанный на идее перехода от дискретного описания больших объемов данных, получаемых при анализе экономических систем, к континуальному описанию в специальном многомерном пространстве, а также на введении системы аксиом о движении многомерных континуумов в этом пространстве.

Обоснование модели

Одним из основных элементов экономических систем является рынок продаж, состоящий из покупателей товаров, продавца товаров и самих товаров. Если число покупателей весьма велико, то применение точных методов прогнозирования информации о покупаемых товарах становится

неэффективным из-за необходимости обработки больших объемов информации и необходимости получения достоверной информации о каждой покупке, что в реальных задачах, как правило, затруднено. В связи с этим представляет интерес применение методов континуального анализа, хорошо разработанных в механике сплошных сред [2], для создания математических моделей рынка массовых продаж. Как и классические сплошные среды, большие экономические системы типа рынка массовых продаж обладают достаточно выраженными закономерностями коллективного поведения покупателей, поэтому на рынке возникают кластеры покупателей, которые движутся согласованным образом в некотором пространстве товаров. Роль физических связей в таких кластерах покупателей выполняют информационные связи между отдельными покупателями и общие социальные закономерности поведения покупателей, проявляющиеся у покупателей одних и тех же социальных групп.

Многомерное евклидово пространство товаров

Пусть имеется продавец, продающий номенклатуру из n товаров, и большое число N покупателей ($N \gg 1$). Введем n -мерное евклидово пространство E_n товаров следующим образом: обозначим x^i — суммарное количество i -го товара, покупаемого или планируемого к покупке одним покупателем за время t , где $i = 1, \dots, n$, а $x^i \in \mathbf{R}$. "Планирование товара к покупке" позволяет перейти к вычислению количества товара в вещественных числах $x^i \in \mathbf{R}$. Совокупность значений $\mathbf{x} = \{x_1, \dots, x_n\}$ представляет собой элемент (точку) n -мерного пространства E_n , каждая точка $\mathbf{x}(t)$ из E_n трактуется как совокупность "количеств" товаров, покупаемых одним покупателем за время t . Операция сложения двух точек $\mathbf{x}_{(1)} = \{x_{(1)}^1, \dots, x_{(1)}^n\}$ и $\mathbf{x}_{(2)} = \{x_{(2)}^1, \dots, x_{(2)}^n\}$ из E_n — корректна: их сумма $\mathbf{x}_{(1)} + \mathbf{x}_{(2)} = \{x_{(1)}^1 + x_{(2)}^1, \dots, x_{(1)}^n + x_{(2)}^n\}$ есть совокупность "количеств" товаров двух покупателей. В E_n можно ввести ортонормированный декартов базис \mathbf{e}_i и декартову систему координат $O\mathbf{e}_i$, тогда радиус-вектор точки M в этой системе координат можно представить следующим образом: $\mathbf{x} = \vec{OM} = \sum_{i=1}^n x^i \mathbf{e}_i$. Скалярное произведение в E_n введем стандартным образом:

$$\mathbf{x}_{(1)} \cdot \mathbf{x}_{(2)} = \sum_{i=1}^n x_{(1)}^i x_{(2)}^i = x_{(1)}^i x_{(2)i}^i,$$

тогда E_n можно рассматривать как точечно-евклидово пространство (аксиома 1 — существование многомерного евклидова пространства товаров E_n).

Каждому отдельному покупателю в пространстве E_n в момент времени t соответствует некоторая точка \mathbf{x} , а совокупности из N покупателей — множество точек в этом пространстве, которое будем называть *кластером*. Поскольку в рассматриваемой модели предполагается, что число покупателей достаточно велико в каждом кластере, то можно перейти к пределу — для каждого кластера ввести континуальную область V в пространстве E_n . Таким образом, введем аксиому 2 — о существовании континуальных областей в пространстве товаров-покупателей E_n , которые также будем называть кластерами.

Введем лагранжево-эйлерово описание движения кластеров в E_n с помощью закона накопления покупок отдельным покупателем (это закон движения кластера в E_n): $\mathbf{x} = \mathbf{x}(X^i, t)$, где $\mathbf{x} = \{x^1, \dots, x^n\}$ — декартовы (эйлеровы) координаты точки; $X^i = X^i(x_i^0)$ — лагранжевы координаты покупателя, сохраняющиеся у покупателя во все время движения; x_i^0 — декартовы координаты покупателя в начальный момент времени. Здесь и далее $i = 1, \dots, n$.

Движение континуума V во времени в пространстве E_n характеризует градиент деформации \mathbf{F} , который связывает элементарные радиус-векторы $d\mathbf{x}^0$ и $d\mathbf{x}$, характеризующие локальные группы покупателей, в начальный момент времени и в момент времени t : $d\mathbf{x} = \mathbf{F} \cdot d\mathbf{x}^0$ (существование абсолютного времени — аксиома 3). Полагая функции $\mathbf{x} = \mathbf{x}(X^i, t)$ дифференцируемыми, стандартным образом [3] вводятся локальные векторы базиса $\mathbf{r}_i = \frac{\partial \mathbf{x}(X^i, t)}{\partial X^i}$ и вектор частоты покупок (вектор скорости) $\mathbf{v} = \frac{\partial \mathbf{x}(X^i, t)}{\partial t}$, метрическая матрица $g_{ij} = \mathbf{r}_i \cdot \mathbf{r}_j$, обратная метрическая матрица $g^{ij} g^{jk} = \delta_i^k$, взаимные векторы базиса $\mathbf{r}^i = g^{ij} \mathbf{r}_j$, символы Кристоффеля — $\frac{\partial \mathbf{r}_i}{\partial X^j} = \Gamma_{ij}^k \mathbf{r}_k$. Ковариантное дифференцирование в E_n вводим стандартным образом [3] — с помощью набла-оператора: $\nabla = \mathbf{r}^i \frac{\partial}{\partial X^i}$.

Для того чтобы ввести понятия объема, нормали, ориентированной площадки, векторного произведения, введем ориентацию пространства E_n и

для базисов одного класса рассмотрим в E_n полилинейные формы k -го порядка

$$\begin{aligned} \mathbf{P}(\mathbf{a}_1, \dots, \mathbf{a}_k) &= \\ &= \sqrt{g} e_{i_1 \dots i_n} a_1^{i_1}, \dots, a_k^{i_k} \mathbf{r}^{i_{k+1}} \otimes \dots \otimes \mathbf{r}^{i_n}, \end{aligned} \quad (1)$$

где $g = \det g_{ij}$, $a_1^{i_1} \dots a_k^{i_k}$ — компоненты векторов $\mathbf{a}_1 \dots \mathbf{a}_k$ в базисе \mathbf{r}_k ; \otimes — знак операции тензорного умножения [3]; $e_{i_1 \dots i_n}$ — n -мерные символы Леви—Чивиты [4]. Объем V параллелепипеда, построенного на векторах $\mathbf{a}_1 \dots \mathbf{a}_n$, определяется как полилинейная форма n -го порядка: $V = \mathbf{P}(\mathbf{a}_1 \dots \mathbf{a}_n) = \sqrt{g} e_{i_1 \dots i_n} a_1^{i_1} \dots a_n^{i_n}$. Ориентированная площадка определяется как полилинейная форма $(n-1)$ -го порядка:

$$n\Sigma = \mathbf{P}(\mathbf{a}_1 \dots \mathbf{a}_{n-1}) = \sqrt{g} e_{i_1 \dots i_n} a_1^{i_1} \dots a_{n-1}^{i_{n-1}} \mathbf{r}^{i_n},$$

т. е. это вектор, ортогональный к гиперплоскости S , представляющей собой линейную оболочку, натянутую на $\mathbf{a}_1 \dots \mathbf{a}_{n-1}$, здесь $\Sigma = |\mathbf{P}(\mathbf{a}_1 \dots \mathbf{a}_{n-1})|$ — длина этого вектора, а $\mathbf{n} = \mathbf{P}(\mathbf{a}_1 \dots \mathbf{a}_{n-1})/\Sigma$ — вектор нормали к гиперплоскости S .

Векторное произведение двух векторов $\mathbf{a}_1, \mathbf{a}_2$ из E_n определим как полилинейную форму 1-го порядка

$$\begin{aligned} \mathbf{a}_1 \times \mathbf{a}_2 &\equiv \mathbf{P}(\mathbf{a}_1 \mathbf{a}_2) = \\ &= \sqrt{g} e_{i_1 \dots i_n} a_1^{i_1} a_2^{i_2} \mathbf{r}^{i_3} \otimes \dots \otimes \mathbf{r}^{i_n}. \end{aligned} \quad (2)$$

С помощью тензора Леви—Чивиты

$${}^n \mathfrak{E} = \sqrt{g} e_{i_1 \dots i_n} \mathbf{r}^{i_1} \otimes \dots \otimes \mathbf{r}^{i_n}, \quad (3)$$

векторное произведение (2) можно представить как свертку тензора 2-го ранга $\mathbf{a}_2 \times \mathbf{a}_1$ с тензором ${}^n \mathfrak{E}$:

$$\mathbf{a}_1 \times \mathbf{a}_2 = (\mathbf{a}_2 \otimes \mathbf{a}_1) \cdot {}^n \mathfrak{E}, \quad (4)$$

(\cdot обозначает двойное скалярное произведение [3]), в чем можно убедиться непосредственно.

Введем элементарный объем dV на элементарных радиус-векторах $d\mathbf{x}_\alpha = \mathbf{r}_\alpha dX^\alpha$ (по греческим индексам суммирование нет), ориентированных по координатным линиям X^α , и ориентированную элементарную площадку:

$$dV = \mathbf{P}(d\mathbf{x}_1 \dots d\mathbf{x}_n) = \sqrt{g} dX^1 \dots dX^n; \quad (5)$$

$$nd\Sigma = \mathbf{P}(d\mathbf{x}_{i_1} \dots d\mathbf{x}_{i_{n-1}}) = \sqrt{g} e_{i_1 \dots i_n} \mathbf{r}^{i_n}.$$

Индивидуальным покупателем будем называть точку \mathbf{x} из E_n , а элементарный объем $dV = dx^1 dx^2 \dots dx^n$, построенный с помощью элементарного радиус-вектора $d\mathbf{x}$, будем называть ло-

кальной группой покупателей. В пространстве E_n для гладких в области V тензорных полей $\mathbf{A}(\mathbf{x}, t)$ имеет место обобщенная теорема Гаусса—Остроградского [12]

$$\int_V \nabla \cdot \mathbf{A} dV = \int_{\Sigma} \mathbf{n} \cdot \mathbf{A} d\Sigma.$$

Закон сохранения числа покупателей и закон изменения частоты покупок

Для кластера V , состоящего из одних и тех же покупателей, аксиоматически введем общее число покупателей в кластере (назовем его массой кластера) M и вектор силы \mathbf{f} взаимодействия рассматриваемого кластера с другими кластерами. Подобно трехмерной механике сплошных сред [2], сформулируем законы сохранения: закон сохранения числа покупателей в кластере и закон изменения частоты покупок в кластере {аксиомы 4 и 5}:

$$\frac{d}{dt} \int_V \rho dV = 0; \quad (6)$$

$$\frac{d}{dt} \int_V \rho \mathbf{v} dV = \int_V \rho \mathbf{f}_m dV + \int_{\Sigma} \mathbf{t}_{\Sigma} d\Sigma, \quad (7)$$

где $\rho = dM/dV$ — плотность кластера (число покупателей в локальной группе); $\mathbf{f}_m = d\mathbf{f}/dM$ — вектор плотности внешней массовой силы, изменяющей частоту покупок в локальной группе; $\mathbf{t}_{\Sigma} = d\mathbf{f}/d\Sigma$ — вектор поверхностных сил.

Закон изменения момента частоты покупок

Принципиально новым моментом развиваемой теории является введение в n -мерном пространстве обобщенной операции векторного произведения (2). Закон изменения момента частоты покупок (аксиома 6) с учетом операции (2) постулирует, что изменение вектора ${}^{n-2}\bar{\mathbf{m}}$ момента частот покупок кластера определяется моментом внешних сил ${}^{n-2}\boldsymbol{\mu}$:

$$\frac{d}{dt} ({}^{n-2}\bar{\mathbf{m}}) = {}^{n-2}\boldsymbol{\mu}, \quad (8)$$

где ${}^{n-2}\bar{\mathbf{m}}$ и ${}^{n-2}\boldsymbol{\mu}$ — тензоры $(n-2)$ -ранга:

$${}^{n-2}\bar{\mathbf{m}} = \int_V \rho \mathbf{x} \times \mathbf{v} dV; \quad (9)$$

$${}^{n-2}\boldsymbol{\mu} = {}^{n-2}\boldsymbol{\mu}_m + {}^{n-2}\boldsymbol{\mu}_{\Sigma}, \quad (10)$$

$${}^{n-2}\boldsymbol{\mu}_m = \int_V \rho \mathbf{x} \times \mathbf{f}_m dV; \quad {}^{n-2}\boldsymbol{\mu}_{\Sigma} = \int_{\Sigma} \mathbf{x} \times \mathbf{t}_{\Sigma} d\Sigma. \quad (11)$$

Модель жесткого кластера

Рассмотрим модель жесткого кластера в пространстве E_n , расстояния между отдельными точками которого полагаются постоянными во все

время движения. Закон накопления покупок (закон движения) жесткого кластера имеет вид

$$\dot{x}^j = x^{0j}(t) + Q_j^i(t)X^i,$$

где X^i — Лагранжевы координаты; $x^{0j}(t)$ — координаты вектора поступательного движения центра масс, $Q_j^i(t)$ — компоненты ортогональной матрицы поворота.

В векторном представлении закон движения будет переписан в следующем виде:

$$\mathbf{x} = \mathbf{x}^0(t) + \mathbf{Q} \cdot \tilde{\mathbf{x}}', \quad (12)$$

где $\mathbf{Q} = Q^{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ — тензор поворота; $\tilde{\mathbf{x}}' = X^i \mathbf{e}_i$ — локальный радиус-вектор точек; $\mathbf{e}_i' = \mathbf{Q} \cdot \mathbf{e}_i$ — подвижный базис, который движется вместе с телом.

Дифференцируя (12) по t , получаем обобщение на n -мерный случай классической формулы Эйлера:

$$\mathbf{v} = \mathbf{v}^0 + \tilde{\mathbf{x}} \cdot \mathbf{W}. \quad (13)$$

Здесь обозначен относительный радиус-вектор $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}^0$ в подвижной системе отсчета, введена скорость движения центра вращения кластера $\mathbf{v}^0 = \dot{\mathbf{x}}^0$ и кососимметричный тензор вращения тела

$$\mathbf{W} = \mathbf{Q} \cdot \mathbf{Q}^T. \quad (14)$$

Если подвижная система отсчета $M_0 \mathbf{e}_i'$ выбрана таким образом, чтобы центр вращения M_0 совпадал с центром тяжести (его будем называть типовым покупателем кластера): $\mathbf{x}^0 = \frac{1}{M} \int_V \rho \mathbf{x} dV$, то

средняя скорость тела совпадает с $\bar{\mathbf{v}} = \mathbf{v}^0$: $\bar{\mathbf{v}} = \frac{1}{M} \int_V \rho \mathbf{v} dV = \mathbf{v}^0$.

Подставляя формулу (13) в формулу (7), с учетом того, что $\frac{1}{M} \int_V \rho \tilde{\mathbf{x}} dV = \frac{1}{M} \int_V \rho \mathbf{x} dV - \mathbf{x}^0 = 0$, получаем следующее уравнение, описывающее движение центра масс (центроида) кластера:

$$M \frac{d\mathbf{v}^0}{dt} = \mathbf{f}; \quad (15)$$

$$\mathbf{f} = \int_V \rho \mathbf{f}_m dV + \int_{\Sigma} \mathbf{t}_{\Sigma} d\Sigma. \quad (16)$$

Подставим формулу (13) в выражение (9) для тензора моментов

$$\begin{aligned} {}^{n-2}\bar{\mathbf{m}} &= \int_V \rho \mathbf{x} \times \mathbf{v} dV = \int_V \rho \mathbf{x} \times \mathbf{v}^0 dV + \\ &+ \int_V \rho \mathbf{x} \times (\tilde{\mathbf{x}} \cdot \mathbf{W}) dV = \mathbf{x}^0 \times \mathbf{v}^0 + \\ &+ \int_V (\mathbf{W}^T \cdot (\tilde{\mathbf{x}} \otimes \mathbf{x}^0)) \cdot {}^{n-2}\boldsymbol{\mu} dV + \int_V (\mathbf{W}^T \cdot (\tilde{\mathbf{x}} \otimes \tilde{\mathbf{x}})) \cdot {}^{n-2}\boldsymbol{\mu} dV, \quad (17) \end{aligned}$$

здесь использована формула (4). Поскольку \mathbf{x}^0 и \mathbf{W} не зависят от координат, то можно привести формулу (17) к окончательному виду:

$$n-2\bar{\mathbf{m}} = M\mathbf{x}^0 \times \mathbf{v}^0 + (\mathbf{W}^T \cdot \mathbf{I}) \cdot n\mathfrak{A}, \quad (18)$$

где $\mathbf{I} = \int_V \rho \tilde{\mathbf{x}} \otimes \tilde{\mathbf{x}} dV = I^{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ — тензор моментов инерции многомерного тела.

