

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

1(188)  
2012

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ  
Издательство "Новые технологии"

## СОДЕРЖАНИЕ

### ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

- Васенин В. А., Шундеев А. С. Эволюция технологии Грид . . . . . 2  
Богатырев В. А. Отказоустойчивые кластеры дублированных вычислительных комплексов . . . . . 9

### МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Владимирский Э. И., Исмаилов Б. И. Синхронизация в управлении хаотическими системами . . . . . 16  
Зак Ю. А. Адаптивные методы глобального случайного поиска в решении одного класса многоэкстремальных задач в условиях ограничений . . . . . 20

### КОМПЬЮТЕРНАЯ ГРАФИКА

- Малыхин В. М., Малыхина Г. Ф., Меркушева А. В. Метод регуляризации для улучшения качества восстановления изображения. . . . . 28

### КОДИРОВАНИЕ И ОБРАБОТКА СИГНАЛОВ

- Мандрикова О. В., Полозов Ю. А. Критерии выбора вейвлет-функции в задачах аппроксимации природных временных рядов сложной структуры . . . . . 31  
Кухаренко Б. Г., Пономарев Д. И. Использование метода Прони для оценки временного масштаба при обнаружении паттернов во временных рядах. . . . . 37  
Браиловский И. В., Соломец Н. А. Моделирование качества для видеокодирования 42

### ОРГАНИЗАЦИОННО-ИНФОРМАЦИОННЫЕ СИСТЕМЫ

- Сенотрусов И. А., Нешков П. А. Анализ и заполнение пропусков данных в базах данных информационно-аналитических систем . . . . . 49

### ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

- Ващенко Г. В., Капулин Д. В. Возможности интеграции бизнес-процессных моделей и ERP-систем . . . . . 52  
Димитриенко Ю. И., Димитриенко О. Ю. Континуальное моделирование динамики кластеров экономических данных в условиях внешних кризисных воздействий . 55

### Журнал в журнале

### НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ

- Мышев А. В. Информационная модель нейросети в технологиях вычислительно-го интеллекта и формах реализации компьютеринга. . . . . 63  
Кулик С. Д., Никонец Д. А. Автоматизация классификационно-диагностических почерковедческих исследований с помощью нейронных сетей. . . . . 70  
Нгуен Виет Хунг. Сжатие изображения с использованием нейронной сети в преобразованном пространстве. . . . . 76  
Contents . . . . . 78  
Приложение. Котенко И. В., Коновалов А. М., Шоров А. В. Исследовательское моделирование бот-сетей и механизмов защиты от них

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.  
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://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.

УДК 004.75+004.41

**В. А. Васенин**, д-р физ.-мат. наук, проф.,  
Механико-математический факультет МГУ  
им. М. В. Ломоносова  
e-mail: vassenin@msu.ru

**А. С. Шундеев**, канд. физ.-мат. наук, вед. науч. сотр.,  
НИИ механики МГУ им. М.В. Ломоносова,  
e-mail: shundeev@msu.ru

## Эволюция технологии Грид

*Архитектура и концептуальные положения Грид были разработаны в период 2001—2006 годов. Этот этап связан с бурным развитием технологии больших веб-сервисов, положенной в основу архитектуры Грид. После 2006 года возник ряд альтернативных подходов как к построению распределенных программных систем (RESTful веб-сервисы), так и к решению задач, которые традиционно относятся к области Грид (облачные вычисления). В этой связи актуальными являются исследования, направленные на анализ истории и перспективы развития технологии Грид. Результаты такой работы представлены в настоящей статье.*

**Ключевые слова:** Грид-технология, облачные вычисления, промежуточное программное обеспечение

### Введение

Активное развитие инфраструктуры метасети Интернет и веб-технологий [1], их востребованность практикой заложили благоприятные предпосылки для создания распределенных прикладных систем, механизмы которых позволяют эффективно и надежно разделять сосредоточенные на сетевой среде информационно-вычислительные ресурсы. Так, в середине 1990-х годов появились первые метакластерные технологии, средства и системы, которые стали именоваться Грид (*grid* (англ.) — решетка). Пройдя многолетний путь от простейших локальных систем до инфраструктур национального и транснационального масштаба, эта методология получила признание в мире. Однако и сегодня она находится в стадии активного изучения [2] и динамично развивается.

В настоящее время исполняется десять лет с момента выхода в свет фундаментальной работы [3], определившей направление развития Грид-технологии на последующее десятилетие. В этой связи актуальным представляется исследование, направленное на анализ тех задач, которые в явной или неявной форме были поставлены в работе [3]. Особый

интерес в этом контексте представляет анализ выполнения этих задач в рамках выбранных и развиваемых впоследствии подходов (моделей, методов и средств) их решения.

Работа [3] содержит концептуальное описание введенного авторами класса распределенных программных систем, получивших название *Грид-системы* и функционирующих в рамках так называемых *виртуальных организаций*. Понятие виртуальной организации является одним из ключевых и изначально имело следующую трактовку. В рамках проведения распределенных, как правило, междисциплинарных исследований научные коллективы (реально существующие организации) на договорных началах объединяют имеющиеся у них информационные и вычислительные ресурсы в целях решения поставленных перед ними задач. При этом выработываются технологии совместного, разделяемого и безопасного использования объединенных ресурсов. Автоматизация процессов подобного использования ресурсов возлагается на Грид-систему. Впоследствии понятие виртуальной организации было обобщено и вышло за рамки научных коллективов и выполняемых ими исследований.

Свое описание Грид-систем авторы работы [3] назвали "анатомическим" и построили его вокруг введенного стека протоколов Грид-системы. С позиций общей архитектуры Грид-системы эти протоколы расположены на пяти уровнях и обеспечивают следующие механизмы: взаимодействия Грид-системы с предоставляемыми в рамках виртуальной организации ресурсами нижнего уровня (*уровень адаптации*); взаимодействия отдельных компонентов Грид-системы друг с другом (*уровни связи, ресурсов, кооперации*); взаимодействия Грид-системы с внешними потребителями (*уровень приложений*). Необходимо отметить, что выделенные уровни в работе были только обозначены и качественно описаны. Их детальное описание и реализация должны были быть предметом последующих работ. В работе [3] при описании архитектуры используется метафора — песочные часы. Песочные часы имеют широкое основание и широкий верх, а также узкое горло. Основанием в случае протоколов Грид-системы является уровень адаптации, а верхом — уровень приложений. В рамках разработки архитектуры Грид-систем выдвигается минимальный набор требований к этим двум уровням. Цель такого подхода заключалась в том, чтобы избежать ограничений как на потенциальные прикладные области и задачи, в которых может применяться

данная технология, так и на типы вычислительных и информационных ресурсов, которые могут быть задействованы в рамках виртуальных организаций. В то же время узкое горло, в которое попадали уровни связи, ресурсов и кооперации, планировалось проработать более детально.

Следующей по важности является вышедшая через год работа тех же авторов [4], в которой дается "физиологический" взгляд на Грид-системы. Если анатомический взгляд касался протоколов взаимодействия распределенных компонентов Грид-системы, то физиологический взгляд направлен на внутреннюю архитектуру подобных компонентов. Вводится понятие динамически создаваемого Грид-сервиса, а также принципы работы с ним. Другое важное решение касалось уровня связи. Было принято принципиальное решение реализовывать его на основе так называемых больших веб-сервисов, детали описания которых представлены в работе [1]. Следует отметить, что в то время выбор такого подхода был далеко не очевиден. Например, в предшествующей работе [3] в качестве промежуточного программного обеспечения рассматривалась технология CORBA [5]. Более того, в дальнейшем при проектировании открытой инфраструктуры Грид-сервисов [6] были заложены механизмы совместности дескрипторов Грид-сервисов со ссылками протокола ПОР, на котором базируется взаимодействие распределенных объектов CORBA.

Работа [4] дала толчок созданию открытой инфраструктуры Грид-сервисов [6, 7], в рамках которой были разработаны механизмы, относящиеся к узкому горлу архитектуры протоколов Грид. Разработка открытой инфраструктуры и сопутствующих инструментальных средств позволила выполнить ряд значимых проектов по созданию прикладных Грид-систем. Опыт выполнения этих проектов был обобщен и проанализирован [8], что позволило оформить результаты этой деятельности в виде набора типовых сценариев — стандартных вариантов использования Грид-технологии. Прделанная за эти годы работа позволила сформулировать основные положения открытой архитектуры Грид-сервисов [9, 10]. В рамках этой архитектуры были выделены группы Грид-сервисов с позиций их функциональных возможностей и решаемых с их помощью задач, а также концептуально описаны общие сценарии взаимодействия между Грид-сервисами при решении тех или иных типовых прикладных задач.

Необходимо отметить, что все рассмотренные работы относятся к 2001—2006 годам. Этот период можно охарактеризовать как "эпоху" больших веб-сервисов [1], которая ознаменовалась рождением, бурным ростом, а затем и определенной стагнацией этой технологии. Учитывая, что большие веб-сервисы являются технологическим фундаментом для Грид-систем, особую актуальность обретают следующие вопросы, ответить на которые и пыта-

ются авторы в данной работе. Насколько технология больших веб-сервисов подходит для создания и развития Грид-систем? Имеются ли адекватные альтернативы для ее замены, насколько перспективны предлагаемые решения? Для решения поставленных задач авторы будут использовать результаты, полученные ими в работе [1], касающиеся анализа влияния веб-технологий на развитие *промежуточного программного обеспечения (middleware)*.

## 1. Ресурсная модель в области информационных технологий

Исходные положения концепции Грид-технологии, сформулированные в главе Computational Grids работы [11], несколько отличаются от подхода, описанного в работе [3]. Можно с уверенностью сказать, что в работе [11] были впервые изложены основные принципы использования *ресурсной модели* [12] в области применения информационных технологий.

Ресурсная модель подразумевает организацию поставок информационно-вычислительных услуг аналогично тому, как предприятия-поставщики коммунальных услуг поставляют своим потребителям электричество, воду, природный газ. Примером реализации ресурсной модели может служить организация работы специализированного оператора, предоставляющего абонентам услуги сетей связи общего пользования. Другой пример связан с функционированием центров обработки данных. В этом случае ресурсной моделью описывается процесс предоставления услуг по хранению данных, размещению оборудования, предоставлению вычислительных мощностей.

Сформулируем рабочее определение ресурсной услуги. Под предоставлением ресурсных услуг будем понимать тип коммерческой деятельности, предлагающий "стандартизованный" перечень услуг, поставка которых выполняется настолько безотказно, что потребители воспринимают эти услуги как должное. Ресурсные компании (компании-поставщики) предоставляют свои услуги через постоянно доступные контролируемые каналы поставки (сеть распространения). Финансирование компании осуществляется потребителями услуг посредством оплаты выставяемых компанией счетов за пользование услугами, сумма которых отражает ценность и стоимость потребляемой услуги.

Отметим ключевые моменты функционирования ресурсной компании:

- каждую услугу, предлагаемую компанией-поставщиком ресурсных услуг, нужно создать, развернуть, обеспечить ее эксплуатацию (поставку потребителям на постоянной основе), учет ее использования;
- создание ресурсной услуги предусматривает формулировку услуги, создание инфраструктуры поставки (сети распространения) и механизма для предоставления услуги потребителям, учета

потребления, выставления счетов и технического обслуживания инфраструктуры;

- для работы в ресурсном режиме необходимо контролировать поставку услуги, обнаруживать возникающие проблемы и вести запись данных об использовании в целях учета потребления.

В настоящее время ресурсная модель получила наибольшее распространение в виде так называемых *облачных вычислений (cloud computing)*, которые будут рассмотрены далее. Забегая вперед, отметим, что современные тенденции в области построения Грид-систем также тяготеют к ресурсной модели. Косвенным свидетельством тому может служить работа [13], содержащая требования (рекомендации), которые следует учитывать при построении современных Грид-систем.

## 2. Открытая инфраструктура Грид-сервисов

Как было показано, реализация основных архитектурных концепций Грид приходится на 2001—2006 годы. Данный период отмечен бурным развитием больших веб-сервисов [1], которые были положены в основу инструментальных средств построения Грид-систем. Это была единственная "зрелая" по тем временам платформа для построения распределенных систем. Однако в настоящее время ситуация изменилась. Появились и завоевали популярность новые альтернативные решения. Кроме того, произошло переосмысление самой методики применения больших веб-сервисов. В связи с изложенными выше соображениями закономерным является вопрос: возможно ли выбрать иную, отличную от больших веб-сервисов, платформу построения Грид-систем? Такой вопрос в том или ином контексте активно обсуждается специалистами-экспертами в течение последних 2—3 лет. Попробуем изложить свои соображения по этому поводу, основываясь на собственном опыте и результатах исследований [2], а также на основе анализа работ других авторов.

В программной инженерии принято разделять этапы концептуального и логического моделирования. Проясним эти понятия на примере. Блок-схема алгоритма является его концептуальной моделью, в то время как реализация алгоритма в виде программы на конкретном языке программирования является его логической моделью. Каждой концептуальной модели может одновременно соответствовать несколько логических моделей, привязанных к конкретным средам выполнения.

В работе [6] описывается открытая инфраструктура Грид-сервисов, определяющая архитектурные решения, которым должны удовлетворять все Грид-сервисы. Эти решения относятся к узкому горлу песочных часов упоминавшейся ранее архитектуры протоколов Грид. Фактически они представляют собой некоторую обобщенную логическую модель реализации приложений в среде Грид, опи-

санную на языке больших веб-сервисов. Попытаемся выделить соответствующую ей концептуальную модель.

Наиболее распространенные способы [1] построения распределенных программных систем сводятся к организации обмена сообщениями между ее компонентами. Подобный обмен сообщениями можно интерпретировать как выполнение запросов на сервере, инициированное клиентом. Выделяют синхронную и асинхронную модели обмена сообщениями, схематично представленные на рис. 1.

Синхронная модель подразумевает блокировку клиента на время выполнения запроса на сервере (на рис. 1 запрос представлен как вызов метода `perform`, получающий в качестве аргумента входные данные `inputData` и возвращающий результат `outputData`). В некоторых случаях это является сильным ограничением. Запрос может выполняться продолжительное время. Ответ может формироваться частями, и у клиента должна быть возможность получать эти части по мере их формирования. Кроме того, может потребоваться временное хранение ответа на сервере и неоднократный доступ к нему. В перечисленных случаях появляется необходимость в реализации асинхронной модели выполнения запросов.

В простейшем случае асинхронная модель выглядит следующим образом. После начала выполнения запроса сервер возвращает клиенту *дескриптор (handle)* временной сущности, созданной на сервере для хранения состояния выполнения запроса. Впоследствии, используя этот дескриптор, клиент

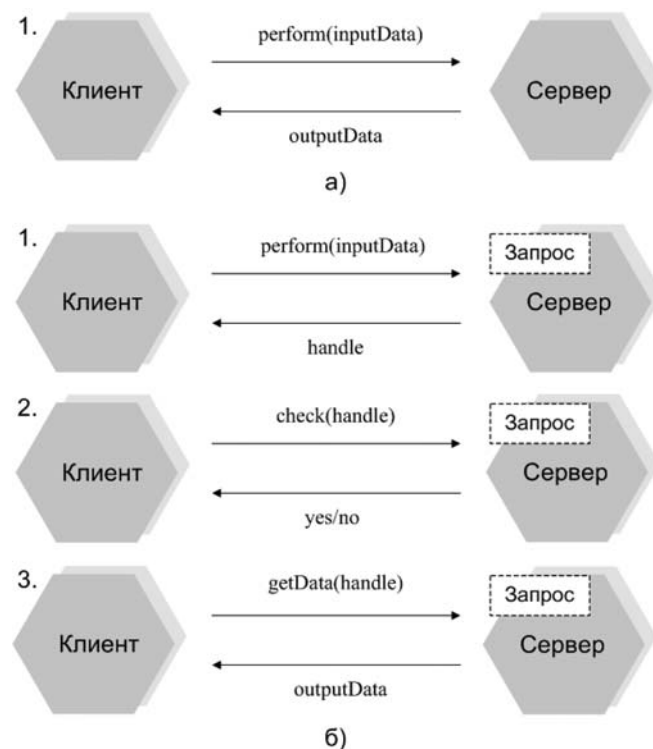


Рис. 1. Модель выполнения запросов: а — синхронная; б — асинхронная

может выполнить проверочный запрос на сервере (на рис. 1 проверочный запрос представлен как вызов метода `check`, вызов метода `getData` используется для получения ответа на исходный запрос). Когда запрос будет выполнен, клиент с помощью дескриптора получит ответ.

Простейшая модель может быть дополнена механизмом уведомлений (*callback*). В случае завершения выполнения запроса клиент получает уведомительное сообщение. Представленные рассуждения можно считать отправной точкой для выработки концептуальной модели Грид-сервисов, схематично представленной на рис. 2.

Изначально выделяются два типа Грид-сервисов — фабрики и динамически создаваемые временные экземпляры. В дальнейшем под Грид-сервисами будем подразумевать подобные экземпляры. На начальном этапе клиент, выбрав соответствующую фабрику, создает временный Грид-сервис (соответствует вызову метода `create` на рис. 2). После создания Грид-сервису присваивается временная метка — период его существования. После завершения этого периода Грид-сервис уничтожается. Клиенту, при необходимости, предоставляется возможность увеличить этот период.

После начала выполнения запроса клиент может управлять его ходом и, в частности, осуществлять мониторинг состояния процесса выполнения. Через дескриптор Грид-сервиса клиент может запросить и/или изменить это состояние. Он может также реагировать на входящие информационные сообщения (на рис. 2 это представлено как вызов метода `hotify`). В этом случае Грид-сервис выступает

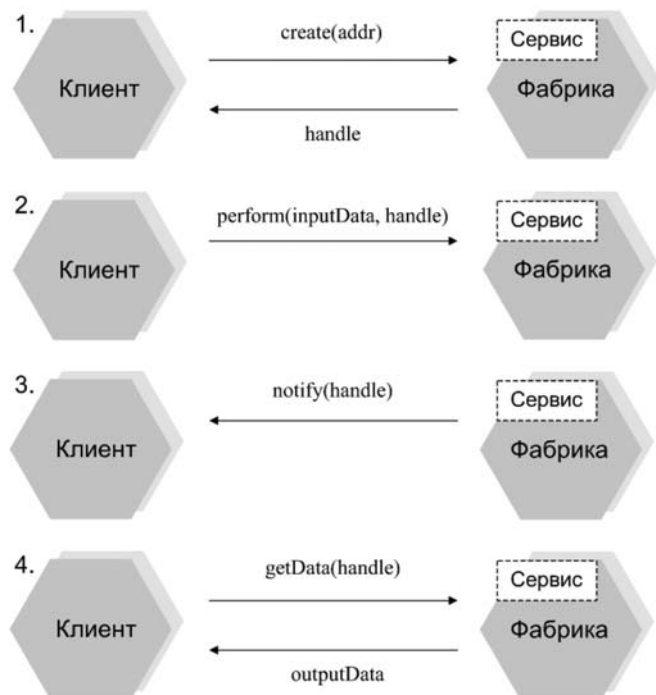


Рис. 2. Концептуальная модель Грид-сервиса

в качестве *источника (source)* событий об изменении своего состояния, а клиент — *подписчика (sink)*.

Другая представляющая интерес особенность Грид-сервиса связана с разделением понятий *дескриптор* и *ссылка (reference)* на Грид-сервис. Дескриптор является уникальным идентификатором Грид-сервиса на протяжении всего его жизненного цикла. В то же время ссылка может изменяться с течением времени и, более того, одновременно может поддерживаться несколько действующих ссылок. На основе дескриптора, используя соответствующий информационный сервис, клиент может получить действующую ссылку на Грид-сервис. В свою очередь, используя ссылку, клиент может осуществлять непосредственное взаимодействие с Грид-сервисом. Подобное, на первый взгляд усложненное, решение позволяет обеспечивать миграцию Грид-сервиса с одной фабрики на другую, а также поддерживать их репликацию. Это обстоятельство обеспечивает устойчивость и масштабируемость в работе Грид-систем.

Подводя итог изложенному выше, можно сделать вывод, что концептуальная модель Грид-сервиса базируется на совместной реализации следующих архитектурных шаблонов:

- фабрика и временные экземпляры объектов;
- управление жизненным циклом экземпляра объекта;
- миграция объектов;
- подписка на источник событий и уведомления.

### 3. Открытая архитектура Грид-сервисов

После выработки основополагающих решений в рамках создания открытой инфраструктуры Грид-сервисов был осуществлен ряд практически значимых проектов по построению Грид-систем. Полученный опыт был осмыслен и систематизирован в виде набора сценариев (вариантов использования) Грид-сервисов и систем [8]. Данная работа послужила отправной точкой для создания открытой архитектуры Грид-сервисов [9, 10]. В рамках открытой архитектуры были выделены наборы Грид-сервисов в соответствии с их функциональными возможностями и назначением. Схематично подобное разделение представлено на рис. 3.

Сервисы первой группы, получившие название *инфраструктурных (infrastructure)* сервисов, на самом деле не являются отдельными Грид-сервисам. Они представляют собой интерфейсы и механизмы, которые должны поддерживаться другими Грид-сервисами.

Основные функциональные возможности целевых Грид-систем реализуются двумя другими группами сервисов: *Сервисами управления выполнением (Execution Management Services, EMS)* и *сервисами данных (Data Services)*. Эти два набора сервисов будут далее проанализированы отдельно. Имеются также четыре вспомогательные группы сервисов: *сервисы*

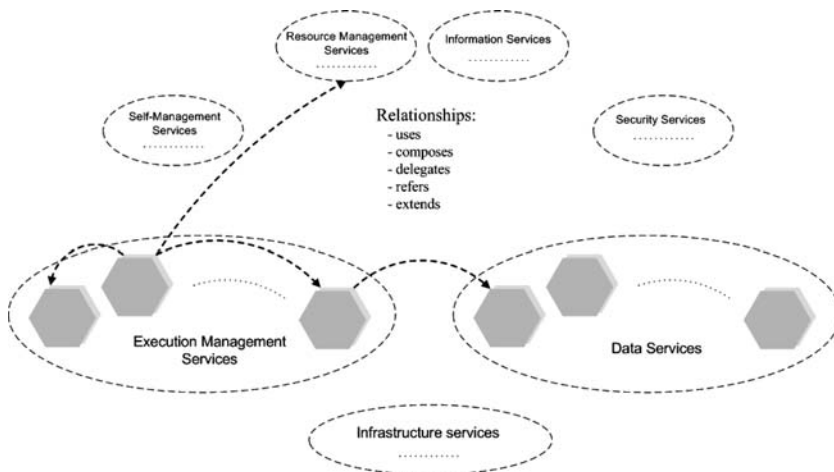


Рис. 3. Открытая архитектура Грид-сервисов

безопасности (*Security Services*); информационные сервисы (*Information Services*); сервисы управления ресурсами (*Resource Management Services*); сервисы самоуправления (*Self-Management Services*).

Отметим, что сервисы внутри выделенных групп и между группами функционируют неизолированно друг от друга. В рамках открытой архитектуры были выделены типовые отношения, которые могут возникать между Грид-сервисами в процессе их работы. Например, было введено отношение "передачи на рассмотрение" (*refers*). Так, сервис управления выполнением перед началом своей работы может отправить запрос, полученный от клиента, сервису безопасности для валидации (проверки соответствия предъявляемым к нему требованиям).

Перечислим оставшиеся отношения: "использование" (*uses*) — Грид-сервис может использовать в своей работе другой Грид-сервис; "расширение" (*extends*) — Грид-сервис может расширять функциональные возможности другого Грид-сервиса; "делегирование" (*delegates*) — Грид-сервис может делегировать выполнение запроса другому Грид-сервису (например, обеспечение взаимодействия между разными виртуальными организациями); "быть композитным" (*composes*). Естественно, что Грид-сервисы одновременно могут удовлетворять нескольким из введенных отношений.

### 3.1. Сервисы управления выполнением

Сервисы EMS предназначены для решения всего спектра задач, касающихся вопросов запуска и управления выполнением, вплоть до завершения некоторых *единиц работы (units of work)*. При этом само понятие единицы работы имеет широкую трактовку. В качестве такой единицы работы может быть запрос к Грид-сервису или выполнение некоторого стороннего приложения. Наряду с понятием единицы работы вводится понятие *точки размещения (location)* — конкретный информационно-вычислительный ресурс, который будет использован для выполнения единицы работы.

Более формально общий сценарий работы EMS-сервисов можно разделить на следующую последовательность шагов. На первом шаге осуществляется поиск кандидатов (список точек размещения), которые по формальным критериям подходят (подходящая операционная система, требуемый набор библиотек и т. д) для выполнения единицы работы. Второй шаг связан с выбором конкретной точки размещения из списка кандидатов для выполнения единицы работы. На этом шаге могут быть использованы разнообразные оптимизационные алгоритмы для обеспечения равномерной загрузки всех имеющихся информационно-вычислительных ресурсов. Третий шаг

связан с проведением всех подготовительных действий перед запуском. Он может включать сборку, конфигурирование и развертывание необходимых исполняемых файлов и библиотек, а также размещение исходных данных. Четвертый шаг заключается в запуске единицы работы на выполнение, а пятый шаг — управление ходом ее выполнения. Последний шаг включает: постоянный мониторинг; реагирование на возникающие проблемы; создание контрольных точек; принятие решения о смене точки размещения для выполнения единицы работы.

В рамках EMS предполагается выделить три класса Грид-сервисов: *контейнеры ресурсов (resources service container)*; *сервисы управления заданиями (job management)*; *сервисы выбора ресурсов (resource selection services)*. Функциональные возможности этих сервисов следуют из их названия. Проясним только понятие *задания (job)*. С одной стороны, под заданием понимается совокупность всей необходимой информации для управления ходом выполнения единицы работы, включая текущее состояние, исходные требования к аппаратно-программному окружению, информацию по результатам мониторинга и ряд других параметров. С другой стороны, задание представляет интерфейс для управления выполнением единицы работы.

Необходимо, однако, отметить, что изначально в технологии Грид механизмы, соответствующие контейнерам ресурсов, были хорошо проработаны. В конце 1990-х годов был разработан протокол GRAM (Grid Resource Access and Management) [14], реализация которого с самого начала входила в состав инструментальных средств Globus Toolkit [15]. Несмотря на то что GRAM представляет собой унифицированный интерфейс загрузки заданий для широкого спектра вычислительных сред, вопросы, связанные с управлением заданиями в рамках технологии Грид, были исследованы недостаточно.

Технология управления вычислительными заданиями активно развивалась в рамках проекта Condor [16]. В частности, был развит подход к управ-

лению композитными заданиями, получивший название DAGMap. К 2002 году была осуществлена успешная попытка соединения технологий Грид и Condor, выполненная в проекте по созданию программного средства Condor-G [17]. В рамках данного проекта осуществлена интеграция протокола GRAM в систему Condor.

Открытая архитектура Грид-сервисов обозначила необходимость реализовать в рамках сервисов управления заданиями и выбора ресурсов технологию, аналогичную технологии Condor. Однако эта задача не была решена. Текущее состояние выполнения работ по этому направлению отражено в спецификации [18].

### 3.2. Сервисы данных

Открытая архитектура Грид-сервисов декларирует необходимость использования разнообразных типов источников данных, включая коллекции плоских файлов, потоки, системы управления базами данных (реляционные, XML, объектно-ориентированные). Выделяются также сценарии работы с данными, в том числе — удаленный доступ, временное размещение, репликация, федеративные представления, создание представлений.

Изначально в рамках технологии Грид был разработан ряд эффективных, хотя и "низкоуровневых" механизмов работы с данными, примером которых может служить GridFTP [19]. Однако потребности в работе с данными, обозначенные в рамках открытой архитектуры Грид-сервисов, удовлетворены не были. Исключением может служить проект OGSA DAI [20], в рамках которого был реализован сценарий построения федеративных представлений над источниками данных. Следует отметить, что в настоящее время этот проект позиционируется отдельно от инструментальных средств Грид.

## 4. Облачные вычисления

Начиная с 2006 года стал развиваться новый подход к построению распределенных программных систем на основе использования внешних информационных и вычислительных ресурсов. Подход получил название *облачные вычисления* (*cloud computing*). Учитывая схожесть задач, решаемых Грид-системами, и задач, которые решаются в рамках облачных вычислений, подробнее остановимся на данном подходе. С этой целью проанализируем структуру и принципы реализации сервисов данных [21] компании *Amazon*, а также бизнес-модель их представления потребителям.

Базовые сервисы данных компании *Amazon* получили название S3 (Simple Storage Service). Схема их использования представлена на рис. 4. Эти сервисы используют очень простую модель данных, состоящую только из двух сущностей: *корзина* (в оригинале используется слово *bucket*) и *объект*. Объекты

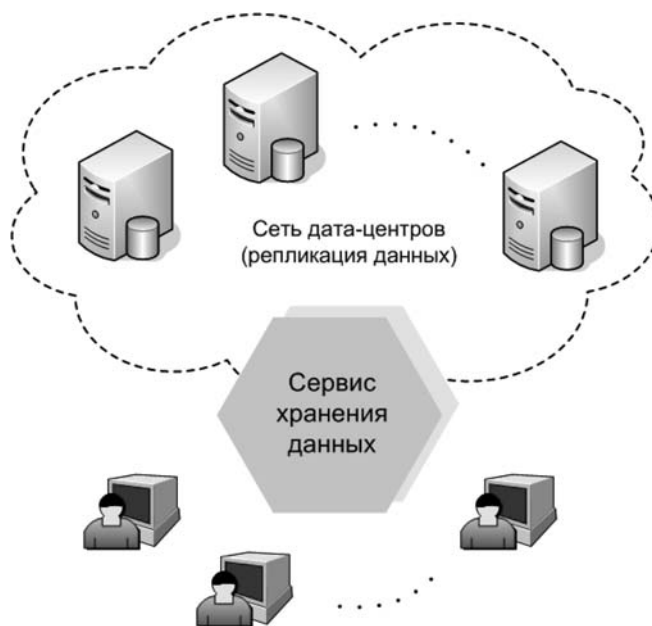


Рис. 4. Сервисы данных в облачных вычислениях

представляют собой некоторые поименованные данные и метаинформацию о них, а корзина — контейнер объектов. Имеется прямая аналогия с файловой системой. Корзина соответствует директории, а объект — файлу. Отметим, что в рамках модели поддерживается только один уровень вложенности. Корзина не может содержать другие корзины. Однако это ограничение легко обойти, если самостоятельно придерживаться специальной нотации имен объектов, имитирующей вложенные друг в друга директории.

Пользователю предоставляется несколько вариантов взаимодействия с сервисом: через свободно распространяемые программные библиотеки, реализующие клиента сервиса в объектно-ориентированном стиле; через интерфейс RESTful веб-сервиса [22]; через интерфейс большого веб-сервиса. Реализация RESTful веб-сервиса несколько отличается от "канонической" и включает использование четырех методов протокола HTTP — GET, HEAD, PUT, DELETE. Метод PUT одновременно используется и для создания, и для модификации соответствующего ресурса. Метод POST задействован только в реализации интерфейса большого веб-сервиса, а также используется для загрузки данных через веб-браузер.

Компания *Amazon* использует территориально-распределенную сеть центров обработки данных для репликации данных, обеспечивая таким образом их надежное хранение и высокую доступность. Выход из строя одного из таких центров не принесет ущерба самим данным, а также не скажется на работе "внешних" программных систем, взаимодействующих с сервисами S3. Однако платой за такое решение является отказ от *строгой непротиворечивости* (*strict consistency*) данных в пользу *потенциаль-*

ной непротиворечивости (*eventual consistency*) [23]. Этот факт означает, что проведенные изменения данных будут "видны" пользователю в той же последовательности, в которой выполнялись соответствующие операции изменения, однако не обязательно сразу. Как правило, изменения вступают в силу в течение одной секунды с момента их появления, в то же время никаких гарантий на этот счет не существует.

Сервисы S3 являются эффективным инструментальным средством, положительно зарекомендовавшим себя во многих практически важных приложениях. Необходимо отметить, что на основе S3 были получены более высокоуровневые сервисы, реализующие функциональные возможности систем управления базами данных.

## 5. Сравнение облачных вычислений с Грид-системами

Одной из задач [6], которая традиционно относится к технологии Грид, является построение коммерческих центров обработки данных. Как было показано выше, данная задача нашла эффективное решение в рамках другой технологии — технологии облачных вычислений.

Ранее был обозначен ряд сценариев работы с данными, которые являются важными с позиций открытой архитектуры Грид-сервисов. Сценарии репликации данных, их временного размещения, а также использования структурированных и неструктурированных данных реализованы в сервисах компании *Amazon* в виде, "идеологически" (концептуально) близком технологии Грид. Кроме того, реализована идея построения высокоуровневых сервисов, использующих для своей работы низкоуровневые сервисы и расширяющие их функциональные возможности.

Следует, однако, отметить, что имеется ряд различий между технологиями Грид и облачных вычислений. Эти различия прежде всего касаются бизнес-моделей, лежащих в основе их использования. Технология Грид направлена на построение в рамках межкорпоративного взаимодействия интегрированных программных систем. В то же время в основе облачных вычислений лежит целенаправленное накопление информационно-вычислительных ресурсов с целью предоставления в виде стандартизованного перечня услуг. Вместе с тем необходимо еще раз подчеркнуть, что в основе этих технологий лежат схожие идеи. Это прежде всего относится к использованию временных программных сервисов, обладающих состоянием. Если продолжительность существования Грид-сервиса задается при его создании и впоследствии может увеличиваться пользователем, то время жизни сервисов данных компании *Amazon* определяется периодически вносимой оплатой за их использование.

В связи с изложенными выше соображениями закономерным является вопрос, в чем причина успеха облачных вычислений и относительной неудачи технологии Грид? Первое, что бросается в глаза, — это тот факт, что сервисы данных компании *Amazon* не так жестко привязаны к технологии используемого промежуточного программного обеспечения. Одновременно поддерживается несколько API, между которыми можно переключаться в соответствии с собственными предпочтениями. Второе, на что следует обратить внимание, — в архитектуре сервисов компании *Amazon* отсутствуют "лишние" детали. Их архитектура усложняется постепенно и ее сложность зависит исключительно от сложности (нетривиальности) реализуемой услуги.

В случае технологии Грид все обстоит иначе. Каждый Грид-сервис вне зависимости от сложности и важности представляемой им услуги должен поддерживать базовый набор механизмов, соответствующий узкому горлу песочных часов протоколов Грид. Последнее обстоятельство является сознательным решением основателей данной технологии [3]. Можно отметить, что в некоторых случаях это может стать причиной задержки в развитии технологии или, во всяком случае, одного из ее направлений.

## Заключение

В качестве заключения сделаем следующие выводы.

1. Инструментальные средства построения Грид-систем должны базироваться на некотором промежуточном программном обеспечении. В период 2001—2006 годов альтернативы большим веб-сервисам в качестве промежуточного программного обеспечения не было.

2. Формулировка основных концептуальных и архитектурных решений технологии Грид приходится на период 2001—2006 годов. Излагая инженерным языком, в этот период было осуществлено эскизно-техническое проектирование данной технологии, результатом которого стала разработка открытой инфраструктуры и архитектуры Грид-сервисов. При этом в качестве базового инструмента были выбраны большие веб-сервисы.

3. После 2006 года начался новый период в развитии веб-технологий. Этот период ознаменовался определенной стагнацией технологии больших веб-сервисов, пересмотром методики их использования и де-факто отказом от некоторых избыточных решений. Этот период связан с возникновением подхода к построению распределенных систем на базе RESTful веб-сервисов.

4. После 2006 года возникла и начала активно развиваться технология облачных вычислений. В рамках облачных вычислений были предложены решения задач, традиционно относящихся к технологии Грид. Другой отличительной чертой об-



ланных вычислений стало использование гибридных веб-сервисов.

5. Архитектурные решения, изложенные в открытой инфраструктуре и архитектуре Грид-сервисов на языке больших веб-сервисов, жестко не привязаны к последним. В этой связи актуальным представляется проведение работ, направленных на исследование возможности построения Грид-систем на базе гибридных веб-сервисов и использования в них решений, выработанных в рамках технологии облачных вычислений.

#### Список литературы

1. **Васенин В. А., Шундеев А. С.** Развитие веб-технологий и промежуточного программного обеспечения // Информационные технологии. 2011. № 12. С. 2–9
2. **Васенин В. А., Инюхин А. В., Шевелев М. В.** Вычислительный Grid-полигон: состояние, идеи, решения // Информационные технологии. 2009. № 7. Приложение. 32 с.
3. **Foster I., Kesselman C., Tuecke S.** The Anatomy of the Grid: Enabling Scalable Virtual Organizations // International J. Supercomputer Applications. 2001. V. 15, N 3. P. 200–222.
4. **Foster I., Kesselman C., Nick J., Tuecke S.** The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration // Global Grid Forum. 2002.
5. **Henning M.** The Rise and Fall of CORBA // ACM Queue. 2006. V. 4, N 5. P. 28–34.
6. **Banks T.** et al. Open Grid Service Infrastructure Primer // Global Grid Forum. 2004. V. I-031.
7. **Tuecke S.** et al. Open Grid Services Infrastructure (OGSI) Version 1.0 // Global Grid Forum Draft Recommendation. 2003.
8. **Foster I., Gannon D., Kishimoto H., Von Reich J.** Open Grid Services Architecture Use Cases // Global Grid Forum. 2004. V. I-029.
9. **Foster I.** et al. The Open Grid Services Architecture, Version 1.0 // Proc. of Global Grid Forum. 2005. V. I-030.
10. **Foster I.** et al. The Open Grid Services Architecture, Version 1.5 // Open Grid Forum. 2006. V. I-080.
11. **Foster I., Kesselman C.** The Grid: Blueprint for a New Computing Infrastructure. Morgan-Kaufman, 1998. 675 p.
12. **Банкер Г., Коултер Т., Харт Ч., Маркус И., Сигер М., Стир Б.** От центра затрат к центру прибыльности. Переход к ресурсной модели в информационных технологиях. Изд. Veritas Software Corporation, 2004. 202 с.
13. **Subramaniam R.** et al. Guidelines of Requirements for Grid Systems. V 1.0 // Open Grid Forum. 2009. V. I-145.
14. **Czajkowski K.** et al. A resource management architecture for metacomputing systems // Proc. of the IPPS/SPDP Workshop on Job Scheduling Strategies for Parallel Processing. 1998. P. 62–82.
15. **Foster I., Kesselman C.** Globus: A metacomputing infrastructure toolkit // International Journal of Supercomputer Applications. 1997. V. 11, N 2. P. 115–128.
16. **Thain D., Tannenbaum T., Livny M.** Distributed Computing in Practice: the Condor Experience // Concurrency and Computation: Practice and Experience. 2005. V. 17, N 2–4. P. 323–356.
17. **Frey J., Tannenbaum T., Foster I., Livny M., Tuecke S.** Condor-G: A computation management agent for multi-institutional grids // Cluster Computing. 2002. V. 5. P. 237–246.
18. **Foster I.** et al. OGSA Basic Execution Service Version 1.0 // Open Grid Forum. 2005. V. R-108.
19. **Allcock W., Chervenak A., Foster I., Kesselman C., Tuecke S.** Protocols and services for distributed data-intensive science // Proc/ of Advanced Computing and Analysis Techniques in Physics Research. 2000. P. 161–163.
20. **Antonioletti M.** et al. The design and implementation of Grid database services in OGSA-DAI // Concurrency and Computation: Practice and Experience. 2005. V. 17, N 2. P. 357–376.
21. **Murty J.** Programming Amazon Web Services: S3, EX2, SQS, FPS and SimpleDB. 1 edition. O'Reilly Media, 2008. 608 p.
22. **Richardson L., Ruby S.** RESTful Web. Services. O'Reilly. 2007. 448 p.
23. **Таненбаум Э., ван Стеен М.** Распределенные системы. Принципы и парадигмы. СПб.: Питер, 2003. 877 с.

УДК 004.052.3

**В. А. Богатырев**, д-р техн. наук, проф.,  
Санкт-Петербургский государственный  
университет информационных технологий,  
механики и оптики  
e-mail: vladimir.bogatyrev@gmail.com

## Отказоустойчивые кластеры дублированных вычислительных комплексов

*Рассмотрены варианты объединения дублированных компьютерных комплексов в кластерные системы с организацией дублированных вычислений при анализе влияния структуры кластеров и организации межкомплексного обмена на надежность, отказоустойчивость и среднее время пребывания запросов в системе.*

**Ключевые слова:** отказоустойчивость, дублированный комплекс, надежность, кластер

## Введение

Обеспечение высокой надежности управляющих компьютерных систем требует введения избыточности [1] как на уровне вычислительных узлов, в качестве которых используются дублированные вычислительные комплексы (ДВК), так и на уровне их объединения в систему (кластер). В системах ответственного назначения ДВК, как правило, работают в режиме дублирования вычислений в полуконфликтах (ПК) при сравнении результатов, что снижает вероятность выдачи неправильных результатов.

В статье исследуется эффективность вариантов объединения ДВК в кластерные системы с организацией дублированных вычислений. Для решения этой задачи рассматриваются:

- варианты межкомплексного обмена при дублировании вычислений, в том числе комбинирующие внутрикомплексную и межкомплексную передачи через магистрали при различных дисциплинах доступа;

- влияние структуры кластеров ДВК и организации межкомплексного обмена данными на надежность, отказоустойчивость и среднее время пребывания запросов;
- выбор вариантов организации кластеров ДВК с учетом надежности, отказоустойчивости и времени межкомплексного обмена.

### Структура ДВК

В качестве базовой рассмотрим структуру ДВК, представленную на рис. 1. Каждый из полукомплексов ДВК содержит процессор (P), модуль памяти (M) и модуль контроля (S), реализующий сравнения результатов вычислений в полукомплексах.

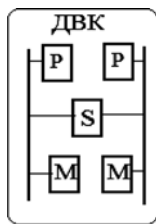


Рис. 1. Базовая структура дублированного комплекса

### Варианты объединения ДВК в кластер

Рассматриваемые варианты объединения ДВК в кластер приведены на рис. 2 и 3. Для организации обмена используются общие магистрали ( $M_1, M_2, M_3$ ), подключение к которым ДВК осуществляется через сетевые адаптеры (CA). Рис. 2 представляет варианты объединения ДВК без введения локальных выделенных средств обмена внутри ДВК, а рис. 3 — с введением таких средств. Рис. 2, *a* соответствует использованию для обмена одной магистрали, рис. 2, *b* — дублированной магистрали, а рис. 2, *в* — дублированной магистрали для межкомплексных обменов и дополнительной общей магистрали для внутрикомплексных обменов. В качестве локальных средств обмена внутри ДВК могут использоваться пары связанных между собой сетевых адаптеров (рис. 3, *a*) или адаптеры внутрикомплексного обмена (A) (рис. 3, *b*). При локализации обмена внутри каждого комплекса при введении адаптера A он может быть реализован с учетом специфики внутрикомплексного обмена (отсутствие множественного доступа, малые расстояния, параллельная передача и др.), что приводит к ускорению обмена.

Будем считать, что общее число ДВК в кластере  $M$ , при этом выделяется  $m$  основных и  $M - m$  резервных ДВК. Число  $m$  ос-

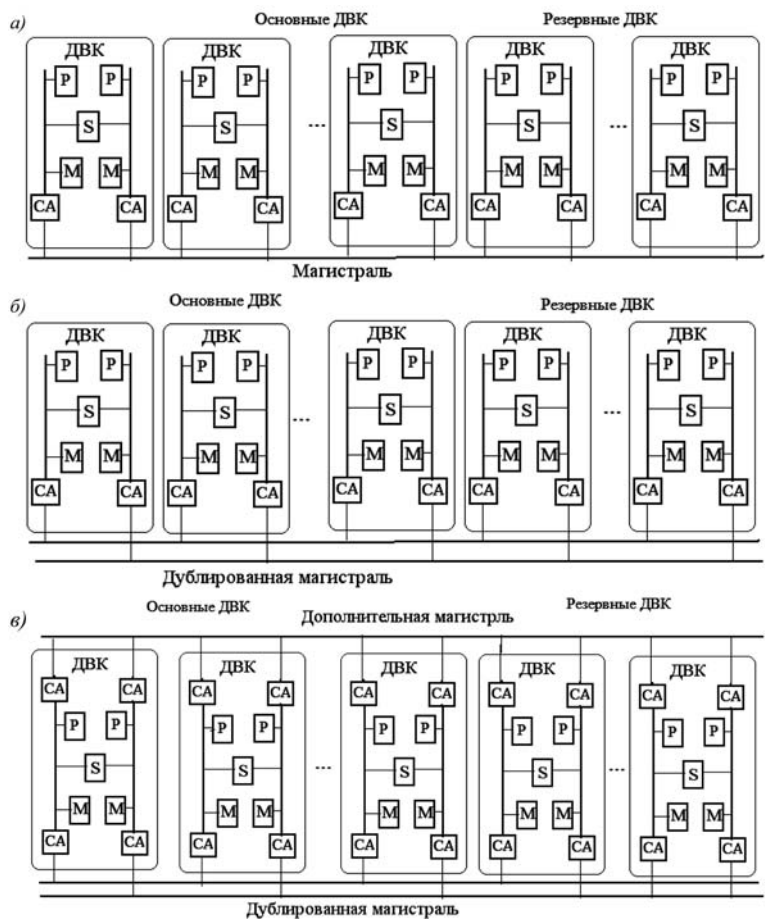


Рис. 2. Варианты кластеров без введения средств обмена внутри ДВК, содержащих одну (*a*), две (*b*) и три (*в*) магистрали обмена

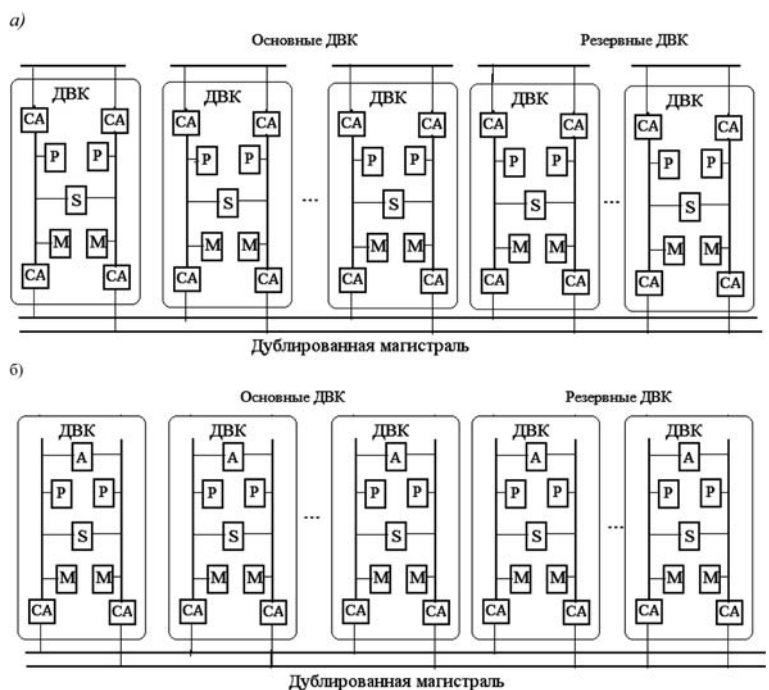


Рис. 3. Варианты кластеров с введением средств обмена внутри ДВК на основе:  
*a* — пары сетевых адаптеров CA; *б* — адаптеров внутрикомплексной связи A

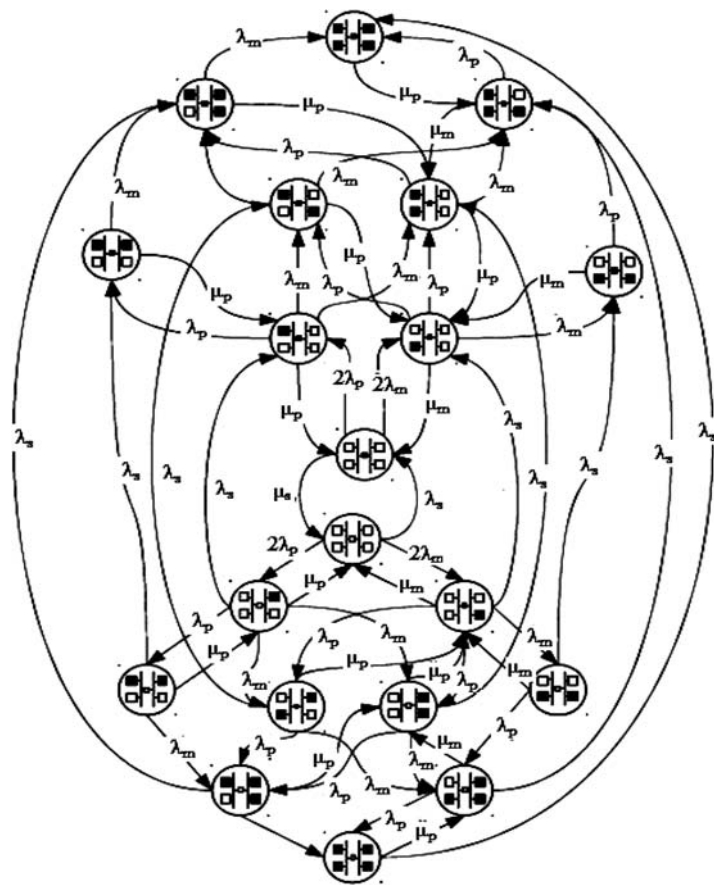


Рис. 4. Граф переходов для марковской модели надежности ДВК

$$\begin{aligned}
 & -P_0 \lambda_s - P_0 2\lambda_p - P_0 2\lambda_m + P_1 \mu_s + P_2 \mu_p + P_3 \mu_m \\
 & -P_1 2\lambda_p - P_1 2\lambda_m - P_1 \mu_s + P_0 \lambda_s + P_4 \mu_p + P_5 \mu_m \\
 & -P_2 \mu_p - P_2 \lambda_p - 2(P_2 \lambda_m) - P_2 \lambda_s + P_0 2\lambda_p + P_9 \mu_p \\
 & -P_3 \mu_m - P_3 \lambda_m - 2(P_3 \lambda_p) - P_3 \lambda_s + P_0 2\lambda_m + P_6 \mu_p + P_7 \mu_p + P_8 \mu_m \\
 & -P_4 \mu_p - P_4 \lambda_p - P_4 \lambda_m - 2 + P_1 2\lambda_p + P_2 \lambda_s + P_{12} \mu_p \\
 & -P_5 \mu_m - P_5 \lambda_m - P_5 \lambda_p - 2 + P_1 2\lambda_m + P_3 \lambda_s + P_{10} \mu_p + P_{11} \mu_m + P_{13} \mu_p \\
 & -P_6 \mu_p - P_6 \lambda_p - P_6 \lambda_s - P_6 \lambda_m + P_2 \lambda_m + P_3 \lambda_p + P_{14} \mu_m + P_{15} \mu_p \\
 & -P_7 \mu_p - P_7 \lambda_p - P_7 \lambda_s - P_7 \lambda_m + P_2 \lambda_m + P_3 \lambda_p \\
 & -P_8 \mu_m - P_8 \lambda_s - P_8 \lambda_p + P_3 \lambda_m \\
 & -P_9 \mu_p - P_9 \lambda_s - P_9 \lambda_m + P_2 \lambda_p \\
 & -P_{10} \mu_p - P_{10} \lambda_p - P_{10} \lambda_m + P_4 \lambda_m + P_5 \lambda_p + P_7 \lambda_s \\
 & -P_{11} \mu_m - P_{11} \lambda_p + P_5 \lambda_m + P_8 \lambda_s \\
 & -P_{12} \mu_p - P_{12} \lambda_m + P_4 \lambda_p + P_9 \lambda_s \\
 & -P_{13} \mu_p - P_{13} \lambda_m - P_{13} \lambda_p + P_4 \lambda_m + P_5 \lambda_p + P_6 \lambda_s + P_{16} \mu_m + P_{17} \mu_p \\
 & -P_{14} \mu_m - P_{14} \lambda_p - P_{14} \lambda_s + P_6 \lambda_m + P_7 \lambda_m + P_8 \lambda_p + P_{18} \mu_p \\
 & -P_{15} \mu_p - P_{15} \lambda_m - P_{15} \lambda_s + P_6 \lambda_p + P_7 \lambda_p + P_9 \lambda_m \\
 & -P_{16} \mu_m - P_{16} \lambda_p + P_{10} \lambda_m + P_{11} \lambda_p + P_{13} \lambda_m + P_{14} \lambda_s + P_{19} \mu_p \\
 & -P_{17} \mu_p - P_{17} \lambda_m + P_{10} \lambda_p + P_{12} \lambda_m + P_{13} \lambda_p + P_{15} \lambda_s \\
 & -P_{18} \mu_p - P_{18} \lambda_s + P_{14} \lambda_p + P_{15} \lambda_m \\
 & -P_{19} \mu_p + P_{16} \lambda_p + P_{17} \lambda_m + P_{18} \lambda_s
 \end{aligned}$$

$Z := \text{rfixd}(P, 0, 100, 1000, D) \quad n := 0..1000$

Рис. 5. Задание системы дифференциальных уравнений в Mathcad-14

новых ДВК определяется числом комплексов, минимально необходимых для решения критичных задач. Резервные ДВК выполняют вспомогательные вычисления, потеря которых не связана с безопасностью и значительными экономическими или иными потерями.

Для критичных вычислений выполнение каждого запроса дублируется в двух полукомплексах ДВК, результаты вычислений сравниваются модулем S и в случае их совпадения выдаются для дальнейших вычислений. При межкомплексном обмене для дублированных вычислений передаваемые данные должны быть занесены в модули памяти M двух полукомплексов адресуемого ДВК. При неспособности ДВК к дублированным вычислениям, например, из-за отказов одного из процессоров для обслуживания критичных запросов подключается один из резервных ДВК. После отказов может начаться восстановление комплекса, причем во время восстановления одного из полукомплексов другой может выполнять некритичные вычисления (ранее выполняемые резервными ДВК).

### Марковская модель надежности ДВК

При построении марковской модели восстанавливаемого комплекса (см. рис. 1) будем считать известными интенсивности отказов  $\lambda_p, \lambda_m, \lambda_s$  и восстановлений  $\mu_p, \mu_m, \mu_s$  процессоров P, памяти M и модуля контроля S. Восстановление проводится одним ремонтником. Во время ремонта одного полукомплекса ДВК другой (исправный) может участвовать в выполнении некритичных задач.

Граф переходов для марковской модели надежности ДВК приведен на рис. 4, на котором отказавшие узлы затемнены.

Решение дифференциальных уравнений, составленных по графу переходов на рис. 4, позволяет найти вероятности всех состояний комплекса и в результате суммирования работоспособных состояний определить нестационарный коэффициент готовности комплекса [1]. Решение системы дифференциальных уравнений проведено в системе компьютерной математики Mathcad-14 по методу Рунге—Кутты, представление системы дифференциальных уравнений по графу на рис. 4 приведено на рис. 5. Для нахождения стационарного коэффициента готовности комплекса составляется система алгебраических уравнений (производные в правой части приравниваются нулю). Суммируя найденные вероятности работоспособных состояний, вычисляем коэффициент готовности (стационарный) комплекса.

Если требуется найти коэффициент готовности ДВК по критичным приложениям (с кон-

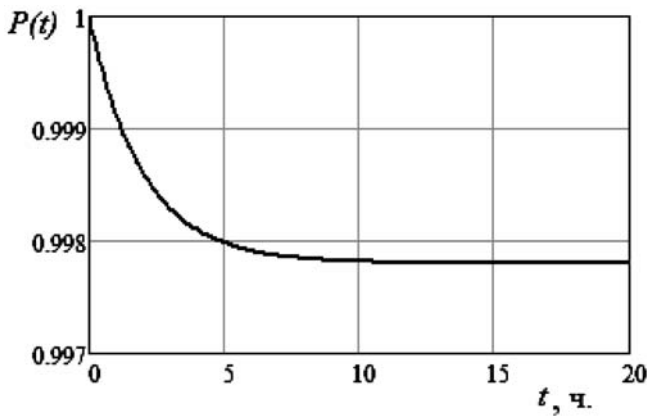


Рис. 6. Нестационарный коэффициент готовности ДВК

тролем дублированных вычислений), то к работоспособным относится только исходное состояние (без отказов).

Результаты определения нестационарного коэффициента готовности ДВК при  $\lambda_p = 0,0003 \text{ ч}^{-1}$ ,  $\lambda_m = 0,0002 \text{ ч}^{-1}$ ,  $\lambda_s = 0,0001 \text{ ч}^{-1}$ ,  $\mu_p = \mu_m = \mu_s = 0,5 \text{ ч}^{-1}$  представлены на рис. 6.

### Варианты организации межкомплексных обменов

Рассмотрим варианты организации обмена между ДВК при дублировании вычислений, когда необходимо занесение передаваемых данных в модули памяти М двух полукомплексов адресуемого ДВК.

Для систем с дублированием магистралей [2—6] и возможностью обмена внутри ДВК (рис. 3) выделим следующие варианты передачи от  $i$ -го к  $j$ -му ДВК.

**Вариант В1.** Первый и второй полукомплексы  $i$ -го ДВК выставляют запросы на доступ соответственно к первой  $M_1$  и ко второй  $M_2$  магистральям, после предоставления которых осуществляется передача  $CA_{1i} \rightarrow M_1 \rightarrow CA_{1j}$ ,  $CA_{2i} \rightarrow M_2 \rightarrow CA_{2j}$ , в результате чего данные  $i$ -го ДВК заносятся в модули памяти М обоих полукомплексов  $j$ -го ДВК.

**Вариант В2.** Множественный доступ реализуется только к одной магистрали, а вторая предоставляется вместе с первой, после чего осуществляется передача  $CA_{1i} \rightarrow M_1 \rightarrow CA_{1j}$ ,  $CA_{2i} \rightarrow M_2 \rightarrow CA_{2j}$ .

**Вариант В3.** Первый либо второй полукомплексы  $i$ -го ДВК выставляют запросы на предоставление доступа соответственно к первой или второй магистралям  $M_1$  и  $M_2$  (закрепление за полукомплексами магистралей проводится так, чтобы обеспечить их равномерную загрузку). После предоставления магистрали  $M_g$  ( $g = 1, 2$ )  $i$ -му ДВК осуществляется передача  $CA_{gi} \rightarrow M_g \rightarrow CA_{gj}$ . В результате такой связи данные заносятся в модуль памяти М  $g$ -го полукомплекса  $j$ -го ДВК. Из модуля памяти  $g$ -го полукомплекса данные передаются в модуль памяти другого полукомплекса либо через два выделенных для внутрикомплексного обмена СА (рис. 3, а), либо

через адаптер внутренней связи А (рис. 3, б). Передача через адаптер А может быть параллельной.

Возможна модификация варианта В3, при которой запрос от ДВК выставляется одновременно к разным магистралям. После предоставления доступа к одной из магистралей запрос на доступ к другой магистрали снимается на аппаратном уровне взаимодействия пары СА, что требует их модификации.

Структура кластера, соответствующая рис. 2, в, предполагает при варианте В3 выделение третьей магистрали для внутрикомплексных обменов на основе множественного доступа, что приводит к замедлению обмена.

Для кластеров без резервирования магистралей по рис. 2, а возможны следующие варианты обмена:  $(CA_{1i} \rightarrow M_1 \rightarrow CA_{1j}, CA_{2i} \rightarrow M_1 \rightarrow CA_{2j})$ ;  $(CA_{1i} \rightarrow M_1 \rightarrow CA_{1j}, CA_{1j} \rightarrow M_1 \rightarrow CA_{2j})$ ;  $(CA_{2i} \rightarrow M_1 \rightarrow CA_{1j}, CA_{2j} \rightarrow M_1 \rightarrow CA_{1j})$ ;  $(CA_{1i} \rightarrow M_1 \rightarrow CA_{1j}, CA_{1i} \rightarrow M_1 \rightarrow CA_{2j})$ ;  $(CA_{2i} \rightarrow M_1 \rightarrow CA_{1j}, CA_{2i} \rightarrow M_1 \rightarrow CA_{2j})$ .

### Оценка среднего времени межкомплексного обмена

Обмен через каждую магистраль представим системой массового обслуживания типа М/М/1 [13]. Для структуры кластера, представленного на рис. 3, в исходном состоянии (без отказов) среднее время пребывания запросов на межкомплексный обмен (от формирования запроса до передачи данных в модули памяти обоих полукомплексов адресуемого ДВК) при вариантах обмена В1, В2, В3 равно соответственно:

$$T_1 = \frac{v}{1 - \lambda v} + d; T_2 = \frac{v}{1 - \lambda v}; T_3 = \frac{v}{1 - 0,5\lambda v} + \alpha v.$$

Здесь  $\lambda$  — интенсивность запросов на межкомплексный обмен при дублировании передачи;  $v$  — среднее время передачи через магистраль при обмене между полукомплексами;  $\alpha$  — коэффициент, устанавливающий соотношение скорости передачи через СА и через адаптер внутрикомплексной передачи А;  $d$  — определяет разницу по времени предоставления доступа к разным магистралям от одного ДВК (из-за неидентичности по времени процесса множественного доступа к разным магистралям). При детерминированных методах множественного доступа, в частности, при маркерном методе (если логическое кольцо передачи полномочий для обеих магистралей совпадает и через магистрали осуществляется только дублированная передача)  $d$  близко к нулю. Для случайных методов доступа значение  $d$  может быть существенным.

Зависимости среднего времени пребывания запросов на межкомплексный обмен  $T(\lambda)$  от интенсивности их формирования при  $v = 1 \text{ мс}$  представлены на рис. 7. Кривые 1 и 2 соответствуют варианту обмена В1 при  $d = 0,5 \text{ мс}$  и В2 (кроме того,

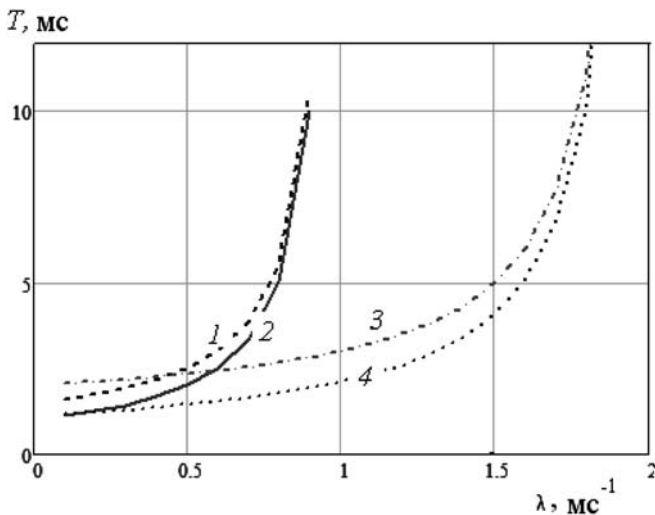


Рис. 7. Среднее время пребывания запросов на межкомплексный обмен

кривая 1 соответствует также варианту обмена В1 при  $d = 0$  мс), а кривые 3–4 — обмену по варианту В3 при  $\alpha = 1$  и  $\alpha = 0,1$ .

Проведенные расчеты показывают явное преимущество варианта обмена В3, причем по мере роста интенсивности запросов  $\lambda$  эта эффективность возрастает. Кроме того, вариант В3 позволяет увеличить предельную интенсивность запросов, выдерживаемую системой в условиях стационарности. Сравнение кривых 3 и 4 показывает целесообразность использования для внутрикомплексной связи высокоскоростных средств обмена. При малой интенсивности запросов вариант В3 может оказаться менее эффективным, но при этом следует отметить, что для практики более важно сократить время пребывания запросов при их большой интенсивности.

### Надежность кластера

Сравним надежность кластеров ДВК с резервированием магистралей межкомплексного обмена при реализации в ДВК дублированных вычислений.

Для вариантов обмена В1 и В2 взаимосвязь в кластере между любой парой не менее  $m$  ДВК с дублированием вычислений поддерживается при исправности пары СА не менее чем в  $m$  ДВК.

При варианте обмена В3 для кластера, представленного на рис. 3, б, взаимосвязь между любой парой не менее  $m$  ДВК поддерживается при тех же условиях, что и для предыдущего варианта обмена. Дополнительно работоспособность при варианте В3 обмена обеспечивается, когда не менее чем  $m$  исправных ДВК с исправными адаптерами внутренней связи А подключаются через  $m$  исправных СА к первой магистрали (либо ко второй магистрали).

Для варианта обмена В3 для кластера на рис. 3, а и рис. 2, в требование исправности адаптеров А не менее чем в  $m$  исправных ДВК заменяется на ис-

правность в них пары СА, выделенных для внутрикомплексных обменов.

Для кластера с вариантом обмена В1 (В2) коэффициент готовности кластера

$$P_1 = \sum_{i=m}^M C_M^i (pp_{ca}^2)^i (1 - pp_{ca}^2)^{M-i},$$

где  $C_M^i = m!/i!(M-i)!$ ,  $m! = 1 \cdot 2 \cdot \dots \cdot m$ .

Для кластера с вариантом обмена В3 с учетом дополнительных возможностей связи коэффициент готовности равен

$$P_3 = P_1 + D, D = \sum_{i=m}^M C_M^i g^i (1-g)^{M-i} 2G_i (1-G_i);$$

$$G_i = \sum_{s=m}^i C_i^s (p_{ca})^s (1-p_{ca})^{i-s},$$

причем для структур по рис. 3, б имеем  $g = pp_a$ , а по рис. 3, а  $g = (pp_{ca})^2$ .

В вышеприведенных формулах  $p$  — вероятность работоспособных состояний ДВК, обеспечивающих дублированные вычисления (определяется на основе приведенной выше марковской модели);  $p_{ca} = \mu_{ca}/(\mu_{ca} + \lambda_{ca})$ ,  $p_a = \mu_a/(\mu_a + \lambda_a)$  — коэффициенты готовности СА и адаптера внутрикомплексной связи, причем  $\lambda_a$ ,  $\mu_a$ ,  $\lambda_{ca}$ ,  $\mu_{ca}$  — интенсивности отказов и восстановлений адаптеров внутрикомплексной связи и СА.

$$\text{При } M = m \quad P_1 = (pp_{ca}^2)^m,$$

$$P_3 = (pp_{ca}^2)^m + 2(pp_a)^m p_{ca}^m (1 - p_{ca}^m).$$

Расчеты проведем при  $\lambda_a = \lambda_{ca} = 0,0001 \text{ ч}^{-1}$ ,  $\mu_a = \mu_{ca} = 0,5 \text{ ч}^{-1}$  и коэффициенте готовности базового оборудования ДВК (см. рис. 1)  $p = 0,9978$ . Полученные зависимости коэффициента готовности от числа ДВК представлены на рис. 8 при  $M = m$ .