Подставляя выражение (18) в уравнение моментов (8), получаем следующее дифференциальное уравнение для тензора вращения \mathbf{W} :

$$\frac{d}{dt} (\mathbf{W}^T \cdot \mathbf{I}) \cdot n\mathfrak{A} = n-2\tilde{\boldsymbol{\mu}}; \quad (19)$$

$$n-2\tilde{\boldsymbol{\mu}} = \int_V \rho \tilde{\mathbf{x}} \times \mathbf{f}_m dV + \int_{\Sigma} \tilde{\mathbf{x}} \times \mathbf{t}_{\Sigma} d\Sigma. \quad (20)$$

Уравнения движения жесткого тела в подвижном базисе

Представим тензоры \mathbf{W} и \mathbf{I} в подвижном базисе \mathbf{e}'_i : $\mathbf{W} = \mathbf{W}^{ij} \mathbf{e}'_i \otimes \mathbf{e}'_j$ и $\mathbf{I} = I^{ij} \mathbf{e}'_i \otimes \mathbf{e}'_j$. Скорость подвижного базиса определяется тензором вращения: $\dot{\mathbf{e}}'_i = \dot{\mathbf{Q}} \cdot \mathbf{e}_i = \dot{\mathbf{Q}} \cdot \mathbf{Q}^T \cdot \mathbf{e}'_i = -\mathbf{W} \cdot \mathbf{e}'_i$. Скорость изменения тензоров 2-го ранга можно представить с помощью производной Яуманна $\mathbf{I}^J = \frac{dI^{ij}}{dt} \mathbf{e}'_i \otimes \mathbf{e}'_j$ [2]:

$$\begin{aligned} \frac{d}{dt} \mathbf{I} &= \mathbf{I}^J - \mathbf{W} \cdot \mathbf{I} + \mathbf{I} \cdot \mathbf{W}^T; \quad \frac{d}{dt} \mathbf{W}^T = \\ &= \mathbf{W}^{TJ} - \mathbf{W} \cdot \mathbf{W}^T + \mathbf{W}^T \cdot \mathbf{W}, \end{aligned} \quad (21)$$

$$\frac{d}{dt} (\mathbf{W}^T \cdot \mathbf{I}) = \mathbf{W}^{TJ} \cdot \mathbf{I} - \mathbf{W} \cdot \mathbf{W}^T \cdot \mathbf{I} + \mathbf{W}^T \cdot \mathbf{I} \cdot \mathbf{W}, \quad (22)$$

поскольку $\mathbf{I}^J = 0$. Подставляя формулу (22) в формулу (19), получаем

$$(\mathbf{W}^{TJ} \cdot \mathbf{I} - \mathbf{W} \cdot \mathbf{W}^T \cdot \mathbf{I}) \cdot n\mathfrak{A} = n-2\tilde{\boldsymbol{\mu}}. \quad (23)$$

Здесь учтено, что тензор $\mathbf{W}^T \cdot \mathbf{I} \cdot \mathbf{W}$ симметричен и, следовательно, $(\mathbf{W}^T \cdot \mathbf{I} \cdot \mathbf{W}) \cdot n\mathfrak{A} = 0$. Умножая уравнение (23) скалярно на транспонированный тензор Леви—Чивиты $n_{\mathfrak{A}}(n \dots 1) = \frac{1}{\sqrt{g}} e^{i_1 \dots i_n} \mathbf{r}_{i_n} \otimes \dots \otimes \mathbf{r}_{i_1}$ и учитывая, что $n_{\mathfrak{A}} \cdot \dots \cdot n_{\mathfrak{A}} = k_{(n-2)} (\Delta_{II} - \Delta_{III})$, где $\Delta_{II} = \mathbf{r}^{i_1} \otimes \mathbf{r}^{i_2} \otimes \mathbf{r}_{i_1} \otimes \mathbf{r}_{i_2}$ и $\Delta_{III} = \mathbf{e}^{i_1} \otimes \mathbf{e}^{i_2} \otimes \mathbf{e}_{i_2} \otimes \mathbf{e}_{i_1}$ — второй и третий единичные тензоры 4-го ранга [2], $k_{(n-2)} = (-1)^{|n \dots 1|} (n-2)!$, $|n \dots 1|$ — число инверсий перестановки, получаем

$$(\mathbf{W}^{TJ} \cdot \mathbf{I} - \mathbf{W} \cdot \mathbf{W}^T \cdot \mathbf{I}) \cdot \cdot (\Delta_{II} - \Delta_{III}) = \tilde{\boldsymbol{\mu}}, \quad (24)$$

Здесь обозначен кососимметричный тензор 2-го ранга

$$\begin{aligned} \tilde{\boldsymbol{\mu}} &= \frac{1}{k_{(n-2)}} n-2 \tilde{\boldsymbol{\mu}} \cdot \dots \cdot n_{\mathfrak{A}} = \\ &= \left(\int_V \rho (\tilde{\mathbf{x}} \otimes \mathbf{f}_m - \mathbf{f}_m \otimes \tilde{\mathbf{x}}) dV + \right. \\ &\quad \left. + \int_{\Sigma} (\tilde{\mathbf{x}} \otimes \mathbf{t}_{\Sigma} - \mathbf{t}_{\Sigma} \otimes \tilde{\mathbf{x}}) d\Sigma \right). \end{aligned}$$

Принимая во внимание свойства единичных тензоров 4-го ранга [3]: $\mathbf{A} \cdot \cdot \Delta_{II} = \mathbf{A}^T$, $\mathbf{A} \cdot \cdot \Delta_{III} = \mathbf{A}$, уравнение (24) записываем в окончательном виде

$$\mathbf{I} \cdot \mathbf{W}^J + \mathbf{W}^J \cdot \mathbf{I} = \mathbf{I} \cdot \mathbf{W} \cdot \mathbf{W}^T - \mathbf{W} \cdot \mathbf{W}^T \cdot \mathbf{I} + \tilde{\boldsymbol{\mu}}. \quad (25)$$

Тензорное уравнение (25) является обобщением на n -мерный случай динамических уравнений Эйлера. В компонентах в подвижном базисе это уравнение записывается так:

$$\begin{aligned} (I_k^i \delta_l^j + I_k^j \delta_l^i) \frac{dW^{kl}}{dt} = \\ = (I_k^i \delta_s^j - I_s^j \delta_k^i) W^{kl} W_l^s + \tilde{\mu}^{ij}. \end{aligned} \quad (26)$$

Если подвижный базис \mathbf{e}'_i выбрать совпадающим с собственным базисом \mathbf{e}_i тензора моментов инерции \mathbf{I} , то матрица компонент I_k^i тензора моментов инерции в этом базисе диагональная: $I_{\alpha}^i = I_{\alpha} \delta_{\alpha}^i$, где $I_{\alpha} = \text{const} > 0$, $\alpha = 1 \dots n$ — собственные значения, являющиеся вещественными и положительными.

Тогда уравнение (26) примет вид

$$\frac{dW^{\alpha\beta}}{dt} = \frac{I_{\alpha} - I_{\beta}}{I_{\alpha} + I_{\beta}} \sum_{k=1}^n W^{\alpha k} W^{\beta k} + \frac{\tilde{\mu}^{\alpha\beta}}{I_{\alpha} + I_{\beta}}. \quad (27)$$

В таком виде это уравнение известно как уравнение Эйлера—Арнольда [5—7].

Из формулы (13) следуют обобщенные уравнения Пуассона для определения тензора \mathbf{Q} :

$$\dot{\mathbf{Q}} + \mathbf{W} \cdot \mathbf{Q} = 0. \quad (28)$$

Общая система уравнений движения жесткого кластера (15), (27), (28) в компонентном виде вместе с начальными условиями представляет собой задачу Коши и записывается следующим образом:

$$M \frac{dv_{0i}}{dt} = \bar{f}_i;$$

$$\frac{dx_{0i}}{dt} = v_{0i};$$

$$\frac{dW^{\alpha\beta}}{dt} = \frac{I_{\alpha} - I_{\beta}}{I_{\alpha} + I_{\beta}} \sum_{k=1}^n W^{\alpha k} W^{\beta k} + \frac{\tilde{\mu}^{\alpha\beta}}{I_{\alpha} + I_{\beta}}; \quad (29)$$

$$\frac{dQ^{ji}}{dt} + Q_k^j W^{ki} = 0;$$

$$t = 0: x_{0i} = x_{0i}^0, v_{0i} = v_{0i}^0,$$

$$Q^{ji} = Q_0^{ji}, W^{ij} = W_0^{ij}.$$

Модель внешних сил

Плотность внешней силы f_m характеризует воздействие на покупателя внешних по отношению к нему экономических факторов. К таким факторам можно отнести явления, которые вызывают изменение частоты покупки по сравнению с некоторым "стабильным" процессом, обусловленным "естественными" потребностями покупателя, это, прежде всего:

1) маркетинговые мероприятия продавца по отношению к покупателям [8];

2) внешние кризисные экономические процессы, сопровождающиеся нестабильностью процессов покупок.

В задачах экономики поверхностные силы, как правило, играют второстепенную роль по сравнению с массовыми силами, поэтому положим $t_\Sigma = 0$.

Будем говорить, что рассматривается *поток-вая модель* покупок товаров покупателями кластера, если плотность внешней массовой силы f_m , вызывающей изменение частоты покупок, определяется суммарной скоростью изменения частоты приобретения товара покупателем "в запас" и для непосредственного потребления:

$$f_m = \frac{dh^s}{dt}, \quad (30)$$

где h^s — вектор, называемый *вектором частот приобретения товара покупателем* на склад и для непосредственного потребления (производства). Будем говорить, что реализуется *стабильный процесс покупок товаров* покупателями, если вектор h^s является постоянным, не изменяющимся во времени: $h^s = h_0^s = \text{const}$. Для стабильного процесса покупок внешние силы отсутствуют ($f_m = 0$) и покупки осуществляются покупателями "по инерции", т. е. под действием начальных данных, обусловленных "стабильными потребностями в потреблении и запасах".

Из уравнения движения жесткого кластера (15) следует, что

$$v_0 = h_0 = \text{const}; x^0 = x_0^0 + th_0 \quad (31)$$

— средний вектор частот покупок товара в кластере в стабильном процессе является постоянным, а изменение количества покупаемого товара

"типичным" покупателем кластера представляет собой равномерное прямолинейное движение в пространстве товаров E_n .

Будем говорить, что задана модель *нестабильного процесса покупок*, если h^s не является постоянным и определяется изменением только финансового запаса покупателей e (аксиоматически вводимая величина)

$$h^s = h_0^s (1 + A(e - e_0)^m), \quad (32)$$

где A, m — константы модели. Финансовый запас $e(t)$ кластера в данной модели будем считать известной функцией времени, резкие изменения финансового запаса $e(t)$ по отношению к некоторому начальному значению e_0 будем называть кризисными явлениями.

Численные алгоритмы

Для применения разработанной модели в целях анализа экспериментальных данных рынка продаж были разработаны четыре группы численных алгоритмов. Первая группа включает в себя алгоритм обобщающей иерархической кластеризации. В качестве второй группы алгоритмов были предложены четыре способа вычисления многомерных интегралов в матрице инерции

$$I^{ij} = \int_V \rho \tilde{x}^i \tilde{x}^j dV, \text{ различающиеся методом задания}$$

распределения плотности $\rho(\tilde{x}^j)$ в V :

1) плотность ρ считалась равной $\rho_k = \frac{\Delta m_k}{\Delta \bar{V}}$, где

ρ_k — плотность в малой окрестности k ; Δm_k — масса точек в этой малой окрестности; $\Delta \bar{V}$ — средний объем в кластере, который вычислялся как сумма объемов n -мерных шаров радиуса, равного половине расстояния до ближайшей точки кластера ("метод n -мерных шаров");

2) все точки имеют равную массу, независимо от их удаленности от центра масс и других точек кластера ("метод равного веса точек");

3) нормальный закон распределения плотности в зависимости от удаленности центра k -й области от центра масс кластера: $\rho_k = \exp\left(-\frac{r_k}{\sigma}\right)$ ("метод нормального распределения"), где r_k — расстояние от центра k -й области до центроида; \bar{r}_k — среднее расстояние до центроида; σ — отклонение расстояний до центроида

$$\sigma = \sqrt{\frac{(r_k - \bar{r}_k)^2}{N_k - 1}};$$

4) "метод локальных групп" основан на том же принципе, что и метод нормального распределения, однако предполагает, что в кластере могут существовать локальные скопления точек и, чтобы учесть эти скопления более точно, метод задает областям разную плотность в зависимости от расстояния до ближайшего соседа. После вычисления компонент матрицы инерции I^{ij} с использованием метода Якоби проводился поиск собственных значений I_α и собственных векторов $\vec{e}_i^0(0) = Q_i^j(0)\vec{e}_j$, и тем самым $Q_i^j(0)$ — компонент матрицы поворота.

Третья группа включает алгоритмы аппроксимации кластеров n -мерными эллипсоидами с помощью метода условной оптимизации, минимизирующего объем эллипсоида $V_n(r) = (\prod_{i=1}^n r^i) V_n \rightarrow \min$. Для проверки попадания точек внутрь эллипсоида (из условия вхождения 95 % точек кластера в эллипсоид) были использованы полученные ранее значения компонент $Q_i^j(0)$ матрицы поворота.

В четвертой группе алгоритмов были использованы явные и неявные разностные схемы численного решения задачи Коши (29) вместе с процедурой линеаризации на временных шагах. Неявная разностная схема для уравнений Эйлера—Арнольда имела следующий вид:

$$\frac{W_s^{\alpha\beta} - W_{s-1}^{\alpha\beta}}{\Delta t} = \frac{I_\alpha - I_\beta}{(I_\alpha + I_\beta)} \left(\sum_{\substack{k=1 \\ k \neq \beta}}^n W_{s-1}^{\alpha k} W_{s-1}^{\beta k} + W_s^{\alpha\beta} W_{s-1}^{\beta\beta} \right), \quad (33)$$

где $W_s^{\alpha\beta} = W^{\alpha\beta}(t_s)$ — значения функций в узлах разностной сетки.

Результаты численного моделирования

Разработанные модели и вычислительные алгоритмы были применены для анализа экспериментальных данных по продажам автомобилей через Интернет-магазин. Исследования проводились в течение 30 месяцев для пяти ценовых классов автомобилей, поэтому в данной задаче пространство товаров E_n было пятимерным. Данные анализировались через каждые три месяца в моменты времени $t_m, m = 1 \dots 11$. Моменты времени t_4 и t_6 были идентифицированы как первая и вто-

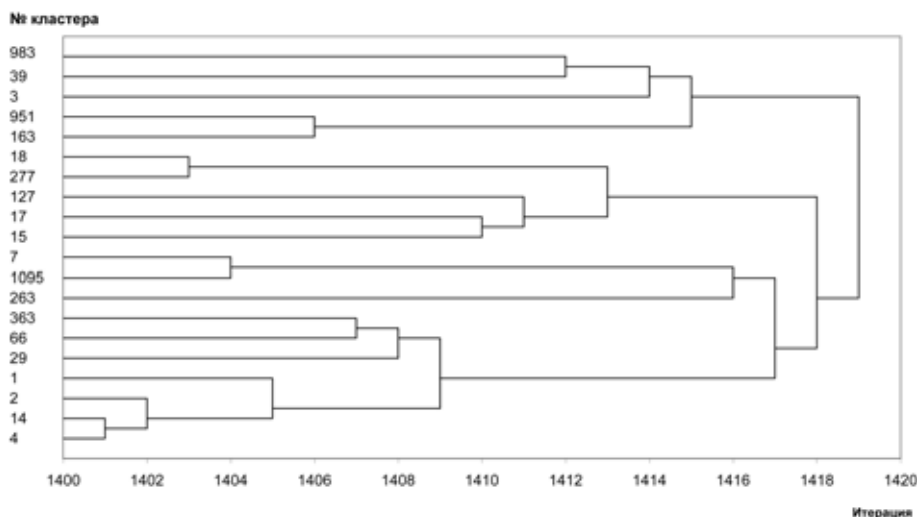


Рис. 1. Дендрограмма кластеризации

рая волна финансового кризиса, который отражался на изменении экономического запаса покупателей согласно модели (32).

Экспериментальные данные по покупкам в каждый момент времени t_m были "континуализованы" в соответствии с желанием покупателей купить автомобиль отношением стоимости автомобиля к ежемесячному доходу и стоимости его предыдущей машины. Новое значение для ненулевой покупки товара вычислялось по следующей формуле:

$$x_i^{(j)} = \min\left(\max\left(\alpha_1 \hat{x}_i^{(j)} - \frac{\alpha_2}{10} - \frac{\alpha_3}{10}; 0\right); 1,8\right),$$

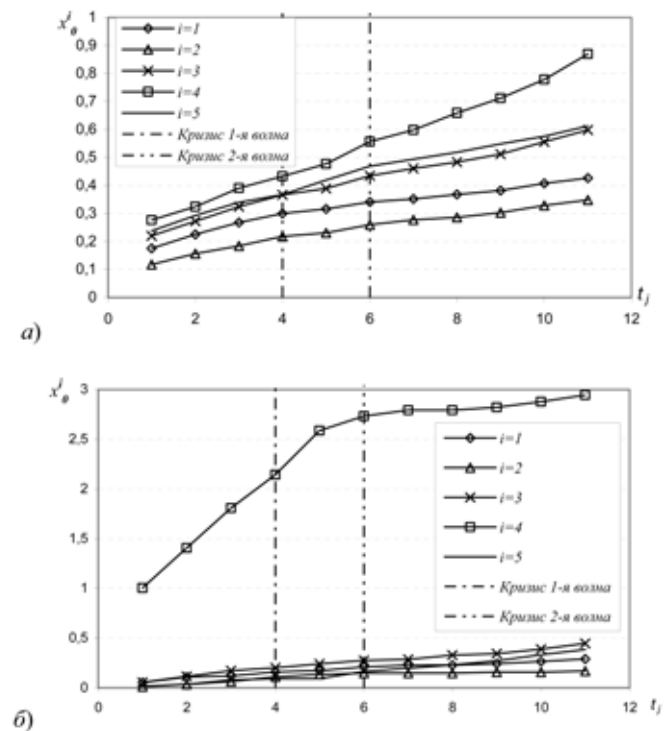


Рис. 2. Динамика центра масс кластеров 1 (а) и 2 (б)

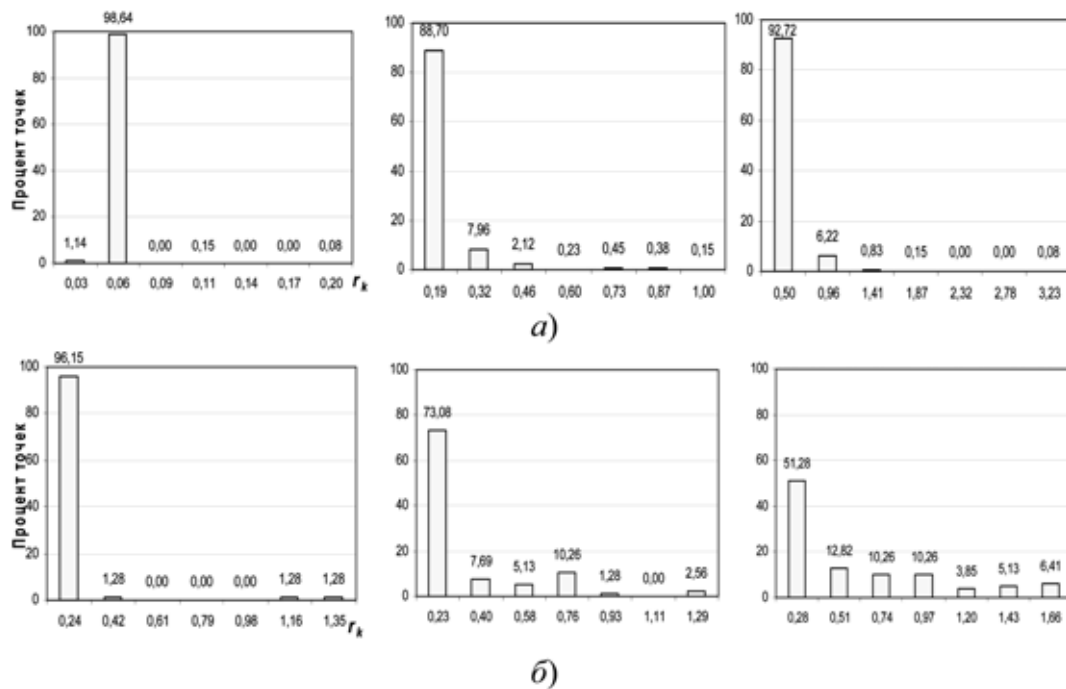


Рис. 3. Распределение точек кластеров 1 (а) и 2 (б) по удаленности от центра масс в моменты t_1 , t_5 и t_{11}

где $\hat{x}_i^{(j)}$ и $x_i^{(j)}$ — числа покупок i -го товара j -м покупателем до и после "континуализации"; $\alpha_1, \alpha_2, \alpha_3$ — эмпирические коэффициенты. Кластеризация данных была проведена для начального момента времени, в качестве которого был выбран момент t_5 .