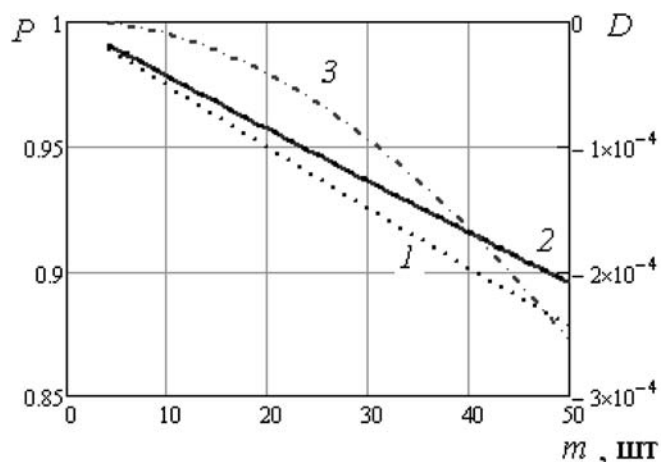


Рис. 8. Коэффициент готовности системы

На рис. 8 кривая 1 соответствует коэффициенту готовности кластера при варианте обмена В1 (В2), а кривая 2 — при варианте обмена В3 (для структур, приведенных на рис. 3). Кривая 3 представляет разность  $D$  коэффициентов готовности структур рис. 3, б и 3, а при варианте обмена В3. Проведенные расчеты показывают, что вариант обмена В3 при дублированных вычислениях позволяет существенно повысить надежность кластера ДВК без увеличения расходов на его реализацию.

### Среднее время пребывания запросов межкомплексного обмена при накоплении отказов

Рассмотрим влияние отказов СА и адаптеров внутрикомплексной связи (А) на среднее время пребывания запросов межкомплексного обмена при дублированных вычислениях. Поскольку вариант обмена В3 предпочтительней, он используется как основной, переход на другой вариант обмена происходит, когда отказы адаптеров не позволяют реализовать вариант обмена В3.

Рассмотрим состояние системы с исправностью СА всех  $m$  ДВК и отказом адаптеров внутрикомплексного обмена для  $i$  ДВК. В этом случае при адресации к ДВК с отказом адаптеров внутрикомплексной связи А реализуется обмен по варианту В1, а при адресации к остальным  $(m - i)$  ДВК — по варианту В3. Если суммарная интенсивность запросов в кластере на межкомплексный обмен  $\lambda$ , то  $(i/m)$ -я их часть требует передачи по двум магистралям, а  $[(i - 1)/m]$ -я часть — по одной из двух магистралей, в результате интенсивность запросов, передаваемых через каждую магистраль, равна

$$\left[ \frac{i}{m} (2\lambda) + \left( \frac{m-i}{m} \right) \frac{\lambda}{2} \right] = \frac{\lambda(i+m)}{2m},$$

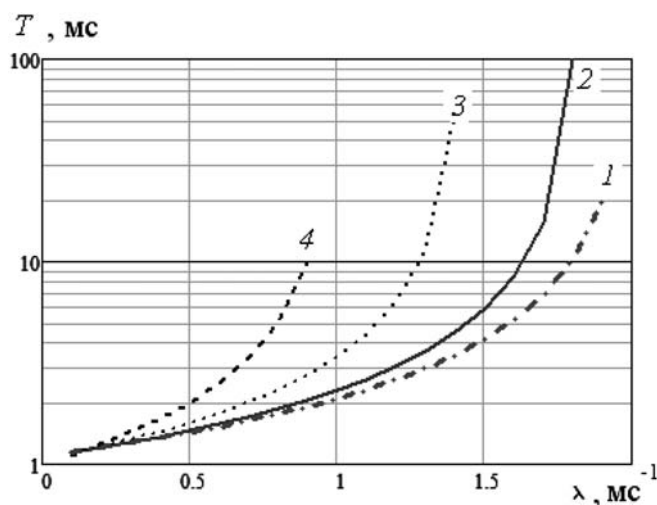


Рис. 9. Зависимость среднего времени пребывания запросов на межкомплексный обмен от их интенсивности при накоплении отказов адаптеров внутрикомплексной связи

а среднее время пребывания запросов на межкомплексный обмен

$$T = [v/(1 - \lambda(i + m)v/2m)] + \alpha v(m - i)/m.$$

Зависимость среднего времени пребывания запросов на межкомплексный обмен от их интенсивности в кластере из  $m = 10$  ДВК при накоплении отказов адаптеров внутрикомплексной связи представлена на рис. 9, на котором кривые 1–4 соответствуют отказу адаптеров внутрикомплексной связи в  $i = 0, 1, 4, 10$  ДВК.

Взаимосвязь через две магистрали между  $m$  ДВК реализуется, если все они подключены хотя бы к одной из магистралей вне зависимости от числа  $i$  отказов СА подключения ДВК ко второй магистрали, влияющего на среднее время пребывания запросов на межкомплексный обмен.

Рассмотрим случай подключения к одной из магистралей (первой) всех  $m$  ДВК, среди которых  $i$  ДВК не подключены ко второй магистрали из-за отказов СА. Если все  $m$  ДВК содержат исправные адаптеры внутрикомплексного обмена А и если при возможности передачи через любую магистраль ее выбор проводится случайно (равновероятно), то доля запросов, передаваемых через первую магистраль, равна

$$\pi_i = \frac{i}{m} + \left( \frac{m-i}{m} \right) \left( \frac{i}{m-1} \right) + \frac{1}{2} \left( \frac{m-i}{m} \right) \left( \frac{m-i-1}{m-1} \right),$$

а среднее время пребывания запросов на межкомплексный обмен при выполнении условия стационарности  $\lambda \pi_i v < 1$  вычисляется по формуле

$$T_i = \pi_i \frac{v}{1 - \lambda \pi_i v} + (1 - \pi_i) \frac{v}{1 - \lambda (1 - \pi_i) v} + \alpha v.$$

В кластере из  $m = 10$  ДВК при накоплении отказов СА зависимость среднего времени пребывания

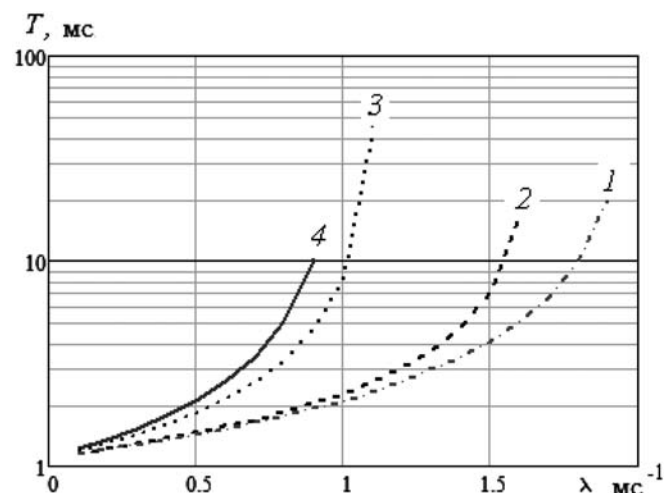


Рис. 10. Зависимость среднего времени пребывания запросов на межкомплексный обмен от их интенсивности при накоплении отказов СА

запросов на межкомплексный обмен от их интенсивности представлена на рис. 10, на котором кривые 1–4 соответствуют отказу СА в  $i = 0, 1, 5, 10$  ДВК.

Проведенные расчеты подтверждают, что при варианте межкомплексного обмена ВЗ кластеры ДВК обладают высокой устойчивостью к отказам адаптеров связи (А и СА) и к сохранению эффективности функционирования по мере накопления отказов.

Дополнительно повысить эффективность обмена по варианту ВЗ при накоплении отказов адаптеров потенциально возможно при балансировке нагрузки на магистрали.

$$\text{Если доля запросов } B_i = \frac{i}{m} + \left(\frac{m-i}{m}\right) \left(\frac{i}{m-1}\right),$$

передаваемых только через магистраль, к которой подключены  $i$  неисправных СА, не превосходит 0,5 (меньше доли запросов, которые могут быть переданы через любую из двух магистралей), то сбалансированность загрузки магистралей может быть достигнута. Если  $B_i < 0,5$ , то загрузка магистралей и, соответственно, время пребывания передаваемых через них запросов будут разными. Таким образом, среднее время пребывания запросов на межкомплексный обмен при реализации балансировки загрузки магистралей можно определить как

$$T_3 = \begin{cases} \frac{v}{1 - 0,5\lambda v} + \alpha v, & \text{если } B_i \leq 0,5; \\ B_i \left( \frac{v}{1 - B_i \lambda v} \right) + (1 - B_i) \left( \frac{v}{1 - (1 - B_i) \lambda v} \right) + \alpha v, & \text{если } B_i > 0,5. \end{cases}$$

В кластере из  $m = 10$  ДВК зависимость среднего времени пребывания запросов на межкомплексный обмен по варианту ВЗ от числа  $i$  отказавших СА представлена на рис. 11. Кривые 1 и 2 представляют обмен с балансировкой и без балансировки загрузки магистралей при интенсивности запросов  $\lambda = 1 \text{ мс}^{-1}$ , а кривые 3 и 4 — при  $\lambda = 1,1 \text{ мс}^{-1}$ .

Расчеты подтверждают эффективность балансировки нагрузки при распределении запросов на межкомплексный обмен по варианту ВЗ.

Таким образом, при объединении ДВК в кластерные системы с организацией дублированных вычислений наиболее эффективен межкомплексный обмен ВЗ при передаче данных адресуемому ДВК по одной из магистралей с дальнейшей их внутри-

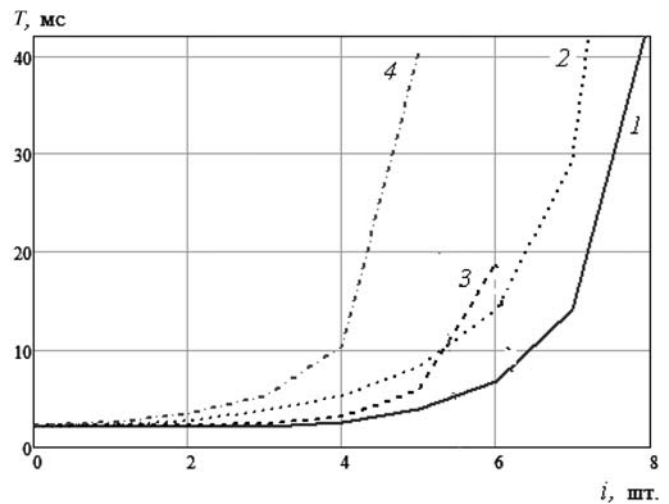


Рис. 11. Зависимость среднего времени пребывания запросов на межкомплексный обмен от числа отказавших СА с балансировкой нагрузки магистралей и без нее

комплексной передачей между модулями памяти полукластеров ДВК. Указанный вариант обмена эффективен как по времени доставки данных, так и по отказоустойчивости кластеров, причем он не требует дополнительных затрат на построение системы, и его эффективность возрастает при увеличении интенсивности межкомплексных обменов.

#### Список литературы

1. **Половко А. М., Гуров С. В.** Основы теории надежности. СПб.: БХВ-Петербург. 2006. 704 с.
2. **Богатырев В. А.** Производительность канала передачи данных с пропорциональным множественным доступом при резервировании неполнодоступных магистралей // Электронное моделирование. 1997. № 4. С. 34–40.
3. **Богатырев В. А.** К оценке среднего времени ожидания в локальных сетях с резервированием магистралей при учете издержек на маркерный доступ // Электронное моделирование. 1997. № 6. С. 31–37.
4. **Богатырев В. А.** Надежность канала взаимосвязи в локальных сетях с резервированием магистралей // Электронное моделирование. 1998. № 3. С. 99–103.
5. **Богатырев В. А.** Об организации и анализе разделения обслуживания очередей узлов между резервированными магистралями // Электронное моделирование. 1998. № 2. С. 107–108.
6. **Богатырев В. А.** К оценке среднего времени ожидания передачи кадров через резервированный канал с ограниченными коммуникационными возможностями магистралей // Автоматика и вычислительная техника. 1998. № 1. С. 77–80.
7. **Клейнрок Л.** Теория массового обслуживания. М.: Машиностроение, 1979. 432 с.

УДК 01.519.7

**Э. И. Владимирский,**  
канд. техн. наук, ст. науч. сотр.,  
e-mail: eduard.vladimirsky@hotmail.com  
**Б. И. Исмаилов,** науч. сотр.,  
Азербайджанская государственная  
нефтяная академия, г. Баку

## Синхронизация в управлении хаотическими системами

*Предложен подход к выбору структуры произвольной пространственно-распределенной хаотической системы. Управление такой системой и идентификация скрытой передачи информации реализуется методом синхронизации в терминах нелинейного рекуррентного анализа.*

**Ключевые слова:** пространственно-распределенная система, перколяция, синхронизация, нелинейный рекуррентный анализ, отображения Хенона, Чирикова, Икеды, скрытая передача информации

Проблема управления пространственно-распределенными хаотическими системами продолжает оставаться актуальной и важной.

Этой проблеме посвящены работы Анищенко В. С. [1], Лоскутова А. Ю. [2, 3], Трубецкова Д. И., Короновского А. А., Храмова А. Е. [4], Фрадкова Ф. Л. [5] и др., в которых рассматривались различные, в том числе решеточные модели, аппроксимирующие поведение пространственно-распределенных систем. Однако принципиальным вопросом является обоснование выбора структуры связей и учет гетерогенности составляющих систему отображений.

В связи с этим в статье предлагается комплексный подход в реализации задачи управления пространственно-распределенной хаотической системой.

Так, на первом этапе произвольная структура пространственно-распределенной системы с элементами гетерогенных хаотических отображений натягивается на квадратную перколяционную решетку (решетку протекания), т. е. выполняется операция вложения в целях формирования структуры связей отображений [6].

На втором этапе структура системы в присутствии шума отображается на квадратную матрицу с одновременным формированием портфеля скрытой передачи информации.

Управление системой и идентификация скрытой передачи информации реализуется методом синхронизации в терминах нелинейного рекуррентного анализа [7].

Предлагаемый в статье подход отличается от подходов, использующих различные традиционные схемы синхронизации как при анализе хаотических систем, так и при передаче скрытой информации [4].

### Вложение пространственно-распределенной системы в перколяционную решетку. Общие положения

**Постановка задачи.** Пусть  $\tilde{S} = \{s_i\}, i = \overline{1, n}$ , — произвольная пространственно-распределенная система с элементами гетерогенных хаотических отображений  $\tilde{S}_i, i = \overline{1, n}$ , типа

$$X_{n+1} = f(X_n), n = 0, 1, 2, \dots; X = (x_1, \dots, x_n);$$

$$f = (f_1, \dots, f_n).$$

Пусть задана перколяционная решетка размерностью  $L \times L$  в пространстве  $E^2$ , рассматриваемая как совокупность узлов  $M = \{m_i\}, i = \overline{1, n}$ , и связей  $C = \{c_j\}, j = \overline{1, k}$ . Имеет место аксиома вложения.

**Аксиома вложения.** Пусть некоторый оператор  $\tau$  вложения  $\tilde{S}$  в  $n$ -мерную решетку  $Z = (M, C)$  имеет вид

$$\tau: G \Rightarrow Z, \quad (1)$$

где 
$$G \Rightarrow Z = \begin{cases} 1 & \text{при } s_i \in C, \\ 0 & \text{в противном случае,} \end{cases} \quad (2)$$

и пусть  $\tilde{S} \Rightarrow Z$  будет рассматриваться по ассоциации с процессами протекания, т. е. как ситуация, возникающая при случайном распространении "жидкости через среды" [8].

Тогда перколяционная решетка  $Z = (M, C)$  будет представлять собой геометрическую и динамическую реализацию структуры системы  $S = \{s_i\}, i = \overline{1, n}$ , в условиях перколяции (протекания).

**Замечание.** Полученная структура характеризует фазовую траекторию системы. Основываясь только на одном измерении, можно восстановить топологию полной фазовой траектории, как если бы мы использовали для этого все компоненты системы. Такой подход впервые был предложен в 1980 г. группой американских ученых [9], показавших, что можно получить удовлетворительную картину странного аттрактора, если вместо переменных  $x$ , входящих в уравнение динамической системы, использовать  $n$ -мерные векторы, получаемые из элементов временного ряда. Это обстоятельство катализировало использование нелинейного рекуррентного анализа [7] в задачах прогнозирования и управления хаотическими системами [10]. ■

Фундаментальным свойством динамических диссипативных систем является рекуррентность (по-



вторяемость) состояний в смысле прохождения последующей траектории достаточно близко к предыдущей. Это свойство было отмечено в 80-х годах XIX века французским математиком Пуанкаре (Poincare) и сформулировано в виде "теоремы рекуррентности" [7].

Данная теорема лежит в основе современного учения о преобразованиях, сохраняющих меру, известного как эргодическая теория. При ее доказательстве Пуанкаре предвосхитил как понятие меры, так и понятие категории.

Рассмотрим ограниченную открытую область  $X$  в  $n$ -мерном пространстве с гомеоморфизмом  $T$  для  $X$  на себя, сохраняющим объем, т. е. для любого открытого множества  $G \subset X$  вводится предположение равенства объемов  $G$  и  $T(G)$ .

При многократном применении  $T$  любая точка  $x$  порождает последовательность  $x, T_x^1, T_x^2, \dots, T_{(x)}^i, \dots$ , называемую положительной полуорбитой  $x$ . Когда  $X \in G$  для бесконечного множества целых положительных значений  $i$ , то говорят о возвращающейся точке  $x$  открытого множества  $G$ .

Тогда на содержательном уровне теорема Пуанкаре гласит, что для любого открытого множества  $G \subset X$  точками, возвращающимися относительно  $G$ , являются все точки  $G$ , кроме некоторого множества первой категории меры нуль.

Итак, рекуррентность состояний в том понимании, что они (состояния) являются повторяющимися по прошествии некоторого времени, есть фундаментальное свойство диссипативных динамических систем.

Формально имеет место следующая теорема.

**Теорема** [7]. Пусть  $T$  — сохраняющее меру преобразование пространства с конечной мерой  $(X, \mu)$  и пусть  $A \subset X$  — измеримое множество. Тогда для любого натурального  $N \in \mathbb{N}$

$$A(\{x \in A: \{T^n(x)\}_{n \geq N} \subset (X \setminus A)\}) = 0, \quad (3)$$

где  $T$  — рекуррентное время;  $X$  — произвольное измеримое множество;  $\mu(\cdot)$  — вероятностная мера;  $x$  — параметр нормализованного ряда;  $N$  — множество натуральных чисел.

Представим краткое доказательство теоремы.

*Доказательство.* Заменяем  $T$  на  $T^n$  в формуле (3) и удостоверимся в достаточности выражения при  $N=1$ .

Множество  $\tilde{A} := \{x \in A | \{T^n(x)\}_{n \in \mathbb{N}} \subset X \setminus A\} = A \cap \left( \bigcap_{n=1}^{\infty} T^{-n}(X \setminus A) \right)$  — измеримо.

Запишем  $T^{-n}(\tilde{A}) \cap \tilde{A} = \emptyset$  для всех  $n$  в предположении, что инверсия  $T^{-n}(\tilde{A}) = B$  с  $B \neq \emptyset$ , подразумеваемая при этом, что  $T^n(B) \subset A$ , к тому же

$$T^{-n}(\tilde{A}) \cap T^m(\tilde{A}) = \emptyset \quad \forall m, n \in \mathbb{N},$$

так как принимается  $T^{-n}(\tilde{A}) \cap T^{-m}(\tilde{A}) = B$  с  $B \neq \emptyset$ ,

и учитывая при этом, что  $T^n(B) \subset \tilde{A}$  и  $T^m(B) \subset \tilde{A}$ .

Не теряя общности, принимается  $m > n$ . Тогда  $T^n(B) = C \subset \tilde{A}$  и  $T^m(B) = T^{m-n}(T^n(B)) = T^{m-n}(C) \subset \tilde{A}$ , которое, с другой стороны, противоречит определению  $\tilde{A}$ .

Кроме того,  $\mu(T^{-n}(\tilde{A})) = \mu(\tilde{A}) \subset T$ , сохраняющим  $\mu$ . Таким образом,  $\mu(\tilde{A}) = 0$ , так как

$$I = \mu(X) = \mu\left(\bigcup_{n=0}^{\infty} T^{-n}(\tilde{A})\right) = \sum_{n=0}^{\infty} \mu(T^{-n}(\tilde{A})) = \sum_{n=0}^{\infty} \mu(\tilde{A}).$$

Отсюда имеем сохраняющее меру преобразование, где траектория при известных обстоятельствах стремится вернуться в режим, достаточно близкий к предыдущему, или проходит через похожие ступени развития.

Однако возможности такого анализа сильно ограничены, так как зачастую системы обладают многомерными фазовыми траекториями, что делает возможным изображать только проекции таких траекторий в двух- или трехмерных пространствах.

В связи с этим использование инструментария, предложенного в работе [9], позволяет работать практически с любыми рядами.

Кроме того, преимущество используемого графического инструмента состоит в том, что он позволяет исследовать  $m$ -размерную траекторию фазового пространства посредством двумерного представления его рекуррентности и, следовательно, видеть динамическую картину в целом.

### Синхронизация. Управление системами в присутствии шума

Известно, что режим обобщенной синхронизации может наблюдаться в системах с диссипативным и недиссипативным типами связей [11].

Пусть для систем с диссипативной связью уравнения, описывающие динамику взаимодействующих систем в присутствии шума, будут записаны в виде [11]

$$\dot{\mathbf{x}}(t) = \mathbf{G}(\mathbf{x}(t), \mathbf{g}_d);$$

$$\dot{\mathbf{u}}(t) = \mathbf{H}(\mathbf{u}(t), \mathbf{g}_r) + \varepsilon \mathbf{A}(\mathbf{x}(t) - \mathbf{u}(t) + D\xi(t)), \quad (4)$$

где  $\mathbf{x}(t)$  и  $\mathbf{u}(t)$  — векторы состояния ведущей и ведомой систем соответственно;  $\xi(t)$  — шумовой сигнал;  $\mathbf{G}$  и  $\mathbf{H}$  определяют векторные поля взаимодействующих систем;  $\mathbf{g}_d$  и  $\mathbf{g}_r$  — векторы управляющих параметров;  $\mathbf{A} = \{\delta_{ij}\}$  — матрица связи,  $\delta_{ii} = 0$  или  $\delta_{ii} = 1$ ,  $\delta_{ij} = 0$  ( $i \neq j$ );  $\varepsilon$  — параметр связи;  $D$  — интенсивность шума.

Пусть ведомая система  $u(t)$  будет представлена как модифицированная система [11]:

$$\dot{\mathbf{u}}_m(t) = \mathbf{H}'(\mathbf{u}_m(t); \mathbf{g}_r, \varepsilon), \quad (5)$$

а под внешним воздействием  $\varepsilon(ax(t) + D\xi(t))$  —

$$\dot{\mathbf{u}}_m(t) = \mathbf{H}'(\mathbf{u}_m(t); \mathbf{g}_r, \varepsilon) + \varepsilon(\mathbf{A}\mathbf{x}(t) + D\xi(t)), \quad (6)$$

где  $\mathbf{H}'(\mathbf{u}(t)) = \mathbf{H}(\mathbf{u}(t)) - \varepsilon\mathbf{A}\mathbf{u}(t)$ . Слагаемое  $\varepsilon\mathbf{A}\mathbf{u}(t)$  вносит дополнительную диссипацию в модифицированную систему (5).

Тогда режим обобщенной синхронизации, возникающий в системе (4), может быть рассмотрен как следствие двух взаимосвязанных процессов, протекающих одновременно: увеличение диссипации в модифицированной системе (5) и возрастание амплитуды внешнего (хаотического и шумового) сигнала [11]. Скрытая информация может быть представлена в виде семантической матрицы (semantic matrices)  $m_s \times m_c$ , сформированной по критерию

$$m_s \times m_c \stackrel{\text{def}}{=} \sup_{sem} \inf_v I, \quad (7)$$

т. е. максимально значащей с семантической точки зрения  $\sup_{sem}$  и минимальной по объему данных  $\inf_v$ .

Таким образом, в контексте вышеизложенного имеет место следующая постановка задачи.

**Постановка задачи.** Пусть  $\mathbf{x}(t)|_1^N$ ,  $\mathbf{Y}(t)|_1^N$ ,  $\mathbf{Z}(t)|_1^N$  —  $m$ -мерные фазовые траектории состояний систем длиной  $N$ ,  $\{\mathbf{x}(t), \mathbf{Y}(t), \mathbf{Z}(t)\}_1^N \in \mathbf{R}^m$ .

Пусть  $\mathbf{x}(t)|_1^N$ ,  $\mathbf{Y}(t)|_1^N$ ,  $\mathbf{Z}(t)|_1^N$  соответствуют временные ряды  $\{\mathbf{u}_i\}_1^{\tilde{N}}$ ,  $\{\mathbf{v}_i\}_1^{\tilde{N}}$ ,  $\{\omega_i\}_1^{\tilde{N}}$  соответственно, где  $\mathbf{u}_i = u(t_i)$ ;  $\mathbf{v}_i = v(t_i)$ ;  $\mathbf{Z}_i = Z(t_i)$ ;  $t_i = i\Delta t$ ;  $\tilde{N}$  — длина ряда (число наблюдений);  $\Delta t$  — интервал выборки.

Пусть заданы [1, 9, 10, 12, 13] следующие отображения.

**Отображение Хенона** представляет собой обратимое двумерное двухпараметрическое отображение вида

$$\begin{aligned} x_{n+1} &= 1 - \alpha x_n^2 + y_n, \\ y_{n+1} &= \beta x_n, \end{aligned}$$

где  $x_n, y_n$  — динамические переменные;  $\alpha, \beta$  — параметры системы.

Отображение Хенона является диссипативным при условии  $0 < \beta < 1$ . При  $\beta = 1$  отображение представляет собой консервативную систему.

**Отображение Чурикова—Тейлора** получено при решении задачи об устойчивости движения заряженных частиц в магнитных ловушках. Представляет собой уравнение вида

$$\begin{aligned} x' &= x - k \sin(y) \pmod{2\pi}, \\ y &= y + k \sin(y) \pmod{2\pi}, \end{aligned}$$

где  $k$  — параметр, определяющий ширину эргодического слоя, отделяющего захваченные частицы от пролетных.

При  $k = 0,9$  отображение имеет вид, представленный на рис. 1 (см. четвертую сторону обложки),

фазовый портрет которого напоминает сэндвич с бесконечным числом чередующихся инвариантных кривых и стохастических слоев.

Красным, голубым, желтым и малиновым цветами отмечены области хаоса.

**Отображение Икеды** возникает при моделировании оптических носителей информации и имеет вид

$$\begin{aligned} x_{n+1} &= 1 + u(x_n \cos t_n - y_n \sin t_n), \\ y_{n+1} &= 1 + u(x_n \cos t_n + y_n \sin t_n), \end{aligned}$$

где  $u$  — параметр, и

$$t_n = 0,4 - \frac{b}{1 + x_n^2 + y_n^2}.$$

Для  $u = 0,9$  система имеет хаотический аттрактор (рис. 1, см. четвертую сторону обложки).

Пусть длины всех модельных рядов указанных отображений составляют  $N = 5000$  точек, длины реальных рядов наблюдений различны.

В качестве взаимодействующих осцилляторов будут выступать два разных отображения, в том числе и с разной размерностью фазового пространства, а в качестве шумовой составляющей будет использоваться отображение Хенона.

Пусть управление динамикой хаотических процессов будет трактоваться в терминах нелинейного рекуррентного анализа (Nonlinear Recurrence Analysis) [7].

Требуется реализовать:

- задачу структуризации пространственно-распределенной хаотической системы в перколяционной решетке протекания (т. е. получить кластер системы);
- рекуррентную диаграмму, характеризующую процесс управления системой с идентификацией скрытой передачи информации.

Чтобы реализовать алгоритм структуризации системы, проводим вложение модели пространственно-распределенной хаотической системы  $\tilde{S} = \{s_i\}$ ,  $i = \overline{1, n}$ , (например, диффузионные, сетевые, иерархические, фрактальные модели) в перколяционную решетку размерностью  $L \times L$  в пространстве  $E^2$ , по правилу (1) и (2). На перколяционной решетке образуется произвольный кластер (рис. 2, см. четвертую сторону обложки). Полученная структура представляет собой последовательно соединенные гетерогенные отображения.

### Реализация алгоритма управления пространственно-распределенной системой

- Отображаем временные ряды  $\{\bar{u}_i\}_1^N$ ,  $\{\bar{v}_i\}_1^N$ ,  $\{\omega_i(t)\}_1^{\tilde{N}}$  соответствующих хаотических процессов на квадратные матрицы  $[\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3] \in \mathbf{R}^2$ :

$$\mathbf{M}_1^3 : \{\mathbf{u}_i, \mathbf{v}_i, \omega_i\}_1^{\tilde{N}} \Rightarrow [\tilde{N} \times \tilde{N}]_1^3, i = \overline{1, \tilde{N}}. \quad (8)$$

- Определяем воздействие помехи как

$$\mathbf{G}_1: \mathbf{M}_1 \rightarrow \mathbf{M}_2. \quad (9)$$

- Результирующее возмущенное отображение определяется как

$$\mathbf{M}_{G1} = \mathbf{M}_1 + \mathbf{M}_2, \mathbf{M}_{G1} \in \mathbf{R}^2. \quad (10)$$

- Формируем подматрицу  $(m_s \times n_c)$  матрицы  $\mathbf{M}_{G1} \in \mathbf{R}^2$  как

$$\sup_{sem} \inf_v I = [X, Y; m_s \times n_c] \subset \mathbf{M}_{G1}, \quad (11)$$

где  $X$  и  $Y$  — координаты верхнего левого угла матрицы  $\mathbf{M}_{G1}$ .

Объединяем в одном квазифазовом пространстве две траектории по правилу

$$\mathbf{M}_{G2} = \mathbf{M}_{G1} \cdot \mathbf{M}_3, \mathbf{M}_{G2} \in \mathbf{R}^2, \quad (12)$$

где  $\mathbf{M}_3: \{\omega_j\}_1^{\tilde{N}} \Rightarrow [\tilde{N} \times \tilde{N}]$ ,  $\mathbf{M}_3 \in \mathbf{R}^2$ .

Сравниваем попадание всех точек второй траектории в  $\varepsilon$ -окрестность всех точек первой траектории по кросс-рекуррентной диаграмме (*cross-recurrence plot*, CRP) [7]:

$$\mathbf{M}_{G2} = \mathbf{C} R_{i,j}^\varepsilon = \theta(\varepsilon - \|x_i - y_j\|, x, y \in R^m, \\ i = \overline{1, N_x}, j = \overline{1, N_y}, \quad (13)$$

где  $N$  — число рассматриваемых состояний  $x_i$ ;  $\varepsilon_j$  — граничное расстояние или расстояние от центра выбранной окрестности (точка  $x(i)$ ) до ее границы;  $\|\cdot\|$  — норма;  $\theta(\cdot)$  — функция Хевисайда.

Здесь попадание точки  $y_j$  в  $\varepsilon$ -окрестности точки  $x_i$  отмечается на диаграмме черной точкой в позиции  $(i, j)$  (см. рис. 1).

Квазисинхронизация имеет место в случае совпадения переменных состояний обеих траекторий (так как находятся в одном фазовом пространстве).

Необходимо отметить, что диагональные структуры CRP отражают подобную эволюцию двух изучаемых траекторий в течение некоторого времени.

Кроме того, критерием существования обобщенной синхронизации является отрицательность старшего условного ляпуновского показателя  $\lambda$  [4] и на этой основе фрактальная (ляпуновская) размерность Каплана—Йорке  $D_{KY}$  [15].

### Идентификация скрытой информации

На рекуррентной диаграмме  $S$  (см. рис. 1) выделяем подматрицу  $m_s \times m_c$  матрицы  $\mathbf{M}_{G1}$ ,  $(m_s \times m_c) \subset \mathbf{M}_{G1}$  по кодированной метке  $l$  (*label*).

Таким образом, рис. 1 демонстрирует процесс квазисинхронизации связанных гетерогенных хаотических систем типа Чириков—Хенон с отобра-

жением Икеды и операции идентификации скрытой передачи информации  $\{X, Y; m_s \times m_c\} \subset \mathbf{M}_{G1}$ .

\* \* \*

В заключение отметим, что режим квазисинхронизации противодействует хаотическому движению и используется как стабилизирующий и управляющий факторы в пространственно-распределенной системе. Показано, что принудительная квазисинхронизация позволяет переводить систему из хаотического режима работы в регулярный. В работе в качестве функциональной зависимости между ведущей и ведомой системами использована кросс-рекуррентная функция (13). Важно также отметить нетрадиционный подход к идентификации скрытой передачи информации. Предложенный алгоритм реализован в программной среде MATLAB.

### Список литературы

1. Анищенко В. С., Вадивасова Т. Е., Астахов В. В. Нелинейная динамика хаотических и стохастических систем. Саратов: Изд-во Саратовского ун-та, 1999. 368 с.
2. Лоскутов А. Ю., Прохоров А. К., Рыбалко С. Д. Динамика неоднородных цепочек связанных квадратичных отображений // Теор. и математ. физика. 2002. Т. 132, № 1. С. 105—125.
3. Лоскутов А. Ю., Михайлов А. С. Основы теории сложных систем. М.: Ижевск: Институт компьютерных исследований, 2007. 620 с.
4. Короновский А. А., Москаленко А. И., Трубецков Д. И., Храмов А. Е. Обобщенная синхронизация и синхронизация, индуцированная шумом, — единый тип поведения связанных хаотических систем // Доклады Академии наук. 2006. Т. 407, № 6. С. 1—5.
5. Андриевский Б. Р., Матвеев А. С., Фрадков Ф. Л. Управление и оценивание при информационных ограничениях: к единой теории управления, вычислений и связи // АИТ. 2010. Вып. 4. С. 34—99.
6. Владимирский Э. И. Фрактальность алгебраических структур // Тр. IV Международной конф. "Идентификация систем и задачи управления" SICPRO'05, Москва. 25—28 января 2005. С. 492—501.
7. Marwan N., Romano M. C., Thiel M., Kurths J. Recurrence plots for the analysis of complex systems // Physics Report 437. 2007. P. 237—329.
8. Федер Е. Фракталы. М.: Мир, 1991. 254 с.
9. Eckmann J. P., Kamphorst S. A., Ruelle D. Recurrence plots of dynamical systems Europhys // Lett. 1987. Vol. 4, N 9. P. 973—977.
10. Владимирский Э. И., Исмаилов Б. И. Синергетические аспекты нелинейного рекуррентного анализа хаотической информации // Тез. докл. Международной научно-практической конф. "Информационные технологии и компьютерная инженерия", г. Винница, Украина, 19—21 мая 2010. С. 96—97.
11. Короновский А. А., Москаленко А. И., Храмов А. Е. Скрытая передача информации на основе режима обобщенной синхронизации в присутствии шумов // ЖТФ. 2010. Т. 80. Вып. 4. С. 1—8.
12. Морозов А. Д., Драгунов Т. Н. Визуализация и анализ инвариантных множеств динамических систем. Москва—Ижевск: Институт компьютерных исследований, 2003. 304 с.
13. Кузнецов А. П., Савин А. В., Савин Д. В. Особенности динамики почти консервативного отображения Икеды // Письма в ЖТФ. 2007. Т. 33. Вып. 3. С. 57—63.
14. Spott J. C. Time-Series Analysis. Oxford University Press, 2003. 528 p.
15. Владимирский Э. И., Тагиев Ф. К. Синергетический подход к формированию интегральных размерностей в интеллектуальных информационно-измерительных системах // Информационные технологии. 2010. № 6 (166). С. 62—67.