Дендрограмма этой кластеризации представлена на рис. 1. Для момента t_5 1420 точек (покупателей) всего множества были объединены в пять кластеров, из которых для дальнейшего анализа были взяты два больших кластера, объединяющих в пространстве E_5 1319 и 78 покупателей соответственно.

Исследование движения центров масс x_0^j показало, что оно носит линейный характер по всем осям товаров в пространстве E_5 (рис. 2), без воздействия внешних сил, как это и следует из разработанной модели (29). Наличие внешних воздействий (кризисных явлений) приводит к изменению углов наклона линейной траектории движения кластеров, что также полностью соответствует предложенной модели.

Сохранение единства найденных кластеров во времени можно охарактеризовать, например, динамикой распределения удаленности точек r_k от центра кластера x_0 . Такие распределения для кластеров 1 и 2 представлены на рис. 3. Кластер 1, содержащий больше точек, демонстрирует большее единство и меньшее удаление точек от центра масс x_0 , чем кластер 2 с меньшим числом точек.

Вращательная динамика кластеров, представленная значениями независимых компонент $Q_i^j(t)$ матрицы поворота (наддиагональные элементы), рассчитанных по экспериментальным значениям с помощью матрицы инерции $I_i^j(t)$, без решения системы (29) для всех моментов вре-

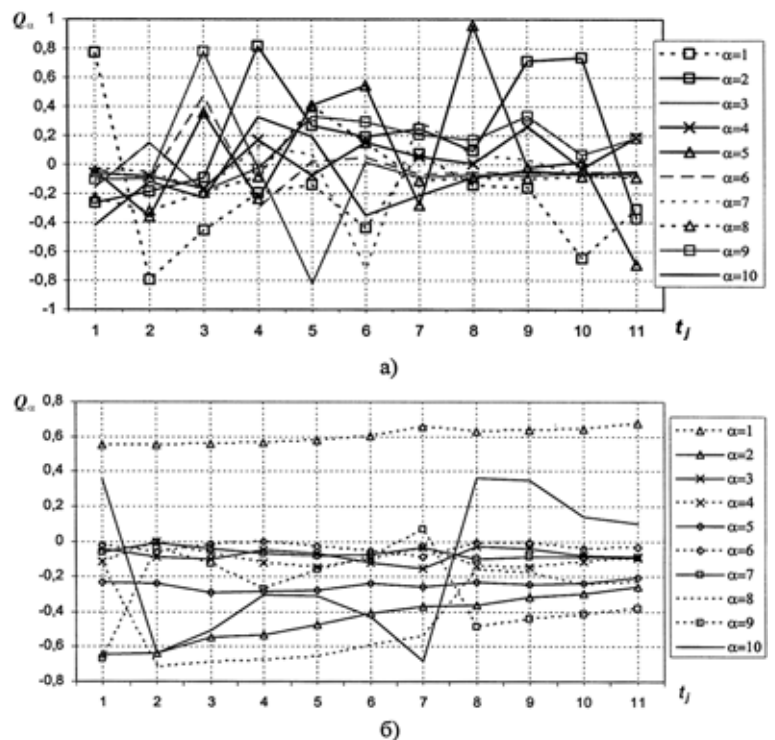


Рис. 4. Изменения компонент матрицы поворота $Q_i^j = Q_\alpha$, $\alpha = 1...10$ кластера 1, полученные поточечной аппроксимацией с использованием метода n -мерных шаров (а) и метода равного веса (б)

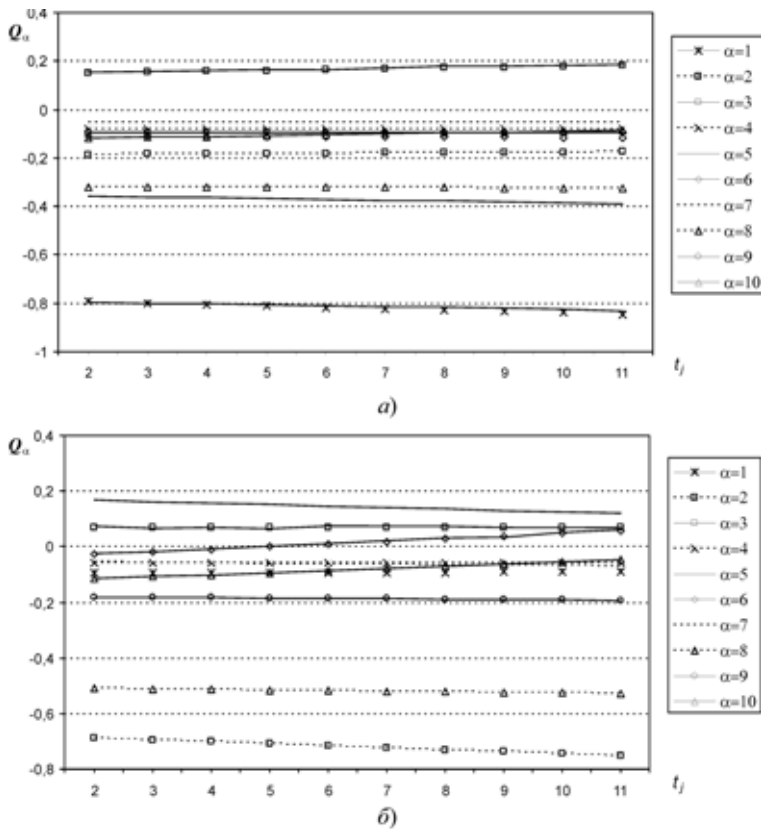


Рис. 5. Изменения компонент матрицы поворота $Q^{ij} = Q_\alpha$ кластеров 1 (а) и 2 (б), рассчитанные по модели жесткого кластера с использованием метода n -мерных шаров

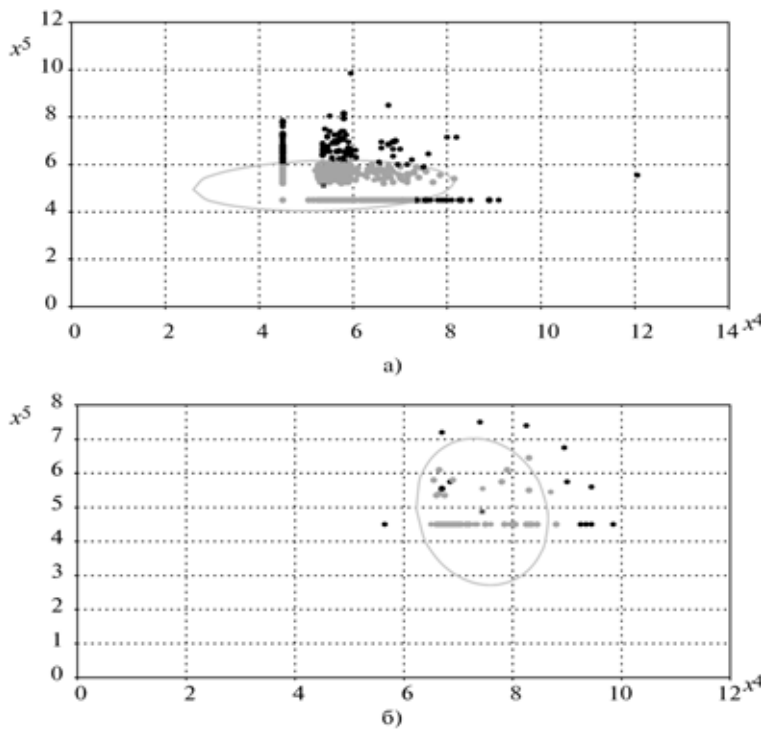


Рис. 6. Сечение n -мерного эллипсоида, построенного методом n -мерных шаров на основе модели жесткого кластера для покупателей кластеров 1 (а) и 2 (б) при t_{11}

мени показана на рис. 4 для кластера 1. Используются следующие обозначения: $Q^{12} = Q_1$, $Q^{13} = Q_2$, $Q^{14} = Q_3$, $Q^{15} = Q_4$, $Q^{23} = Q_5$, $Q^{24} = Q_6$, $Q^{25} = Q_7$, $Q^{34} = Q_8$, $Q^{35} = Q_9$, $Q^{45} = Q_{10}$. Анализируя полученные значения компонент $Q_i^j(t)$ матрицы поворота, можно сказать, что такая прямая аппроксимация экспериментального движения кластеров приводит к хаотичному, нестабильному его движению. Поэтому для получения стабильных результатов, которые могут стать основой прогноза движения кластеров, была далее использована разработанная модель жесткого кластера, основанная на решении системы (29).

Результаты расчетов динамики вращения с использованием в качестве начальных данных матриц $Q(0)$ и $I(0)$, рассчитанных методом n -мерных шаров, показаны на рис. 5. Было установлено, что изменение независимых компонент матрицы поворота $Q(t)$ подчиняется линейному закону.

Рис. 6 показывает сечение n -мерного эллипсоида для момента t_{11} , построенного по модели жесткого кластера для кластеров 1 и 2 с использованием метода n -мерных шаров для расчета инерции. Светлым на рисунках обозначены точки, попадающие внутрь моделируемого эллипсоида, а черным — точки, не попадающие в него. На рис. 7 приведены графики, показывающие процент попадающих точек в жесткий n -мерный эллипсоид для разных моментов времени и разных способов аппроксимации начальных данных при расчетах $Q_i^j(0)$ и $I_i^j(0)$.

Пунктирными линиями приведены данные, полученные по модели жестких кластеров, а сплошными — с помощью аппроксимации экспериментальных данных эллипсоидом для каждого момента времени, без решения системы (29). Кластер 1 с большим числом точек показывает большее расхождение в точности методов аппроксимации, однако само абсолютное значение точности модели для кластера 1 выше, и максимальная точность составляет 85 % для метода жесткого кластера в последней контрольной точке, в то время как для меньшего кластера 2 максимальная точность составляет 82 %. Резкие скачки точности модели наблюдаются в контрольных точках, для которых свойственно уменьшение длин полуосей эллипсоида — тогда при фиксированном объеме в эллипсоид попадает больше точек. Такую закономерность можно наблюдать в точке t_5 для обоих кластеров. Наибольшую точность да-

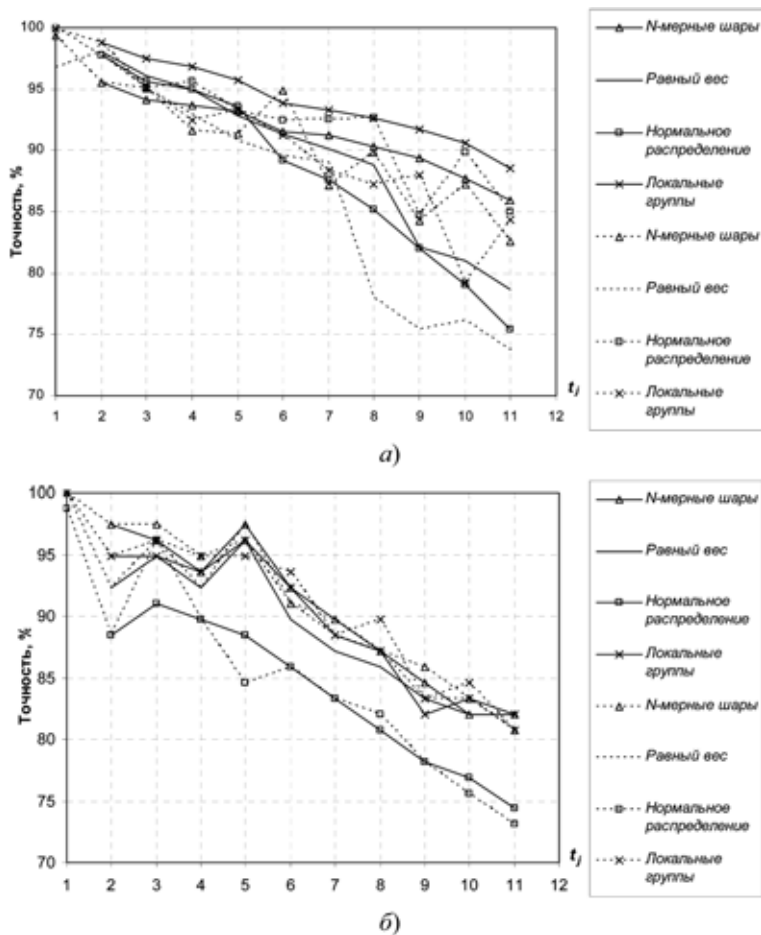


Рис. 7. Изменение точности модели для кластеров 1 (а) и 2 (б)

ют методы расчета матрицы инерции, учитывающие геометрию точек — метод n -мерных шаров и метод локальных групп. Метод нормального распределения, задающий плотность точек в окрестности V , работает лучше на кластерах с большим числом точек: точность на более многочисленном первом кластере больше, чем на втором. Однако хорошие результаты этот метод дает только для расчетов по эмпирическим данным, расчеты же по модели жесткого кластера для этого метода дают наименьшую точность. Следует также отметить, что точность модели жесткого кластера не сильно расходитсся с точностью поточечной аппроксимации экспериментальных данных для каждого момента времени. Таким образом, можно сделать вывод, что модель жесткого кластера можно использовать для целей прогноза динамики кластеров покупателей.

Выводы

Разработана кластерно-континуальная модель рынка массовых продаж, включающая методику построения кластеров с большим числом покупа-

телей, переход к континуальному описанию и обобщенную модель движения континуумов в многомерном пространстве.

Разработана модель жестких кластеров покупателей, получены основные дифференциальные уравнения, описывающие динамику движения кластеров на рынке продаж. Построена модель внешних экономических воздействий на рынок продаж.

Разработаны вычислительные алгоритмы решения задач динамики жестких кластеров, основанные на модели n -мерных эллипсоидов.

Проведен численный анализ экспериментальных данных типового рынка продаж и осуществлена верификация разработанной модели, показавшая, что разработанная модель и вычислительные алгоритмы обеспечивают высокую точность описания динамики движения кластеров.

Разработанная модель открывает возможность создания эффективного математического "инструмента" для решения ряда практических задач в экономике:

- прогнозирования динамики развития рынков массовых продаж как в условиях стабильного функционирования, так и в условиях возникновения кризисных процессов;
- оценки влияния маркетинговых мероприятий, осуществляемых продавцом и производителем товаров на динамику рынка продаж;
- управления рынком массовых продаж путем построения оптимальных экономических и маркетинговых стратегий.

СПИСОК ЛИТЕРАТУРЫ

1. Петров А. А., Поспелов И. Г., Шанин А. А. Опыт математического моделирования в экономике. М.: Энергоатомиздат. 1996. 544 с.
2. Димитриенко Ю. И. Нелинейная механика сплошной среды. М.: Физматлит. 2009. 624 с.
3. Димитриенко Ю. И. Тензорное исчисление. М.: Высшая школа. 2001. 575 с.
4. Ефимов Н. В., Розендорн Э. Р. Линейная алгебра и многомерная геометрия. М.: Физматлит. 2004. 464 с.
5. Арнольд В. И. Гамильтоновость уравнений Эйлера динамики твердого тела и идеальной жидкости // УМН. 1969. Т. 24. № 3. С. 225—226.
6. Манаков С. В. Замечание об интегрировании уравнений Эйлера динамики n -мерного твердого тела // Функциональный анализ и его приложения. 1976. Т. 10. Вып. 4. С. 93—94.
7. Јовановић В. Some multidimensional integrable cases of non-holonomic rigid body dynamics // Regular and chaotic dynamics. 2003. V. 8. N 1. P. 125—132.
8. Димитриенко О. Ю. Мультиагентная модель для стратегического маркетингового планирования // Актуальные проблемы фундаментальных наук. Сб. тр. Третьей научно-метод. конф. аспирантов и молодых исследователей, февраль 2009. М.: НИИ РЛ МГТУ им. Н. Э. Баумана. 2009. С. 9—12.

А. А. Штрик, д-р техн. наук, проф., гл. науч. сотр.,
ГУ "Российский НИИ информационных
технологий и систем автоматизированного
проектирования",
e-mail: shtrik@land.ru

Анализ воздействия кризиса 2009 года на состояние и перспективы развития ИТ-рынка России

Анализируется воздействие кризиса 2009 года на состояние и перспективы развития ИТ-рынка России. Проводится оценка влияния кризиса 1998 года в России на развитие рынка компьютерного оборудования. Даны оценки ведущих экспертов по текущему состоянию российского ИТ-рынка до и в начале кризиса 2009 года. Представлен тренд изменения объема ИТ-рынка России в докризисный период 1 кв. 2008 — 1 кв. 2009 г., а также оценки воздействия кризиса на ИТ-рынок в 2009 году.

Ключевые слова: анализ влияния кризиса, информационные технологии, инфраструктура, ИТ-рынок, развитие ИТ-рынка, рынок компьютеров, состояние ИТ-рынка, тренды развития рынка, финансовый кризис

Для проведения прогнозов поведения российского ИТ-рынка в период действующего сейчас мирового кризиса полезно обратиться к анализу поведения этого рынка в ходе предыдущего кризиса 1998 года, даже несмотря на явно выраженную разницу между ними. Эта разница, по мнению многих ведущих мировых экспертов, состоит в том, что экономический кризис 1998 года в России был классическим бюджетным и валютным кризисом, тогда как кризис 2008 года начался глобальным кризисом частного сектора, который был спровоцирован чрезмерными заимствованиями частного сектора. В то время как в 1998 году финансовое положение российского госсектора было устойчивым, настоящий кризис сопровождается тремя глобальными и серьезными проблемами: условия внешней торговли, отток капитала и ужесточение условий внешних заимствований (http://www.rian.m/trend/economy_Russia_11022009).

Если в 1998 году по данным Всемирного Банка российская экономика сократилась на 8,8 и 9,1 % в третьем и четвертом квартале соответственно, то уже ко второму кварталу 1999 года экономический рост достиг 3,1 %, а к концу года — 12 %, продемонстрировав таким образом быстрое оживление. Это же относилось и к ИТ-рынку, который восстановился достаточно быстро (темпы роста ИТ-рынка практически за 1,5 года вернулись к

докризисным). На этом фоне последствия текущего кризиса, как прогнозируют эксперты, будут ощущаться гораздо дольше — по крайней мере до тех пор, пока не восстановится мировая экономика и на мировом рынке не начнут расти цены на основные статьи российского экспорта — на энергоресурсы и сырье.

Основная цель настоящей статьи состоит в оценке влияния происходящего экономического кризиса на российский ИТ-рынок по его состоянию на конец 2009 года путем сопоставления фактических показателей с показателями, спрогнозированными для его бескризисного развития. При этом, в порядке ретроспективы, сначала дадим оценку влияния кризиса 1998 года в России на развитие рынка компьютерного оборудования, а затем приведем оценки ведущих экспертов по текущему состоянию российского ИТ-рынка до и в начале кризиса 2009 года. Для сопоставления ожидаемых и фактических показателей ИТ-рынка представлен тренд изменения объема ИТ-рынка России без учета кризиса на период до 2015 года, на который наложены фактические показатели поведения этого рынка в докризисный и кризисный периоды до конца 2009 года. Поскольку статья готовилась еще в конце 2008 — начале 2009 годов, в ней присутствуют прогнозные оценки поведения ИТ-рынка на 2009 год. Их сравнение с фактическими значениями дает возможность оценить точность прогнозов.

1. Оценка влияния кризиса 1998 года в России на развитие рынка компьютерного оборудования

Для проведения количественного анализа влияния кризиса 1998 года была использована статистика авторитетной и профессиональной организации *European Information Technology Observatory* — ЕИТО (<http://www.eito.com>). Эта организация начиная с 1993 года выпускает ежегодный отчет под тем же названием, в котором содержится подробная статистическая информация о рынке ИКТ большинства европейских стран с подробной разбивкой на отдельные сектора. В этом отчете вплоть до 2002 года публиковались данные и о российском рынке ИКТ (после 2002 года эксперты ЕИТО сосредоточились только на странах ЕС, включая все вновь принятые страны). Однако для решения задачи анализа влияния кризиса 1998 года на российский ИТ-рынок этой статистической базы вполне достаточно.

Анализ влияния кризиса 1998 года в России был проведен на примере одного из самых мощных секторов ИТ-рынка — сектора компьютерно-

Таблица 1

Фактические значения показателей изменения российского рынка компьютерного оборудования в период 1992—2002 гг.

Год	Рынок компьютеров до и после кризиса, млн. евро	Годовое число компьютеров до и после кризиса, тыс. штук	Год	Темпы изменения рынка компьютеров до и после кризиса, %
1993	530	408	1994/93	82
1994	1076	642	1995/94	16
1995	1250	837	1996/95	25
1996	1569	1039	1997/96	29
1997	1980	1374	1998/97	-36
1998	2286	935	1999/98	-23
1999	1747	1174	2000/99	8
2000	1888	1309	2001/00	9
2001	2063	1462	2002/01	8
2002	2217	1633		

го оборудования. Для анализа были выбраны три главные характеристики:

- объем рынка компьютеров (млн евро);
- объем рынка компьютеров (единиц компьютеров);
- темпы ежегодного изменения объема рынка (%).

Для оценки влияния кризиса сначала строится диаграмма (график) фактического изменения соответствующего показателя в период, предшествующий наступлению кризиса, т. е. до середины 1998 года, а затем по этому графику формируется его тренд. Полученный тренд сравнивается с фактическим графиком изменения показателей в период после кризиса, т. е. начиная с середины 1998 года и далее. Соответствующие исходные данные по фактическому состоянию рынка компьютерного оборудования России в период 1993—2002 гг. приведены в табл. 1.

На рис. 1 показан график изменения фактических (т. е. измеренных) показателей объема рынка компьютерного оборудования в млн евро за период 1993—2002 годы, а также прогнозируемый тренд изменения этого показателя, который имел бы место в условиях отсутствия кризиса.

График показывает, что в следующий после кризиса год (1999) объем рынка сразу упал в 1,5 раза — вместо прогнози-

вавшегося объема 2,6 млрд евро было зафиксировано 1,75 млрд евро. В последующие 2000—2002 годы объем рынка компьютеров начал увеличиваться почти параллельно росту тренда рынка, т. е. темпы его роста приблизились к докризисным (изменение темпов см. ниже на рис. 3).

На рис. 2 показаны аналогичный график изменения фактических показателей объема рынка компьютерного оборудования — поступивших на рынок за год компьютеров (в тысячах штук) за период 1993—2002 годы, а также прогнозируемый тренд изменения этого показателя в условиях отсутствия кризиса.

Видно, что общий характер изменения фактических и прогнозных показателей аналогичен представленным на предыдущем рисунке.

График годовых темпов изменения объема рынка компьютерного оборудования в России за период 1993—2002 гг. и соответствующий тренд представлены на рис. 3 (см. третью сторону обложки).

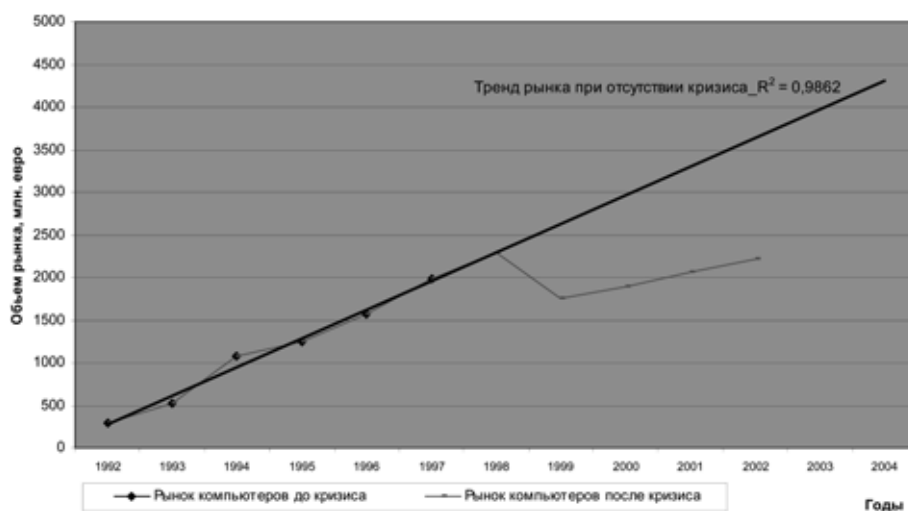


Рис. 1. Объем рынка компьютерного оборудования России, млн евро

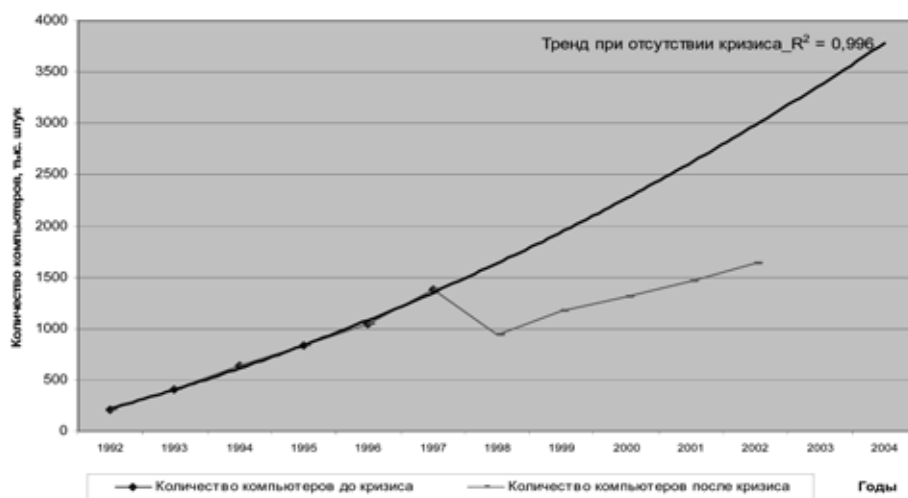


Рис. 2. Объем рынка компьютерного оборудования России, годовое поступление в тыс. штук

На период кризиса отмечено резкое падение годовых темпов изменения объема рынка (до минус 36 %), однако уже в 2000 году этот показатель практически вышел на прогнозируемый тренд, т. е. темпы развития рынка компьютеров вернулись на докризисный период времени. Таким образом, приведенные графики подтверждают вывод о том, что российский ИТ-рынок восстановился достаточно быстро: согласно графикам рис. 3 темпы роста ИТ-рынка за 1,5—2 года вернулись к докризисным значениям.

2. Оценки ведущих экспертов по текущему состоянию российского ИТ-рынка до и в начале кризиса 2009 года (1 кв. 2008 — 1 кв. 2009 гг.)

Исследования, связанные с кризисом, с оценкой текущего состояния и с различными прогнозами по ИТ-отрасли, в том числе и по России, часто появляются на страницах прессы, а также и в научной литературе. Приводимые ниже материалы содержат краткий обзор результатов исследований ряда наиболее авторитетных международных и отечественных организаций, которые в первую очередь характеризуют текущее состояние российского ИТ-рынка. Очевидно, что на полное совпадение результатов, представленных различными группами исследователей, рассчитывать нельзя, но в то же время наличие радикальных расхождений может свидетельствовать либо о наличии методологических погрешностей при проведении исследований, либо о каких-либо иных причинах, которые следовало бы локализовать и объяснить.

При анализе текущего состояния российского ИТ-рынка в докризисный и начальный период кризиса были использованы данные, представленные различными организациями, которые проводят мониторинг или исследование ИТ-рынка и его отдельных сегментов: IDC, Forrester,

ITResearch, eMarketer, PMR, Gartner, Минкомсвязь России, Минэкономразвития России, ЛИНЭКС, OCS и ряд других. При этом охвачен диапазон фактических данных, отражающих состояние ИТ-рынка или его секторов в интервале от 2001 до 2008 года, а также прогнозы их изменения в основном на период 2009—2010 гг. (Напомним, что статья готовилась в конце 2008 — начале 2009 гг.).

Основные данные, полученные экспертами из семи различных организаций и компаний [1—12], показаны в табл. 2, где содержатся количественные оценки фактических значений объемов российского ИТ-рынка на 2007—2008 гг., а также их прогнозы на 2009 год.

Сопоставление конкретных цифр, содержащихся в проанализированных источниках, а также в табл. 2, позволяет утверждать, что оценки, приводимые официальными ведомствами (Минкомсвязь России и Минэкономразвития России), как правило, существенно отличаются от оценок, приводимых международными экспертными компаниями (IDC, ITResearch), а также от оценок, полученных независимыми российскими организациями (ЛИНЭКС, OCS). В частности:

- оценки, приводимые компанией IDC в разные моменты претерпевают значительное изменение — для 2007 года изменение достигает 37 %; у компании PMR в 2008 году расхождение составило 16 %, прогноз для 2009 года — 5 %; в то же время у компании ITResearch оценки практически совпадают;
- максимальное отклонение от среднего значения составляет 20 % (оценка IDC в \$ 16 млрд для 2007 года); в основном отклонение составляет 10—12 %;
- отклонение оценок Минэкономразвития России от среднего значения составляет около 250 % (для 2007 года, [3]); прогноз Минкомсвязи России для 2009 имеет отклонение 165 % [11].

Таблица 2

Оценки объема российского ИТ-рынка от разных разработчиков

Источник	Разработчик статистики	Объем ИТ-рынка 2007, млрд	Объем ИТ-рынка 2008, млрд	Объем ИТ-рынка 2009, млрд
1	Минэкономразвития России	467 руб	580 руб	
2,4	IDC	\$16		
9	IDC	\$18,2-21,8		
12	IDC	\$22	\$24	\$13,2
3	Минэкономразвития России	1185 руб		330 руб
4	Минкомсвязь России	\$17,6		
8,11	Минкомсвязь России	450 руб	580 руб	770 руб
5	PMR	492,9 руб	462,3 руб	487,4 руб
12	PMR	492,8 руб	538,6 руб	463,2 руб
6	Cartner		\$17,23	\$17,69-18,21
7,8	ITResearch	\$18,49	\$20,30	
12	ITResearch	\$18,35	\$20,30	
10	ЛИНЭКС	\$21,736	\$24,496	\$17,637

3. Тренд изменения объема ИТ-рынка России и прогнозные оценки воздействия кризиса

Особое значение имеет статистика, охватывающая интервал в несколько лет, поскольку непрерывная последовательность фактических значений позволяет прогнозировать динамику их изменения в перспективе. Такие последовательности данных представлены в исследованиях компаний ITResearch [7], IDC [10], PMR [5].

Соответствующие тренды по данным этих компаний рассматриваются ниже. Наличие данных по развитию ИТ-рынка России в период, предшествующий кризису, т. е. до 2008 года, позволяет построить тренд без учета влияния кризиса и затем дает возможность оценить воздействие кризиса на фактические показатели российского ИТ-рынка путем их сопоставления с трендами, построенными для условий отсутствия кризиса. В табл. 3 приведены последовательности данных по объемам ИТ-рынка России и годовым темпам их изменения от трех упоминавшихся выше компаний, а также данные Минкомсвязи России [11].

На рис. 4 (см. третью сторону обложки) показаны годовые темпы изменения объема российского ИТ-рынка по данным указанных четырех экспертных организаций.

Как следует из приведенных графиков, замедление темпов роста объема российского ИТ-рынка началось еще в 2007 году. В 2008 году темпы роста упали еще больше, в 2009 году, согласно разработанным тогда прогнозам трех экспертных организаций, этот показатель ожидался в зоне отрицательных значений. В то же время согласно прогнозам Минкомсвязи России темпы роста объема ИТ-рынка России в 2009 году должны были не только не снизиться, но и, наоборот, даже возрасти по сравнению с предыдущим 2008 годом.

На рис. 5 (см. третью сторону обложки) показаны графики изменения объема российского ИТ-рынка по данным четырех экспертных орга-

низаций, а также тренд изменения объема ИТ-рынка России до 2015 года. Этот тренд построен по наиболее представительной последовательности данных от ITResearch.

Приведенные графики показывают, что до 2008 года объем ИТ-рынка постоянно увеличивался, хотя по абсолютным значениям оценки различаются. В 2009 году согласно представленным тогда прогнозам IDC и PMR объем ИТ-рынка должен был существенно уменьшаться, но уже в следующем 2010 году ожидалось его увеличение. На этом же графике представлен прогноз Минкомсвязи России на 2009 год, вызвавший в свое время недоумение (напомним, что этот график построен по данным, приведенным министром Минкомсвязи России 12 апреля 2009 года на заседании коллегии министерства в рамках выставки "Связь-Экспоком 2009"). Этот прогноз оказался выше даже самого оптимистического (экспоненциального) бескризисного тренда, построенного по данным ITResearch.

Примечания, добавленные при окончательной редакции статьи.

1. Справедливости ради необходимо отметить, что на официальном сайте Минкомсвязи России в разделе "Статистика отрасли" объем ИТ-рынка в 2008 году указан в 497,2 млрд рублей; соответственно при принятых прогнозируемых темпах роста в 2009 году в 32,8 % прогноз объема рынка с 770 млрд рублей должен снизиться до 660 млрд рублей.

2. В начале 2010 г. министр Игорь Щеголев назвал прошлогодний прогноз ведомства на 2009 год "технической ошибкой". Он также заявил, что падение доходов ИТ-компаний в 2009 г. по сравнению с 2008 г. составило 8 % [13].

4. Состояние ИТ-рынка России и фактические оценки воздействия кризиса в течение 2009 г.

Ниже кратко анализируются данные, отражающие фактическое поведение российского ИТ-рын-

Таблица 3

Динамика изменения ИТ-рынка России по годам

Год	IDC		ITResearch		PMR		Минкомсвязи России	
	Темпы, %	Объем, млрд руб.	Темпы, %	Объем, млрд \$ (млрд руб.)	Темпы, %	Объем, млрд руб.	Темпы, %	Объем, млрд руб.
2001				3,68 (92)				
2002			20,2	4,42 (110)				
2003			26,3	5,59 (140)				
2004			39,7	7,80 (195)				
2005			34,8	10,52 (263)		355		
2006		410	35,9	14,29 (357)	16,7	415		
2007	35,0	550	29,3	18,49 (462)	18,9	493		450
2008	10,5	600	9,8	20,30 (507)	9,3	537	28,9	580
2009	-22,1	330			-14,2	462	32,8	770
2010	9,0	359			5,4	487		

ка в течение всего 2009 года и которые были получены уже в процессе окончательной редакции статьи. Реальные показатели объема ИТ-рынка за весь 2009 год могут быть сопоставлены с приведенным на рис. 5 трендом изменения ИТ-рынка России.

Согласно данным аналитического центра REAL-IT Лиги независимых экспертов —ЛИНЭКС (<http://www.linex.ru/league/press/releases/?press=24>) объем ИТ-рынка России в 2009 году составил 458,9 млрд рублей против 578,2 млрд рублей в 2008 году (соответственно, падение объема ИТ-рынка составило 20,6 %). В то же время тренд развития ИТ-рынка (см. рис. 5), построенный по данным компании ITRResearch, дает на 2009 год прогноз около 650 млрд рублей (напомним, что этот тренд был построен до появления кризиса). Таким образом, можно утверждать, что в результате кризиса падение объема российского ИТ-рынка в течение 2009 года в рублях составило 29,4 %.