Ю. А. Зак, д-р техн. наук, науч. консультант,  
Аахен, Германия  
e-mail: yuriy\_zack@notmail.com

## Адаптивные методы глобального случайного поиска в решении одного класса многоэкстремальных задач в условиях ограничений

*Исследованы свойства допустимых планов для некоторого достаточно широкого класса квазимоноотонных функций цели и ограничений задачи и введены оценки нижней границы оптимальных решений. Предложенные два алгоритма позволяют на основе исключения областей, не содержащих допустимых и оптимальных решений, сократить расходы на поиск, повысить его эффективность и улучшить оценки качества полученного результата.*

**Ключевые слова:** глобальный случайный поиск, многоэкстремальные задачи в условиях ограничений, квазимоноотонные функции

### Введение

В методах глобального случайного поиска на каждой итерации процесса выбор нового решения осуществляется не детерминировано, а в результате реализации некоторого случайного процесса. Выбираемая стратегия может зависеть от предыстории поиска. На основе реализации некоторого случайного процесса на каждой итерации алгоритма определяется значение вектора переменных задачи  $X^k$ . Затем устанавливается принадлежность  $X^k$  к заданной области ограничений  $G$ , и в случае, если  $X^k \in G$ , вычисляется значение функции цели  $F(X^k)$ . В результате строится последовательность независимых случайных реализаций значений векторов переменных задачи  $X^k, r = 1, 2, \dots, K$ , распределенных по некоторому закону в области  $G$ . Для векторов  $X^k \in G$  рассчитываются значения  $F(X^k)$ , которые сравниваются с наилучшим значением функции цели, полученным за предыдущие  $(k - 1)$  шагов —  $F^*(\bar{X}^{k-1})$ , и, применительно к решению задачи минимизации и целевой функции, выполняются следующие вычисления:

$$F^*(\bar{X}^k) = \begin{cases} F(X^k), & \text{if } X^k \in G \ \& \ [F(X^k) < F^*(\bar{X}^{k-1})], \\ F^*(\bar{X}^{k-1}), & \text{if } X^k \notin G \ \vee \ [F(X^k) \geq F^*(\bar{X}^{k-1})]; \end{cases}$$

$$\bar{X}^k = \arg F^*(\bar{X}^k) = \begin{cases} X^k, & \text{if } F^*(\bar{X}^k) = F(X^k), \\ \bar{X}^{k-1}, & \text{if } F^*(\bar{X}^k) = F^*(\bar{X}^{k-1}). \end{cases}$$

Следовательно, сохраняются только одно допустимое значение вектора переменных и соответствующее ему значение целевой функции, наилучшее из всех предыдущих. Эти методы обеспечивают при  $K \rightarrow \infty$  попадание с некоторой наперед заданной вероятностью в окрестность глобального экстремума, в которой  $[F^*(\bar{X}^k) - F(X^*)]/F(X^*) \leq \varepsilon$ , т. е. в некоторую область

$$\tilde{G}_k = \left\{ X \mid X \in G, \frac{F(X) - F(X^*)}{F(X^*)} \leq \varepsilon \right\}.$$

Здесь  $F(X^*)$  — значение функции цели в оптимальном решении.

Данные методы можно применять как при решении задач локальной оптимизации, так и при поиске глобального экстремума.

В алгоритмах направленного случайного поиска результаты вычислений, выполненных на предыдущих итерациях, используются для выбора направлений дальнейших испытаний и изменений вектора оптимизируемых переменных.

Основным методом случайного поиска глобального экстремума многомерных функций является мультистарт [1, 3]. При его использовании из множества  $X^k \in G$  случайно или детерминировано выбирается некоторое подмножество из  $K$  точек. Из каждой  $k$ -й случайной начальной точки  $X^k$  делается локальный спуск в ближайший минимум  $\bar{X}^k$  любым локальным методом поиска. За глобальный минимум  $\bar{X}^*$  принимается тот, для которого показатель качества минимален

$$\bar{X}^* = \arg \min_{1 \leq s \leq S} F(\bar{X}^k).$$

Очевидно, что при  $K \rightarrow \infty$  вероятность того, что вектор  $\bar{X}^*$  является значением глобального минимума, стремится к единице. Однако при  $K \neq \infty$  вероятность получения глобального минимума отлична от нуля, и при остановке алгоритма необходимо оценить точность полученного приближенного решения.

Если область ограничений задачи имеет достаточно сложную структуру, то для реализации алгоритма ее нужно вписать в некоторую область  $Q$ , где  $G \subset Q$ , большего размера и более простой структуры, представляющей собой, например,  $n$ -мерный параллелепипед.

При реализации равномерного распределения вектора  $X$  в области  $Q$  с границами изменения переменных  $x_i \in [h_{1i}, h_{2i}]$  координаты вектора  $X$  могут быть выбраны по формулам  $x_i = h_{1i} + (h_{2i} - h_{1i})\xi_i$ ,  $i = 1, \dots, n$ , где  $\xi_i$  — независимые случайные числа,

равномерно или согласно другому закону распределенные на отрезке  $[0, 1]$ .

Большинство алгоритмов глобального случайного поиска состоит из двух этапов — глобального и локального. Основным достоинством данного метода является его простота и универсальность. Достоинства таких методов заключаются в следующем:

- структура алгоритмов проста, их можно легко реализовать в виде программ;
- легко построить схему метода, гарантирующую сходимость процесса к точке глобального минимума при бесконечном числе реализаций вектора случайных значений переменных;
- эффективность применения метода не снижается в случае наличия в задаче как непрерывных, так и дискретных переменных.

Методам случайного поиска, имеющим наибольшую практическую эффективность, присущи недостатки, характерные для эвристических методов, которые заключаются в следующем:

- четко не определен класс задач, для которых применение этого метода является эффективным;
- трудности реализации и существенное увеличение затрат на поиск в условиях наличия ограничений в виде систем неравенств и равенств и даже получения случайной реализации вектора переменных задачи, удовлетворяющего этой системе ограничений;
- как правило, они требуют очень больших объемов вычислений; скорость сходимости крайне медленная;
- отсутствуют оценки точности полученного приближенного решения;
- отсутствуют четкие рекомендации о выборе параметров методов.

Эти алгоритмы преимущественно используются либо в простых случаях, когда программирование специального метода займет гораздо больше времени, чем поиск решения этим медленным методом, либо в сложных случаях, когда даже неизвестно, каким алгоритмом можно решить эту задачу.

При реализации методов глобального случайного поиска в случае сложной конфигурации области ограничений  $G$ , определяемой системой нелинейных и невыпуклых равенств и неравенств, возникают серьезные вычислительные трудности, связанные с генерацией векторов  $X \in (G \cap Q)$ . В ряде случаев область  $G$  может иметь достаточно сложную конфигурацию, и оптимизационная задача может оказаться многоэкстремальной или задачей математического программирования в несвязанных областях. Это обстоятельство значительно увеличивает затраты на поиск и в большинстве случаев существенно снижает его эффективность.

В работе предлагаются и обосновываются методы, позволяющие на основе установленных авто-

ром свойств допустимых решений задачи и введенных оценок нижней границы оптимальных решений как для всей системы ограничений задачи, так и для отдельных подобластей допустимых планов, существенно сократить затраты на поиск и повысить его эффективность.

Рассматриваются два алгоритма реализации метода. Первый алгоритм предусматривает разбиение области  $G$  на  $R$  подобластей, которые могут быть вписаны в  $R$   $n$ -мерных параллелепипедов минимальных размеров. Для каждой из подобластей вычисляется оценка нижней границы критерия оптимальности. Случайный поиск осуществляется в каждой из подобластей. Области, для которых значение нижней границы больше или равно наилучшему из полученных допустимых значений, исключаются из дальнейшего рассмотрения. Во втором алгоритме в процессе каждой реализации случайного вектора  $X$  проводится отсечение областей, не содержащих допустимых планов, т. е. уменьшение диапазонов изменения подмножества еще не сгенерированных значений переменных. Вычисление оценок и сравнение их с уже полученным значением критерия оптимальности в наилучшем из допустимых значений позволяет остановить данную генерацию случайного вектора.

### 1. Класс рассматриваемых функций цели и ограничений задачи.

#### Сокращение областей поиска. Вычисление оценок

Рассматривается следующая многоэкстремальная задача невыпуклого математического программирования в условиях ограничений:

$$f_0^*(X) = \min_{X \in (H \cap G)} f_0(X); \quad (1)$$

$$G = \{X | f_i(X) \leq b_i, i = 1, \dots, m\}; \quad (2)$$

$$H = \{X = (x_1, \dots, x_n) | h_j^1 \leq x_j \leq h_j^2, j = 1, \dots, n\}. \quad (3)$$

Функции  $f_i(X)$ ,  $i = 0, 1, \dots, m$ , могут быть невыпуклыми, недифференцируемыми и даже разрывными. На некоторые или все компоненты вектора переменных задачи могут накладываться условия целочисленности.

Рассмотрим модификации методов адаптивного случайного поиска для функций, обладающих следующими свойствами.

Функцию  $f(x_1, \dots, x_j, \dots, x_n)$ , заданную на  $n$ -мерном параллелепипеде (3), будем называть сепарабельно квазимоноотонной по переменной  $x_j$  (в дальнейшем будем обозначать ее  $QMF(j)$ ), если для любых фиксированных значений  $\bar{y}_k \in H_k$ ,  $k = 1, \dots, n$ ,  $k \neq j$ , всех остальных переменных и любых двух зна-

чений переменной  $x_j$ , равных  $x_j = g_j$  и  $x_j = l_j$ , где  $h_j^1 \leq g_j < l_j \leq h_j^2$ , выполняется одно из пары условий:

$$\begin{aligned} f(\bar{y}_1, \dots, \bar{y}_{j-1}, g_j, \bar{y}_{j+1}, \dots, \bar{y}_n) &\leq \\ &\leq f(\bar{y}_1, \dots, \bar{y}_{j-1}, l_j, \bar{y}_{j+1}, \dots, \bar{y}_n), \end{aligned} \quad (4)$$

$$\begin{aligned} f(\bar{y}_1, \dots, \bar{y}_{j-1}, g_j, \bar{y}_{j+1}, \dots, \bar{y}_n) &\geq \\ &\geq f(\bar{y}_1, \dots, \bar{y}_{j-1}, l_j, \bar{y}_{j+1}, \dots, \bar{y}_n). \end{aligned} \quad (5)$$

Если  $QMF(j)$  — неубывающая функция по переменной  $x_j$ , то справедливо выражение (4), а если  $QMF(j)$  — невозрастающая функция, то справедливо выражение (5). Далее рассматриваются функции  $f(x_1, \dots, x_j, \dots, x_n)$ , которые являются  $QMF(j)$ -функцией по каждой из переменных. В работе автора [6] показано, что класс таких функций достаточно широк, и приведены многочисленные примеры таких функций.

Рассмотрим поведение и свойства таких функций в системе  $n$ -мерных параллелепипедов  $H_s$ . При  $S \rightarrow \infty$   $H_s$  стягиваются в точку:

$$\begin{aligned} H_s &= H_{1s} \times H_{2s} \times \dots \times H_{js} \times \dots \times H_{ns}, \\ \text{где } H_{js} &= \{x_j: h_{js}^1 \leq x_j \leq h_{js}^2\}, H_{js} \subseteq H_{j,s-1}, \\ & \quad s = 1, \dots, S. \end{aligned} \quad (6)$$

Из условий (2) и (3) следует, что для любых значений  $\bar{y}_k \in H_{ks}$ ,  $k = 1, \dots, n$ ,  $k \neq j$ ,

$$\begin{aligned} \min_{x_j \in H_{js}} f(\bar{y}_1, \dots, \bar{y}_{j-1}, x_j, \bar{y}_{j+1}, \dots, \bar{y}_n) &= \\ = f(\bar{y}_1, \dots, \bar{y}_{j-1}, x_{js}^-, \bar{y}_{j+1}, \dots, \bar{y}_n); \end{aligned} \quad (7)$$

$$\begin{aligned} \max_{x_j \in H_{js}} f(\bar{y}_1, \dots, \bar{y}_{j-1}, x_j, \bar{y}_{j+1}, \dots, \bar{y}_n) &= \\ = f(\bar{y}_1, \dots, \bar{y}_{j-1}, x_{js}^+, \bar{y}_{j+1}, \dots, \bar{y}_n), \end{aligned} \quad (8)$$

где

$$x_{js}^- = \begin{cases} h_{js}^1 \\ h_{js}^2 \end{cases}, x_{js}^+ = \begin{cases} h_{js}^1 \\ h_{js}^2 \end{cases}. \quad (9)$$

Систему параллелепипедов (6), в которых некоторое подмножество переменных может принимать только целочисленные значения, назовем системой частично целочисленных параллелепипедов (*ChP*).

**Теорема 1.** Аргумент минимума и максимума сепарабельно квазимонотонных функций, обладающих свойствами (4), (5), (7)—(9) на  $n$ -мерном па-

раллелепипеде (6), достигается при значении переменных, принадлежащих множеству  $\Lambda_s$ , т. е.

$$\begin{aligned} \arg \min_{(x_1, x_2, \dots, x_n) \in H_s} f(x_1, x_2, \dots, x_j, \dots, x_n) \cap \hat{\Lambda}_s &= \emptyset, \\ \arg \min_{(x_1, x_2, \dots, x_n) \in H_s} f(x_1, x_2, \dots, x_j, \dots, x_n) \cap \hat{\Lambda}_s &= \emptyset. \end{aligned} \quad (10)$$

Доказательство теоремы приведено в работе автора [6].

**Следствия теоремы 1.**

Экстремальные значения сепарабельно квазимонотонных функций следует искать среди вершин  $n$ -мерного параллелепипеда (6), что сводится к перебору и вычислению значений функции в  $2^n$  точках множества  $\Lambda_s$ . Однако для задач большой размерности такой объем вычислений может оказаться значительным.

Особый интерес представляют функции  $f(x_1, \dots, x_j, \dots, x_n)$ , сепарабельные свойства которых по всем переменным не зависят от значений других переменных. Обозначим эти функции  $SQMF$ . Минимум и максимум таких функций по каждой из переменных  $x_j$  не зависит от значения остальных переменных  $x_k \in H_{ks}$ ,  $k = 1, \dots, n$ ,  $k \neq j$ , и всегда достигается при одном и том же граничном значении области определения этой переменной, т. е. при значении  $h_{js}^1$  или  $h_{js}^2$ . Примерами таких функций являются линейные функции, функции произведения переменных при условии, что каждая из переменных не может принимать отрицательных значений, и многие другие функции.

Заметим, что на класс  $SQMF$  и  $QMF$  функций не накладывается никаких других ограничений: они могут быть невыпуклыми, недифференцируемыми, разрывными, целочисленными как по отдельным группам, так и по всем переменным, не заданными аналитически, а вычисленными в результате решения некоторых вспомогательных задач или цифрового моделирования.

В работе автора [6] доказан ряд утверждений и теорем, обосновывающих итеративный процесс выделения областей, не содержащих допустимых планов для систем нелинейных неравенств с двухсторонними ограничениями, функции левых частей которых  $f_i(X)$ ,  $i = 1, \dots, m$ , — суть функции класса  $SQMF$ :

$$\begin{aligned} y_{jis}^{\min} &= \arg \min_{x_j \in H_{js}} \{f_i(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n) \mid \\ & \quad x_k \in H_{ks}, k = 1, \dots, n, k \neq j\}, j = 1, \dots, n. \end{aligned} \quad (11)$$

Обозначим:

$J_i^+$  и  $J_i^-$  — соответственно подмножества индексов  $j \in J$ , для которых минимальное значение функции по данной переменной, т. е. значение

$y_{jis}^{\min}$ , достигается при значениях  $y_{jis}^{\min} = h_{js}^1$ , т. е. по данной переменной функция является неубывающей, и  $y_{jis}^{\min} = h_{js}^2$ , т. е. по данной переменной функция является невозрастающей;

$I_j^+$  и  $I_j^-$  — соответственно подмножества индексов функций  $f_i(X)$ ,  $i = 1, \dots, m$ , являющихся неубывающими и невозрастающими функциями по переменной  $x_j$ ,  $j = 1, \dots, n$ .

**Алгоритм 1.** Введем счетчик числа шагов  $M$ , в течение которых  $H_{s+1} = H_s$ , т. е. не происходит изменения  $n$ -мерного параллелепипеда ни по одной из переменных. В начале процесса положим  $s = 0$ ,  $M = 0$ ,  $H_{j,0} = H_j$ ,  $j = 1, \dots, n$ .

На каждом  $s$ -м шаге итеративного процесса в цикле для каждой переменной  $j = 1, \dots, n$  выполняем вычисления (11) при подстановке значений остальных переменных  $x_k = y_{kis}^{\min}$ ,  $k = 1, \dots, n$ ,  $k \neq j$ . Если выполняется хотя бы одно из неравенств вида

$$f_i(y_{1is}^{\min}, \dots, y_{jis}^{\min}, \dots, y_{nis}^{\min}) > b_i, \quad i = 1, \dots, m, \quad (12)$$

то система ограничений задачи в пределах  $X \in H_s$  является несовместной и алгоритм завершает работу, т. е.  $H_s = \emptyset$ .

Определим значения:

$$t_{jis}^{l,+} = \arg \max_{x_j \in H_{js}} \{f_i(\bar{t}_{1, is}^l, \dots, \bar{t}_{j-1, is}^l, x_j, \bar{t}_{j+1, is}^{l-1}, \dots, \bar{t}_{n, is}^{l-1}) \leq b_i\}, \quad i \in J_i^+, \quad l = 1, \dots, L; \quad (13)$$

$$t_{jis}^{l,-} = \arg \min_{x_j \in H_{js}} \{f_i(\bar{t}_{1, is}^l, \dots, \bar{t}_{j-1, is}^l, x_j, \bar{t}_{j+1, is}^{l-1}, \dots, \bar{t}_{n, is}^{l-1}) \leq b_i\}, \quad i \in J_i^-, \quad l = 1, \dots, L; \quad (14)$$

$$\bar{t}_{js}^{l,+} = \min_{j \in J_i^-} t_{jis}^{l,-}, \quad j \in I_j^+, \quad \bar{t}_{js}^{l,-} = \max_{j \in J_i^+} t_{jis}^{l,+}, \quad j \in I_j^-, \quad l = 1, \dots, L. \quad (15)$$

Здесь

$$\bar{t}_{js}^0 = y_{jis}^{\min}, \quad j \in J, \quad s = 1, \dots, S. \quad (16)$$

Если на каком-то шаге процесса  $l = 1, \dots, L$  хотя бы для одного индекса  $j \in J$  справедливо неравенство  $\bar{t}_{js}^{l,-} > \bar{t}_{js}^{l,+}$ , то в пределах  $n$ -мерного параллелепипеда  $X \in H_s$  не содержится допустимых планов. В противном случае полагаем

$$\bar{t}_{jis}^{(l+1)} = \begin{cases} \bar{t}_{jis}^{l,+}, & \text{if } j \in J_i^+, \\ \bar{t}_{jis}^{l,-}, & \text{if } j \in J_i^-, \end{cases} \quad j = 1, \dots, n, \quad (17)$$

и переходим к следующему  $(l + 1)$ -му шагу итеративного процесса.

Алгоритм удаления областей, не содержащих допустимых решений, завершает свою работу, если на некотором шаге описанного выражениями итеративного процесса получим соотношения

$$\begin{aligned} \bar{t}_{js}^{l,-} &\leq \bar{t}_{js}^{l,+}, \quad \bar{t}_{js}^{l,-} = \bar{t}_{js}^{(l-1),-}, \\ \bar{t}_{js}^{l,+} &= \bar{t}_{js}^{(l-1),+}, \quad j = 1, \dots, n; \end{aligned} \quad (18)$$

$n$ -мерный параллелепипед  $H_s$  преобразуется в параллелепипед  $\bar{H}_s \in H_s$  меньших размеров

$$\begin{aligned} \bar{H}_s &= \{x_j, j = 1, \dots, n | \bar{h}_{js}^1 = \\ &= \bar{t}_{js}^{l,-} \leq x_j \leq \bar{h}_{js}^2 = \bar{t}_{js}^{l,+}\}. \end{aligned} \quad (19)$$

Если на некотором  $s_1$ -м шаге справедливо  $H_{s_1+1} = H_{s_1}$  (т. е.  $\bar{H}_{js_1} = \emptyset$ ), то значение  $M$  увеличиваем на 1. Если  $M = n$ , то полученное подмножество  $H_{s_1+1}$  не может быть подвергнуто дальнейшему изменению (сужению) и итеративный процесс останавливается, т. е. дальнейшее исключение областей, не содержащих допустимых планов, на основе описанного алгоритма невозможно.

Если  $H_{s_1+1} \subset H_{s_1}$  и при этом  $M > 0$ , то полагаем  $M = 0$  и переходим к следующему шагу процесса.

Ясно, что через конечное число шагов процесса мы либо придем к факту о несовместности исходной системы ограничений, либо получим  $n$ -мерный параллелепипед  $\tilde{H} = H_{s+1} \neq \emptyset$ , который не может быть в дальнейшем подвергнут изменениям.

Назовем описанный выше итеративный процесс Ит\_1.

Отметим, что успешное завершение этого итеративного процесса и получение в результате этого  $n$ -мерного параллелепипеда  $\tilde{H} = H_{s+1} \neq \emptyset$  не гарантирует, что система ограничений исходной задачи совместна, а является только необходимым условием существования допустимых решений.

Пусть на некотором шаге реализации вектора случайных значений переменных  $X \in H_s$  определены значения  $z_{js}$  для подмножества индексов переменных  $\Gamma_s \in J$ . Область допустимых значений другой части переменных (подмножества индексов  $j \in V_s = (J/\Gamma_s)$ ) может быть определена в результате итеративного процесса исключения подобластей,

не содержащих допустимых планов, аналогичного (13)—(18), где

$$y_{kis}^{\min} = \begin{cases} z_{ks}, & \text{if } k \in \Gamma_s, \\ \bar{t}_{kis}^{l-}, & \text{if } k \in (J/\Gamma_s) \text{ \& } k \in J_k^-, \\ \bar{t}_{kis}^{l+}, & \text{if } k \in (J/\Gamma_s) \text{ \& } k \in J_k^+, \end{cases} \quad (20)$$

а значения  $\bar{t}_{js}^{l-}$  и  $\bar{t}_{js}^{l+}$ ,  $j \in V_s = (J/\Gamma_s)$ , вычисляются по формуле (17), критерии остановки итеративного процесса определяются согласно (18), граничные значения для подмножества переменных  $j \in V_s$  — по формулам, аналогичным (19). Если на некотором шаге этого итеративного процесса будет установлено  $\bar{t}_{js}^{l-} > \bar{t}_{js}^{l+}$ , то подмножество значений переменных  $x_{js} = z_{js}$ ,  $j \in V_s$ , не содержит допустимых решений. Назовем этот итеративный процесс Ит\_2.

## 2. Алгоритмы адаптивного глобального случайного поиска

Пусть могут быть построены некоторые мажоранты функций

$$f_i(X), \varphi_i(X) \geq f_i(X) \quad \forall X \in Q, \quad i = 1, \dots, m, \quad (21)$$

а также миноранта целевой функции  $F_0(X) \leq f_0(X) \quad \forall X \in Q$ , обладающие следующими свойствами. Функции  $\varphi_i(X)$ ,  $i = 1, \dots, m$ , являются сепарабельно квазимонотонными функциями по каждой переменной  $x_j$  (в дальнейшем будем обозначать ее  $SQMF(j)$ ). Аргументы минимума и максимума таких функций по каждой из переменных на  $n$ -мерном параллелепипеде не зависят от значения всех остальных переменных и лежат на одном и том же конце отрезка ее возможных значений. Функция  $F_0(X)$  относится либо к классу  $QMF(j)$  общего вида, и минимум ее по каждой переменной достигается в одной из вершин  $n$ -мерного параллелепипеда, либо к классу функций, для которых минимальное значение в  $n$ -мерном параллелепипеде может быть определено путем простых вычислений.

На предварительном шаге алгоритма в результате реализации итеративного процесса Ит\_1 для системы неравенств  $\varphi_i(X) \leq b_i$ ,  $i = 1, \dots, m$ ,  $\forall X \in Q$ , проводится исключение областей, не содержащих допустимых решений, и вписывание области  $G$  в  $n$ -мерный параллелепипед  $H \subseteq Q$  меньших размеров. Если выполняются условия (21) и в процессе выполнения итеративного процесса Ит\_1 будет установлена несовместность системы ограничений задачи, то исходная система ограничений задачи является несовместной и процесс решения задачи останавливается. В противном случае на основе простых вычислений (одно вычисление функции

для функций класса  $QMF(j)$ ,  $j = 1, \dots, n$ , и не более чем  $2^n$  вычислений функции для функций класса  $SQMF(j)$ ) находим значение

$$F_0(X^*) = \min_{X \in H} F_0(X),$$

которое является нижней границей оптимального решения задачи.

Если  $X^* \in G$ , то получено решение задачи и нет необходимости в реализации методов глобального случайного поиска. В противном случае переходим к реализации одного из описываемых ниже алгоритмов.

### Алгоритм А1

1. Разобьем исходный  $n$ -мерный параллелепипед  $H$  ( $\bar{H}$  или  $H$ ) на  $W$  параллелепипедов  $H^w$ ,  $w = 1, \dots, W$ , таких, что

$$H^{w_1} \cap H^{w_2} = \emptyset, \quad w_1, w_2 = 1, \dots, W, \quad w_1 \neq w_2, \\ \bigcup_{w=1}^W H^w = H. \quad (22)$$

Этот процесс разбиения может быть выполнен, например, одним из следующих способов.

1.1. Область изменения одной переменной  $H_j = \{h_j^1 \leq x_j \leq h_j^2\}$  разобьем на две подобласти:

$$H_j^1 = \{h_j^1 \leq x_j \leq g_j, \quad g_j < h_j^2\}; \\ H_j^2 = \{g_j < x_j \leq h_j^2\}. \quad (23)$$

Определим каждое из альтернативных подмножеств следующим образом:

$$H^1 = H_1 \times H_2 \times \dots \times H_{j-1} \times H_j^1 \times H_{j+1} \times \dots \times H_n; \quad (24)$$

$$H^2 = H_1 \times H_2 \times \dots \times H_{j-1} \times H_j^2 \times H_{j+1} \times \dots \times H_n; \quad (25)$$

Затем область изменения какой-либо другой переменной  $H_l = \{h_l^1 \leq x_l \leq h_l^2\}$ ,  $l \neq j$ , разобьем на две подобласти

$$H_l^3 = \{h_l^1 \leq x_l \leq g_l, \quad g_l < h_l^2\}; \quad H_l^4 = \{g_l < x_l \leq h_l^2\} \quad (26)$$

аналогично (23) и определим

$$H^3 = H_1^1 \times H_2^1 \times \dots \times \\ \times H_{l-1}^1 \times H_l^3 \times H_{l+1}^1 \times \dots \times H_n^1; \quad (27)$$

$$H^4 = H_1^1 \times H_2^1 \times \dots \times \\ \times H_{l-1}^1 \times H_l^4 \times H_{l+1}^1 \times \dots \times H_n^1; \quad (28)$$



$$H^5 = H_1^2 \times H_2^2 \times \dots \times H_{l-1}^2 \times H_l^3 \times H_{l+1}^2 \times \dots \times H_n^2; \quad (29)$$

$$H^6 = H_1^2 \times H_2^2 \times \dots \times H_{l-1}^2 \times H_l^4 \times H_{l+1}^2 \times \dots \times H_n^2. \quad (30)$$

Этот процесс разбиения продолжаем до тех пор, пока число параллелепипедов не станет равным  $W \leq 2n$ .

1.2. Область изменения одной переменной разбивается на  $W$  подобластей:

$$H_j^w = \{g_j^{w-1} \leq x_j \leq g_j^w\}, w = 1, \dots, W, \quad (31)$$

где  $g_j^0 = h_j^1, g_j^W = h_j^2$ .

Тогда каждая из подобластей определяется согласно выражению

$$H^w = H_1 \times H_2 \times \dots \times H_{j-1} \times H_j^w \times H_{j+1} \times \dots \times H_n, \quad w = 1, \dots, W. \quad (32)$$

2. Выполним для каждого из этих параллелепипедов  $H^w$  описанный выше итеративный процесс Ит\_1, в результате которого получим  $n$ -мерные параллелепипеды  $\bar{H}^w$  (для некоторых индексов  $w$  может выполняться условие  $\bar{H}^w = \emptyset$ ). Для вновь построенных параллелепипедов справедливы следующие соотношения:

$$\bigcup_{w=1}^W \bar{H}^w \subseteq \bar{H} \subseteq H, \quad (33)$$

$$\min_{x \in (H \cap G)} f^0(x_1, \dots, x_j, \dots, x_n) = \min_{1 \leq w \leq W} \min_{x \in (\bar{H}^w \cap G)} f^0(x_1, \dots, x_j, \dots, x_n). \quad (34)$$

Такое представление области глобального поиска в виде  $W$   $n$ -мерных параллелепипедов существенно меньшего размера в ряде случаев является очень эффективным и позволяет существенно сократить область допустимой вариации значения переменных.

3. Для каждого из этих параллелепипедов может быть определена нижняя граница значения  $f^0(X)$ , т. е. значения

$$\xi^{w,-}[f_0(X)] = \min_{X \in \bar{H}^w} \{F_0(X) | \bar{H}^w \neq \emptyset\}, \quad (35)$$

а также нижняя граница значения критерия оптимальности в оптимальном решении

$$\xi^-[f_0(X)] = \min_{1 \leq w \leq W} \min_{X \in \bar{H}^w} \{F_0(X) | \bar{H}^w \neq \emptyset\}. \quad (36)$$

4. Реализация случайного вектора осуществляется в каждом из этих параллелепипедов. Частота обращения к  $w$ -му параллелепипеду может быть выбрана в виде величины, обратно пропорциональной значениям, которые определяются на основе следующих простых вычислений:

$$\mu^w = \frac{\sum_{w=1}^W \xi^{w,-}[f_0(X)]}{\xi^{w,-}[f_0(X)]}, \lambda^w = \frac{\mu^w}{\sum_{w=1}^W \mu^w}, w = 1, \dots, W. \quad (37)$$

Если для какой-то области  $w$  получено наилучшее допустимое решение  $\bar{X}^w \in H^w \cap G$ , а следовательно,  $\bar{X}^w \in H \cap G$ , со значением функции цели, лучшим, чем граничные значения для некоторых подобластей, т. е. справедливы соотношения:

$$f_0(\bar{X}^w) \leq \xi^{r,-}[f_0(X)], r = 1, \dots, \rho, \quad (38)$$

то это подмножество подобластей  $H^r, r = 1, \dots, \rho$ , не содержит оптимальных решений и может быть исключено из рассмотрения. Это, в свою очередь, еще сокращает область поиска оптимальных решений.

5. Для каждой из подобластей фиксируем два числа  $M(w)$ : суммарное число реализаций случайного вектора значений переменных в области  $H^w$ ;  $M_1(w)$  — число реализаций, удовлетворяющих системе ограничений  $G$ . Если отношение  $0 \leq \frac{M_1(w)}{M(w)} \leq 1$  —

большая величина, то в целях сокращения потерь на поиск подобласть  $H^w$  может быть разбита на две или несколько подобластей, для каждой из которых вычислены соответствующие границы оптимальных решений, исключены подобласти, не содержащие допустимых планов Ит\_1. Этот процесс увеличения числа подобластей, исключения подобластей, не содержащих допустимых решений, уточнения значений  $\xi^{w,-}[f_0(X)]$  и  $\xi^-[f_0(X)]$ , а также проверка выполнения условий (38) и связанные с этим исключения из рассмотрения бесперспективных подобластей и пересчет вероятностей обращения к каждой из подобластей проводятся периодически в процессе реализации алгоритма.

6. Остановка процесса решения задачи может быть выполнена при выполнении условий

$$\frac{\min_{1 \leq w \leq W} f_0(\bar{X}^w) - \xi^-[f_0(X)]}{\xi^-[f_0(X)]} \leq \varepsilon, \quad (39)$$

где  $\varepsilon$  — заданная точность получения оптимального решения.

При этом в качестве решения задачи выбирается вектор  $X^* = \arg \min_{1 \leq w \leq W} f_0(\bar{X}^w)$ .