По данным, представленным Минэкономразвития России (<http://www.ibusiness.ru/markets/464431/> или <http://www.itbestsellers.ru/statistics/detail.php?ID-15991>), объем российского ИТ-рынка в 2009 году составляет 496,5 млрд рублей (падение по отношению к 2008 году определено в 12,7 %). Сопоставляя объем ИТ-рынка в 2009 году с трендом рис. 3 легко получить, что в результате кризиса падение ИТ-рынка в 2009 году составило 23,6 %.

Что касается Минкомсвязи России, то на момент подготовки статьи к печати обобщенные данные по ИТ-рынку России за 2009 год на официальном сайте министерства еще не появились. Однако согласно информации директора департамента экономики и финансов Минкомсвязи России А. Нарукавникова (<http://www.cnews.ru/news/top/index.shtml?2010/02/09/379034>) министерство оценило объем ИТ-рынка в 2009 году около 500 млрд руб.; в предыдущем 2008 году по данным ведомства, этот показатель был равен 580 млрд руб., т. е. падение по сравнению с 2008 годом составило 13,8 %. В то же время падение ИТ-рынка в 2009 году по сравнению с трендом, приведенным на рис. 5, составило 23,1 %. Заметим, что оценки Минкомсвязи России и Минэкономразвития России достаточно близки между собой, однако они отличаются от оценок независимых экспертов ЛИНЭКС примерно на 6–6,5 %.

5. Основные тенденции и особенности развития ИТ-рынка в 2009 году

1. *Четвертый квартал 2009 года — начало выхода ИТ-рынка из кризиса.* Анализ поквартальной динамики изменения российского ИТ-рынка в течение 2009 года, проведенный экспертами ЛИНЭКС (http://www.real-it.ru/research/ritm_Q4-09.pdf), позволил им сделать вывод о том, что пик

кризиса (максимальное падение рынка) пройден в 3 квартале 2009 года и в 4 квартале наметилась тенденция к улучшению ситуации. Действительно, сопоставление показателей поквартального фактического падения ИТ-рынка в 2009 с соответствующими периодами 2008 года выглядят следующим образом: 1 квартал падение 18 %, 2 квартал — 24 %, 3 квартал — 36 % и 4 квартал — 20 %.

2. *Нефтяная зависимость.* Аналитики IDC выявили наличие высокой корреляции между объемом рынка ИТ в России и ценами на нефть, сопоставив среднегодовое значение барреля нефти и объем российского ИТ-рынка за период с 1994 по 2008 годы. Полученный ими график [14] показывает, что рост объема ИТ-рынка строго следует за увеличением цены на нефть и наоборот. По мнению президента объединения компаний-разработчиков программного обеспечения России RUSSOFT В. Макарова это означает, что российское государство следует политике, согласно которой выделение средств для развития ИТ осуществляется по остаточному принципу. Подтверждением этому служат приведенные там же [14] оценки аналитиков IDC по затратам на ИТ во всем мире и отдельно в России для 2009 года. Согласно этим оценкам во всем мире расходы на ИТ сократятся по сравнению с 2008 годом на 5,2 %, тогда как сокращение расходов на ИТ в России в долларовом выражении составит 45,7 %. Это показывает, что роль ИТ в российской экономике и российском обществе невысока по сравнению с тем, что они значат для стран с развитой экономикой и высокими социальными стандартами.

3. *Взаимосвязь ИТ-рынка и ВВП.* Согласно исследованиям экспертов ЛИНЭКС (http://www.real-it.ru/research/ritm_Q4-09.pdf) существует тесная взаимосвязь между динамикой изменения ВВП и динамикой расходов на ИТ. При этом тренд изменения ИТ-рынка запаздывает на 3–4 месяца по сравнению с изменением ВВП. Так, максимум падения ИТ-рынка в 38 % был отмечен в августе 2009 года, тогда как наибольший спад ВВП (около 17 %) имел место в мае 2009 года. При этом величина падения ИТ-рынка превышает величину спада ВВП. За первое полугодие 2009 года среднее падение ВВП составило 10,2 %, а падение ИТ-рынка — 21 %; во втором полугодии эти показатели были равны соответственно 6,9 и 28,0 %. Это также косвенно подтверждает ранее высказанное утверждение о вторичности ИТ для российской экономики и общества.

Список литературы

1. **Аналитики:** ИТ-рынок России резко сократится // *Technograd.com*. Навигатор в мире техники. 11.02.2009. URL: <http://www.technograd.com/news.16753>.

2. **Игорь Королев.** ИТ России и финансовый кризис. Антикризисные меры // Ведомости. 03.02.2009, № 18 (2288). URL: <http://www.mforum.ru/t4/forum/uk17hm?goto = 94267>.
3. **Прогноз.** ИТ-рынок России в 2020 году. — Silicon Taiga. Альянс разработчиков программного обеспечения. 08.08.2008. URL: <http://www.silicontaiga.ru/home.asp?artId = 8634>.
4. **Министерство** Эльвиры Набиуллиной подготовило два варианта долгосрочного развития России // Silicon Taiga. Альянс разработчиков программного обеспечения. 08.08.2008. URL: <http://www.silicontaiga.ru/home.sap?artId = 8634>.
5. **Российский** ИТ-рынок 2009: прогноз PMR // CRN/RE. 06.04.2009. URL: <http://www.crn.ru/news/detail.php?ID = 26665>.
6. **Призрачный** рост // Ежедневная деловая газета РБК daily. 20.03.2009. URL: <http://www.rbcdaily.ru/print.shtml?2009/03/20/media/406875>.
7. **ITResearch** выступила с докладом на IT-Summit 2009. 07.04.2009. URL: <http://www.itresearch.ru/press-center/news/detail.php?ID = 13110>.

8. **Атака** на Щеголева. Минкомсвязи надеется на лучшее // Ежедневная деловая газета РБК daily. 13.05.2009. URL: <http://www.rbcdaily.ru/2009/05/13/media/413811>.
9. **Российский** рынок ПО // PC Week/RE. № 45 (651). 08.12.2008.
10. **Реальный** ИТ-рынок. 1-й квартал 2009. — ЛИНЭКС — LINEX | ЛИГА НЕЗАВИСИМЫХ ЭКСПЕРТОВ. URL: http://www.real-it.ru/teleconf/teleconf_090204.pdf.
11. **Прогноз** Минкомсвязи: взлет ИТ или крах рубля? // PC Week/RE. № 19 (673). 26.05.2009.
12. **В конце** "тучных" лет. Итоги 2008 года // CRN-RE № 2 (52), 30.04.2009. URL: <http://www.crn.ru/numbers/spec-numbers/detail.php?ID = 27274>.
13. **Щеголев И.** ИТ-рынок упал. Министерство добивается льгот для компаний // CNews.ru. 13.01.10. URL: <http://www.cnews.ru/news/top/index.shtml?2010/01/13/376085>.
14. **Фактор** развития ИТ-рынка в РФ // PC WEEK/RE № 47 (701). 15.12.2009.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В МЕДИЦИНЕ

УДК 519.72:612.172.2 + 53.083.71

А. А. Кузнецов,
канд. физ.-мат. наук, доц.,
Владимирский госуниверситет,
e-mail: artemi-k@mail.ru

Информационные технологии донозологической диагностики

Информационные технологии донозологической диагностики представлены методом оценки вариабельности сердечного ритма (ВСР), амплитудным методом, интерквантильным методом и ярусной технологией. Показано, что ВСР, представленная в форме функциональных зависимостей параметров ВСР, может служить фактором функционального состояния организма. Амплитудным методом показано, что столбцы гистограммы диаграммы ритма сердца (ДРС) даже здорового человека не являются равноценными по информации о формирующих их диапазонах частот. Представлены возможности параметрического анализа топологической структуры ярусной ДРС от шаговых переходов и минимальных межъярусных расстояний до интегральных макроструктур низкочастотных сигналов.

Ключевые слова: электрокардиограмма (ЭКГ), информационные технологии, диаграмма ритма сердца (ДРС), вариабельность сердечного ритма (ВСР), функциональное состояние организма (ФСО), ярусная технология

Постановка задачи

Современная система здравоохранения и ее составная часть — клиническая медицина основаны на учении о болезни (нозологии) и не обладают средствами дальнего прогноза развития заболева-

ний. Поэтому они пассивны для всех условно здоровых людей и не всесильны для больных.

Предметом донозологической диагностики является не развитие патологий, а специфические особенности и характеристики защитных реакций, препятствующих их развитию, механизмы перехода от здоровья к болезни и пограничные донозологические состояния, обратимые адаптационными возможностями организма. Функциональные резервы организма определяют способность организма адаптироваться к непрерывным вариативным стимулам внешней среды. Снижение адаптационных возможностей является основной причиной развития заболевания [1, 2].

Общепризнанные критерии нормы и предболезни в донозологической диагностике отсутствуют, поэтому в качестве прогнозного признака пользуются вероятностным понятием "риск развития заболевания", определяемого внешними и внутренними факторами риска [2].

Цель работы — создание информационных технологий донозологической диагностики на основе методов исследований устойчивых физиологических закономерностей в ритме сердца и определения адекватных физических и математических критериев текущего функционального состояния организма (ФСО). Объектами исследования являются реализации ритма сердца в формах цифровых рядов последовательностей RR-интервалов на электрокардиограммах (ЭКГ) и диаграмм ритма сердца (ДРС).

Материалы и методы

Регистрации (ЭКГ) проводились на базе лабораторий Владимирского госуниверситета (ВлГУ) для условно здоровых обследуемых (УЗО) и паци-

ентов отделения реанимации областной клинической больницы. Группу УЗО составляли студенты ВлГУ обоих полов в возрасте 18—24 лет. Возраст второй группы больных варьировался в пределах 30—70 лет. Измерения с длиной записи около 20 мин проводили монитором Холтера комплекса амбулаторной регистрации электрокардиосигнала "AnnA-Flash3000" в программе "EScreen" [3, 4]. При регистрации биопотенциалов применяли двухполюсные отведения (по Небу): первый электрод располагали во втором межреберье у правого края грудины, второй — в области верхушки сердца. Такое расположение электродов позволяет записать переднее грудное отведение (A-*anterior*), соответствующее стандартному отведению II с максимальной амплитудой зубцов.

Процедура "сканирующего окна" [3—5] для выделения R-зубца требует его максимальной выраженности и является чувствительной к помехе. Предложено два варианта повышения эффективности выделения координат R-зубцов на ЭКГ дублированием синхронной записью: 1) по другому отведению ЭКГ [6]; 2) по фотоплетизмограмме [7, 8]. Каждый временной ряд последовательности электрических потенциалов (на ЭКГ) конвертировался в цифровой ряд RR-интервалов с сохранением в текстовом формате.

Информационные технологии донозологической диагностики представлены следующими методами обработки и анализа ритма сердца: методом оценки вариабельности сердечного ритма (ВСР); амплитудным методом; интерквантильным методом и структурно-топологическим методом.

Метод оценки вариабельности сердечного ритма

Информативным методом количественной оценки вегетативной регуляции сердечного ритма признан метод оценки вариабельности сердечного ритма (ВСР) [1, 2, 9]. Термин ВСР применяют в диагностике регуляторных систем и процессов, когда взаимодействия в системе "регуляция — сердце" не нарушены или отклонения от нормы обратимы [9]. Метод ВСР фиксирует уровень превалирования всех внесердечных влияний на ритм сердца над внутрисердечной регуляцией в форме отклонений значений параметров от условной нормы и не оценивает ни усредненный уровень состояния вегетативного тонуса, ни параметры колебаний вегетативных влияний на сердце. В такой общепринятой форме вариабельность ритма сердца, будучи фактором оценки отклика на внешние влияния, не является фактором ФСО, а тем более фактором прогноза состояния организма [9].

Анализ параметров ВСР был проведен в форме поиска парных параметрических функциональных зависимостей ДРС УЗО: 1) по оконным выборкам одной ДРС; 2) по групповым данным, полученным в течение трех недель; 3) по серийным

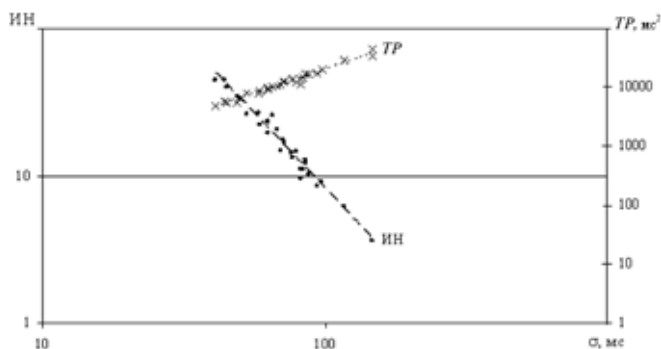


Рис. 1. Графики зависимостей ИН(σ) и TP(σ) по серии регистраций УЗО студента К. Уравнения линий тренда $ИН = 95294\sigma^{-2,03}$ и $TP = 11,046\sigma^{1,63}$ при достоверности аппроксимации $R^2 = 0,975$ и $R^2 = 0,974$, соответственно

данным регистраций ДРС одного и того же молодого человека.

На рис. 1 в логарифмическом масштабе приведены графики зависимостей индекса напряжения (ИН) регуляторных систем и полной спектральной мощности (*Total Power*, или TP) от среднего квадратичного отклонения (σ), построенные по серии ежедневных регистраций в течение 35 дней. Регистрации этой серии проводили с разной длиной записи ЭКГ (20, 40, 60, 80 мин) в вечернее время суток в состоянии покоя.

Оказалось, что параметры ВСР ДРС группы УЗО (32 человека) или серии измерений одного человека вне зависимости от пола имеют функциональные зависимости, названные "формулами ФСО" (см. рис. 1). Колебания размера n цифрового ряда RR-интервалов не влияют на формулу ФСО. Рост n при переходе от коротких записей к длинным записям ЭКГ лишь увеличивает интервал распределения за счет смещения правой границы.

Пространственно-временная комплектация параметров ВСР показала полную идентичность "формул ФСО", поэтому в предложенной форме анализа вариабельность ритма может быть не только фактором ФСО, но и фактором прогноза ФСО при проведении серийных регистраций.

Непрерывно работающий механизм процесса "сердце — регуляция" является причиной изменчивости как отдельных значений RR-интервалов, так и полигона экспериментального распределения. Это может означать осмысленное формирование топологической структуры ДРС, встроенной в постоянную структуру хаотической составляющей ритма [10]. Состояния стресса и предболезни могут характеризоваться ДРС с многомодальным распределением [9—11]. В таких случаях в вариабельности RR-интервалов обнаруживается дискретность по амплитуде и для анализа ДРС применяют методы нелинейной динамики и теории хаоса [5, 9—13], описывающие в основном разные стороны ее топологической структуры.

Амплитудный метод анализа ДРС

Для обработки и анализа ДРС был разработан и применен амплитудный метод с параллельным анализом спектров плотности мощности и фазовых портретов ДРС в программе САДР [6, 14]. Фазовый портрет (ФП) ДРС представлялся в координатах значений ДРС и соответствующей дифференциальной ДРС. Площадь ФП определялась как площадь S_3 эллипса рассеяния значений изображающей точки.

Вариационный возрастающий ряд площадей ФП ДРС по всем опытам графически представляет собой кривую в условной форме "лестницы с тремя ступеньками", на которой "верхняя ступенька" заканчивается крутым экспоненциальным ростом. По данным анамнеза каждая ступенька представляет три основных уровня ФСО: донологический, переходный, нозологический. Устойчивое ухудшение ФСО сопровождается: 1) структурным преобразованием ФП в простые геометрические фигуры; 2) снижением спектральной мощности в диапазоне низких частот и ее ростом в диапазоне высоких частот [6]. Уже это делает возможным проведение первичной оценочной сертификации по уровню ФСО с выделением границ обозначенных участков и назначением граничных значений S_3 .

При сравнении ДРС здоровых молодых людей и пациентов реанимационного отделения выяснилось, что у первых ДРС характеризуется распределением значений RR -интервалов, близким к нормальному [6, 10, 12]. На ДРС больных людей значения RR -интервалов стремятся перераспределиться по фиксированным горизонтальным интервальным уровням [6, 10, 11]. Переход от преобладающего хаотического распределения к структурированному может означать, что ритм сердца формируется не одним энергетическим источником, а несколькими. Поэтому эти интервальные уровни были названы амплитудными (АУ) [6].

В таких случаях построение гистограммы резонно проводить с назначением ширины столбцов по естественному характеру разбивки кумулятивной кривой вариационного ряда. Цифровой ряд RR -интервалов ранжируется в статистический вариационный ряд. Каждый уровень представляет собой горизонтальную полосу с шириной захвата близких по значению RR -интервалов (высота ступеньки). Эти уровни вбирают в себя участки исходного цифрового ряда и составляют собственные выборки (СВ). Такая техника комплектаций названа автором амплитудным методом обработки ДРС [5, 6, 10]. Процедуру обработки выполняет программный модуль оцифровки графических изображений (рис. 2) программы САДР [14], позволяющий в ручном режиме выделять АУ с формированием собственной выборки значений.

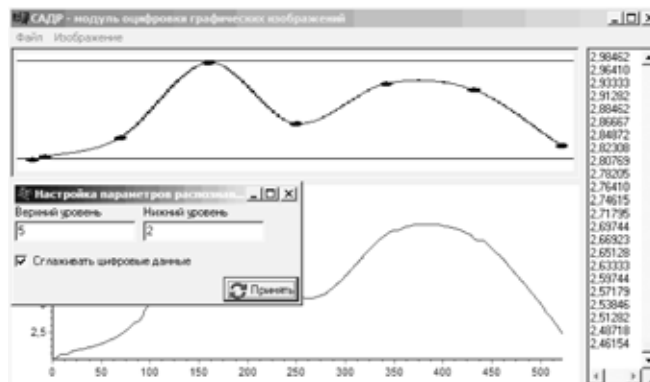


Рис. 2. Интерфейс и окно настройки расчетного модуля распознавания и оцифровки графических изображений

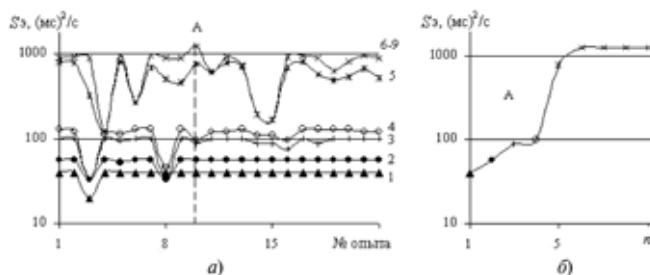


Рис. 3. Динамика значений площадей S_3 эллипса рассеяния по номерам (n) АУ для 20 УЗО (а) и развертка сечения A для опыта 10 (б)

Процедура "файл" принимает *.bmp файлы в верхнее окно, где выполняется обозначение границ графика горизонтальными линиями. В окно настройки параметров распознавания заносятся значения верхней и нижней границ. Процедура "Принять" прорисовывает в нижнем окне график полученного цифрового ряда в принятом масштабе. В правой колонке появляется последовательность значений сформированного цифрового ряда СВ, сохраняемого в текстовом файле. Процедура "изображение" сохраняет график в буфере обмена и чистит окна для принятия новых данных.

При использовании амплитудного метода анализа ДРС преобразование ФП всех СВ описывалось изменением площади S_3 эллипса рассеяния (рис. 3, а, б) и топологической структуры. При сопоставительном анализе данных обработки по номерам АУ в направлении "от нижнего — к верхнему" на ДРС получены следующие результаты:

1. Рост спектральной мощности в диапазоне высоких частот.
2. Упрощение структуры ФП и увеличение его вертикального размера.
3. Амплитудные уровни (столбцы гистограммы) ДРС даже здорового человека не являются равноценными в части информации о формирующих их диапазонах частот. Вероятно, ответственность за диапазоны частот от очень низких до высоких распределена в указанном направлении.

4. Число информативных АУ по критерию S_3 на ДРС больных людей ограничено девятью, а для здоровых людей — пятью-шестью.

Результаты параллельного анализа ФП и спектров ДРС по пунктам 1 и 2 при устойчивом ухудшении ФСО (см. выше) и при амплитудном анализе полностью совпадают. Этот факт и остальные результаты применения амплитудного метода к анализу ДРС позволяют связать ухудшение ФСО с ростом значений RR -интервалов перераспределением отсчетов на верхние АУ. Причиной этого, вероятно, является усиление интенсивности внешних влияний, вызывающих рост частоты разрядов генераторного потенциала. Спектр плотности мощности при этом теряет мощность в диапазоне низких частот и приобретает ее в диапазоне высоких частот.

Комбинированный интерквантильный метод анализа ДРС

Автоматизированной комбинированной технологией параллельного анализа спектров и фазовых портретов ДРС совместно с применением амплитудного метода является интерквантильный метод исследования ритма сердца [15]. ДРС здоровых людей имеет одно значение моды, поэтому критерий выделения АУ исчезает, но его индивидуальные характеристики остаются. Под АУ теперь понимается интерквантильный интервал (интерквантильная широта), равный ширине столбца гистограммы и характеризуемый разностью функций вероятности квантилей.

Интерквантильный (ИК) метод анализа ДРС реализован в программном комплексе САДР [5, 14] в среде С++ [16]. Интерфейс программного модуля ИК анализа RR -интервалов (рис. 4) формируется составляющими анализа: ДРС (рис. 4, а), гистограмма (рис. 4, б) вариационный ряд (рис. 4, в); фазовый портрет (рис. 4, г); автокорреляционная функция АСФ (рис. 4, д); структура сигнала по выделенному ИК-фрагменту ДРС (рис. 4, е); спектр плотности мощности (рис. 4, ж).

Назначение "ячейки" (рис. 4) — столбца гистограммы приводит к построению фазового портрета (ФП), автокорреляционных функций (АКФ) и спектра АКФ СВ.

При применении метода ИК анализа к крайним столбцам рекомендуется использовать ДРС с не менее 10 000 значениями RR -интервалов. ИК метод анализа ДРС позволяет исследовать только амплитудные уровни значений RR -интервалов и распределения внутри них, но не связи между ними. Для исследования микропереходов топологической структуры ДРС ширина кармана гистограммы должна быть минимальной, равной шагу дискретиза-

ции Δx . В этом случае применение статистического метода исследования ДРС становится проблематичным. Метод ярусов является альтернативной технологией анализа, позволяющей проводить структурно-топологические исследования от микропереходов до макросостояний.

Ярусная технология анализа ДРС

При записи ЭКГ приборные измерения придат ДРС структурно-топологические особенности: на частоте сканирования 500 Гц (шаг дискретизации $\Delta x = 1$ мс) точечная ДРС (рис. 5, а) по вертикали приобретает ярусную структуру, визуально напоминающую "нотное письмо" (рис. 5, б).

Под ярусом на ДРС понимается фиксированное приборное значение RR -интервала [5, 6, 17]. В пределах интервала времени регистрации и вариационного размаха ΔX ярусные значения RR -интервалов на ДРС, общим числом $k_j = (\Delta X / \Delta x) + 1$, подразделяют на наполненные — "рабочие" и не наполненные — "не рабочие". Наполнение падает в направлениях от яруса, имеющего значение моды, но фазовое поле "выткано" так, что в проекции на номерную ось не пропущена ни одна точка. Все ячейки яруса "включены" и находятся в "ждушем режиме". Для такой структуры ДРС всегда можно определить количество информации I_j , недостающее для ее полного описания, или ее информационную энтропию I^* [6, 10, 12]. Очевидно, что чем больше ширина кармана гистограммы (меньше число столбцов), представляющего один ярус, тем меньше будет число комбинаций построения ярусной структуры и тем меньше будут величины I_j и I^* .

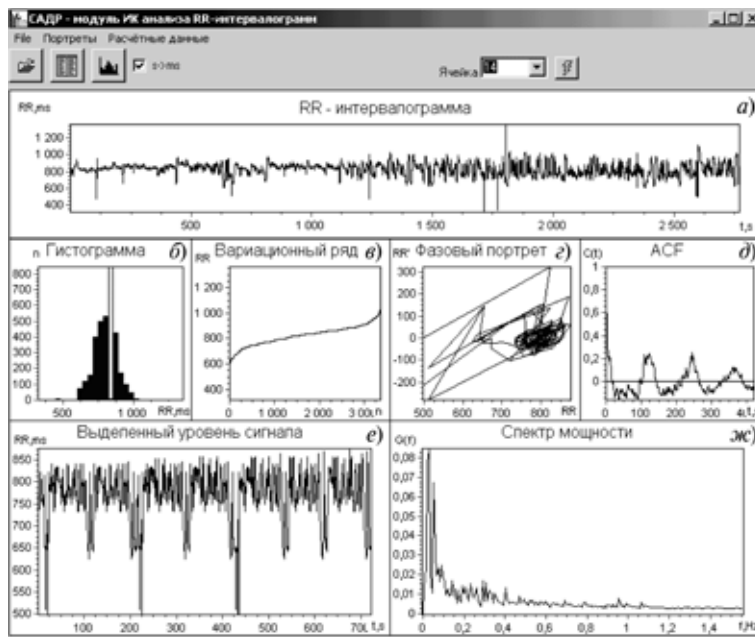


Рис. 4. Модуль ИК анализа ДРС программного комплекса САДР. При анализе АУ соответствующий столбец гистограммы выделяется белым цветом

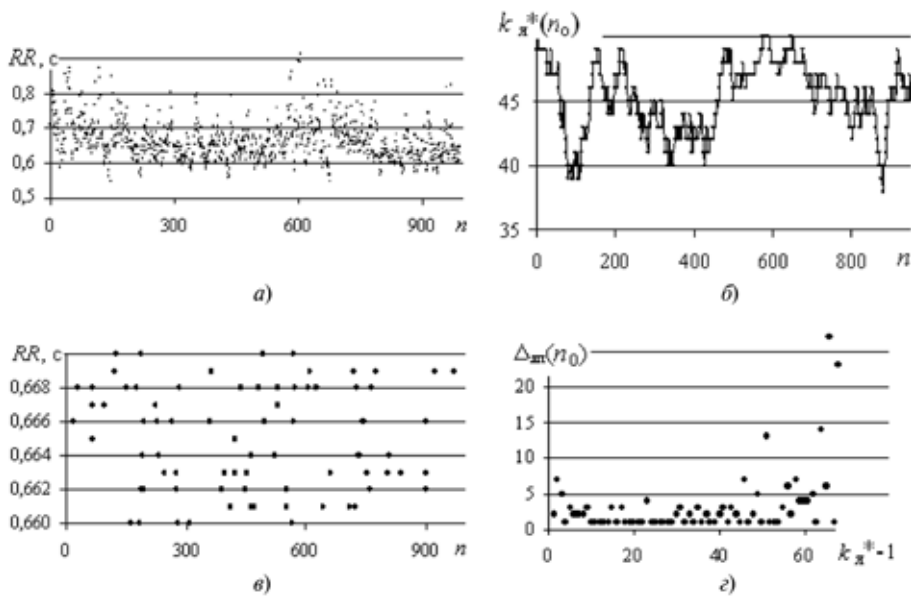


Рис. 5. ДРС УЗО студента К. (а) и ее фрагмент (б), графики оконных функций $k_{я}^*(n_0, n)$ (в) и $\Delta_{яп}(n_0, N_{МЯП})$ (г). Ширина окна $n_0 = 50$ и шаг сдвижки $\Delta = 1$

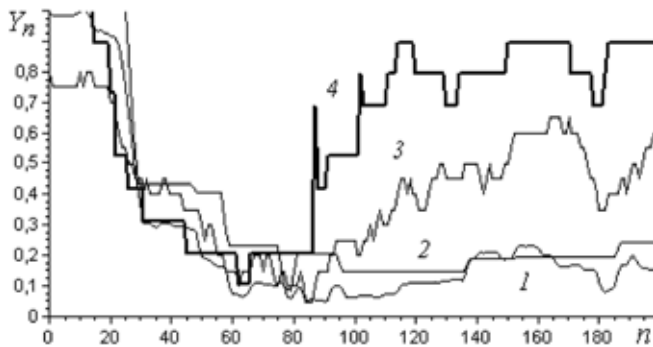


Рис. 6. Фрагмент диаграммы динамики изменений оконных значений σ (1), S_3 (2) [2], $k_{я}^*$ (3) и I_{Σ} (4) ДРС студента К. Показатели нормированы на их максимальные значения

Для определения числа "рабочих" ярусов $k_{я}^*$, межъярусных промежутков $N_{МЯП} = (k_{я}^* - 1)$ и среднего значения межъярусного промежутка (МЯП) $\langle \Delta_{яп} \rangle = \Delta X / N_{МЯП}$ создан программный модуль исследования ярусной структуры программы САДР [10]. Для анализа динамики числа "рабочих" ярусов на ДРС был применен метод "скользящего среднего" с формированием цифровых рядов среднеоконных значений $k_{я}^*(n_0, n)$ (рис. 5, в) и $\Delta_{яп}(n_0)$ (рис. 5, г) [10].

Среднеоконные значения $k_{я}^*(n_0)$ для ДРС условно здоровых людей варьируются в интервале 20...117. Сердечная недостаточность характеризуется увеличением среднеоконного числа ярусов до 200, а для пациентов в коме оно не больше 20, снижаясь фрагментарно на графике динамики числа ярусов до 2.

Для групп УЗО выборочное число межъярусных промежутков $N_{МЯП}$ для ярусной структуры

ДРС в норме линейно зависит от среднего квадратичного отклонения (σ): $N_{МЯП} = 3\sigma + 100$ при достоверности аппроксимации $R^2 = 0,96$. Линейная плотность распределения точечных значений максимальна в центре в интервале $\sigma = 40...80$ мс. Соответствующие значения $\langle \Delta_{яп} \rangle$ имеют нелинейную зависимость.

Суммарное число "фазовых ячеек" на ярусах всегда сильно избыточно в сравнении с числом значений RR -интервалов, поэтому при формировании ДРС предоставляется возможность выбора. Механизм дублирования ячеек цугов разных ярусов позволяет ритму сердца оставаться устойчивым. Распространение этого механизма на всю структуру ДРС приводит хаотическую составляющую

внутреннего управления ритмом. В этом смысле приближение значения $\langle \Delta_{яп} \rangle \rightarrow 1$ следует считать показателем роста представительности хаотической составляющей в ритме сердца, удаление значения $\langle \Delta_{яп} \rangle$ от 1, означает превалирование внешних влияний регуляции и управления.

На рис. 6 приведены графические результаты сравнительного анализа динамики четырех показателей ритма сердца условно здорового студента К. (23 года). При формировании "оконных" выборок одинакового объема использовались "окна" одинакового размера и одинаковый шаг сдвижки Δ .

Динамика показателей ритма качественно одинакова и подчинена "процессу влияния" с периодом $\Delta n = 200...220$ (0,0024...0,0025 Гц) и относится к малоисследованному спектральному диапазону ULF [9]. Показатель числа ярусов наиболее чувствителен и специфичен, а функция $k_{я}^*(n_0)$ наиболее информативна. Наименее чувствительны к изменению ритма сердца статистические показатели.

Выводы

1. Групповые и серийные параметрические функциональные зависимости ДРС УЗО идентичны по форме записи ("формулы ФСО"). Изменения длины записи ЭКГ не влияют на "формулу ФСО", но меняют границы ее действия. "Формулы ФСО", составленные по групповым параметрам ВСР, являются фактором ФСО, а составленные по временной серии регистраций — фактором прогноза ФСО.

2. Столбцы гистограммы ДРС даже здорового человека не являются равноценными по информации о формирующих их диапазонах частот. Для

больного человека их число ограничено девятью, для здорового — тремя-шестью. Наблюдаемые по мере ухудшения ФСО рост масштаба и упрощение топологии ФП ДРС, резкое уменьшение низкочастотной энергетической составляющей спектра ДРС полностью повторяются при росте номера АУ на ДРС для здоровых людей в меньшей степени, а для больных людей в большей степени. Роль детерминированной составляющей ритма на ДРС здоровых людей ограниченно возрастает при росте номера АУ, начиная с уровня моды. Роль хаотической составляющей ритма растет в обратном направлении.

3. Ярусная технология предоставляет возможности параметрического анализа топологической структуры ярусной ДРС от шаговых переходов и минимальных межъярусных расстояний до интегральных макроструктур низкочастотных сигналов. Рост относительного числа "рабочих" ярусов k_y^*/k_y сопровождается снижением величины $\langle \Delta_{яп} \rangle$. Приближение значения $\langle \Delta_{яп} \rangle \rightarrow 1$ следует считать показателем роста представительности хаотической составляющей в ритме сердца и уменьшением регуляции ритма извне. Сравнительный анализ оконной динамики четырех показателей статистического, нелинейного и структурно-топологического анализа ритма сердца определил, что наибольшую чувствительность и специфичность при оценке изменений ритма сердца имеют параметры ярусной технологии. Результаты работы могут быть полезны в клинике для функциональных диагностов и при автоматизированном мониторинге больших групп людей в случае необходимости их сертификации по уровню ФСО.

Список литературы

1. Баевский Р. М., Берсенева А. П. Оценка адаптационных возможностей организма и риск развития заболеваний. М.: Медицина. 1997. 235 с.

2. Баевский Р. М., Берсенева А. П. Введение в донозологическую диагностику. М.: Слово, 2008. 176 с.

3. Прилуцкий Д. А., Кузнецов А. А., Плеханов А. А. Накопитель ЭКГ "AnnAFlash2000" // Методы и средства измерений физических величин. Н. Новгород: НГТУ, 2006. С. 31.

4. Medical Computer Systems, Zelenograd, Moscow. URL: <http://www.mks.ru>.

5. Кавасма Р. А., Кузнецов А. А., Сушкова Л. Т. Автоматизированный анализ и обработка электрокардиографических сигналов. Методы и система / Под ред. Л. Т. Сушковой. М.: Сайнс-пресс, 2006. 144 с.

6. Кузнецов А. А. Методы анализа и обработки электрокардиографических сигналов: Новые подходы к выделению информации. Владимир: Изд-во ВлГУ. 2008. 140 с.

7. Кузнецов А. А., Плеханов А. А., Сушкова Л. Т. Алгоритм выделения RR-интервалов из фотоплетизмограмм // Биомедицинская радиоэлектроника. М.: Радиотехника. 2007. № 12. С. 27—29.

8. Кузнецов А. А., Плеханов А. А., Сушкова Л. Т., Шарьпов А. В. Программно-аппаратный комплекс синхронной регистрации электрокардиограмм и фотоплетизмограмм // Медицинская техника. 2009. № 2 (254). С. 41—43.

9. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology // European Heart Journal. 1996. V. 17. P. 354—381.

10. Кузнецов А. А. Энтропия ритма сердца: монография. Владимир: Изд-во Владим. гос. ун-та. 2009. 172 с.

11. Кузнецов А. А. Фазовая структура ритма сердца при физических нагрузках // Биомедицинская радиоэлектроника. 2009. № 3. С. 3—8.

12. Эйдукайтис А., Варонекас Г., Жемайтис Д. Применение теории хаоса для анализа сердечного ритма в различных стадиях сна у здоровых лиц // Физиология человека. 2004. Т. 30, № 5. С. 56—61.

13. Goldberger A. L. Nonlinear dynamics, fractals, cardiac physiology and sudden death // Temporal disorder in human oscillatory systems / Ed. L. Rensing. N. Y., 1986. P. 118.

14. Кузнецов А. А., Плеханов А. А., Сушкова Л. Т. Программный комплекс системного анализа динамики сердечного ритма // Вестник новых медицинских технологий. 2007. Т. XIV, № 3. С. 214.

15. Кавасма Р. А., Кузнецов А. А., Сушкова Л. Т. Энергетический и интерквантильный методы анализа электрокардиоинтервалов // Вестник новых медицинских технологий. 2005. Т. XII, № 3—4. С. 30—33.

16. Джосьютис Н. С++ Стандартная библиотека. Для профессионалов. СПб.: Питер. 2004. 730 с.

17. Кавасма Р. А., Кузнецов А. А., Сушкова Л. Т. Ярусный метод анализа RR-интервалограмм // Биомедицинская радиоэлектроника. 2007. № 12. С. 62—64.



IX ВСЕРОССИЙСКАЯ КОНФЕРЕНЦИЯ "ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ. РЕГИОНАЛЬНЫЕ АСПЕКТЫ"

7—11 сентября 2010 г.