### Алгоритм А2

Алгоритм А2 реализует стратегию выбора на каждом  $t$ -м шаге реализации случайного значения некоторой  $j$ -й компоненты вектора  $X^t$  из диапазона ее возможных значений,  $h_j^{1t} \leq x_j^{2t} \leq h_j^{2t}$ . Затем в результате реализации итеративного процесса ИТ\_2 удаляются подобласти, не содержащие допустимых планов, и уменьшаются диапазоны возможных значений еще не выбранных значений вектора переменных. Если в процессе выполнения итеративного процесса ИТ\_2 для какого-либо индекса будет установлен факт, что  $h_l^{1t} > h_l^{2t}$ , то при выбранных ранее других компонентах вектора  $X^t$  (т. е. значениях  $x_j^t, j \neq l$ ) не существует допустимых решений задачи. Процесс случайной реализации остальных компонент вектора приостанавливается и должен быть возобновлен заново. Если при генерации  $j$ -й компоненты вектора  $X^t$  не установлен факт несовместности системы ограничений, то вычисляется значение нижней границы функции цели  $\xi^- [f_0(X^t|j)]$  для данной реализации вектора  $X^t$ . Если

$$\xi^- [f_0(X^t|j)] \geq f_0(Z^r) = \min_{1 \leq r \leq (t-1)} f_0(\bar{X}^r),$$

где  $\bar{X}^r \in G$  — допустимое значение вектора переменных задачи, полученное на  $r$ -м шаге реализации случайного вектора значений переменных, то процесс генерации  $X^t$  прекращается и выполняется сначала. Процесс решения останавливается, когда получено  $T$  значений векторов  $\bar{X}^t \in G, t = 1, \dots, T$ , т. е.  $r = T$ , либо когда выполняется условие

$$\frac{f_0(Z^r) - \xi^- [f_0(X)]}{\xi^- [f_0(X)]} \leq \varepsilon.$$

Обозначим  $\hat{J}_L^t = \{j_1^t, \dots, j_l^t, \dots, j_L^t\}$  — множество индексов переменных, значения которых на  $t$ -м шаге реализации вектора еще не выбраны;  $L$  — число переменных, значения которых еще не выбраны;  $h_{j_l}^{1t} \leq x_{j_l}^{2t} \leq h_{j_l}^{2t}$  — диапазоны допустимых значений этих переменных;  $z_k^t$  — выбранные значения переменных подмножества индексов  $\Gamma_L^t = J/\hat{J}_L^t$ .

Значение нижней границы критерия оптимальности на данном этапе вычислений определяется по формуле

$$\xi^- [f_0(X^t|j)] = F_0[z_k^t | k \in \Gamma_L^t, (y_j^{\min, 0} | j \in \hat{J}_L^t)]. \quad (40)$$

На каждом  $t$ -м шаге реализации случайного вектора  $X^t$  выполняется следующий объем вычислений.

1. Положим в начале выполнения  $t$ -го шага  $L = 0, \hat{J}_L^t = \emptyset, \Gamma_L^t = J$ , а значения  $h_{j_l}^{1t} = \bar{h}_{j_l}^1, h_{j_l}^{2t} = \bar{h}_{j_l}^2$ , т. е. значения, полученные после реализации итеративного процесса ИТ\_1.

2. Разобьем отрезок  $\theta \in [0, 1]$  на  $L$  участков  $\theta_l$ , длина каждого из которых обратно пропорциональна значению  $(h_{j_l}^{2t} - h_{j_l}^{1t})$  и вычисляется по формуле

$$\theta_l = \frac{(h_{j_l}^{2t} - h_{j_l}^{1t})}{\sum_{l=1}^L (h_{j_l}^{2t} - h_{j_l}^{1t})}, \quad l = 1, \dots, L.$$

3. Реализуем случайное число  $\psi_L^t$ , равномерно распределенное в диапазоне  $\psi \in [0, 1]$ .

Индекс выбираемой переменной  $\lambda$  определяется в зависимости от того, какому интервалу  $\theta_l$  принадлежит число  $\psi_L^t$ .

4. Реализуем случайное число  $\bar{\psi}_\lambda^t$ , равномерно распределенное в диапазоне  $\bar{\psi} \in [0, 1]$ . Значение переменной  $z_\lambda^t$  определяется из следующих условий:

- для непрерывных значений переменной  $z_\lambda^t = h_\lambda^{1t} + \bar{\psi}_\lambda^t (h_\lambda^{2t} - h_\lambda^{1t})$ ;
- для дискретных значений переменных выбирается ближайшая меньшая или большая величина из ряда допустимых ее значений.

5. Полагаем  $\hat{J}_L^t = \hat{J}_L^t \cup j_\lambda, L := L + 1, \Gamma_L^t = \Gamma_L^t / j_\lambda$ . Если  $\hat{J}_L^t = J$ , т. е.  $L = n$ , то реализация случайного вектора переменных на  $t$ -м шаге (вектора  $\bar{X}^t$ ) завершена. Если  $\bar{X}^t \in G$ , то переходим к п. 6. В противном случае выполняем п. 7 алгоритма.

6. Вычисляем значение  $f_0(\bar{X}^t)$ . Если  $f_0(\bar{X}^t) < f_0(Z^{(t-1)})$ , то полагаем  $f^0(Z^t) = f^0(\bar{X}^t)$  и переходим к выполнению  $(t+1)$ -го шага (п. 1 алгоритма). Если  $\bar{X}^t \in G$  и  $f_0(\bar{X}^t) \geq f_0(Z^{(t-1)})$ , то переходим к выполнению  $(t+1)$ -го шага (п. 1 алгоритма). Если  $\bar{X}^t \notin G$ , то переходим к п. 1 и вновь выполняем п. 1 алгоритма.

7. Выполняем итеративный процесс ИТ\_2 исключения областей, не содержащих допустимых решений и сокращения диапазонов допустимого выбора переменных подмножества  $\hat{J}_L^t$ . Если в процессе выполнения ИТ\_2 будет установлен факт несовместности системы ограничений, то вновь выполняем п. 1 алгоритма. В противном случае после завершения ИТ\_2 вычисляем нижнюю границу  $\xi^- [f_0(X^t | j_\lambda)]$  в соответствии с выражением (40).

Если  $\xi^- [f_0(X^t | j_\lambda)] < f_0(Z^t)$ , то продолжаем генерацию оставшихся компонент вектора  $X^t$  и переходим к п. 2.

### Заключение

Разработано программное обеспечение и проведены вычислительные эксперименты в целях исследования эффективности предложенных алгоритмов. Было выполнено решение тестовых задач размерностью до 10 переменных в условиях 3–5 ограничений, левые части которых выбирались в виде функций вида  $\sum_{k=1}^K a_l^i \prod_{l \in J_l} x_j \leq b^i, i = 1, \dots, m,$   
 $0 \leq h_j^1 \leq x_j \leq h_j^2$ , а критерий оптимальности — различные функции класса *SQMF*. Выполненные эксперименты показали:

- 60...75 % реализации случайных векторов значений переменных попадает в область ограничений задачи;
- в среднем предлагаемыми в работе алгоритмами решение задач с той же точностью может быть получено при объеме вычислений в 1,7–2,3 раза

меньше, чем при использовании стандартных алгоритмов глобального случайного поиска;

- на каждом шаге алгоритма по значению отклонения от нижней границы оптимального решения задачи может быть оценена точность наилучшего из полученных решений, что позволяет остановить процесс при получении эффективного с точки зрения лица, принимающего решение, результата и более объективно оценить его эффективность.

Предлагаемые в работе подходы для рассматриваемого класса функций (разбиение пространства поиска на несколько подобластей, использование оценок и выделение областей, не содержащих допустимых и оптимальных решений) могут быть развиты также при разработке генетических алгоритмов, что также позволит повысить их эффективность.

### Список литературы

1. Жиглявский А. А., Жилинскас А. Г. Методы поиска глобального экстремума. М.: Наука, 1991.
2. Гурии Л. С., Дымарский Я. С., Меркулов А. Д. Задачи и методы оптимального распределения ресурсов. М.: Сов. радио, 1968. 464 с.
3. Чипига А. Ф., Колков Д. А. Анализ методов случайного поиска глобальных экстремумов многомерных функций // Фундаментальные исследования. 2006. № 2. С. 24–26.
4. Зак Ю. А. Об одном классе многоэкстремальных задач и методах их решения // Автоматика и телемеханика. Рига. 1979. № 5. С. 41–47.
5. Зак Ю. А. Исследование статистических алгоритмов многоэкстремальной адаптации в задачах распределения ресурсов вычислительной системы // Автоматика и телемеханика, Рига. 1982. № 1. С. 59–66.
6. Зак Ю. А. Методы многоэкстремальной оптимизации в условиях ограничений для сепарабельно квазимонотонных функций // Известия РАН. Теория и системы управления. 2011. № 3.

## ИНФОРМАЦИЯ

**V Всероссийская научно-техническая конференция "Проблемы разработки перспективных микро- и нанoeлектронных систем"** состоится в октябре 2012 года. Она посвящена актуальным вопросам автоматизации проектирования МЭС, систем на кристалле, IP-блоков и новой элементной базы микро- и нанoeлектроники. Эти вопросы были и остаются актуальными для науки и техники, о чем свидетельствует тематика соответствующих крупнейших ежегодных международных конференций по САПР и разработке микро- и нанoeлектронной аппаратуры.

Сборник трудов конференции МЭС включен в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук.

Полная информация о конференции на сайте: [www.mes-conference.ru](http://www.mes-conference.ru)

УДК 621.391;519.21;519.245

**В. М. Малыхин**, канд. техн. наук, доц.,  
**Г. Ф. Малыхина**, д-р техн. наук,  
**А. В. Меркушева**, канд. техн. наук, инж.-исслед.  
 Санкт-Петербургский государственный  
 политехнический университет,  
 e-mail: g\_f\_malychina@mai.ru

## Метод регуляризации для улучшения качества восстановления изображения

*Регистрируемое изображение имеет искажения, связанные с некоторым видом его размытия и наличием шума. Устранение этих дефектов называется восстановлением изображения (ВИ) и осуществляется на основе минимизации функционала, представляющего суммарный квадрат ошибки регистрируемого изображения относительно истинного. Поскольку процедура ВИ является обратной задачей, неустойчивой к неточности исходных данных (т. е. плохо обусловленной задачей), то применяется метод регуляризации. При ВИ применение этого метода приводит к некоторому сглаживанию изображения и (что особенно недопустимо) к уменьшению резкости локальных границ изображения. Рассмотрена особая форма регуляризации, которая избирательна к ориентации градиента интенсивности изображения таким образом, что значительно уменьшает сглаживание в области границ контрастности изображения.*

**Ключевые слова:** изображение, модель формирования, критерий восстановления, функционал регуляризации, интенсивность, ориентация, адаптация

Реально получаемому изображению присущи искажения за счет размытия и за счет шума. При моделировании формирования изображения предполагается линейная и инвариантная к сдвигу функция размытия точки (ФРТ) и диффузно распределенный по полю изображения шум. Таким образом, для регистрируемого изображения используется модель в виде

$$g(x, y) = \int_D d(s, t) f(x - s, y - t) ds dt + n(x, y),$$

$$(x, y) \in \Omega, \quad (1)$$

где  $f(\cdot)$  — исходное изображение (наблюдаемая "сцена");  $g(\cdot)$  — регистрируемое наблюдаемое изображение;  $d(\cdot)$  — функция, отражающая размытие

каждого элемента регистрации точки первичного изображения (функция размытия точки);  $n(\cdot)$  — шум наблюдения;  $\Omega$  — область, занимаемая регистрируемым изображением.

Задача восстановления изображения (ВИ) состоит в получении истинного распределения  $f(\cdot)$  плотности исходного изображения по данным  $g(\cdot)$  и  $d(\cdot)$ . Эта задача связана с решением интегрального уравнения (1), т. е. с процедурой операции, обратной относительно свертки. Выполнение процедуры сталкивается с трудностями, характерными для задач, относящихся к классу обратных [1, 2] и имеющих большую размерность. Важным фактором задачи является плохая обусловленность обратного оператора, определяющего формирование размытия (так называемого оператора "блэра"). Существующие методы для ВИ построены на основе использования процедуры регуляризации. При этом за критерий качества восстановления изображения принимается суммарная квадратичная ошибка (разность истинного и восстановленного изображений), и при регуляризации критерий расширяется на определенный функционал. Этот функционал накладывает на решение дополнительные условия и делает обращение оператора размытия лучше обусловленным, т. е. значительно более устойчивым к неточности исходных данных. Первоначально в области ВИ использовалась пространственно инвариантная регуляризация, т. е. применение этой операции однородно на всей площади изображения. В результате осуществлялось ВИ с некоторым уровнем шума в зонах изображения с малым изменением интенсивности, и возникали значительные искажения в областях резких изменений интенсивности<sup>1</sup>.

Для устранения этих трудностей Лагендийком, Бьмондом и Боки (*Lagendijk, Biemond, Boeke*) [3] предложено использование функционала регуляризации с пространственной адаптацией к локальной интенсивности изображения. Тем не менее регуляризация адаптировалась только к уровню интенсивности, но оставалась равномерной по всем направлениям в зоне с близкими значениями интенсивности. Как правило, такая регуляризация также приводит к снижению резкости контуров с высокой интенсивностью — так называемых границ в поле изображения.

<sup>1</sup> Возникновение таких артефактов, связанных с использованием регуляризации, равномерной по всему изображению, в работе [3] трактовалось как избыточная регуляризация.

Для восстановления изображения наиболее эффективным является метод построения критерия качества ВИ, включающего функционал регуляризации, который не только адаптируется к локальной интенсивности, но дополнительно обладает свойством избирательности к ориентации. При использовании такой регуляризации, наряду с характерным сглаживанием, интенсивность ослабляется в направлении высокого значения ее градиента, т. е. при подходе к граничным контурам изображения. По сути, это метод анизотропной регуляризации.

**Восстановление изображения с использованием метода анизотропной регуляризации.** Для восстановления изображения используется минимизация критерия в виде общего функционала

$$L(\hat{f}) = \frac{1}{2} \int_{\Omega} e^2(x, y) dx dy + \lambda \int_{\Omega} B(|\nabla \hat{f}(x, y)|) dx dy \quad (2)$$

при дополнительных ограничениях на диапазон интенсивности изображения:

$$a \leq \hat{f}(x, y) \leq b; (x, y) \in \Omega.$$

В первом слагаемом функционала (2)  $e(\cdot)$  — это ошибка в ВИ, которая отражает степень правильности восстановленного изображения  $\hat{f}(x, y)$  относительно представления наблюдаемого изображения  $g(x, y)$ :

$$e(x, y) = g(x, y) - \int_D d(s, t) \hat{f}(x - s, y - t) ds dt. \quad (3)$$

Непосредственная минимизация первого слагаемого в выражении (2) привела бы к избыточному усилению шума, что связано с плохой обусловленностью решаемой задачи ВИ. Поэтому введено второе слагаемое, накладывающее ограничение на гладкость, где  $|\nabla \hat{f}(x, y)|$  — значение градиента изображения, а  $B(\cdot)$  — возрастающая функция, такая, что минимизация второго слагаемого в функционале (2) представляет уменьшение градиента изображения (операцию сглаживания);  $\lambda$  — параметр регуляризации, который контролирует сбалансированность степени доверия к наблюдению и меру сглаженности  $\hat{f}(\cdot)$ . Первое слагаемое в (2) выпуклое, как интеграл квадратичной функции, а второе слагаемое — тоже выпуклое ввиду требования выпуклости у функции  $B(\cdot)$  [4], [5]. Поэтому  $L(\hat{f})$  — также выпуклая функция своего аргумента, а следовательно, задача минимизации хорошо определена, т. е. имеется единый минимум и отсутствуют локальные.

Для минимизации  $L(\hat{f})$  используется градиентный метод, т. е. после вычисления градиента  $\nabla L(\hat{f})$

изменение  $\hat{f}(\cdot)$  проводится в противоположном ему направлении согласно соотношению (5)<sup>2</sup>:

$$\nabla L(\hat{f}) = - \int_{\Omega} e(u, v) d(u - x, v - y) dudv - \lambda \operatorname{div} \left( B'(|\nabla \hat{f}|) \frac{\nabla \hat{f}}{|\nabla \hat{f}|} \right); \quad (4)$$

$$\partial \hat{f} / \partial t = -\nabla L(\hat{f}). \quad (5)$$

Второе слагаемое в соотношении (5) после использования  $\nabla L(\hat{f})$  в представлении (4) соответствует методу анизотропной регуляризации Перроны и Малика (Perona, Malik) [4] и Йоу, Ксу (You, Xu) [5].

**Анизотропная регуляризация.** Для полноты представления метода ВИ с использованием критерия квадратичной ошибки, расширенного функционалом анизотропной регуляризации, целесообразно проследить его избирательность к ориентации направления в поле регистрируемого изображения. Для этого следует рассмотреть ортогональное разложение операции регуляризации в выражении (4), которое получено Йоу, Ксу, Танненбаумом и Кавэ (You, Xu, Tannenbaum, Kaveh) [5], в виде

$$\operatorname{div} \left( B'(|\nabla \hat{f}|) \frac{\nabla \hat{f}}{|\nabla \hat{f}|} \right) = \lambda_1(|\nabla \hat{f}|) D_{\perp(\text{град})} + \lambda_2(|\nabla \hat{f}|) D_{\parallel(\text{град})}, \quad (6)$$

где  $\lambda_1(|\nabla \hat{f}|) = \frac{B'(|\nabla \hat{f}|)}{|\nabla \hat{f}|}$  и  $\lambda_2(|\nabla \hat{f}|) = B''(|\nabla \hat{f}|)$  — собственные значения для матрицы Гессе (т. е. для матрицы вторых производных от функции  $B(|\nabla \hat{f}(x, y)|)$  по вектору  $\nabla \hat{f}(x, y)$ );  $D_{\perp(\text{град})}$  и  $D_{\parallel(\text{град})}$  — вторые производные от  $\nabla \hat{f}(x, y)$  в направлениях соответственно ортогональном к вектору градиента  $\nabla \hat{f}(x, y)$  и параллельном этому вектору.

Поскольку в качестве  $B(\cdot)$  принимается возрастающая функция, то значения  $\lambda_1$  — неотрицательны. Поэтому первое слагаемое в выражении (6) представляет вырожденное сглаживание (регуляризацию) в направлении, перпендикулярном к градиенту, что позволяет сохранить границы изображения без искажений (граница в изображении ортогональна градиенту). Второе слагаемое в выражении (6) представляет вырожденное сглаживание в направлении, перпендикулярном к границам. Если  $\lambda_2(|\nabla \hat{f}|) > 0$ ,

<sup>2</sup> Для производной  $L(\hat{f})$  по  $\hat{f}$  использовано дифференцирование по Гато [6], [7].

то вырожденное сглаживание идет по направлению вперед, и это может привести к размытию границ или получению в этой области характерного искажения, называемого звенящим артефактом.

Вместе с тем, если  $\lambda_2(|\nabla \hat{f}|) < 0$ , то регуляризация происходит в обратном направлении (усиливая контраст), и это нежелательно, потому что возникает плохая обусловленность. Но это не служит препятствием в особом случае при требовании, чтобы  $B(\cdot)$  было выпуклым (так как для выпуклой функции  $B(\cdot)$  выполняется условие  $B'(x) \geq 0$ ).

Если в качестве примера принять  $B'(x) = x$ , то операция регуляризации имеет вид:

$$\operatorname{div}\left(B'(|\nabla \hat{f}|) \frac{\nabla \hat{f}}{|\nabla \hat{f}|}\right) = \operatorname{div}(\nabla \hat{f}) = \nabla^2 \hat{f} \quad (7)$$

и представляет просто лапласиан.

Если же рассмотреть в качестве  $B(\cdot)$  функцию, представляемую соотношением

$$B'(x) = \begin{cases} x/T & \text{при } x < T \\ 1 & \text{при } x \geq T, \end{cases} \quad \text{где } T \text{ — порог,} \quad (8)$$

то при этом в гладкой области изображения, где  $|\nabla \hat{f}| \geq T$ , контуры и границы остаются без размытия (сглаживания за счет регуляризации). Сглаживание и будет только в касательном к границе направлении, так как  $\lambda_2(|\nabla \hat{f}|) = 0$ . Порог  $T$  может быть подстроен в соответствии с уровнем шума, однако можно считать, что достаточно приемлемым является значение  $T = 1$ , так как уровни шума в приложениях, связанных с ВИ, невелики.

Отметим также, что, как и в исследовании Лагендикка [3], в анализируемом методе степень регуляризации также адаптируется к локальной пространственной интенсивности изображения, поскольку

$$\lambda_1(|\nabla \hat{f}|) = \frac{1}{|\nabla \hat{f}|}.$$

**Заключение.** Восстановление изображения (ВИ) по регистрируемому его образу связано с устранением искажений, которые выражены некоторым размытием, обусловленным несовершенством оптической системы, диффузной непрозрачностью среды, микродвижением объекта или регистрирующих сенсоров, а также наличием шума. Устранение этих дефектов, необходимое для ВИ, осуществляется на основе минимизации функционала, представляющего суммарный квадрат ошибки регистрируемого изображения относительно истин-

ного. Однако процедура ВИ является обратной задачей, неустойчивой к неточности исходных данных (т. е. плохо обусловленной задачей), поэтому применяется метод регуляризации (по А. Н. Тихонову), который состоит в установлении некоторых ограничений на вид решения (по его гладкости) и реализуется введением в критерий дополнительного функционала. Увеличивая робастность к шуму и размытию изображения, метод регуляризации в непосредственном применении осуществляет сглаживание участков изображения с плавным изменением интенсивности, но ухудшает контрастность границ и контуров изображения. Поэтому основной проблемой является оптимизация метода регуляризации, ориентированной именно на особенности проблемы ВИ.

На модели формирования изображения общего вида (с учетом искажения шумом и размытием) проанализирован метод регуляризации с адаптацией к интенсивности, показаны его недостатки и рассмотрен новый метод Перрона, Малика и Йоу, Ксуа. В основе этого метода лежит форма функционала регуляризации, учитывающая ориентацию направления, т. е., по сути, этот метод осуществляет анизотропную регуляризацию.

Функционал регуляризации включает возрастающую функцию от градиента восстановленного изображения, благодаря которой нивелируется сглаживание интенсивности (связанное с регуляризацией) в направлении, перпендикулярном к границам и контурам. Таким образом, введенный метод анизотропной регуляризации для восстановления изображений полностью сохраняет контрастность внутренних границ изображения.

#### Список литературы

1. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.: Наука, 1979. 281 с.
2. Тихонов А. Н., Гончарский А. В., Степанов В. В., Ягола А. Г. Численные методы решения некорректных задач. М.: Наука, 1990. 226 с.
3. Legendijk R. L., Biemond J., Boeke D. E. Regularized interactive restoration with ringing reduction // IEEE Transaction on Acoustic, Speech and Signal Processing. 1988. V. 36, N 12. P. 1874—1887.
4. Perona P., Malik J. Scale-space and edge detection using anisotropic diffusion // IEEE Transaction on Pattern Analysis and Machine Intellect. 1990. V. 12, N 7. P. 529—639.
5. You Y. L., Xu W., Tannenbaum A., Kaveh M. Behavioral analysis of anisotropic diffusion in image processing // IEEE Transaction on Image Processing, 1996. N 5. P. 1539—1553.
6. Голуб Дж., Ван Лоун У. Матричные вычисления. М.: Мир, 1999. 548 с.
7. Smith P. Convexity methods in variational calculus. London: Research Studies, 1985. 300 p.

УДК 519.6

**О. В. Мандрикова**<sup>1,2</sup>,

д-р техн. наук, проф., вед. науч. сотр.,

**Ю. А. Полозов**<sup>1</sup>, мл. науч. сотр.,

<sup>1</sup>Институт космических исследований и распространения радиоволн ДВО РАН, с. Паратунка, Камчатская обл.,

<sup>2</sup>Камчатский государственный технический университет, г. Петропавловск-Камчатский  
e-mail: oksanam1@mail.kamchatka.ru

## Критерии выбора вейвлет-функции в задачах аппроксимации природных временных рядов сложной структуры

*Определены критерии выбора вейвлет-функции и построена схема преобразования временного ряда, позволяющая минимизировать число аппроксимирующих слагаемых и погрешность аппроксимации. С использованием этой технологии обработаны данные критической частоты ионосферы и выявлена связь между процессами, протекающими в ионосфере и в литосфере.*

**Ключевые слова:** вейвлет-преобразование, временные ряды сложной структуры, параметры ионосферы, аномальные особенности

### Введение

Предметом исследований являются ряды наблюдений геофизических параметров, включающие в себя компоненты различной структуры, локальные аномальные эффекты, связанные с активностью Солнца, возникающие накануне сейсмических явлений, или другими процессами различной природы [1–3]. Эти локальные особенности несут важную информацию для исследователя, и при обработке стоит задача их выявления и идентификации. Сложности решения данной задачи связаны с большим разнообразием, достаточно сложной формой особенностей и отсутствием адекватных математических моделей.

Традиционные методы построения моделей таких временных рядов [4, 5] основаны на процедуре сглаживания, что влечет существенную потерю информации. Эффективным методом для их описания

является вейвлет-преобразование [2, 3, 6]. В процессе создания систематизированной теории вейвлет-преобразования построены ортонормированные регулярные вейвлет-базисы с компактными носителями и разработаны кратномасштабные аппроксимирующие схемы представления данных, что послужило широкому распространению этой теории в различные сферы деятельности.

В то же время качество работы конечных систем, основанных на вейвлет-преобразовании, определяется технологией его применения [6, 7]. Существенную роль играет используемая конструкция разложения, выбор вейвлет-функции, процедура построения аппроксимирующего базиса и др. В данной работе на основе теории непрерывных временных вейвлетов и их связи с дискретными схемами преобразования определены критерии выбора вейвлет-функции и разработана схема преобразования временного ряда, позволяющие минимизировать число аппроксимирующих слагаемых и погрешность аппроксимации.

### Выделение локальных особенностей функции на основе вейвлет-преобразования

Так как вейвлет-функция  $\Psi$  имеет нулевое среднее значение, вейвлет-преобразование функции  $f \in L^2(R)$  ( $L^2(R)$  — пространство Лебега) [6, 8]

$$Wf(s, u) = \int f(t) \frac{1}{\sqrt{s}} \Psi\left(\frac{t-u}{s}\right) dt,$$

где  $s, u \in R, s \neq 0$ ;  $t$  — время;  $u$  — параметр сдвига;  $s$  — параметр масштаба;  $R$  — множество действительных чисел, измеряет изменение функции  $f$  в окрестности точки  $u$ , размер которой пропорционален  $s$ . При стремлении масштаба  $s$  к нулю вейвлет-коэффициенты характеризуют свойства функции  $f$  в окрестности  $u$ .

В работах [6, 8] показано, что если вейвлет-функция  $\Psi$  имеет  $l$  нулевых моментов, т. е.

$$\int_{-\infty}^{+\infty} t^k \Psi(t) dt = 0, \quad k = \overline{0, l-1},$$

и  $n$  производных, то для функции  $f$ , удовлетворяющей равномерному условию Липшица  $\alpha$ ,  $\alpha \leq 1$ , на  $[a, b]$ , существует  $A > 0$  такое, что

$$\forall (s, u) \in [a, b] |Wf(s, u)| \leq As^{\alpha + 1/2}. \quad (1)$$

Обратно, если  $f$  ограничена и  $Wf(s, u)$  удовлетворяет неравенству (1) для нецелого  $\alpha < 1$ , тогда  $f$  удовлетворяет равномерному условию Липшица  $\alpha$  на  $[a + \varepsilon, b - \varepsilon]$  при любом  $\varepsilon > 0$ .

Неравенство (1) является условием асимптотического убывания абсолютных значений вейвлет-коэффициентов  $|Wf(s, u)|$  при  $s \rightarrow 0$ . Таким образом, убывание амплитуды вейвлет-преобразования в зависимости от масштаба связано с гладкостью Липшица функции.

Если функция  $f(t)$  имеет сложную структуру, характеризовать гладкость  $f$  в некоторой точке  $t = v$  бывает достаточно трудно, потому что  $f$  может иметь различные виды особенностей, присутствующих в окрестностях точки  $v$ . На рис. 1 в качестве примера показан результат расчета локальных экстремумов абсолютных значений вейвлет-коэффициентов данных критической частоты ионосферы, подтверждающий сложную структуру этой функции и показывающий наличие различных видов особенностей в каждой ее окрестности.

Будем считать, что функция  $f$  в окрестности точки  $v$  имеет локальную особенность, если  $|Wf(s, u)|$  не удовлетворяет условию (1) в окрестности точки  $v$ .

Обработка дискретных последовательностей данных базируется на дискретной схеме вейвлет-разложения. Полагают, неполная выборка параметров преобразования полностью определяет функцию, если она может быть полностью восстановлена. Теория вейвлетов имеет условия точного восстановления функции, которые накладывают ограничения на используемые семейства вейвлет-функций. Семейства таких функций определяют дискретные вейвлеты, схему разложения по таким базисам называют *дискретным вейвлет-преобразованием* [6, 8].

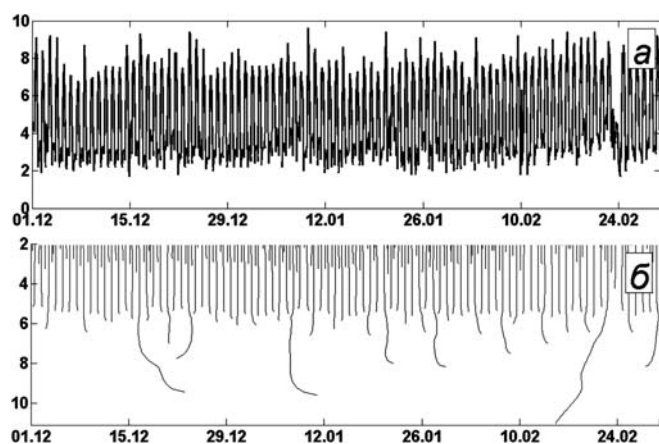


Рис. 1. Значения критической частоты ионосферы в МГц (данные за период декабрь—февраль 1973 г.), по оси  $ox$  — время  $t$ , по оси  $oy$  — значения критической частоты (а); моменты возникновения локальных экстремумов абсолютных значений вейвлет-коэффициентов критической частоты ионосферы, по оси  $ox$  — время  $t$ , по оси  $oy$  —  $\log_2 s$ , где  $s$  — параметр масштаба (б)

Если  $\Psi$  порождает базис Рисса [8] с нормой выбора  $u_0 = 1$ , то  $\Psi$  называется  $R$ -функцией:

$$\Psi_{j,n}(t) := 2^{j/2} \Psi(2^j t - n),$$

где  $j$  — параметр масштаба;  $n$  — параметр сдвига, и дискретное вейвлет-преобразование определяется следующим образом [6, 8]:

$$f(t) = \sum_{j,n \in Z} c_{j,n} \Psi_{j,n}(t),$$

где  $c_{j,n} = \langle f, \Psi_{j,n} \rangle$ ;  $Z$  — множество целых чисел.

Эти бесконечные ряды называются *вейвлет-рядами* и сходятся в  $L^2(R)$ .

### Вейвлет-образ природного временного ряда

Формальная модель природного временного ряда  $f(t)$  имеет вид [7]

$$f(t) = f_1(t) + f_2(t) + \dots + f_s(t), \quad (2)$$

где  $f_j(t)$  — разномасштабные ортогональные компоненты временного ряда, имеющие различную структуру.

Поскольку компоненты  $f_j(t)$  в формуле (2) имеют различную структуру, то с учетом локального характера содержащихся в них локальных особенностей, их разномасштабности и разнообразия по форме наиболее подходящим пространством для их представления является вейвлет-пространство:

$$f_j(t) = \sum_n c_{j,n} \Psi_{j,n}(t),$$

где  $f_j(t) \in L^2(R)$ ,  $\{\Psi_{j,n}(t)\}_{j,n \in Z}$  — ортогональный вейвлет-базис пространства  $L^2(R)$ . Здесь вейвлет-коэффициенты  $c_{j,n} = \langle f, \Psi_{j,n} \rangle$  рассматриваются как результат отображения  $f(t)$  в пространство с разрешением  $j$ , который определяется свойствами используемой вейвлет-функции  $\Psi$ . В зависимости от задачи возможны различные способы формирования критериев выбора вейвлет-функции  $\Psi$ . Решая задачу аппроксимации функции в базисе ортогональных функций в качестве основных *критериев* логично определить:

- **минимизацию числа аппроксимируемых слагаемых;**
- **минимизацию погрешности аппроксимации.**

В целях возможности построения адаптивных аппроксимирующих вейвлет-схем будем использовать нелинейные преобразования [6]:  $f_M(t)$  — проекция  $f(t)$  на  $M$  векторов, индексы которых содержатся в  $I_M$  ( $I_M$  — множество индексов):

$$f_M(t) = \sum_{j,n \in I_M} \langle f, \Psi_{j,n} \rangle \Psi_{j,n}(t). \quad (3)$$



В этом случае приближение функции  $f$  выполняется  $M$  векторами, зависящими от ее структуры.

Погрешность аппроксимации есть сумма оставшихся коэффициентов:

$$\varepsilon[M] = \|f - f_M\|^2 = \sum_{j, n \notin I_M} |\langle f, \Psi_{j,n} \rangle|^2.$$

Как отмечено в работе [6], чтобы минимизировать эту погрешность, индексы в  $I_M$  должны соответствовать  $M$  векторам, имеющим скалярные произведения с наибольшей амплитудой  $|\langle f, \Psi_{j,n} \rangle|$ . Поэтому минимизация числа аппроксимируемых слагаемых и минимизация погрешности аппроксимации могут быть достигнуты выбором базиса

$$\beta^* = \{\Psi_{j,n}\}_{j, n \in I_M},$$

обеспечивающего как можно большее число пренебрежимо малых коэффициентов  $c_{j,n} = \langle f, \Psi_{j,n} \rangle$ .

Предполагая, что компонента

$$e(t) = \sum_{j, n \notin I_M} \langle f, \Psi_{j,n} \rangle \Psi_{j,n} \quad (4)$$

является следствием влияния шумового фактора и что она является некоррелированной и аддитивной

$$y(t) = f(t) + e(t),$$

на основе формул (3), (4) получаем **представление в вейвлет-пространстве случайного временного ряда:**

$$\begin{aligned} y(t) &= \sum_{j, n \in I_M} \langle f, \Psi_{j,n} \rangle \Psi_{j,n}(t) + \sum_{j, n \notin I_M} \langle f, \Psi_{j,n} \rangle \Psi_{j,n}(t) = \\ &= \sum_{j, n \in I_M} c_{j,n} \Psi_{j,n}(t) + e(t). \end{aligned} \quad (5)$$

Каждая  $\Psi_{j,n}$  определяет свой частотно-временной прямоугольник в частотно-временной плоскости с временным  $n$  и масштабным  $j$  параметрами. Важность такого представления состоит в том, что различные составляющие временного ряда рассортировываются и хранятся в различных пространствах  $W_j$  с разрешением  $j$ :

$$L^2(R) = \bigoplus_{j \in Z} W_j := \dots \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \dots,$$

где  $W_j := \text{clos}_{L^2(R)}(\Psi_{j,n}; n \in Z)$ .

При этом наименьший анализируемый масштаб ограничен шагом выборки дискретного временного ряда.

**Имея представление временного ряда в виде формулы (5), операция выделения локальных особенностей временного ряда можно построить на основе свойства (1) путем определения наибольших значе-**

**ний функции  $|\langle f, \Psi_{j,n} \rangle|$  на малых масштабах  $j$ . Реализуем ее на основе проверки условия**

$$|\langle f, \Psi_{j,n} \rangle| \leq T_j, \quad (6)$$

где  $T_j$  — пороговое значение, определяющее наличие в компоненте временного ряда масштаба  $j$  локальной особенности.

Путем выбора порогового значения  $T'$  мы можем реализовать выделение и идентификацию особенности в природном временном ряде.

### Критерии выбора вейвлет-функции

В задачах аппроксимации временного ряда и выделения и идентификации локальных особенностей определяющими при выборе вейвлет-функции будут следующие ее характеристики: *число нулевых моментов, размер носителя, гладкость*. Покажем это.

**Число нулевых моментов.** Вейвлет  $\Psi$  имеет  $l$  нулевых моментов, т. е.

$$\int_{-\infty}^{+\infty} t^k \Psi(t) dt = 0, \quad k = \overline{0, l-1}. \quad (7)$$

Как показано выше (см. свойство (1)), число нулевых моментов  $m$  характеризует способность вейвлет-функции выявлять особенность вида  $\alpha \leq 1$ . Также из свойства (7) следует, что  $\Psi$  ортогональна любому многочлену степени  $l-1$ , т. е. "пропускает" полиномы степени выше  $l-1$ . Поэтому в окрестностях, где  $f$  — гладкая функция, большое число нулевых моментов  $\Psi$  обеспечит большое число пренебрежимо малых вейвлет-коэффициентов. В этом случае мы получим аппроксимирующую схему с минимальным числом слагаемых при заданном уровне точности аппроксимации.

**Размер носителя.** Вейвлет-преобразование порождает искусственные "скачки" на краях функции  $f$ , находящие отражения в коэффициентах разложения. Поскольку размер окрестности на масштабе  $j$ , содержащей возникающий при этом краевой эффект, зависит от размера носителя вейвлет-функции  $\Psi$  и определяется по формуле

$$h_j = 2^j q,$$

где  $q$  — размер носителя используемой базисной функции, получаем: **чем меньше размер носителя  $q$ , тем меньшую погрешность мы имеем на краях  $f$ .**

Если  $f$  имеет локальную особенность в некоторой точке  $v$  и если  $v$  находится внутри носителя  $\Psi_{j,n}$ , то, как показано выше, вейвлет-коэффициенты имеют большую амплитуду. Чем больше носитель, тем большее число таких коэффициентов на каждом масштабном уровне мы получаем. **Чтобы минимизировать их число, мы должны использовать вейвлет-функции с наименьшим размером носителя  $\Psi$ .**

**Гладкость вейвлета.** Аналогично числу нулевых моментов, гладкость вейвлет-функции характеризует ее способность выявлять особенность вида  $\alpha \leq 1$ . Что касается погрешности аппроксимации, то, как отмечено в работе [8], гладкость вейвлет-функции  $\Psi$  оказывает в основном "косметическое" влияние на нее: если  $\Psi$  — гладкая, то и погрешность будет гладкая.

Таким образом, при выборе вейвлет-функции мы приходим к выбору между числом нулевых моментов и размером носителя. Для ортогональных вейвлет-функций в работе [8] Добеши доказала, что если вейвлет  $\Psi$  имеет  $l$  нулевых моментов, то его наименьший носитель равен  $2l - 1$ . Поэтому, если функция  $f$  имеет несколько локальных особенностей и очень гладкая между этими особенностями, необходимо использовать вейвлет с большим числом нулевых моментов; если число особенностей нарастает, лучше уменьшить размер носителя ценой уменьшения числа нулевых моментов. Также в работе [8] доказано, что **ортогональные вейвлеты Добеши с компактным носителем — это единственное семейство базисных вейвлет-функций, которые имеют минимальный размер носителя при заданном числе нулевых моментов.**

Известным фактом является то, что число нулевых моментов и гладкость вейвлет-функции связаны друг с другом, но характер связи может быть различным в зависимости от вида рассматриваемого семейства базисов. В работах [6, 8] показано, что, например, для вейвлетов Добеши характерно следующее свойство: **гладкость вейвлета возрастает с возрастанием числа нулевых моментов.**

В большинстве практических задач необходимо, чтобы крупномасштабные аппроксимирующие компоненты позволяли получить наилучшее приближение функции, а с помощью мелкомасштабных составляющих идентифицировались отдельные локальные особенности. Наилучшее приближение аппроксимируемой функции  $f$  в крупномасштабных компонентах обеспечивает скэйлинг-функция  $\phi$  [8] с достаточным числом нулевых моментов  $l$ . Поэтому в этом случае важно, чтобы нулевые моменты имели не только  $\Psi$ , но и  $\phi$ . Семейство ортогональных базисов, удовлетворяющее этому требованию и имеющее носитель наименьшего размера, — базисы Койфлета [8]:

**Базисы Койфлета** порядка  $m$  удовлетворяют условиям:

$$\int_{-\infty}^{+\infty} t^k \Psi(t) dt = 0, \quad k = \overline{0, l}; \quad \int_{-\infty}^{+\infty} \phi(t) dt = 1,$$

$$\int_{-\infty}^{+\infty} t^k \phi(t) dt = 0 \quad \text{при } k = \overline{1, l}.$$

Если функция  $f$  принадлежит  $C^k$  ( $C^k$  — пространство функций,  $k$  раз непрерывно дифференцируемых) в окрестности  $2^J n$  с  $k \leq l$ , то

$$2^{J/2} \langle f, \phi_{J,n} \rangle \approx f(2^J n) + O(2^{-(k+1)J}). \quad (8)$$

Порядок приближения возрастает с ростом  $l$ , результирующий базис Койфлета имеет носитель размера  $3l - 1$  вместо  $2l - 1$  для вейвлета Добеши.

### Эксперименты с модельными данными

Модельные временные ряды были построены на основе трех компонент. Первая компонента временного ряда содержит соответствующие каждому часу медианные значения критической частоты ионосферы, рассчитанные в скользящем временном окне по данным предыдущих 30 дней. Вторая компонента — белый шум, третья компонента — аномалии. Первая компонента характеризует средний уровень часовых флуктуаций критической частоты ионосферы (рис. 2). Вторая компонента — белый шум, отражающий влияние внешнего фактора (вспышки на Солнце и др.). Третья компонента — аномалии, по форме характерные для критической

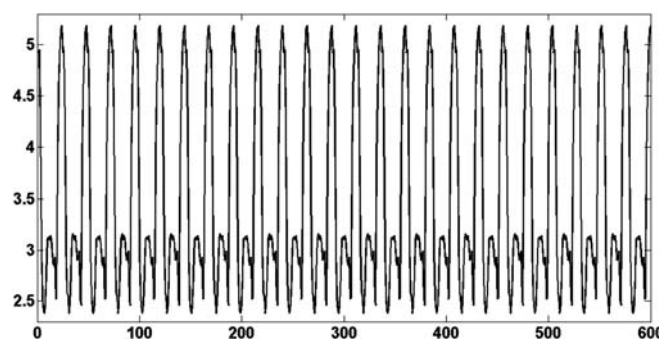


Рис. 2. Медианные значения критической частоты ионосферы

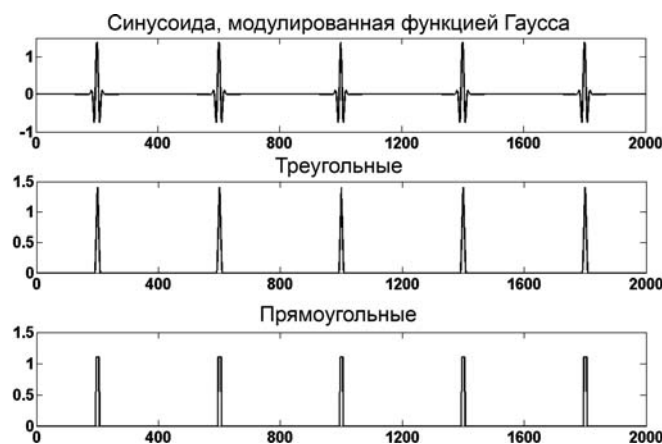


Рис. 3. Аномалии трех видов (синусоида, модулированная функцией Гаусса, треугольный импульс, прямоугольный импульс)

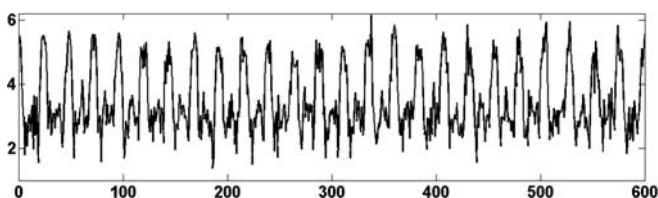


Рис. 4. Модельный временной ряд с внесенной шумовой составляющей

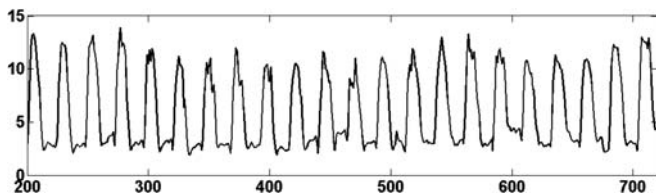


Рис. 5. Значения критической частоты ионосферы за январь 1982 г.

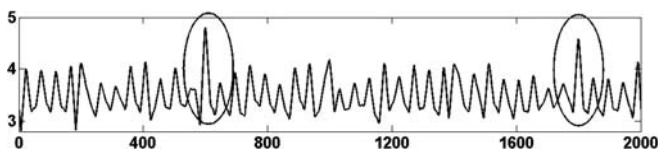


Рис. 6. Результаты обнаружения аномалий треугольного вида в модельном сигнале при использовании вейвлет-функции Добеши порядка 3. Овалами отмечены обнаруженные аномалии

частоты ионосферы и имеющие следующий вид: синусоида, модулированная функцией Гаусса; треугольный импульс; прямоугольный импульс (рис. 3). Длительность внесенных аномалий составляла от 10 до 80 отсчетов, амплитуда в 1,3 раза превышала фоновый уровень. Сумма первой и второй компонент показана на рис. 4. Для сравнения на рис. 5 показаны данные критической частоты ионосферы за 1982 г.

Обнаружение внесенных аномалий выполнялось на основе операции (6). Аномалия считалась обнаруженной, если она определялась на каждом масштабе  $j$ . Для обнаружения аномалий использовались вейвлет-функции Добеши порядка 3 и Койфлета порядка 2. Результат обнаружения аномалий на масштабе  $j = 3$  представлен на рис. 6.

Результаты экспериментов по выявлению аномалий вейвлет-функциями Добеши порядка 3 и Койфлета порядка 2

Вейвлет-функция	Вид аномалии, длительность аномалии (в отсчетах)	Число выявленных аномалий, %	Число ложных аномалий, %
Добеши	Треугольная, 10	54	4
Койфлет	Треугольная, 10	48	0
Добеши	Треугольная, 15	68	3
Койфлет	Треугольная, 15	59	0
Добеши	Функция Гаусса, 30	58	2
Койфлет	Функция Гаусса, 30	94	4

Эксперименты показали, что эффективность выявления аномалий зависит от ее вида и длительности (см. таблицу):

- базис Койфлета порядка 2 более эффективно выявляет гладкие особенности вида синусоиды, модулированной функцией Гаусса;
- базис Добеши порядка 3 более эффективно выявляет особенности треугольного и прямоугольного вида.

### Результаты обработки и анализа данных критической частоты ионосферы

В работе рассматривались данные критической частоты ионосферы  $f_oF2$ , полученные автоматической ионосферной станцией, расположенной в с. Паратунка ( $\varphi = 52,97^\circ$  с. ш.,  $\lambda = 158,25^\circ$  в. д., п-ов Камчатка). Обработка выполнялась с использованием вейвлет-функций Добеши порядка 3. Регистрация данных происходит один раз в час. Для экспериментов были взяты результаты измерений за период 1968—2006 г. На фоне регулярных изменений в ионосферных данных могут возникать локальные аномальные особенности длительностью от десятков минут до нескольких часов, имеющие разную физическую природу. Они возникают на фоне мощных ионосферных возмущений, которые обусловлены активностью Солнца. Проводимые в последнее время исследования процессов в средней и верхней атмосфере позволяют обнаружить связи в развитии некоторых явлений в литосфере и ионосфере [1, 2]. Данные особенности могут интерпретироваться как предвестники сильных сейсмических событий на Камчатке. Задача выделения и идентификации таких особенностей является крайне важной.

Путем обработки данных  $f_oF2$  на основе операции (5) и с учетом фонового уровня процесса и величины флуктуаций для каждого масштаба  $j$  в пределах скользящего временного окна, равного 720 отсчетов данных (что соответствует временному интервалу в один месяц), определялись пороговые значения  $T_j$  по следующей формуле:

$$T_j = \text{med}_n (|c_{j,n}|) + 2,5St_j,$$

где  $\text{med}$  — медиана,  $St_j = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (|c_{j,n}| - \bar{c}_{j,n})^2}$ ;

$\bar{c}_{j,n}$  — среднее значение;  $N$  — длина временного ряда, коэффициент 2,5 был определен статистически.

Данные пороговые значения далее использовались для выполнения процедуры обнаружения аномалий в данных критической частоты  $f_oF2$  на основе операции (6). На рис. 7, 8 показаны результаты обработки данных за 1973 и 1982 г. Выделенные аномальные окрестности показаны на рис. 7, 8 чер-

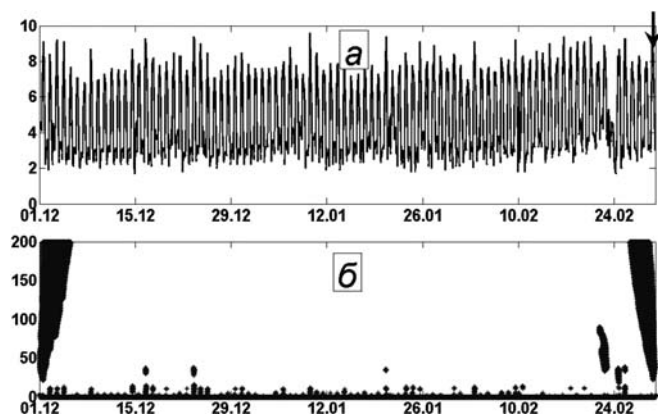


Рис. 7. Результаты выявления аномалий в данных критической частоты  $f_0F_2$ :

*a* — данные критической частоты  $f_0F_2$  за декабрь—февраль 1973 г.; *b* — результат применения операции (6)

ным цветом. Стрелкой отмечены моменты произошедших в этот период землетрясений (рассматривались землетрясения в радиусе  $\sim 300$  км от места регистрации данных с энергетическим классом  $E > 12,5$ ). Результаты обработки и анализа данных критической частоты  $f_0F_2$  показали, что накануне землетрясений в данных могут наблюдаться разно-масштабные аномалии длительностью от 5—6 ч до 2—3 суток, которые возникают за 1—2,5 недели до землетрясений.

### Выводы

В работе предложена схема преобразования природного временного ряда сложной структуры на основе вейвлетов и разработаны критерии выбора вейвлет-функции, обеспечивающие минимизацию числа аппроксимирующих слагаемых и погрешности аппроксимации.

С использованием данной технологии обработаны данные критической частоты  $f_0F_2$  и выявлена связь между процессами, протекающими в ионосфере и литосфере.

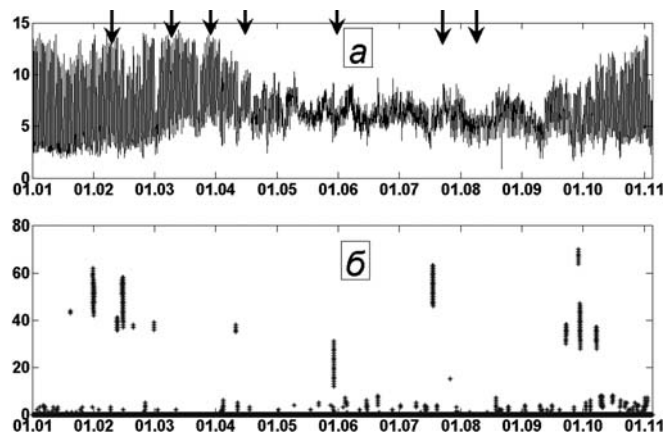


Рис. 8. Результаты выявления аномалий в данных критической частоты  $f_0F_2$ :

*a* — данные критической частоты  $f_0F_2$  за 1982 г.; *b* — результат применения операции (6)

*Данные сейсмического каталога любезно предоставлены Камчатским филиалом геофизической службы РАН (г. Петропавловск-Камчатский).*

*Работа поддержана грантом Президента Российской Федерации МД-2199.2011.9.*

### Список литературы

1. Демин М. Г. Ионосфера Земли. Плазменная гелиогеофизика. М.: Физматлит, 2008. Т. II. С. 92—163.
2. Богданов В. В., Гешенер В. В., Мандрикова О. В. Моделирование нестационарных временных рядов геофизических параметров со сложной структурой. С.-Петербург: ЛЭТИ, 2006. 107 с.
3. Мандрикова О. В. Моделирование геохимических сигналов на основе вейвлет-преобразования. Владивосток: Дальнаука, 2007. 123 с.
4. Брандт З. Анализ данных. Статистические и вычислительные методы для научных работников и инженеров: Пер. с англ. М.: Мир, ООО "Издательство АСТ", 2003. 668 с.
5. Марпл.-мл. С. Л. Цифровой спектральный анализ и его приложения: Пер. с англ. М.: Мир, 1990. 265 с.
6. Mallat S. A Wavelet tour of signal processing: Пер. с англ. М.: Мир, 2005. 671 с.
7. Мандрикова О. В., Горева Т. С. Метод идентификации структурных компонентов сложного природного сигнала на основе вейвлет-пакетов. // Цифровая обработка сигналов. 2010. № 1. С. 45—50.
8. Daubechies I. Ten Lectures on Wavelets: Пер. с англ. Ижевск: НИЦ "Регулярная и хаотическая динамика". 2001. 464 с.

**Б. Г. Кухаренко**, канд. физ.-мат. наук,  
ст. науч. сотр., вед. науч. сотр.,

Институт машиноведения РАН, г. Москва,  
e-mail: kukharenko@imash.ru

**Д. И. Пономарев**, аспирант,

Московский физико-технический институт (ГУ),  
e-mail: ponomarev-102@mail.ru

## Использование метода Прони для оценки временного масштаба при обнаружении паттернов во временных рядах

*Одночастотная аппроксимация временного ряда по методу Прони оценивает его характерный период колебаний — временной масштаб для поиска его почти совпадающих последовательностей (паттернов). В этом временном масштабе почти совпадающие последовательности идентифицируются с использованием алгоритмов обнаружения паттернов. В качестве примера рассматриваются записи сигналов дистанционного манипулятора с прецизионным акселерометром. Паттерны записи управляющего сигнала манипулятора представляют произвольные движения — жесты руки оператора.*

**Ключевые слова:** временные ряды, паттерн, аппроксимация, метод Прони, алгоритм обнаружения паттернов, дистанционный манипулятор

### Почти совпадающие последовательности (паттерны) временных рядов

Обнаружение паттернов (задача разведывания данных — Data Mining) временных рядов состоит в поиске почти совпадающих последовательностей элементов во временном ряду  $x[\overline{1, N_0}]$ ,  $N_0 \gg 1$ . Для заданного временного ряда  $x[\overline{1, N_0}]$  длиной  $N_0$  последовательность  $S$  — это выборка длиной  $M \ll N_0$ , являющаяся непрерывной частью  $x[\overline{1, N_0}]$ , т. е.

$$S_k[\overline{1, M}] = \{x[k], \dots, x[k + M - 1]\},$$

$$k = \overline{1, (N_0 - M + 1)}. \quad (1)$$

Все последовательности  $S_k[\overline{1, M}]$ ,  $k = \overline{1, (N_0 - M + 1)}$  (1), получаются из временного ряда  $x[\overline{1, N_0}]$  сдвигом временного окна длины  $M$ . Подобие двух последовательностей  $S_k, S_l$  (1) с заданными  $k, l = \overline{1, (N_0 - M + 1)}$  определяется некоторой мерой расстояния  $\text{dist}(S_k, S_l)$ . Последовательности

$S_k, S_l$  (1) считаются почти совпадающими, т. е. они представляют паттерн, если  $\text{dist}(S_k, S_l) < R$ , где  $R$  — значение порога [1]. Как правило, для последовательностей элементов из временных рядов используется Евклидова мера расстояния в пространстве размерности  $M$ , равной длине этих последовательностей [1]. Перед вызовом функции расстояния  $\text{dist}(\dots, \dots)$  каждая последовательность нормируется, чтобы иметь нулевое среднее и стандартное отклонение  $\sigma^2 = 1$ , поскольку нет смысла сравнивать последовательности с различными смещениями и амплитудами.

Сформулируем задачу перечисления  $K$  наиболее значимых часто повторяющихся паттернов (frequent pattern — motif).

Для заданного временного ряда  $x[\overline{1, N_0}]$  длиной  $N_0$  и значения порога  $R$  наиболее значимый паттерн (обозначаемый  $1 - \text{Motif}(N_0, R)$ ) представляет некоторая последовательность  $S^{(1)}[\overline{1, M}]$  (1), которая имеет наивысшее число нетривиальных совпадений среди последовательностей  $S_k[\overline{1, M}]$ ,  $k = \overline{1, (N_0 - M + 1)}$  (1).  $K$ -й по очереди наиболее значимый паттерн (обозначаемый  $K - \text{Motif}(N_0, R)$ ) представляет последовательность  $S^{(K)}[\overline{1, M}]$ , которая имеет наивысшее число нетривиальных совпадений среди последовательностей  $S_k[\overline{1, M}]$ ,  $k = \overline{1, (N_0 - M + 1)}$  (1), и удовлетворяет условию

$$\text{dist}(S^{(K)}, S^{(i)}) > 2R, \quad i = \overline{1, K}. \quad (2)$$

Наконец,  $K$ -й по очереди наиболее  $d$ -значимый паттерн (обозначаемый  $K - \text{Motif}(N_0, R, d)$ ) — это последовательность  $S^{(K)}[\overline{1, M}]$ , которая имеет наивысший счетчик нетривиальных совпадений и удовлетворяет условию (2), причем  $d$  элементов (возможно, не смежных) могут быть проигнорированы при вычислении расстояний  $\text{dist}(S^{(K)}, S^{(i)})$ ,  $i = \overline{1, K}$ . Как правило,  $d \ll M \ll N_0$ . Принятие решения о том, какие элементы будут проигнорированы, является простым. Поскольку минимизируется вычисление расстояния  $\text{dist}(S^{(K)}, S^{(i)})$ , индексы  $k$  сортируются по возрастанию  $|S^{(K)}[k] - S^{(i)}[k]|$  и игнорируются первые  $d$  значений.

Когда временной масштаб  $M \ll N_0$  для обнаружения паттернов временного ряда  $x[\overline{1, N_0}]$  не известен, их поиск требует варьирования длины  $M$  временной последовательности [2, 3]. Это многократно увеличивает фактические временные затраты по сравнению с оценками временных затрат по применению конкретного алгоритма обнаружения

паттернов при фиксированной длине последовательностей.

Для временного ряда присутствие паттернов, т. е. приблизительно повторяющихся последовательностей элементов этого временного ряда, предполагает его колебательное изменение. То есть существует некоторый характерный средний период колебательных изменений временного ряда, определяющий временной масштаб, в котором осуществляется обнаружение паттернов. Этот характерный период (временной масштаб) может быть определен в результате (одночастотной) аппроксимации более крупных сегментов временного ряда по методу Прони [4].

### Аппроксимация сегментов временных рядов по методу Прони

Считается, что полный временной ряд  $x[\overline{1, N_0}]$  имеет постоянный интервал дискретизации времени. Для простоты приводимых формул этот интервал дискретизации времени  $\Delta t = 1$ . Последовательные сегменты фиксированной длины  $M \ll N \ll N_0$  имеют вид

$$y_i[\overline{1, N}] = x[\overline{(N(i-1)+1), Ni}], \quad i = \overline{1, \text{round}(N_0/N)}. \quad (3)$$

Таким образом, считается, что анализируемые сегменты имеют длину гораздо больше, чем длина паттернов, присутствующих в полном временном ряду  $x[\overline{1, N_0}]$ . Декомпозиция Прони сегмента (3) (для краткости индекс сегмента  $i = \overline{1, \text{round}(N_0/N)}$  ниже опускается) имеет вид

$$y[k] = \sum_{l=1}^p r[l](z[l])^{k-1} + n[k], \quad k = \overline{1, N}, \quad (4)$$

где  $p$  — число определяемых полюсов сегмента временного ряда;  $z[l] = \exp(\delta[l] + j2\pi f[l])$ ,  $l = \overline{1, p}$ , — полюса, которые определяются для сегмента (2), где  $\delta[l]$  и  $f[l]$  — соответственно фактор демпфирования (логарифмический декремент) и частота;  $r[l] = \alpha[l]\exp(j\varphi[l])$ ,  $l = \overline{1, p}$ , — вычеты в этих полюсах, где  $\alpha[l]$  и  $\varphi[l]$  — соответственно амплитуда и фаза;  $n[k]$ ,  $k = \overline{1, N}$ , — аддитивный шум [4]. То есть метод Прони использует модель локально переходного изменения, которая не столь широко распространена для описания нестационарных временных рядов [5].

Матричный алгоритм метода Прони использует Ганкелеву матрицу данных  $\mathbf{Y}$  размерности  $(N-L) \times (L+1)$  [6, 7]. Матрица данных  $\mathbf{Y}$  строится на основе сегмента  $y[\overline{1, N}]$  временного ряда с применением сдвигов:

$$Y[j; \overline{1, (L+1)}] = y[\overline{j, (j+L)}], \quad j = \overline{1, (N-L)}. \quad (5)$$

На основе матрицы  $\mathbf{Y}$  (5) определяются две усеченные матрицы

$$\begin{aligned} \mathbf{Y}^{(0)} &= \mathbf{Y}[\overline{1, (N-L)}; \overline{1, L}] \\ \mathbf{Y}^{(1)} &= \mathbf{Y}[\overline{1, (N-L)}; \overline{2, (L+1)}]. \end{aligned} \quad (6)$$

При отсутствии шума ( $n[k] = 0$ ,  $k = \overline{1, N}$ , в (4)) полюса  $z[l]$ ,  $l = \overline{1, p}$ , определяются как обобщенные собственные числа усеченных матриц  $\mathbf{Y}^{(1)}$  и  $\mathbf{Y}^{(0)}$  (6):

$$\mathbf{Y}^{(1)}\mathbf{q} = z\mathbf{Y}^{(0)}\mathbf{q}, \quad (7)$$

где  $\mathbf{q}$  — соответствующий  $z$  обобщенный собственный вектор [6, 7]. При наличии шума  $n[k] \neq 0$ ,  $k = \overline{1, N}$ , в (4) ранг  $p$  матриц  $\mathbf{Y}^{(v)}$ ,  $v = \overline{0, 1}$ , (6) и их псевдообращений  $\mathbf{Y}^{(v)+}$ ,  $v = \overline{0, 1}$ , фиксируется посредством разложения сингулярных чисел

$$\mathbf{Y}^{(v)} = \mathbf{U}^{(v)}\mathbf{S}^{(v)}(\mathbf{V}^{(v)})^T, \quad v = \overline{0, 1}, \quad (8)$$

здесь  $\mathbf{S}^{(v)} = \text{diag}\{s^{(v)}[1], s^{(v)}[2], \dots\}$ , где  $s^{(v)}[1] > s^{(v)}[2] > \dots > 0$

и

$$(\mathbf{U}^{(v)})^T\mathbf{U}^{(v)} = \mathbf{I}_{(N-L) \times (N-L)}, \quad (\mathbf{V}^{(v)})^T\mathbf{V}^{(v)} = \mathbf{I}_{(L+1) \times (L+1)},$$

верхний индекс "T" обозначает транспонирование и  $\mathbf{I}_{(N-L) \times (N-L)}$ ,  $\mathbf{I}_{(L+1) \times (L+1)}$  — тождественные матрицы размерности  $(N-L) \times (N-L)$  и  $(L+1) \times (L+1)$  соответственно. Определяются усеченные диагональные матрицы из  $p$  главных сингулярных чисел

$$\mathbf{S}_p^{(v)} = \text{diag}\{s^{(v)}[1], \dots, s^{(v)}[p], 0, \dots, 0\}, \quad v = \overline{0, 1}, \quad (9)$$

и обратные им матрицы

$$\mathbf{S}_p^{(v)-1} = \text{diag}\{(s^{(v)}[1])^{-1}, \dots, (s^{(v)}[p])^{-1}, 0, \dots, 0\}, \quad v = \overline{0, 1}. \quad (10)$$

Подстановка  $\mathbf{S}^{(v)} \rightarrow \mathbf{S}_p^{(v)}$  в формулу (8) дает аппроксимацию  $\mathbf{Y}_p^{(v)}$ ,  $v = \overline{0, 1}$ , ранга  $p$  для матриц  $\mathbf{Y}^{(v)}$ ,  $v = \overline{0, 1}$  (6). Формула

$$\mathbf{Y}_p^{(v)+} = \mathbf{V}^{(v)}\mathbf{S}_p^{(v)-1}(\mathbf{U}^{(v)})^T, \quad v = \overline{0, 1}, \quad (11)$$

дает псевдообращение ранга  $p$  для матриц  $\mathbf{Y}^{(v)}$ ,  $v = \overline{0, 1}$  (6). После этого  $p$  обобщенных собственных чисел  $z$  задачи (7) определяются как  $p$  собственных чисел

$$(\mathbf{Y}_p^{(0)+} \mathbf{Y}_p^{(1)} - z\mathbf{I}_{p \times p})\mathbf{q} = 0, \quad (12)$$

где  $\mathbf{q}$  — соответствующий  $z$  собственный вектор. После определения полюсов  $z[\overline{1, p}]$  сегмента  $y[\overline{1, N}]$  из формулы (12) вычеты  $r[\overline{1, p}]$  в этих полюсах находятся из (4) по методу наименьших квадратов [7].

В преобразовании Прони

$$y[\overline{1, N}] \rightarrow \{z[\overline{1, p}], r[\overline{1, p}]\}$$

или

$$y[\overline{1, N}] \rightarrow \{f[\overline{1, p}], \delta[\overline{1, p}], \alpha[\overline{1, p}], \phi[\overline{1, p}]\}$$

размерность  $p$  частотного пространства всегда меньше, чем размерность  $N$  пространства сегментов (3). Возможность варьирования в методе Прони размерности  $p$  частотного пространства посредством ограничения ранга матриц  $\mathbf{Y}_p^{(v)}$ ,  $v = \overline{0, 1}$ , и  $\mathbf{Y}_p^{(v)+}$ ,  $v = \overline{0, 1}$ , (11) дает следующее преимущество. Линейный или нелинейный тренд в переходном сегменте  $y[\overline{1, N}]$  (3) в его декомпозиции (4) представляется составляющими с низкими частотами порядка  $N^{-1}$ .

После удаления среднего значения  $\bar{y} = \frac{1}{N} \sum_{k=1}^N y[k]$ , т. е.

$$y[k], k = \overline{1, N} \rightarrow (y[k] - \bar{y}), k = \overline{1, N},$$

либо удаления линейного или нелинейного тренда  $\tilde{y}(i\Delta t)$  в результате регрессии

$$\sqrt{\sum_{k=1}^N (y[i] - \tilde{y}(i\Delta t))^2} \rightarrow \min$$

и преобразования

$$y[k], k = \overline{1, N} \rightarrow (y[i] - \tilde{y}(i\Delta t)), k = \overline{1, N}, \quad (13)$$

для преобразованного сегмента  $y[\overline{1, N}]$  по методу Прони может быть определена одночастотная аппроксимация [8, 9]. Она соответствует минимальному числу  $p = 2$  полюсов в (4) и собственных чисел в (12). Частота  $f[1] = \text{Imlog}(z[1])$  (и  $f[2] = -f[1]$ ) определяет характерный период  $M = 1/f[1]$ , ( $M \ll N \ll N_0$ ) колебательного изменения преобразованного сегмента  $y[\overline{1, N}]$  (13). Оцененный по всем последовательным сегментам полного временного ряда  $x[\overline{1, N_0}]$  этот период  $M \ll N_0$  дает временной масштаб при обнаружении совпадающих временных последовательностей (паттернов) временного ряда  $x[\overline{1, N_0}]$ .