г. Сочи, отель "Рэдиссон САС Лазурная"

Подробнее о конференции на сайте: www.vipforum.ru

Е. И. Митрушкин, д-р техн. наук,
 уч. секретарь института,
 e-mail: 89629785042@mail.ru,
С. Ю. Робский, гл. специалист,
 ФГУП "НИИАА", Москва,
 e-mail: fgupniiaa@list.ru

Оценка вклада информационной системы в эффективность медицинского учреждения

Предложены направления информатизации медицинских учреждений, многоуровневая структура эффективности системы здравоохранения страны и комплексный критерий эффективности деятельности медицинского учреждения. Для оценки критерия разработаны сеть массового обслуживания (СеМО) для медицинского учреждения и приближенное математическое выражение. Представлен пример оценки вклада информационной системы в эффективность медицинского учреждения.

Ключевые слова: информационная система, медицинское учреждение, эффективность деятельности, комплексный критерий, сеть массового обслуживания, время пребывания в сети, загрузка сети

Направления информатизации медицинских учреждений

Здоровье нации является одной из основ национальной безопасности Российской Федерации [1], что определяет важность проводимого в жизнь приоритетного национального проекта "Здоровье".

Эффективной реализации этого проекта препятствует целый ряд факторов, и в том числе недостаточный уровень внедрения передовых информационных технологий в деятельность медицинских учреждений различных уровней [2].

Спектр направлений внедрения этих технологий включает:

- информатизацию отдельных служб и процессов медицинского учреждения;
- создание взаимосвязанной информационной системы медицинского учреждения в целом;
- интеграция этой системы в единую взаимосвязанную информационно-управляющую систему Министерства здравоохранения и социального развития РФ;
- последующее ее слияние с медицинскими информационными системами других стран для создания глобальной структуры.

Концептуальная модель медицинского учреждения

Для качественной реализации указанных направлений предложена концептуальная модель медицинского учреждения — набор множества понятий предметной области с заданными на них отношениями:

- внешних категорий (цели и задачи, условия, ресурсы и критерии деятельности медицинского учреждения);
- внутренних категорий (организация процессов и результатов деятельности учреждения).

Концептуальная модель медицинского учреждения приведена на рис. 1.

При использовании медицинской информационной системы (МИС) модель дополняется ее категориями: составом, условиями, процессами, ресурсами, критериями и всеми видами организации построения и функционирования.

Концептуальная модель обеспечивает определение эффективности деятельности медицинского учреждения с помощью расширенного кортежа:

$$\mathcal{E} = \langle \mathcal{U}, \mathcal{Y}, \mathcal{R}, \mathcal{K}, \mathcal{S}, \mathcal{P}, \mathcal{D} \rangle. \quad (1)$$

Раскрытие составных категорий кортежа приведено на рис. 1.

Анализ этой модели позволяет сделать важный вывод [3], что эффективность МИС (как и любых средств автоматизации) не может быть всеобъемлюще оценена автономно и что она характеризуется лишь вкладом в общую эффективность деятельности медицинского учреждения.

Эффективность деятельности медицинского учреждения

Для анализа деятельности медицинского учреждения предлагается трехуровневая структура эффективности сложной социально-экономической системы здравоохранения, представленная на рис. 2. Многообразие и сложность процессов системы отражено совокупностью следующих показателей:

- социальной эффективности, отражающей степень обеспечения граждан гарантированным государством объемом медицинской помощи;

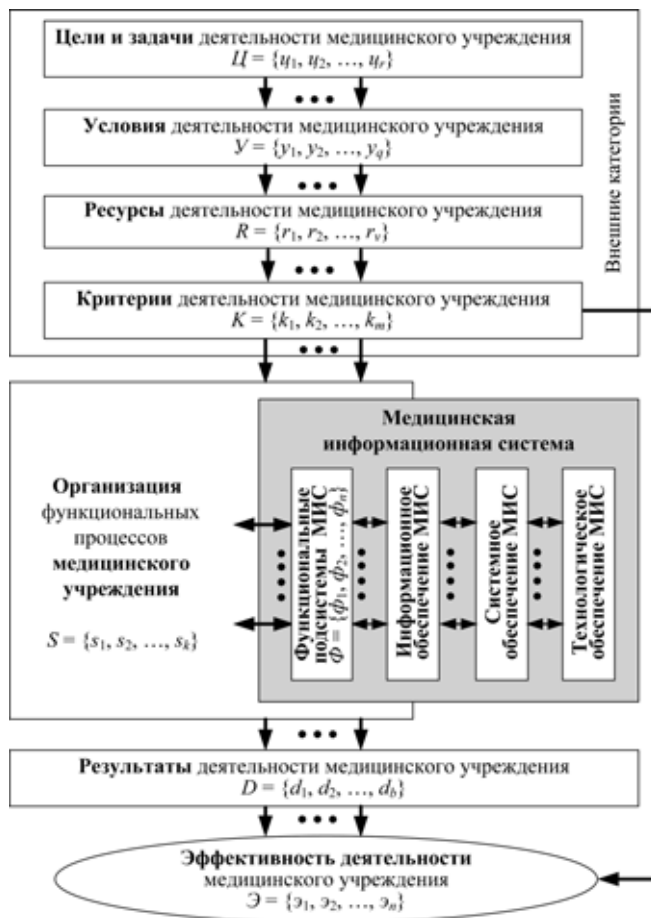


Рис. 1. Концептуальная модель медицинского учреждения



Рис. 2. Трехуровневая структура эффективности системы здравоохранения

- функциональной эффективности, характеризующей качество и результативность медицинских мероприятий;
- эффективности управления, оценивающей своевременность, целесообразность и адекватность решений и последствий их реализации;
- экономической эффективности, отражающей вклад отрасли в экономический потенциал страны.

Интегральный учет всех этих показателей является архисложной проблемой. Поэтому целесообразно использовать важнейшие частные показатели, объединенные в комплексный критерий [3].

Одним из показателей является комплексный критерий функциональной эффективности — среднее время обследования и лечения пациентов в медицинском учреждении. Он обобщает времена ожидания пациентом приема врача, обследования врачом пациента, проведения процедур, ведения медицинской документации, поиск нужной информации, составления отчетных документов и т. д.

В свою очередь, этот критерий может являться составной частью вектора критерия эффективности, дополнительно учитывающего показатели других видов эффективности (качество и стоимость обслуживания и т. д.).

Граф сети массового обслуживания для медицинского учреждения

Для оценки предложенного критерия проведен анализ типовых сценариев обследования и лечения пациентов в медицинском учреждении.

На основе результатов анализа и аппарата теории массового обслуживания разработана модель **сети массового обслуживания** (СеМО), объединяющая типовые фазы обследования и лечения пациентов в медицинском

учреждении. Граф и показатели этой сети представлены на рис. 3.

Предложенная модель имеет следующую специфику:

- модель является многофазной (последовательной) для простейшего сценария обслуживания пациента, а в общем случае — сетевой (разветвленной);
- сумма интенсивностей потоков, входящих в каждый узел графа, равна сумме интенсивностей исходящих потоков;
- отдельные узлы реализуют одноканальное обслуживание, а отдельные — многоканальное;
- сеть является разомкнутой и др.

Модель является типовой и может быть адаптируема для конкретного вида медицинского учреждения.

1. Для социального медицинского учреждения снижается вероятность углубленного обследования пациента (т. е. использования узлов 3, 4, 5 и 6 графа), что может приводить к появлению не указанных на рисунках переходов с определенными вероятностями, например, от узла 2 к узлам 7, 9 и 10.

2. В обычной поликлинике отсутствует госпитальное лечение пациентов, что приводит к исключению узла 8 с входящими и исходящими связями и т. д.

Описание переходов между узлами этого графа можно представить в виде квадратной матрицы 10×10 :

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & P_{45} & P_{46} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{67} & P_{68} & 0 & P_{60} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & P_{79} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & P_{89} & P_{80} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (2)$$

Каждый узел сети характеризуется следующими показателями:

- P_i — вероятность поступления заявок в i -й узел;
- λ_i — интенсивность входного потока заявок в i -й узел;
- τ_i — среднее время обслуживания заявок в i -м узле;
- n_i — число каналов обслуживания в i -м узле;
- $\rho_i = \lambda_i \tau_i$ — среднее число заявок, поступивших в i -й узел за время обслуживания одной заявки;
- $R_i = \lambda_i \tau_i / n_i$ — коэффициент пропускной способности i -го узла.

На основе анализа графа получены выражения для расчета вероятности поступления пациента (заявки) в каждый узел, приведенные в табл. 1.

Оценка времени пребывания в каждом узле сети

Для моделирования сети каждый узел интерпретируется как $M/M/n_i/\infty$ [4], т. е. приняты следующие условия:

- входящий поток пациентов — простейший;

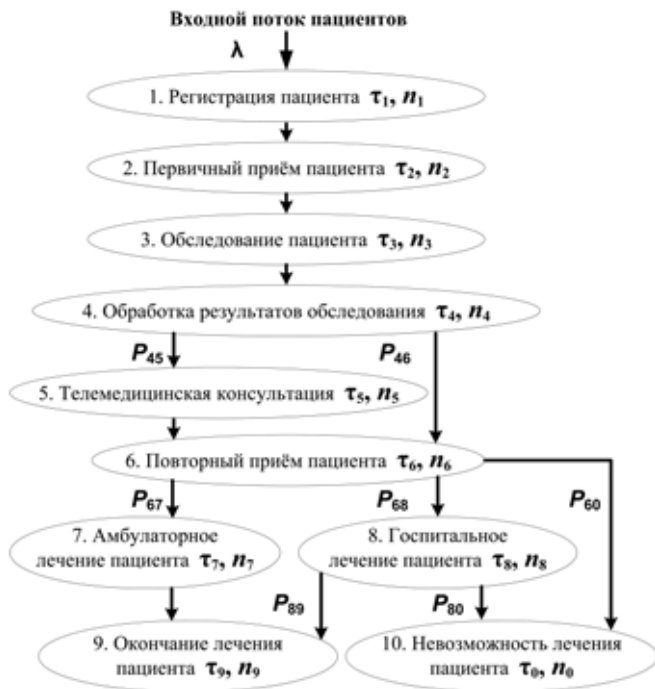


Рис. 3. Граф сети массового обслуживания для медицинского учреждения

- длина очереди каждого узла и время ожидания — бесконечные, т. е. интенсивность покидания очередей "нетерпеливыми" пациентами равна нулю;
- закон времени (интенсивность) обслуживания в узле — показательный;
- марковский случайный процесс;
- стационарный режим для каждого узла.

При принятых условиях значение вероятности отказа заявке в обслуживании в каждом узле равно нулю, исходящий поток каждого узла является пуассоновским [5], а вся сеть — показательной (экспоненциальной) [4].

Среднее время пребывания (ожидания и обслуживания) заявки в каждом узле можно определить с помощью выражения [5]:

$$\tilde{T}_i = \tau_i \left[1 + \frac{\rho^{n_i}}{(n_i - 1)! \cdot (n_i - \rho_i)^2 \cdot \left(\sum_{j=0}^{n_i-1} \frac{\rho^j}{j!} + \frac{\rho^{n_i}}{n_i!} \cdot \frac{n_i}{n_i - \rho} \right)} \right]. \quad (3)$$

Однако это выражение является очень громоздким, особенно при больших значениях n .

Поэтому для этих условий предложена достаточно точная инженерная оценка математического ожидания времени пребывания (ожидания и обслуживания) заявки в каждом i -м узле сети с помощью компактного выражения:

$$\tilde{T}_i \approx \tau_i \cdot [1 - (\lambda_i \tau_i / n_i)]^{n_i}. \quad (4)$$

Результаты вычислений по этим выражениям сведены в табл. 2.

Сопоставление оценок показывает, что в указанных диапазонах параметров сети приближенное выражение дает погрешность от 0 до — 4,7 %.

Оценка среднего времени обследования и лечения пациентов

Для получения оценки математического ожидания времени (среднего времени) обследования и лечения пациентов в медицинском учреждении использованы следующие три положения:

1. Математическое ожидание времени пребывания заявки в **последовательных узлах** сети (по схеме "И") равно сумме математических ожиданий времени пребывания в каждом узле.

2. Математическое ожидание времени пребывания заявки в **параллельных узлах** сети (по схеме "ИЛИ") равно сумме произведений математического ожидания времени пребывания заявки в каждом узле на вероятность поступления заявки в этот узел (см. табл. 1).

3. Математическое ожидание времени пребывания заявки в **сети в целом** (по комбинации схем "И" и "ИЛИ", т. е. первого и второго положения) равно сумме произведений математического ожидания времени каждого возможного частного сценария обследования и лечения пациентов на вероятность этого сценария.

Для представленного графа возможны восемь сценариев лечения пациентов (прохождения через узлы графа), характеристики которых сведены в табл. 3.

Перемножив значения второго и третьего столбца каждого сценария и сложив восемь произведений, в результате получим выражение:

$$\begin{aligned} \tilde{T}_{\text{сети}} = & \tilde{T}_1 + \tilde{T}_2 + \tilde{T}_3 + \tilde{T}_4 + \tilde{T}_5 \cdot P_{45} + \tilde{T}_6 + \\ & + \tilde{T}_7 \cdot P_{67} + \tilde{T}_8 \cdot P_{68} + \tilde{T}_9 \cdot (P_{67} + P_{68} \cdot P_{89}) + \\ & + \tilde{T}_0 (P_{60} + P_{68} \cdot P_{80}). \end{aligned} \quad (5)$$

Из анализа этого выражения ясно, что каждая его составляющая — это произведение среднего времени пребывания заявки в каждом узле на вероятность поступления заявки в этот узел (см. табл. 1).

Таким образом, для представленного графа справедливо выражение

$$\tilde{T}_{\text{сети}} = \sum_i \tilde{T}_i P_i. \quad (6)$$

Подставляя в него выражение (4) для оценки \tilde{T}_i — математического ожидания времени пребывания (ожидания и обслуживания) заявки в каждом i -м узле сети, получим

$$\tilde{T}_{\text{сети}} = \sum_i \frac{\tau_i P_i}{1 - (\lambda \tau_i P_i / n_i)^{n_i}}. \quad (7)$$

Это компактное выражение обеспечивает достаточно точную инженерную оценку математического ожидания времени пребывания (ожидания и обслуживания) заявки в сети в целом.

Следовательно, оно обеспечивает оценку математического ожидания времени (среднее время) обследования и лечения пациентов в медицинском учреждении.

Для получения более точных оценок необходимо совместное использование выражений (6) и (3).

Таблица 1

Номер узла	1	2	3	4	5	6	7	8	9	0(10)
Значение P_i	1	1	1	1	P_{45}	1	P_{67}	P_{68}	$P_{67} + P_{68} \cdot P_{89}$	$P_{60} + P_{68} \cdot P_{80}$
Проверка									$P_9 + P_0 = 1$	

Таблица 2

Число каналов в i -м узле	Оценка	Относительное среднее время пребывания (ожидания и обслуживания) заявки в каждом узле			
		$R = 0$	$R = 0,25$	$R = 0,5$	$R = 0,75$
$n = 1$	Точная Приближенная	1 1	1,333 1,333	2 2	4 4
$n = 2$	Точная Приближенная	1 1	1,067 1,067	1,333 1,333	2,287 2,287
$n = 3$	Точная Приближенная	1 1	1,020 1,016	1,158 1,143	1,759 1,730
$n = 4$	Точная Приближенная	1 1	1,007 1,004	1,087 1,067	1,513 1,463
$n = 5$	Точная Приближенная	1 1	1,006 1,001	1,052 1,032	1,376 1,311

Таблица 3

Возможный сценарий	Длительность сценария	Вероятность сценария
А	$T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_7 + T_9$	$P_{45} \cdot P_{67}$
Б	$T_1 + T_2 + T_3 + T_4 + T_6 + T_7 + T_9$	$P_{46} \cdot P_{67}$
В	$T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_8 + T_9$	$P_{45} \cdot P_{68} \cdot P_{89}$
Г	$T_1 + T_2 + T_3 + T_4 + T_6 + T_8 + T_9$	$P_{46} \cdot P_{68} \cdot P_{89}$
Д	$T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_8 + T_0$	$P_{45} \cdot P_{68} \cdot P_{80}$
Е	$T_1 + T_2 + T_3 + T_4 + T_6 + T_8 + T_0$	$P_{46} \cdot P_{68} \cdot P_{80}$
Ж	$T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_0$	$P_{45} \cdot P_{60}$
З	$T_1 + T_2 + T_3 + T_4 + T_6 + T_0$	$P_{46} \cdot P_{60}$
Проверка	—	$\sum P_{\text{сценариев}} = 1$

Оценка загрузки сети

Для практического применения представляет также интерес оценка пропускной способности сети. Для этого предлагается следующий порядок:

- определяется коэффициент пропускной способности каждого узла R_i ;
- выявляется узел сети, имеющий минимальное значение R_i (min), т. е. "узкое место" сети; в общем случае таких узлов может быть несколько;
- оценивается пропускная способность сети в целом с помощью выражения

$$R_{\text{сети}} = R_i(\text{min}). \quad (8)$$

Условием стационарности всей сети (т. е. всех сценариев) является выполнение всех неравенств $R_i < 1$. Однако даже при нестационарности сети в целом могут быть стационарными отдельные узлы и цепи (сценарии) сети.

Пример применения модели

Иллюстрация эффективности внедрения медицинской информационной системы по предложенному комплексному критерию функциональной эффективности представлена на рис. 4.

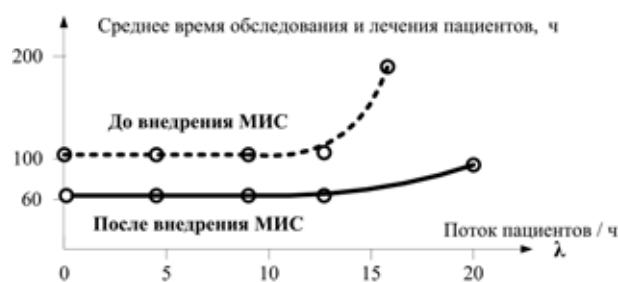


Рис. 4. Зависимости среднего времени обследования и лечения пациентов в медицинском учреждении от его загрузки

При построении графика приняты следующие исходные данные:

$P_{45} = 0,05$; $P_{46} = 0,95$; $P_{67} = 0,80$; $P_{68} = 0,15$; $P_{60} = 0,05$; $P_{89} = 0,70$; $P_{80} = 0,30$; $n_1 = 2$; $n_2 = 3$; $n_3 = 20$; $n_4 = 1000$; $n_5 = 10$; $n_6 = 3$; $n_7 = 6$; $n_8 = 1500$; $n_9 = 2$; $n_0 = 1$. До внедрения МИС: $\tau_1 = 0,05$ ч; $\tau_2 = 0,167$ ч; $\tau_3 = 1$ ч; $\tau_4 = 48$ ч; $\tau_5 = 72$ ч; $\tau_6 = 0,167$ ч; $\tau_7 = 0,25$ ч; $\tau_8 = 336$ ч; $\tau_9 = 0,1$ ч; $\tau_0 = 0,1$ ч.