### Обнаружение паттернов в записях сигналов дистанционного манипулятора с прецизионным акселерометром

В последнее время широкое распространение получают дистанционные манипуляторы, передающие движения руки оператора в трехмерном пространстве. В настоящей работе исследуются сигналы разрабатываемого авторами манипулятора (рис. 1), чувствительным элементом которого является прецизионный трехосевой MEMS-акселерометр (Micro-Electro-Mechanical System) [10]. Акселерометр об-



Рис. 1. Дистанционный манипулятор с прецизионным акселерометром

ладает чувствительностью к земной гравитации, поэтому изменение положения манипулятора относительно направления ускорения  $g$  силы тяжести приводит к изменению значений проекций этого ускорения на оси акселерометра  $x, y, z$ . Значения этих проекций используются как управляющие сигналы для позиционирования курсора (стандартно управляемого компьютерной мышью) на экране дисплея компьютера.

Типичная запись сигнала акселерометра для позиционирования курсора на экране компьютера представлена на рис. 2. Эта запись представляет изменение ускорения при произвольных движениях (жестах) руки оператора и поэтому содержит характерные для этих жестов повторяющиеся паттерны.

В качестве примера на рис. 3 показана одночастотная аппроксимация по методу Прони сегмента  $x[k = \overline{1000, 3000}]$  записи ускорения, приведенной на рис. 2.

Как видно на рис. 3, оценка по методу Прони характерного периода  $M$  сегментов записи ускорения

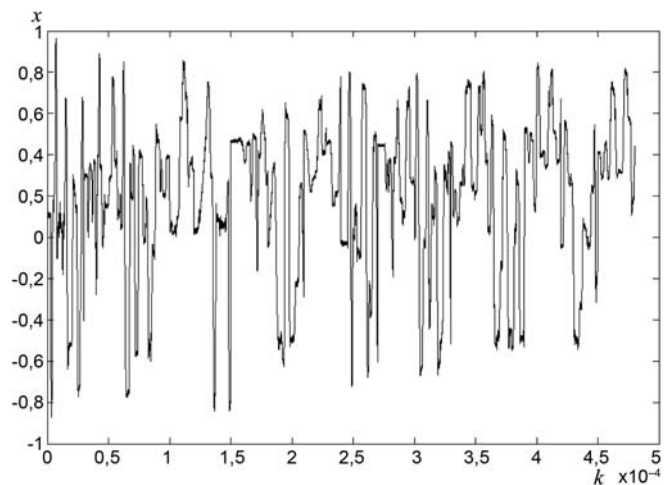


Рис. 2. Запись ускорения  $x$  при произвольных движениях (жестах) руки оператора ( $k$  — индекс временного ряда)

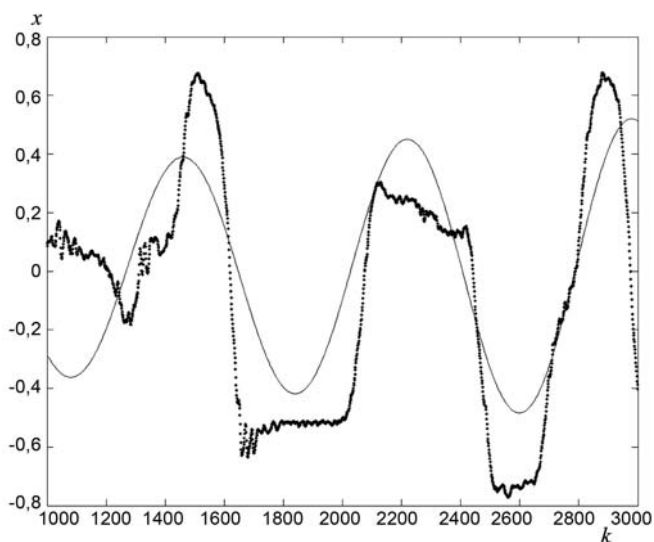


Рис. 3. Одночастотная аппроксимация сегмента  $x[k = \overline{1000, 3000}]$  записи ускорения (рис. 2)

Алгоритм MK Motif Discovery	
<b>Procedure</b> $[L1, L2] = \text{MK\_Motif}(D, K)$	
1	$best\_so\_far = INF$
2	<b>for</b> $i = 1$ to $K$
3	$ref[i] =$ случайным образом выбранная последовательность $D[k]$ из $D$
4	<b>for</b> $j = 1$ to $(N_0 - M + 1)$
5	$Dist[i; j] = \text{dist}(ref[i], D[j])$
6	<b>if</b> $Dist[i; j] < best\_so\_far$
7	$best\_so\_far = Dist[i; j], L1 = k, L2 = j$
8	$S[i] = \text{standard\_deviation}(Dist[i; 1, \text{size}(Dist, 2)])$
9	$[SS, Z] = \text{sort}(S, \text{"descend"})$ сортировка $SS = S[Z]$ , такая, что для индексов $i$ ссылочной переменной ( $ref$ ) $S[Z[i]] \geq S[Z[i+1]]$
10	$[DDist, I] = \text{sort}(Dist[Z; 1, \text{size}(Dist, 2)])$ сортировка $DDist = Dist[Z; I]$ , такая, что для индексов $j$ временных последовательностей (в $D$ ) $Dist[Z[I]; I[j]] \leq Dist[Z[I]; I[j+1]]$
11	$offset = 0, abandon = \text{false}$
12	<b>while</b> $abandon = \text{false}$
13	$offset = offset + 1, abandon = \text{true}$
14	<b>for</b> $j = 1$ to $m$
15	$reject = \text{false}$
16	<b>for</b> $i = 1$ to $K$
17	$lower\_bound =  Dist[Z[I]; I[j]] - Dist[Z[I]; I[j + offset]] $
18	<b>if</b> $lower\_bound > best\_so\_far$
19	$reject = \text{true}, \text{break}$
20	<b>else if</b> $i = 1$
21	$abandon = \text{false}$
22	<b>if</b> $reject = \text{false}$
23	<b>if</b> $\text{dist}(D[I[j]], D[I[j + offset]]) < best\_so\_far$
24	$best\_so\_far = \text{dist}(D[I[j]], D[I[j + offset]])$
25	$L1 = I[j], L2 = I[j + offset]$

Рис. 4. Алгоритм MK Motif Discovery

ния на рис. 2 дает  $M \approx 1000$ . Ниже при поиске паттернов используется длина последовательностей в виде  $2^n$ ,  $n$  — целое, т. е.  $M = 2^{10} = 1024$ .

Теоретическая нижняя граница для поиска ближайшей пары из  $N_0$  точек в  $M$ -мерном пространстве — это  $O(N_0 \log N_0)$  [11]. Однако скрытая константа, сопутствующая этой оценке нижней границы, растет экспоненциально с ростом  $N_0$ . Для временных рядов с  $N_0 \gg 1$  эта константа доминирует в оценке времени выполнения любого алгоритма поиска совпадающих последовательностей временного ряда, и в результате она оказывается не лучше, чем временная оценка для очевидного алгоритма сравнения всех пар последовательностей временного ряда. Поскольку временная оценка для очевидного алгоритма обнаружения паттернов полного временного ряда  $x[\overline{1, N_0}]$  имеет порядок  $O(N_0^2)$ ,

предложено множество приближенных алгоритмов для обнаружения паттернов временного ряда. Большинство этих алгоритмов имеют временную оценку  $O(N_0)$  или  $O(N_0 \log(N_0))$  с очень большими постоянными множителями.

В качестве алгоритма обнаружения паттернов в записи ускорения на рис. 2 (во временном масштабе  $M = 1024$ ) используется алгоритм Mueen—Keogh (МК) Motif Discovery [12, 13]. Это алгоритм реализует поиск совпадающих последовательностей в базе данных  $D$ , представляющей собой неупорядоченный набор временных последовательностей  $D[k] = S_k[\overline{1, M}], k = \overline{1, (N_0 - M + 1)}$  (1).

В начале выполнения алгоритма предполагается, что "до сих пор наилучшее расстояние" ( $best\_so\_far$  distance) для пары почти совпадающих последовательностей (представляющих паттерн) равно бесконечности ( $INF$ ). После этого из базы данных  $D$  случайным образом выбирается "ссылочная последовательность" (reference subsequence), и все остальные последовательности из  $D$  упорядочиваются в одномерном пространстве в соответствии с их расстояниями до этой ссылочной последовательности. Как побочный эффект этого шага, происходит обновление значения "до сих пор наилучшего расстояния" ( $best\_so\_far$  distance) для пары почти совпадающих последовательностей. Следует отметить, что расстояния между последовательностями  $D[k], k = \overline{1, (N_0 - M + 1)}$  в этом одномерном пространстве не являются истинными расстояниями  $\text{dist}(\dots)$  между этими последовательностями в исходном про-



пространстве размерности  $M$ , а представляют собой только "нижние границы" (*lower\_bounds*) для этих истинных расстояний. Ключевая идея алгоритма МК Motif Discovery состоит в том, что одномерное (линейное) упорядочение последовательностей предоставляет полезную эвристическую информацию для поиска паттернов. Действительно, если две последовательности близки в исходном пространстве размерности  $M$ , то они должны быть также близки в линейном упорядочении. Отметим, что обратное неверно. Две последовательности могут быть произвольно близки в линейном упорядочении, но очень далеки в исходном пространстве последовательностей. На следующем шаге алгоритма выполняется просмотр полученного линейного упорядочения последовательностей и определяется истинное расстояние между соседними парами последовательностей. Если находится пара последовательностей, которая имеет расстояние меньше

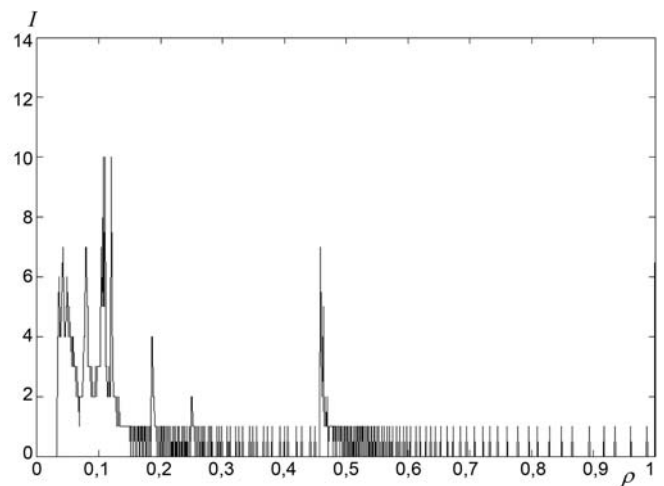


Рис. 7. Распределение значений  $\rho = d^2/2M$

текущего "наилучшего до сих пор расстояния" (*best\_so\_far distance*), проводится обновление значения "наилучшего до сих пор расстояния". Алгоритм МК Motif Discovery представлен на рис. 4.

На рис. 5 показаны две последовательности, почти совпадающие после нормировки, которые представляют первый паттерн в записи ускорения на рис. 2 (штриховая линия на рис. 4) и обнаруженный посредством алгоритма МК Motif Discovery (сплошная линия).

Представляющие этот паттерн последовательности, которые почти совпадают после нормировки, показаны на рис. 6. Отличие двух последовательностей на рис. 6 связано со случайными колебаниями руки оператора и может быть уменьшено посредством линейной или нелинейной Байесовской фильтрации [10, 14, 15].

Определим относительную меру расстояния между двумя последовательностями  $\rho = d^2/2M$ , где  $d$  — расстояние между этими последовательностями,  $M = 1024$  — длина последовательностей. Для паттерна на рис. 5 и 6  $d = 8,3267$ , а следовательно,  $\rho = 0,0339$ . Распределение числа  $I$  обнаруженных пар для значений  $\rho \in [0, 1]$ , предшествующих обнаружению первого паттерна на рис. 5, показано на рис. 7. Для записи, приведенной на рис. 2, среднее расстояние между парами последовательностей, представляющих различные паттерны,  $d = 6,7268$ , т. е.  $\rho = 0,0221$ .

Показанный на рис. 5 и 6 паттерн записи управляющего сигнала манипулятора (рис. 2) представляет произвольное движение — жест руки оператора.

#### Список литературы

1. Chiu B., Keogh E., Lonardi S. Probabilistic discovery of time series motifs // Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 24–27, 2003. DC. Washington: ACM. 2003. P. 493–498.
2. Ferreira P. G., Azevedo P. J., Silva C., Brito R. Mining approximate motifs in time series // Proceedings of the 9th International Conference on Discovery Science. Barcelona: Springer-Verlag. 2006. P. 89–101.

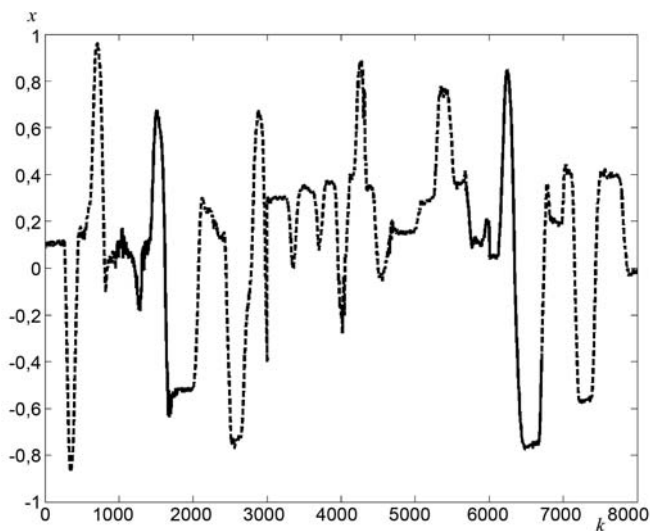


Рис. 5. Паттерн в записи ускорения на рис. 2

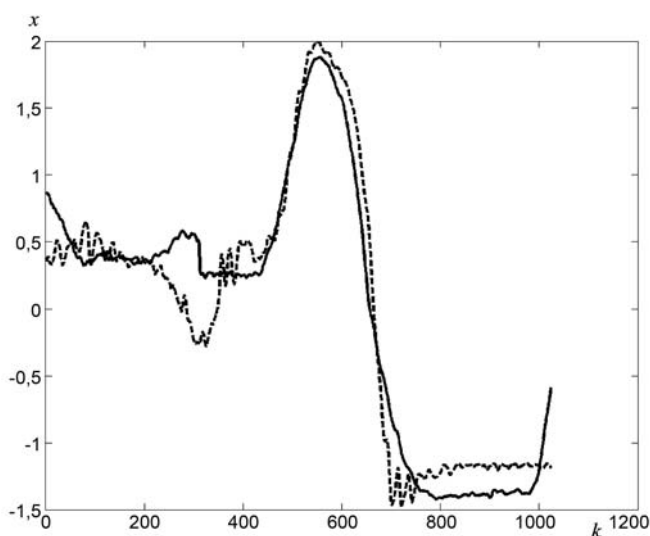


Рис. 6. Две последовательности, приведенные на рис. 5, после нормировки

3. **Castro N., Azevedo P. J.** Multiresolution motif discovery in time series // Proceedings of the SIAM International Conference on Data Mining (SDM 2010). American Statistical Association (ASA). 2010. P. 665–676.
4. **Weiss L., McDonogh R. N.** Prony's method, Z-transform, and Pade approximation // SIAM Review. 1963. V. 9, N 2. P. 145–149.
5. **Porat B., Friedlander B.** Performance analysis of a class of transient detection algorithms — a unified framework // IEEE Transactions on Signal Processing. 1992. V. 40, N 10. P. 2536–2545.
6. **Sarkar T. K., Pereira O.** Using the matrix pencil method to estimate the parameters of a sum of complex exponentials // IEEE Antenna and Propagation Magazine. 1995. V. 37, N 1. P. 48–55.
7. **Кухаренко Б. Г.** Технология спектрального анализа на основе быстрого преобразования Прони // Информационные технологии. 2008. № 4. С. 38–42.
8. **Кухаренко Б. Г.** Исследование по методу Прони динамики систем на основе временных рядов // Труды МФТИ. 2009. Т. 1. № 2. С. 176–192.
9. **Кухаренко Б. Г.** Исследование режимов колебаний на основе численных решений нелинейных моделей // Проблемы машиностроения и надежности машин. 2011. № 2. С. 22–30.
10. **Kukharensko B. G., Ponomarev D. I.** Bayesian filtering of control signal of telerobotic manipulator with precise accelerometer // Проблемы машиностроения и автоматизации. 2011. N 1. С. 72–76.
11. **Cormen T. H., Leiserson C. E., Rivest R. L., Stein C.** Introduction to Algorithms. 2nd Edition. Cambridge, MA: McGraw Hill Book Company, The MIT Press. 2001.
12. **Mueen A., Keogh E., Zhu Q., Cash S., Westover B.** Exact discovery of time series motifs // Proceedings of the SIAM International Conference on Data Mining (SDM 2009). American Statistical Association (ASA). 2009. P. 473–484.
13. **Mueen A., Keogh E. J.** Online discovery and maintenance of time series motifs // Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2010), July 25–28, 2010. Washington, DC: ACM. 2010. P. 1089–1098.
14. **Кухаренко Б. Г.** Байесовская фильтрация в технологии спектрального анализа на основе быстрого преобразования Прони // Информационные технологии. 2010. № 8. С. 36–42.
15. **Кухаренко Б. Г., Пономарев Д. И.** Нелинейная байесовская фильтрация многомерных временных рядов // Информационные технологии. 2011. № 6. С. 33–39.

УДК 004.9;519.25

**И. В. Браиловский**<sup>1</sup>, канд. физ.-мат. наук,  
архитектор медиа и графики,  
**Н. А. Соломеш**<sup>2</sup>, инженер-практикант,  
<sup>1</sup>Intel Corporation, Санта-Клара, США;  
<sup>2</sup>ЗАО "Интел А/О", Москва  
e-mail: Ilya.v.brailovskiy@intel.com,  
Natalya.solomeshch@intel.com

## Моделирование качества для видеокодирования

*Рассматривается вопрос моделирования качества видеокodeков. Предложена нелинейная двумерная параметрическая модель, позволяющая определить качество работы видеокodeка без выполнения кодирования. На основе полученной модели рассмотрена задача классификации исходных видеопоследовательностей.*

**Ключевые слова:** метрики видеокачества, видеокомпрессия, битрейт, функция скорости искажений

### Введение

В задаче разработки алгоритмов видеокodирования часто возникает вопрос предварительной оценки качества работы кодека на некоторых данных. Такая задача появляется либо при автоматическом тестировании видеокodeков в отсутствие эталонного кодека, либо когда работа эталонного кодека затруднена (отсутствует возможность контроля параметров, результаты доступны только на ограниченном числе данных, работа кодека занимает большое время). Сильное расхождение с предсказанными результатами сигнализирует о потенциальной ошибке. При наличии предсказания ка-

чества видеокodирования на уровне кадра его можно использовать в самом алгоритме видеокodирования для определения качества кодека и для соответствующей поправки уровня квантования кадра. Задачу предварительной оценки качества работы кодека решали достаточно много исследователей, и получено много интересных результатов.

Очевидно, что качество работы кодека зависит как от исходной последовательности, так и от самого кодека. В ряде работ, в частности в работах Б. Жирода [1], построены RD (Rate-Distortion) модели для некоторого класса видеокodeков. Жирод строит теоретическую зависимость качества кодирования для различных способов предсказания, применяемых в стандартных видеокodeках, таких как MPEG2 [2], MPEG4 part 2 [3], AVC [4] и др., и приводит оценки улучшения качества кодирования при использовании более сложных алгоритмов определения качества. Эти теоретические результаты важны, однако их сложно использовать в практических целях, поскольку зависимость от последовательности указана неявно и требуется дополнительное исследование на правильность оценки зависимости.

Работы Б. Жирода [1] и ряд других работ [5, 6] указывают на то, что качество закодированной последовательности зависит от сложности самой последовательности. Интуитивно понятно: чем "сложнее" исходная последовательность, тем сложнее ее сжать. Можно предположить, что мерой сложности является энтропия последовательности, и чем эффективнее кодек может находить корреляции в исходной последовательности (чем точнее он оценивает энтропию), тем эффективнее он работает. Исходя из такой логики очевидно, что если взять эффективный энтропийный алгоритм кодирования (такой, например, как RPPM [7]), то он сможет

предсказать качество работы кодеков. Мы провели соответствующие эксперименты и выяснили, что корреляция на самом деле невысока. Это, скорее всего, говорит о том, что подобные алгоритмы сжатия мало приспособлены для нахождения корреляции в видеопоследовательностях. Например, проведенный нами эксперимент показал, что при использовании энтропийного алгоритма предсказания сложности (основанного на PPM-кодировании) последовательность Stefan оказалась сложнее видеопотока Matrix примерно в 3 раза (характеристики этих последовательностей приведены в табл. 1). Объективные и субъективные эксперименты показали противоположные результаты. Это лишь один из многих противоречивых результатов, который получается при использовании энтропийного алгоритма для предсказания сложности видеокодирования.

Практиками часто используется подход так называемого золотого, или эталонного, кодека. Запускается некоторый работающий кодек (например, имплементация кодека стандартизационных комитетов), и его результаты считаются эталонными. Это достаточно эффективный метод, который, к сожалению, не масштабируется на разные кодеки или даже на разные конфигурации одного кодека. Например, кодек с включенными и выключенными V кадрами может давать значительно различающиеся

результаты по качеству, результаты работы AVC могут сильно отличаться от работы MPEG2. Другой пример, когда желательна использование оценки, — это кодеки компании Микрософт WMV [8] или RealVideo [9], референсная реализация которых недоступна или доступна с ограниченным набором опций. Наличие модели, описывающей качество кодеков с различными свойствами, позволило бы не зависеть от поиска удачного эталонного кодека.

В настоящей работе мы предлагаем вычислять некоторые характеристики исходного видеопотока для оценки качества работы кодека и определять функциональную модель зависимости качества кодека от исходной последовательности. Следует отметить, что такой подход в целом не является новым.

Австрийские авторы (Pues и др.) предложили в работе [5] двухэтапный подход к решению этой проблемы. Они провели классификацию контента низкого разрешения, основываясь на пространственной или временной сложности, а также на количестве движения и количестве зеленого цвета в последовательности. Далее, с помощью регрессионного анализа они выявили зависимость между MOS [10] и параметрами исходного видеопотока, такими как битрейт и число кадров в секунду, и классом видеопоследовательности. Этот подход интересен, но применим только к фиксированному битрейту, поскольку пространственная и временная сложность в общем случае слабо зависят друг от друга и не могут быть взаимозаменяемыми.

Другой метод классификации, предложенный Гасталдо и Роветтом [6], основан на создании нейронной сети. Авторы предложили нейронную сеть и обучили ее, сопоставляя субъективное качество и характеристики видеопоследовательностей, такие как битрейт, среднее значение компонент в кадре и абсолютное значение векторов движения. Классификация была проведена для последовательностей сжатых MPEG2. Метод очень интересен и эффективен, но требует перенастройки для каждого нового кодека другого стандарта или кодека с другими параметрами. Кроме того, параметры настройки заданы неявно, что затрудняет использование такого подхода на практике, когда необходимо быстро перейти от одного кодека к другому или выбрать другой набор настроек кодека. Для того чтобы сеть хорошо работала на других стандартах, предполагается обучать ее каждый раз заново, а это достаточно продолжительный процесс, особенно если учитывать, что оценка MOS проставляется людьми. Заметим, что имеется целый ряд работ, где в качестве характеристик исходного потока используются промежуточные результаты работы кодека; при таком подходе высока зависимость от конкретного кодека.

В настоящей работе предлагается метод оценки качества кодеков с двумерной линейной зависимостью качества закодированной последовательности от кодека и нелинейно зависящий от ряда характеристик исходного видеопотока. Для описа-

Таблица 1

Параметры последовательностей

Название последовательностей	Ширина, высота и число кадров в потоке	Битрейт
Bus	w 352, h 288, n 150	200, 400, 800, 1600, 2200
Flower	w 352, h 288, n 250	200, 400, 800, 1600, 2200
Foreman	w 352, h 288, n 300	200, 400, 800, 1600, 2200
News	w 352, h 288, n 300	200, 400, 800, 1600, 2200
Stefan	w 352, h 288, n 300	200, 400, 800, 1600, 2200
Soccer	w 640, h 480, n 257	750, 1500, 3000, 6000, 8000
Susi	w 704, h 576, n 374	750, 1500, 3000, 6000, 8000
Bbc	w 704, h 576, n 374	750, 1500, 3000, 6000, 8000
Iceage1	w 720, h 576, n 120	750, 1500, 3000, 6000, 8000
Iceage2	w 720, h 576, n 189	750, 1500, 3000, 6000, 8000
Iceage	w 720, h 576, n 491	750, 1500, 3000, 6000, 8000
Lotr	w 720, h 416, n 116	750, 1500, 3000, 6000, 8000
Simpson_small	w 720, h 480, n 108	750, 1500, 3000, 6000, 8000
Simpson	w 720, h 480, n 265	750, 1500, 3000, 6000, 8000
Smith1	w 720, h 432, n 358	750, 1500, 3000, 6000, 8000
Smith2	w 720, h 432, n 330	750, 1500, 3000, 6000, 8000
Forest_gamp	w 720, h 576, n 1500	750, 1500, 3000, 6000, 8000
Smith	w 720, h 432, n 772	750, 1500, 3000, 6000, 8000
Beach	w 1280, h 720, n 101	2000, 5000, 8000, 11000, 15000
Building	w 1280, h 720, n 102	2000, 5000, 8000, 11000, 15000
Foot-passenger	w 1280, h 720, n 157	2000, 5000, 8000, 11000, 15000
Matrix	w 1920, h 1072, n 115	2000, 5000, 8000, 11000, 15000
Parkrun_1	w 1280, h 720, n 274	2000, 5000, 8000, 11000, 15000
Parkrun_2	w 1280, h 720, n 297	2000, 5000, 8000, 11000, 15000
Troy	w 1920, h 1072, n 139	2000, 5000, 8000, 11000, 15000
Shields_ter	w 1280, h 720, n 500	2000, 5000, 8000, 11000, 15000

ния свойств исходных последовательностей мы использовали следующие метрики:

- пространственную и временную сложность изображения;
- размытость изображения;
- среднее значение яркости в кадре;
- среднеквадратичное отклонение яркости.

В качестве критерия качества мы использовали SSIM [11] как одну из наиболее близких к субъективному восприятию метрик. Мы рассмотрели ряд наиболее популярных кодеков AVC/H.264 [4], MPEG4 [3], MPEG2 [2], VC1 [8] и на основании регрессионной модели оценили параметры нашей модели.

В следующем разделе приводится обзор использованных метрик, далее следует описание условий, при которых проводились эксперименты, затем рассматривается полученная нами метрика качества кодирования EQM (Encoding Quality Metric), а потом следует раздел с полученными результатами и заключение.

### Описание метрик

*Фильтр Собеля* [12] реализуется путем свертки двух блоков размером  $3 \times 3$  пикселя по видеокадру и извлечением квадратного корня из суммы квадратов этих свертков. Пусть  $y = \text{Sobel}(x)$ ; обозначим как  $x(i, j)$  пиксель входного изображения в  $i$ -й строке и  $j$ -м столбце.  $G_v(i, j)$  будет результатом первой свертки:

$$G_v(i, j) = -1 \cdot x(i-1, j-1) - 2 \cdot x(i-1, j) - 1 \cdot x(i-1, j+1) + 0 \cdot x(i, j-1) + 0 \cdot x(i, j) + 0 \cdot x(i, j+1) + 1 \cdot x(i+1, j-1) + 2 \cdot x(i+1, j) + 1 \cdot x(i+1, j+1).$$

Аналогичный результат второй свертки

$$G_h(i, j) = -1 \cdot x(i-1, j-1) + 0 \cdot x(i-1, j) + 1 \cdot x(i-1, j+1) - 2 \cdot x(i, j-1) + 0 \cdot x(i, j) + 2 \cdot x(i, j+1) - 1 \cdot x(i+1, j-1) + 0 \cdot x(i+1, j) + 1 \cdot x(i+1, j+1).$$

Отсюда отфильтрованное выходное изображение в  $i$ -й строке и  $j$ -м столбце будет

$$y(i, j) = \sqrt{[G_v(i, j)]^2 + [G_h(i, j)]^2}.$$

Вычисления проводились для всех  $2 \leq i \leq n-1$  и  $2 \leq j \leq m-1$ , где  $n$  — число строк и  $m$  — число столбцов. Для того чтобы избежать нежелательных граничных эффектов, рассматривалась только часть изображения (рис. 1).

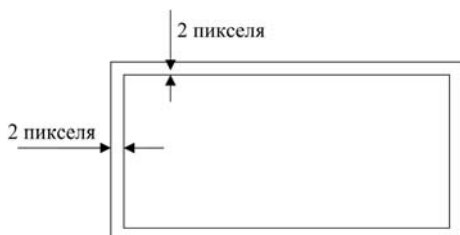


Рис. 1. Рассматриваемая область на видеокадре

*Пространственная сложность изображения.* Метрика  $SI$  (Spatial perceptual Information) [12] основана на фильтре Собеля. Яркостная плоскость каждого видеокадра обрабатывается фильтром Собеля [ $\text{Sobel}(F_n)$ ]. Далее считается среднеквадратичное отклонение по пикселям в каждом отфильтрованном кадре в видеопоследовательности и получается временной ряд пространственной информации сцены. Этот процесс может быть представлен следующей формулой:

$$SI = \{\text{std}_{space}[\text{Sobel}(y(i, j))]\}.$$

*Временная сложность изображения.* Метрика  $TI$  (Temporal perceptual Information) [12] основана на характеристике различий движения. Пусть  $M_n(i, j)$  обозначает разницу между значениями пикселей яркостной плоскости с одинаковыми координатами на последовательных кадрах:

$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j),$$

где  $F_n(i, j)$  является пикселем в строке  $i$ -й и  $j$ -м столбце для  $n$ -го кадра во времени. Затем считается стандартное отклонение по пространству от  $M_n(i, j)$  по всем  $i$  и  $j$ :

$$TI = \{\text{std}_{space}[M_n(i, j)]\}.$$

Большие значения  $TI$  являются результатом интенсивного движения в смежных кадрах.

*Временная сложность изображения с учетом компенсации движения.* Метрика  $TI\_ME$  (Temporal perceptual information with motion estimation) основана на метрике  $TI$  с добавлением алгоритма компенсации движения к метрике  $TI$ . В данном алгоритме поиск осуществляется по среднему значению блока пикселей  $4 \times 4$ . Радиус поиска — 16 пикселей. Далее считается среднеквадратичное отклонение для всего кадра по полученным минимальным значениям разницы блоков:

$$TI\_ME = \{\text{std}_{space}[\min\{F_n(i, j) - F_{n-1}(k, l)\}]\},$$

где  $F_n(i, j)$  — среднее значение блоков  $4 \times 4$ ;  $n$  — номер кадра.

В ходе экспериментов было установлено, что метрика  $TI\_ME$  более точно характеризует временную сложность последовательности, нежели метрика  $TI$ .

*Метрика четкости изображения (Blur)* [13] основана на эффекте второго размытия. Эффект заключается в том, что после повторного размытия изображения разница между яркостными значениями пикселей изменяется мало. Следовательно, на основании разницы между исходным и размытым изображением можно судить о том, насколько размыто исходное изображение.

*Алгоритм подсчета Blur.* Пусть  $F$  — матрица значений яркостной компоненты для кадра размером  $m \times n$  пикселей. Для оценки четкости изображения необходимо размыть исходное изображение по яркостной компоненте. Размытие происходит по горизонтальной и вертикальной плоскостям, усредняется значение девяти соседних пикселей по го-

ризонтали или вертикали соответственно. В итоге получаются два изображения, размытых по разным плоскостям.

Затем вычисляется абсолютное значение разницы между соседними пикселями на изображении  $D_{F_{Ver}}$ ,  $D_{F_{Hor}}$ ,  $D_{B_{Ver}}$  и  $D_{B_{Hor}}$ :

$$D_{F_{Ver}}(i, j) = |F(i, j) - F(i - 1, j)| \quad \forall i = 1, \dots, m - 1; \\ j = 0, \dots, n - 1;$$

$$D_{F_{Hor}}(i, j) = |F(i, j) - F(i, j - 1)| \quad \forall j = 1, \dots, n - 1; \\ i = 0, \dots, m - 1;$$

$$D_{B_{Ver}}(i, j) = |B_{Ver}(i, j) - B_{Ver}(i - 1, j)| \quad \forall i = 1, \dots, \\ m - 1; j = 0, \dots, n - 1;$$

$$D_{B_{Hor}}(i, j) = |B_{Hor}(i, j) - B_{Hor}(i, j - 1)| \quad \forall j = 1, \dots, \\ n - 1; i = 0, \dots, m - 1.$$

Далее необходимо проанализировать изменение яркости на соседних пикселях на исходном и размытом изображениях. Если эта величина будет больше, чем на исходном кадре, то делается вывод о том, что исходное изображение уже было размыто. Это изменение оценивается только для абсолютной разницы и описывается следующими формулами:

$$V_{Ver} = \max(0, D_{F_{Ver}}(i, j) - D_{B_{Ver}}(i, j)) \quad \forall i = 1, \dots, \\ m - 1; j = 1, \dots, n - 1;$$

$$V_{Hor} = \max(0, D_{F_{Hor}}(i, j) - D_{B_{Hor}}(i, j)) \quad \forall i = 1, \dots, \\ m - 1; j = 1, \dots, n - 1.$$

Сравнив по порядку изменения для исходной картинки, высчитываем сумму коэффициентов  $D_{F_{Ver}}$ ,  $D_{F_{Hor}}$ ,  $D_{B_{Ver}}$ ,  $D_{B_{Hor}}$ :

$$s_{F_{Ver}} = \sum_{i,j=1}^{m-1, n-1} D_{F_{Ver}}(i, j);$$

$$s_{F_{Hor}} = \sum_{i,j=1}^{m-1, n-1} D_{F_{Hor}}(i, j);$$

$$s_{B_{Ver}} = \sum_{i,j=1}^{m-1, n-1} D_{B_{Ver}}(i, j);$$

$$s_{B_{Hor}} = \sum_{i,j=1}^{m-1, n-1} D_{B_{Hor}}(i, j).$$

Кроме того, делается нормировка результата в диапазоне от 0 до 1:

$$b_{F_{Ver}} = \frac{s_{F_{Ver}} - s_{B_{Ver}}}{s_{F_{Ver}}}; \quad b_{F_{Hor}} = \frac{s_{F_{Hor}} - s_{B_{Hor}}}{s_{F_{Hor}}}.$$

В заключение полученные величины сравниваются с горизонтальными и вертикальными размытиями. Выбирается максимальное значение, и ему присваивается окончательное значение метрики  $Blur$ :

$$Blur = \max(b_{F_{Ver}}, b_{F_{Hor}}).$$

**Цветовые метрики.** В качестве цветовых характеристик мы использовали две величины:

- среднее значение каждой из компонент в кадре:

$$Mean\_Value\_Y(U, V) = \\ = \frac{1}{width * height} \sum_{i,j=0}^{i=width, j=height} Value\_Y(U, V)(i, j);$$

- среднее квадратичное отклонение компонент по кадру:

$$StdDev\_Y = \frac{1}{width * height} \times \\ \times \sum_{i,j=0}^{i=width, j=height} \sqrt{Value\_Y^2(i, j) - Mean\_Value\_Y^2(i, j)}.$$

**Нормировка метрик.** Для того чтобы расширить область применения разработанной нами методики моделирования качества кодирования (EQM), мы нормировали используемые в ней метрики на две величины.