После внедрения МИС: $\tau_1 = 0,333$ ч; $\tau_2 = 0,133$ ч; $\tau_3 = 0,667$ ч; $\tau_4 = 24$ ч; $\tau_5 = 6$ ч; $\tau_6 = 0,133$ ч; $\tau_7 = 0,20$ ч; $\tau_8 = 240$ ч; $\tau_9 = 0,067$ ч; $\tau_0 = 0,067$ ч.

Из графика видно, что при принятых исходных данных внедрение МИС повышает эффективность деятельности медицинского учреждения по предложенному критерию на 40 %, увеличиваясь с ростом загрузки.

Результативность предложенной модели сети массового обслуживания

Результаты углубленного анализа модели и возможностей ее применения показывают, что разработанная модель СеМО позволяет:

- а) провести оценку показателей эффективности использования МИС;
- б) оценить вклад внедрения конкретной МИС в общую эффективность;
- в) получить целый ряд других количественных оценок (коэффициент загрузки каждой медицинской службы и др.);
- г) сформировать предложения по совершенствованию МИС;
- д) сформировать предложения по оптимизации (балансировке [4, с. 253]) структуры и деятельности медицинского учреждения.

Вывод

На основе обобщения изложенных выше результатов анализа можно утверждать следующее:

1. Разработанные критерий, модель и приближенные выражения позволяют достаточно точно оценить как эффективность деятельности медицинского учреждения и отдельных служб, так эффективность его информатизации.
2. Предложенный подход может быть распространен на широкий класс информационных систем и технологий.

Список литературы

1. **Концепция** национальной безопасности Российской Федерации. Утверждена указом Президента РФ от 10 января 2000 г. № 24.
2. **Робский С. Ю.** Информационные технологии и телемедицина / Здоровье нации. 2006. № 1.
3. **Митрушкин Е. И.** Системотехника. Основы создания автоматизированных систем: учеб. пособие. М.: МИРЭА, 2010. 180 с.
4. **Клейнрок Л.** Вычислительные системы с очередями. М.: Мир, 1979. 600 с.
5. **Гнеденко Б. Д., Коваленко И. Н.** Введение в теорию массового обслуживания. М.: Наука, 1987.

Государственный университет Высшая школа экономики подписал соглашение с IEEE CS

Москва, Россия, 06 мая 2010 г. — Государственный университет Высшая школа экономики (ГУ ВШЭ), один из ведущих национальных исследовательских университетов России, подписал соглашение с компьютерным обществом института инженеров электротехники и электроники IEEE CS (Institute of Electrical and Electronics Engineers — Computer Society) о профессиональной международной аттестации студентов по программе "Сертифицированный специалист по разработке программного обеспечения" (Certified Software Development Associate, CSDA).



Таким образом, ГУ ВШЭ стал еще одним в мире и первым в России высшим учебным заведением, принявшим к действию не только руководство "Совокупность знаний по программной инженерии" (Software Engineering Body of Knowledge, SWEBOOK), но и программу аттестации CSDA IEEE CS.

В соответствии с подписанным соглашением студенты отделения программной инженерии ГУ ВШЭ будут проходить аттестацию по программе CSDA. Подготовка к аттестации будет способствовать приобретению знаний и навыков в области программной инженерии на международном уровне и позволит выпускникам отделения быть конкурентоспособными во всем мире.

"Эта сертификация имеет международное признание и будет подтверждать, что наши выпускники, имеющие такие сертификаты, являются высококвалифицированными специалистами, профессионалами по программной инженерии, способными внести свой вклад в становление инновационной экономики России", — говорит профессор С. Авдошин, руководитель отделения программной инженерии ГУ ВШЭ.



Программа CSDA предназначена для разработчиков программного обеспечения, начинающих свою карьеру после окончания бакалавриата. Она является одной из двух программ сертификации профессиональных разработчиков программного обеспечения, предоставляемых IEEE CS. Вторая программа — CSDP — предназначена для специалистов профессионального уровня. Такую сертификацию смогут пройти выпускники магистерской программы "Управление разработкой программного обеспечения" отделения программной инженерии, поскольку в учебный план магистратуры, начиная с первого курса, входит практика работы в компаниях — лидерах ИТ-индустрии. Подробную информацию по сертификациям можно найти на <http://www.computer.org/getcertified>.

Обе программы (CSDA и CSDP) основаны на руководстве SWEBOOK — всеобъемлющем документе, содержащем профессиональные стандарты отрасли и общепринятые принципы программной инженерии. Впервые разработанное в 2004 г. руководство SWEBOOK в настоящее

время находится в процессе совершенствования. Более полную информацию по руководству SWEBOOK можно найти на <http://www.computer.org/swebok>.

Сертификаты CSDA и CSDP признаются крупнейшими компаниями мира, такими как Accenture, Boeing, Borland Software, Cisco Systems, Ernst & Young LLP, Hewlett Packard, HP Corporation, IBM Corporation, IBM Global Services, Infosys Technologies, Intel Corporation, Microsoft Corporation, Oracle Palm, Samsung, Schlumberger, Siemens и др.

С 1 января 2010 г. приказами Минобрнауки РФ № 542 (бакалавриат), № 543 (магистратура) от 09 ноября 2009 г. введены в действие Федеральные государственные образовательные стандарты (ФГОС) высшего профессионального образования по новому направлению 231000 "Программная инженерия". ГУ ВШЭ получил первую в России лицензию на право подготовки бакалавров и магистров по направлению 231000 "Программная инженерия" (приказ Рособрнадзора № 1093 от 5 мая 2010 г.), и уже в 2010 г. будет проведен прием студентов на новое направление.

"При разработке программ для подготовки бакалавров и магистров мы использовали рекомендации руководства SWEBOOK и международных образовательных стандартов Computing Curricula 2005, Computer Science 2001 и Software Engineering 2004. Это должно способствовать карьере наших выпускников по программной инженерии, — говорит профессор С. Авдошин. — Наши студенты приобретают компетенции в экономике, управлении, компьютерных науках и разработке программного обеспечения, что помогает им стать отличными специалистами в области программной инженерии и руководителями программных проектов".

Учебные планы новой образовательной программы получили признание не только в России, но и за рубежом. Они прошли международную экспертизу с участием официальных представителей IEEE CS на соответствие международным рекомендациям по преподаванию программной инженерии в высших учебных заведениях.

В настоящее время в России специалисты в области программной инженерии весьма востребованы. "Инновационная экономика требует высокообразованных специалистов по программной инженерии в различных отраслях промышленности и науки, — говорит С. Авдошин. — Так как программные продукты разрабатываются на основе международных стандартов, то люди, которые создают эти продукты, должны получать образование, также основанное на международных стандартах. Программа CSDA дает хорошую возможность оценить соответствие уровня подготовки профессиональным международным стандартам".

О Государственном университете Высшая школа экономики

ГУ ВШЭ, национальный исследовательский университет России, находится в Москве и имеет филиалы в Санкт-Петербурге, Нижнем Новгороде и Перми. В настоящее время в университете обучается около 20 000 студентов, в том числе — около 17 000 будущих бакалавров, около 2000 будущих магистров и 576 аспиран-

тов. В числе 1500 профессоров и преподавателей имеются специалисты из сфер бизнеса, исследовательских институтов и государственных учреждений, 345 докторов и 909 кандидатов наук.

ГУ ВШЭ был первым университетом в России, который внедрил принятую во всем мире практику "4 + 2": четыре года для подготовки бакалавров и два года для подготовки магистров.

Отделение программной инженерии было создано в ГУ ВШЭ в 2006 г. Студенты отделения успешно участвуют в конкурсах и конференциях по компьютерным наукам и программной инженерии в России и за рубежом.

О Компьютерном обществе IEEE

Компьютерное общество IEEE (Institute of Electrical and Electronics Engineers — Computer Society) является лидирующей в мире организацией профессионалов в области вычислительной техники, насчитывающей

около 85 000 членов. Основанное в 1946 г. и являющееся самым большим из 39 обществ IEEE компьютерное общество нацелено на развитие теории и практики компьютерных и информационных технологий, является всемирно известным благодаря своей деятельности в области развития компьютерных стандартов.

Компьютерное общество IEEE обслуживает информационные и профессиональные запросы современных исследователей и практиков в области вычислительной техники, издавая журналы, сборники, проводя конференции и публикуя их материалы, предоставляя книги и сетевые курсы. Проводимые обществом аттестации по программам CSDP (для профессионалов) и CSDA (для специалистов) подтверждают уровень навыков и образования тех, кто работает в соответствующей области. Цифровая библиотека Компьютерного общества (CSDL) является отличным инструментом для исследователей, содержит более 250 000 статей и материалов 1600 конференций, а также 26 периодических изданий Компьютерного общества, начиная с 1988 г.

CONTENTS

Karnaugh D. I. <i>Development of Integrated Technological Platform as Infrastructure Basis of e-Government</i>	2
In this article author takes a review of common problems in governmental management automation in Russian Federation. The author also analyzes deficiencies in current e-Government concept realization and gives an approach to e-Government development on basis of integrated information platform.	
The newness of described approach consists in idea to integrate all subjects of governmental management and industry software modules, databases and information sources of various economic sectors in unified information space. This approach will allow to build the infrastructure basis of e-Government and will give opportunity to its constant advance.	
Article results can be applied: to build integrated technology platform as infrastructure basis of e-Government; to form the system of requirements how to develop and integrate industry software modules of various economic sectors and functional types for making conditions of constant advance of e-Government.	
Keywords: governmental management, e-Government, informational technology software, integrated technology platform, economic sector, integration, unified information space	
Norenkov I. P., Uvarov M. Yu. <i>Case-Based Reasoning on the Base of Ontology Clasterization</i>	8
The problem of decision search in case-based reasoning systems is discussed. The suggested approach is based on clasterization of object ontologies and creation of decision search patterns.	
Keywords: intellectual system, decision search, case-based reasoning, ontology clasterization	
Karpov A. A., Tsurulnik L. I., Železný M. <i>Development of a Computer System "Talking Head" for Text-to-Audiovisual-Speech Synthesis</i>	13
The paper presents first high-quality synthesizer of audio-visual Russian speech, which has been developed as the result of integrated interdisciplinary research. This system is an integration of full-fledged Russian TTS-synthesis system with the computer model of human's face that has controllable lips and other facial organs. The system converts an input Russian text into natural speech synchronized with motion of facial organs of the virtual head model. Know-how of the developed system is the specificity of the Russian speech and language, as well as an original model for asynchrony of speech modalities that allows improving both intelligibility of synthesized bimodal speech and naturalness of the virtual talking head.	
Keywords: talking head, audio-visual speech, speech asynchrony modeling, speech interface, TTS, speech synthesis, multimodal user interfaces, artificial intelligence	
Popov S. I., Popov Yu. I., Shalyto A. A. <i>Artificial ant-eater</i>	18
The article considers a task which is an expansion of Artificial ant-3 and was named Artificial ant-eater by authors. In the task we are given two-dimensional torus of arbitrary size with an ant-eater, ants (food for the ant-eater) and apples (food for the ants) placed on it. The ant-eater and the ants are controlled by Mealy machines. The ant-eater has to eat the ants as much as possible for a limited number of moves. Each ant has to eat the maximum number of apples and not be eaten by the ant-eater. An ant must not step on other ants and the ant-eater because he dies.	
The aim is to obtain the Mealy machines with the genetic algorithm. The machines have to control the ant-eater, the ants and successfully solve the task. It is necessary to develop a way of storing individuals, fitness-functions, crossing, mutation and selection operators for the genetic algorithm.	
The authors developed a tool for the study of Artificial ant-eater. It can be used with different values of parameters.	
Keywords: artificial ant, artificial intelligence, coevolution, genetic algorithm, Mealy machine	
Zhuk D. M., Manichev V. B., Ilitskiy A. O. <i>Methods and Algorithms of the Differential-Algebraic Equations Solution for Simulation of Systems and Objects in the Time Domain. Part 2</i>	23
In Part 2 of this article the choice of methods and algorithms for the decision of the linear algebraic equations (SLAE) systems is considered for the simulation of systems and objects in the time domain. Using of LU decomposition method with a choice of the main element is proved as a base method of the SLAE decision and is proved for using of the increased arithmetic operation word length, when necessary, for maintenance of required accuracy for calculation of all unknown of SLAE.	
Keywords: simulation, ordinary differential equations (ODE), differential algebraic equations (DAE), integration methods, time domain, linear algebraic equations (LAE)	
Cheremisina L. D. <i>Logic Level Estimation of Power Dissipation in CMOS Circuits</i>	27
Power estimation techniques for synchronous CMOS circuits are analyzed which are targeted logic level design. These estimations are associated with power consumption of CMOS VLSI circuits and allow getting predictive comparative estimates of synthesis variants with regard to their contributions to CMOS circuit power consumption. When evaluating power dissipation on logic level it is assumed that the basic schematic design solutions such as clock frequency and supply voltage are fixed.	
Keywords: low power design, logic synthesis, VLSI design, CMOS technology	

Kukhareno B. G. <i>Bayesian Filtering in Spectral Analysis Technology Based on the Fast Prony Transform</i>	36
A preliminary analysis of time-series is described extending the time-series dimension by adding the time-series difference, smoothing obtained multi-dimensional time-series by the Ranch method, which is based on an adaptive identification of linear dynamical system and an estimation of state-vector time-series by the Expectation-Maximization algorithm, and performing the Independent Component Analysis of smoothed time-series. It increases exactness and enlarge frequency resolution of the spectral Prony method.	
Keywords: multi-dimensional time-series, Kalman filter, Ranch smoother, linear dynamical systems, Expectation-Maximization algorithm, Independent component analysis, spectral analysis, Prony method	
Agievich S. N. <i>Using Local Spline for Estimating Parameteres of Joint Power Distributions</i>	42
Method of estimating of signals' frequency parameters on the basis of processing of their time-frequency distributions with use of spline analysis is offered. Analytic algorithms of estimating parameters based on use of local splines are designed. Analytic formulas for computing of spline-functions, their derivatives and results of practical experiments are shown.	
Keywords: spline analysis, computing of spline-functions, signal's of estimating parameters, processing time-frequency distributions	
Kravchenko T. K. <i>The Design of Key Performance Indicators Set for Processes in IT Service</i>	48
Comparative assessment of applicability of some existing approaches concerning construction of KPI (Key Performance Indicator) for enterprise's IT service is carried out. The new approach is proposed, that takes into account the specific character of IT service processes, and combines the advantages of two existing approaches: GQM and the Process Approach. The essence of the new approach is to define key indicators using GQM for each process element.	
Keywords: IT service, KPI (Key Performance Indicator); ITIL (Information Technology Infrastructure Library); CobiT (Control Objectives for Information and related Technology); BSC (Balanced Scorecard); Process Approach; GQM (Goal, Question, Metric)	
Dimitrienko Yu. L., Dimitrienko O. Yu. <i>Cluster-Continuum Modeling of Economical Processes Based on Multidimensional Continuum Mechanics Methods</i>	54
The article suggests the mass sales market model based on the clustering of objects in the multidimensional space and also on the constructing of the continuous model of clusters' movements. The system of axioms is elaborated for describing the movement of clusters in the multidimensional space. Using this system the principal equations were deduced for movement in terms of the solid clusters model. The article represents examples of numerical calculation of varying sales market parameters and the comparison with experimental data.	
Keywords: clusters, modeling in economics, multidimensional continuum mechanics, rigid cluster, conservation laws, computational simulation	
Shtrik A. A. <i>Analysis Impact of the Crisis 2009 on Condition and Prospects of Development IT Market Russia</i>	63
This article analyzes the impact of the crisis in 2009 on the status and prospects of development of IT market in Russia. Evaluated the impact of the 1998 crisis in Russia on development of the market of computer equipment. Estimates of the leading experts on the current state of the Russian IT market up to and at the beginning of the crisis of 2009. Submitted by the trend of change in volume of IT market in Russia in the pre-crisis period of 1Q 2008 — 1Q 2009, as well as assessing the impact of the crisis on the IT market in 2009.	
Keywords: analysis of influence of the crisis, information technology, infrastructure, IT market, development of the IT market, the computer market, state IT market trends in the markets, financial crisis	
Kuznetsov A. A. <i>Pre-Nozological Diagnosis Information Technologies</i>	68
Pre-nozological diagnosis information technologies are submitted by a heart rate variability (HRV) estimation method, a peak method and circle technology. The heart rate variability (HRV) estimation method is applied in the form of the analysis of HRV parameters functional dependences from a standard deviation. It is shown, that in such form HRV can serve as an organism functional condition factor. By a peak method it is shown, that heart rhythm diagram (HRD) histograms pockets even the healthy person are not equivalent under the information on ranges of frequencies forming them. Opportunities of the parametrical analysis of topological structure storey HRD from step-by-step transitions and minimal distances between circles up to integrated macrostructures of low-frequency signals are submitted.	
Keywords: electrocardiogram (ECG), heart rhythm diagram (HRD), heart rate variability (HRV), an organism functional condition (OFC), circle technology	
Mitrushkin E. I., Robski S. Yu. <i>Assessing the Contribution of Information System in the Efficiency of Medical Institutions</i>	74
The lines for informatization of medical facilities, multi-tiered structure of health system performance of the country and integrated performance measures of medical institutions. To assess the criterion developed a network of queuing for a medical facility and an approximate mathematical expression. An example of assessing the contribution of information system in the effectiveness of medical institutions.	
Keywords: information system, medical facility, performance, integrated test, the network queuing, time spent in the network, the network load	

Адрес редакции:

107076, Москва, Стромьинский пер., 4

Телефон редакции журнала **(499) 269-5510**

E-mail: it@novtex.ru

Дизайнер *Т.Н. Погорелова*. Технический редактор *О. А. Ефремова*.

Корректор *Т. В. Пчёлкина*

Сдано в набор 07.06.2010. Подписано в печать 13.07.2010. Формат 60×88 1/8. Бумага офсетная. Печать офсетная.

Усл. печ. л. 9,8. Уч.-изд. л. 11,34. Заказ 568. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Отпечатано в ООО "Подольская Периодика"

142110, Московская обл., г. Подольск, ул. Кирова, 15