- На смену сцены. Нормировка на смену сцены необходима, потому что при смене сцены все метрики получают новые значения. Предсказать заранее значения этих изменений невозможно. Поэтому мы умножили используемые метрики на  $K = \log_{10}(N)$ , учитывая тем самым данный эффект.  $N = 1$  на кадре со сменой сцены, на следующем кадре  $N = 2$  и т. д. до 10-го кадра, где  $N = 10$ , далее метрики последующих кадров умножались на  $K = \log_{10}(10) = 1$  до очередной смены сцены (рис. 2). Сегодня известно множество алгоритмов обнаружения смены сцены в видеопоследовательностях [14].
- На отношение битрейта к площади кадра. Для того чтобы корректно провести анализ результатов полученных значений видеопоследовательностей с разными разрешениями и битрейтами, мы нормировали метрики на величины  $E$  и  $S$ :

$$E = \frac{Bitrate}{S};$$

$$S = width * height.$$

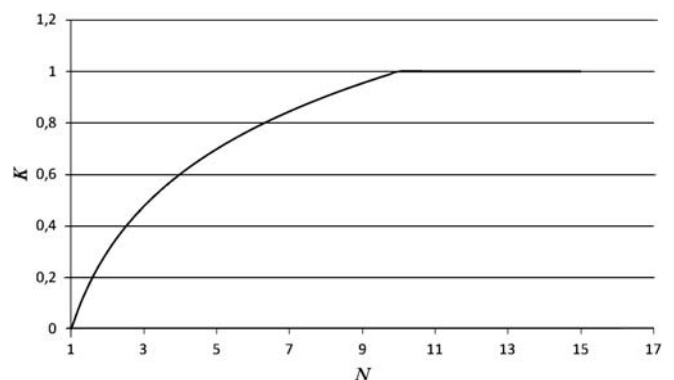


Рис. 2. График множителя нормировочного коэффициента

Используемые в исследовании кодеки и настройки к ним

Некоторые параметры файлов настройки	IPP264_High_quality	IPP264_Fast	X264_High_quality	MPEG4_High_quality	MPEG2_default	VC1_High_quality
Число кадров между кадрами типа I	200	200	200	—	12	—
Число кадров типа B между кадрами других типов	2	—	2	—	2	—

### Условия проведения эксперимента

В рамках нашего исследования мы постарались рассмотреть наиболее часто используемые кодеки и настройки к ним. Были выбраны две реализации AVC/H.264 стандарта, а также MPEG4, MPEG2 и VC1 (табл. 2).

Эксперименты проводились на 26 видеофрагментах с различными разрешениями на пяти битрейтах (см. табл. 1). Число исследованных кадров для каждого кодека превышало 42 тыс. Тестовые данные были разделены на две группы. В первую группу вошли последовательности из каждого типа для трех битрейтов, во вторую — все тестовые данные. На первой группе модель строилась посредством регрессионного анализа. После ее построения была проведена проверка на второй группе. Цель этой проверки — выяснение, как наша модель будет работать с новыми данными. Для того чтобы проанализировать полученные результаты, были построены RD-кривые для каждой последовательности.

### Метрика качества кодирования (EQM)

В результате проведенного исследования нами была предложена метрика EQM, которая позволяет приближать SSIM [11] метрику закодированной последовательности, используя только данные исходной видеопоследовательности.

Таблица 3

Коэффициенты A и B рассмотренных кодеков

Кодеки	Постоянная A	Коэффициент B
IPP264	4,202	1,184
X264	1,034	1,275
IPP264_Low	5,291	0,829
MPEG4	6,5	0,831
MPEG2	2,075	0,537
VC1	-0,480	1,076

Таблица 4

Оценки коэффициентов

Значения коэффициентов	
$c(1)$	1,339
$c(2)$	2,275
$c(3)$	7,356
$c(4)$	2,645
$c(5)$	4,900

Таблица 5

Области значений метрик

Названия метрик	Область значений
SI	[15; 80]
TI_ME_Y	[2,5; 11,5]
Blur	[0,3; 0,6]
StdDev_Y	[30; 90]
MeanValue_Y	[30; 150]

Мы выбрали для оценки качества именно SSIM, потому что она лучше коррелирует с результатами визуального сравнения, чем, например, метрика PSNR [11]. Для того чтобы оценить сложность входной последовательности, мы использовали суперпозицию метрик, которые были описаны выше. Общий вид метрики качества кодирования видео (EQM) следующий:

$$EQM = \frac{1}{1 - SSIM(x, bitrate)} \quad (1)$$

$$EQM = A + B \frac{1}{1 - F(x, bitrate)} \quad (2)$$

где  $F(x, bitrate)$  — это эмпирически подобранная суперпозиция метрик, описывающих сложность входной последовательности;  $A$  и  $B$  — уникальные коэффициенты для каждого рассмотренного кодека (табл. 3), а переменная  $x$  — исходный видеоряд. Общий вид функции  $F(x, bitrate)$  следующий:

$$\frac{1}{1 - F(x, bitrate)} = KE(a + b + c + d + g) \quad (3)$$

$$a = [c(1) \cdot 10^5] e^{-MeanValue\_Y(x)/14}; \quad (4)$$

$$b = [c(2)/10^3] SI(x)^3; \quad (5)$$

$$c = c(3) \sqrt{\frac{S}{Blur(x)}}; \quad (6)$$

$$d = [c(4)/10] e^{StdDev\_Y(x)/10}; \quad (7)$$

$$g = [c(5)/10^3] S \log(TI\_me\_Y(x)); \quad (8)$$

$$K = \log_{10} N,$$

где  $c$  — оценки коэффициентов, их значения приведены в табл. 4.

Проведя большое число экспериментов по выявлению корреляций между исходными параметрами видеопоследовательностей и качеством закодированного видео, мы убедились, что пространственная сложность вносит больший вклад, нежели временная. Поэтому мы использовали быстро растущие функции, определяющие пространственную сложность (формулы (4), (5) и (7)) и медленно растущие функции, определяющие временную сложность (например, формула (8)). В табл. 5 приведены области значений метрик.

Для того чтобы выявить зависимость между качеством кодирования и сложностью последовательности, мы использовали линейную регрессию [15]. Кроме того, мы провели тест на мультиколлинеарность и удостоверились в ее отсутствии. Следова-

Значение величины  $R^2$ 

IPP264_High_quality	IPP264_Fast	X264_High_quality	MPEG4_High_quality	MPEG2_default	VC1_High_quality	H.264_High_quality	H.264
0,71	0,69	0,72	0,69	0,67	0,69	0,71	0,67

тельно, рассмотренные регрессоры не являются сильно скоррелированными.

В регрессионной модели мы проверяли значимость коэффициентов, т. е. мы принимали или отвергали гипотезу  $H_0$  о равенстве нулю коэффициентов  $c_1, c_2, c_3, \dots, c_5$ .

$$y_i = c_1 + c_2 x_{1i} + c_3 x_{2i} + c_4 x_{3i} + c_5 x_{4i} + c_6 x_{5i} + \varepsilon_i.$$

Уровень скоррелированности результатов оценивался по величине  $R^2$ :

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - y_{average})^2}, \quad (9)$$

где  $y_i$  — реальное значение;  $\hat{y}_i$  — предсказанное значение,  $m$  — число экспериментов.

Из формулы (9) видно, что чем ближе  $R^2$  к единице, тем лучше регрессионная модель, т. е. большая доля отклонений была объяснена этими переменными.

### Результаты

В табл. 6 приведены значения величины  $R^2$  для нашей регрессионной модели. Из нее видно, насколько точно предлагаемая модель работает на реальных данных. Мы также привели пример работы метрики  $EQM$ , нарисовав RD-графики для одной из типичных последовательностей, и пример видеоряда, с которым наша метрика справилась хуже всего (рис. 3, 4). Графики приведены для шести кодеков. Полученные результаты позволяют говорить о практической ценности метрики  $EQM$ .

В частности, можно сделать вывод о том, что полученная оценка видеокачества  $EQM$  хорошо приближает реальные значения качества, полученные путем кодирования видеопоследовательностей.

Из рис. 3 и 4 видно, что приближение лучше всего работает на кодеках AVC, а точность приближения для MPEG2 несколько ниже. Хотя в обоих случаях приближение является достаточно точным, чтобы судить о характере качества кодеков. Обычно ошибка  $EQM$  выше на низких битрейтах и мала на высоких.

Очевидно, что найденные эмпирическим путем параметры характеризуют не все возможные варианты конфигурации кодека, однако указанные параметры являются типичными. Для быстрой оценки параметров новой конфигурации кодека можно использовать линейную интерполяцию или экстраполяцию параметров  $A$  и  $B$  из табл. 3. Для более

точной оценки желательно знать, какие алгоритмы качества включены в кодек, например, наличие или отсутствие В кадров, multi reference и т. д.

Также интересно отметить тот факт, что с ростом битрейта вклад оценки движения в общее качество уменьшается. Это объясняется тем, что число ненулевых коэффициентов на высоких битрейтах больше и качество изображения в большей степени определяется тем, насколько хорошо кодек может сохранить оригинальные значения, минимально искажая их при квантовании или при оценке движения.

Мы полагаем, что предложенный метод является хорошим балансом между сложностью вычислений и эффективностью использования. Добавление до-

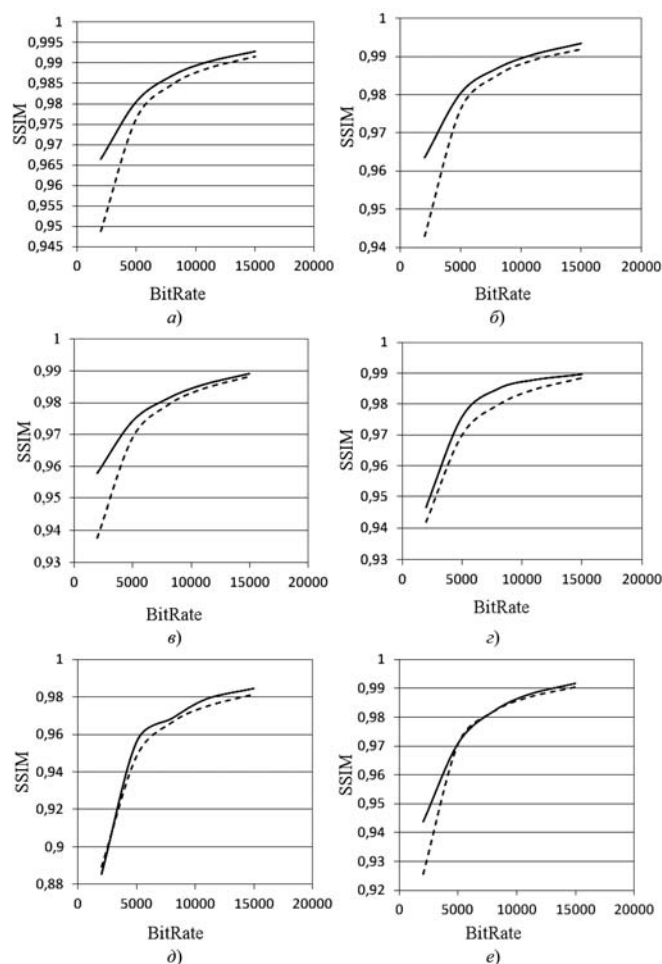
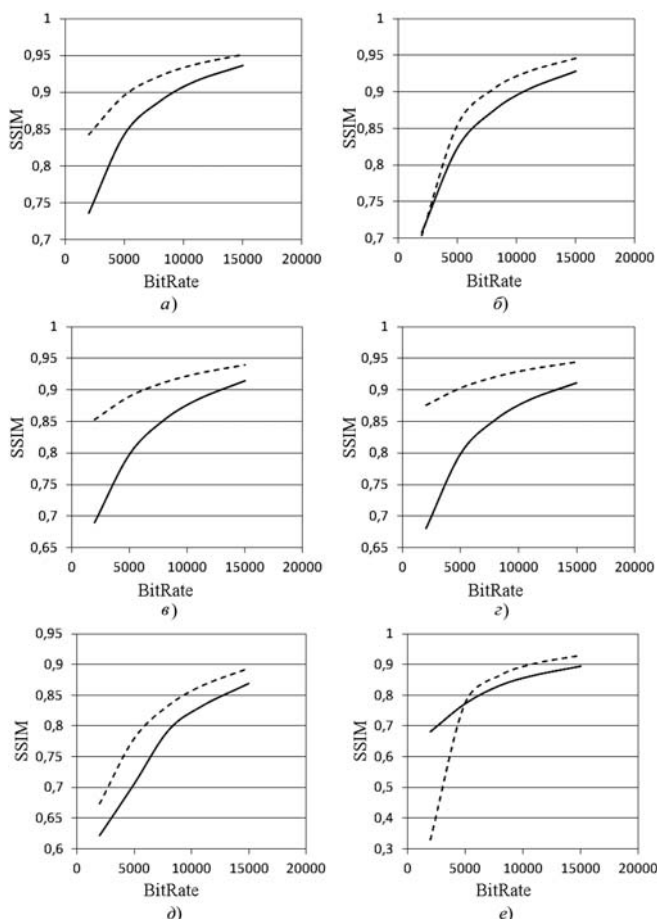


Рис. 3. Пример RD-кривых типичного случая работы метрики  $EQM$  (значения метрики отмечены штриховой линией), последовательность Building. На рисунке кривые для шести кодеков (значения  $SSIM$  для кодеков отмечены сплошной линией): а — IPP264; б — X264; в — IPP264\_Low; г — MPEG4; д — MPEG2; е — VC1



**Рис. 4.** Пример RD-кривых наихудшего случая работы метрики EQM (значения метрики отмечены штриховой линией), последовательность Parkrup\_1. На рисунке кривые для шести кодеков (значения SSIM для кодеков отмечены сплошной линией): а — IPP264; б — X264; в — IPP264\_Low; г — MPEG4; д — MPEG2; е — VC1

полнительных регрессоров увеличивает вычислительную сложность и требует дополнительного анализа увеличения точности. Удаление метрик, по нашим наблюдениям, приводит к существенно-му снижению качества оценки.

Мы не можем утверждать, что функциональные зависимости (1)–(8) являются единственными или оптимальными. Однако для практической работы эти формулы дают достаточно высокую точность оценки качества, чтобы использовать полученные результаты для кодирования потокового видео (например, видеоконференций) или для видео высокого качества (видео для Blu-Ray).

### Заключение

В работе предложена двумерная параметрическая оценка качества видеокодеков без процесса кодирования, причем полученные результаты могут быть применены к широкому классу видеокодеков. С помощью полученной метрики могут быть улучшены алгоритмы кодирования видео, такие как rate control, mode decision и др. Нам не известны

аналогичные результаты, применимые в таком широком диапазоне битрейта и целевых моделях использования, как описанная в настоящей работе метрика EQM.

Можно также утверждать, что полученные метрики качества позволяют классифицировать видеопоследовательности по сложности, что существенно помогает структурировать тестирование видеокодеков. Посчитав EQM, можно заранее классифицировать исходные последовательности по сложности. Выбрать только уникальные видеоряды (убрать дублирующиеся по сложности последовательности), а также обеспечить полноту покрытия разных диапазонов сложности (для хорошего покрытия необходимы как простые последовательности, так и сложные).

Развитием настоящего исследования может быть анализ других метрик качества. Вместо использованной метрики SSIM интересно было бы провести корреляцию с результатами субъективных тестов или других, считающихся более точными, метриками качества видео. Кроме того, можно рассмотреть и другие функциональные зависимости для дальнейшего повышения точности предсказания качества.

### Список литературы

1. Girod B. Efficiency Analysis of Multihypothesis Motion-Compensated Prediction for Video Coding // IEEE Transactions on Image Processing. Feb. 2000. Vol. 9, N 2. P. 173–183.
2. International Standard ISO/IEC 13818. Information Technology, Coding of Moving Pictures and Associated Audio Information, 1994.
3. International Standard ISO/IEC 14496-2. Information Technology, Coding of Audio-Visual Objects, Part 2: Visual, 2001.
4. ITU-T Recommendation H.264|ISO/IEC 14496-10. Audiovisual and Multimedia Systems. Infrastructure of audiovisual services Coding of moving video. Advanced video coding for generic audiovisual services, 2005.
5. Ries M., Crespi C., Nemethova O., Rupp M. Content Based Video Quality Estimation for H.264/AVC Video Streaming // Wireless Communications and Networking Conference IEEE, Kowloon Hong Kong, 2007. P. 2668–2673.
6. Gastaldo P., Rovetta S., Zunino R. Objective quality assessment of MPEG-2 video streams by using CBP neural networks // IEEE Transactions on Neural Networks. 2002. Iss. 4. P. 939.
7. Cleary J., Witten I. Data compression using adaptive coding and partial string matching // IEEE Transactions on Communications. 1984. Vol. 32. N 4. P. 396–402.
8. SMPTE RP 227 VC-1 Bitstream Transport Encodings, 2010.
9. TU-T Recommendation H.263: Transmission of non-telephone signals. Video coding for low bit rate communication. 1996.
10. Winkler S. Digital video quality: vision models and metrics. J. Wiley & Sons, 2005. P. 170.
11. Wang Z., Bovik A., Sheikh H., Simoncelli E. Image quality assessment: From error visibility to structural similarity // IEEE Transactions on Image Processing. 2004. Vol. 13. N 4. P. 600–612.
12. TU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications. 1999.
13. Crété F., Dolmiere T., Ladret P., Nicolas M. The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric // Proceedings of SPIE. Vol. 6492. Human Vision and Electronic Imaging XII. February 2007.
14. Huang C., Liao B. A robust scene-change detection method for video segmentation // IEEE Transactions on Circuits and Systems for Video Technology. 2001. Vol. 11. P. 1281–1288.
15. Montgomery D., Peck E., Vining G. Introduction to Linear Regression Analysis. Wiley, 2001. P. 635.



УДК 004.67

**И. А. Сенотрусов**, канд. техн. наук, доц.,  
**П. А. Нешков**, слушатель,  
 e-mail: pavelnesh@mail.ru  
 Академия федеральной службы охраны  
 Российской Федерации, г. Орел

## Анализ и заполнение пропусков данных в базах данных информационно-аналитических систем

*Рассмотрена проблема повышения эффективности наблюдения за социально-экономическими процессами в регионах Российской Федерации за счет использования процессов анализа и заполнения пропусков данных в базах данных информационно-аналитических систем. Рассматривается применение базового алгоритма ZET заполнения пробелов.*

**Ключевые слова:** информационно-аналитическая система, прогноз, алгоритм ZET заполнения пробелов, компетентные строки и столбцы

### Введение

В связи с проводимыми в Российской Федерации политическими и экономическими преобразованиями особую важность приобретают проблемы

информационной поддержки процессов принятия решений органами государственной власти.

Главную роль при этом играет информация о состоянии социально-экономического развития регионов, вследствие этого одной из приоритетных задач является задача анализа и прогнозирования. Данная задача решается путем обработки социально-экономической информации информационно-аналитическими системами (ИАС) в органах государственной власти. Они созданы в интересах органов государственной власти для наблюдения за социально-экономическими процессами, определения динамики и тенденций выбранных показателей, расчета интегральных индексов развития регионов, проведения межрегионального сравнительного анализа с ранжированием и определением мест каждого региона.

Процесс прогнозирования развития социально-экономической ситуации в регионе представлен на рис. 1. Он состоит из четырех этапов. В блоке задания исходных данных осуществляется выбор социально-экономических показателей и региона для анализа и прогнозирования развития социально-экономической ситуации в регионе, которые потом передаются в блок выбора значений из базы данных для формирования отчета. На его выходе получаем совокупность значений показателей, используемых для прогноза. В блоке обработки значений составляется прогноз значений показателей развития отрасли. На завершающем этапе формируется

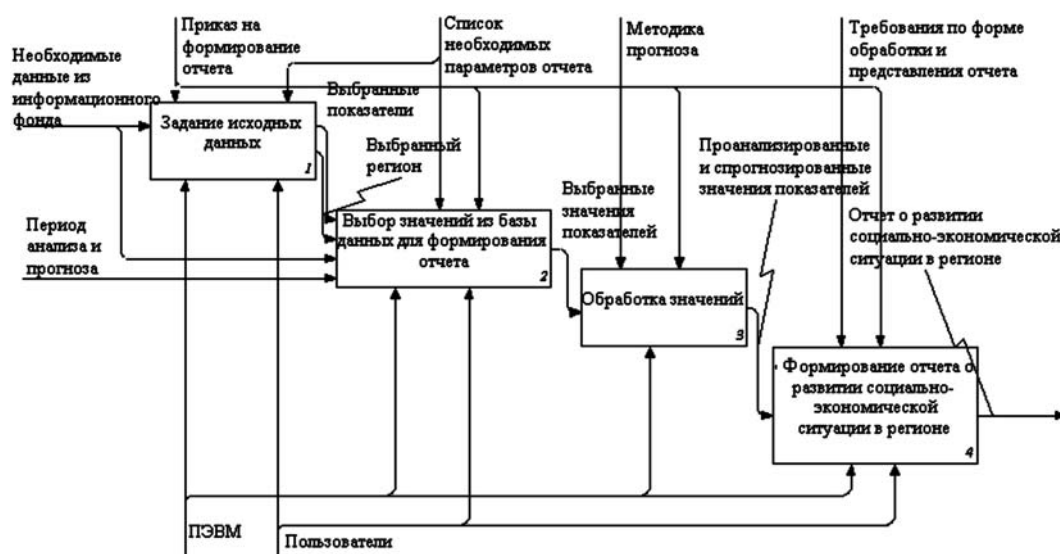


Рис. 1. Модель существующего процесса прогнозирования развития социально-экономических показателей в регионе (AS-IS)

	2004	2005	2006	2007	2008	2009
Промышленность	543			564	663	754
Сельское хозяйство		435	445		375	345
Транспорт	344	564		645	643	598
Финансы	365	354	434	552		654
Макроэкономика		266	345	435	546	543

отчет о развитии социально-экономической ситуации в регионе.

Однако в ИАС существуют пропуски информации — реальные таблицы данных часто содержат пробелы: у некоторых объектов  $a_i$  значение того или иного признака  $x_j$  может отсутствовать. В результате на вход программ анализа данных подается таблица с пропущенными клеточками (см. таблицу).

### Постановка задачи

Для повышения эффективности процессов мониторинга, комплексного анализа и прогнозирования важнейших социально-экономических показателей на основе информации, содержащейся в ИАС, необходимо заполнить пропуски информации. Для этого в модель процесса прогнозирования развития социально-экономической ситуации в регионе необходимо добавить блок заполнения пропусков данных (рис. 2). Он будет осуществлять поиск пробелов в базе данных ИАС и заполнять их по определенной методике.

Существует огромное число методов заполнения пропусков в базах данных. Главная сложность их заключается в том, что для заполнения пропусков в базах данных необходимо выбрать алгоритм заполнения, который бы наиболее точно предсказывал пропущенные величины. Эти данные впоследствии

могут быть применены для прогнозирования социально-экономического состояния региона в будущем.

В данной статье будет рассмотрен базовый алгоритм ZET заполнения пробелов. Преимуществом этого базового алгоритма является определение компетентных строк и столбцов в таблице для подбора таких параметров в формуле предсказания пропущенного элемента, при которых ожидаемая ошибка предсказания достигнет минимума.

### Методы решения задачи

Рассмотрим базовый алгоритм ZET заполнения пробелов. В его основе лежат три предположения. Первое (гипотеза избыточности) состоит в том, что реальные таблицы имеют избыточность, проявляющуюся в наличии похожих между собой объектов (строк) и зависящих друг от друга свойств (столбцов).

Второе предположение (гипотеза локальной компактности) состоит в утверждении, что для предсказания пропущенного элемента  $b_{ij}$  нужно использовать не всю таблицу, а лишь ее "компетентную" часть, состоящую из элементов строк, похожих на строку  $i$ , и элементов столбцов, похожих на строку  $j$ . Остальные строки и столбцы для данного элемента неинформативны.

Третье предположение (гипотеза линейных зависимостей) заключается в том, что из всех возможных видов зависимостей между столбцами (строками) в алгоритме ZET используются только линейные зависимости, зависимости более сложного характера не рассматриваются, так как для их обнаружения требуется большой объем данных.

Под компетентностью  $l$ -й строки по отношению к  $i$ -й понимается величина  $L_{il} = r_{il}t_{il}$ . Здесь  $r_{il} = 1 - \rho_{il}$ ;  $\rho_{il}$  — евклидово расстояние между  $i$ -й и  $l$ -й строками;  $t_{il}$  — коэффициент комплектности,

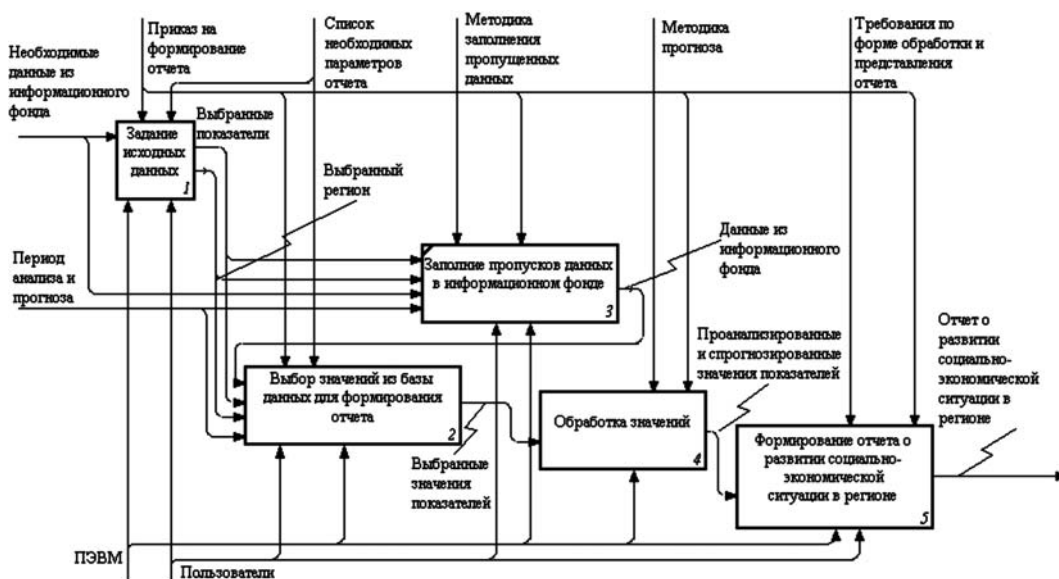


Рис. 2. Концептуальная модель процесса прогнозирования развития социально-экономических показателей в регионе (ТО-ВЕ)

равный числу свойств, значения которых известны как для  $i$ -й, так и для  $l$ -й строки. Компетентная строка не должна содержать пробела в  $j$ -м столбце.

Под компетентностью  $k$ -го столбца по отношению к  $j$ -му понимается величина  $L_{jk} = r_{jk}t_{jk}$ , где  $r_{jk}$  — модуль коэффициента корреляции между  $j$ -м и  $k$ -м столбцами, а  $t_{jk}$  — коэффициент компактности, равный числу объектов, у которых известны как  $j$ -е, так и  $k$ -е свойства. Компетентный столбец не должен иметь пробела в  $i$ -й строке.

В процессе предсказания значения пробела с использованием зависимостей между  $j$ -м и всеми остальными ( $k$ -ми) столбцами вырабатываются "подсказки"  $b_k$ . Для их получения используется уравнение линейной регрессии между  $j$ -м и  $k$ -м столбцами (рис. 3). Если в подматрице было  $(q + 1)$  столбцов, то затем  $q$  подсказок усредняются с весом, пропорциональным компетентности соответствующего столбца. В итоге получается прогнозная величина  $b_q$ , порожденная избыточностью, содержащейся в столбцах

$$b_j = \sum_{k=1}^q b_k L_{jk}^\alpha / \sum_{k=1}^q L_{jk}^\alpha,$$

где  $\alpha$  — коэффициент, регулирующий влияние компетентности на результат предсказания. При малых значениях  $\alpha$  разница в компетентности сказывается мало, при больших  $\alpha$  более компетентные столбцы влияют гораздо больше других. Выбор  $\alpha$  и составляет суть этапа подбора формулы для прогнозирования: все известные элементы  $j$ -го столбца предсказываются при разных значениях  $\alpha$ , и затем выбирается такое значение  $\alpha$ , при котором ошибка прогноза  $\delta_j$  была минимальной. Процедура заполнения пробела с использованием связи между  $i$ -й строкой и всеми  $s$  другими ( $l$ -ми) строками аналогична описанной выше и выполняется по формуле

$$b_i = \sum_{l=1}^s b_l L_{il}^\alpha / \sum_{l=1}^s L_{il}^\alpha.$$

Общий прогноз  $b'_{ij}$  значения пропущенного элемента  $b_{ij}$  получается выбором либо прогноза  $b_j$ , если  $\delta_j < \delta_i$ , либо прогноза  $b_i$ , если  $\delta_i < \delta_j$ . Возможно

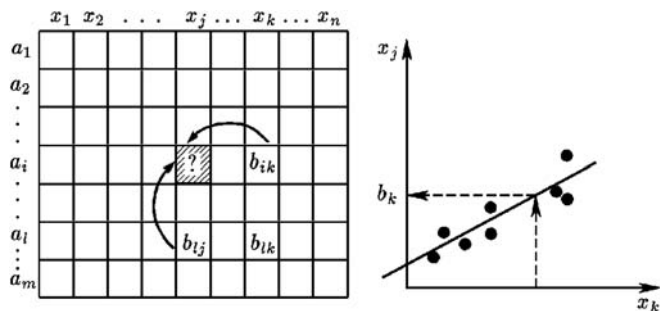


Рис. 3. Определение зависимости между  $j$ -м и  $k$ -м столбцами

и их усреднение с весом, обратно пропорциональным значению ожидаемой ошибки:

$$b'_{ij} = \lfloor b_j / (\varepsilon + \delta_j) + b_i / (\varepsilon + \delta_i) \rfloor / \lfloor (\varepsilon + \delta_j)(\varepsilon + \delta_i) \rfloor 2\varepsilon + \delta_j + \delta_i \rfloor,$$

где  $\varepsilon$  — константа, например, равная 0,01, введенная для предотвращения деления на нуль.

О значении ожидаемой ошибки  $d_{ij}$  можно судить по ошибкам  $\delta_j$  и  $\delta_i$  предсказаний известных элементов  $i$ -й строки и  $j$ -го столбца при наилучшем значении  $\alpha$ . Эксперименты показывают, что корреляция между средним значением этих ошибок  $\delta^* = [\delta_j + \delta_i]/2$  и ошибкой  $d_{ij}$  всегда положительна.

Второй способ определения ожидаемой ошибки основан на оценке дисперсии подсказок. Вычисляется дисперсия значений подсказок  $b_k$  и  $b_l$ , получаемых от всех  $k$  столбцов и  $l$  строк компетентной подматрицы. Большая дисперсия указывает на отсутствие устойчивой закономерной связи между элементом  $(ij)$  и другими элементами подматрицы, т. е. на отсутствие компактности. В этом случае не будет высокой точность предсказания значения  $b_{ij}$ . Эксперименты показали, что коэффициент корреляции между дисперсией и ошибкой предсказания  $d_{ij}$  достигает значения +0,7, т. е. это указывает на наличие сильной корреляционной зависимости.

Кроме алгоритма ZET заполнения пробелов, существуют алгоритмы семейства WANGA. Они, как и алгоритм ZET, основаны на гипотезе локальной компактности, но предназначены для заполнения пробелов в таблицах с разнотипными переменными.

## Заключение

Применяя программно алгоритмы ZET и WANGA, можно заполнить пропуски информации в базах данных информационно-аналитических систем и сделать прогноз социально-экономических показателей, тем самым повышая эффективность контроля за важнейшими социально-экономическими процессами в регионах.

Кроме того, заполнять пробелы в информационном фонде информационно-аналитических систем можно в одном из следующих режимов:

- заполнение всех пробелов;
- заполнение пробелов, ожидаемая ошибка для которых не превышает заданного значения;
- заполнение пробелов только на базе информации, имеющейся в исходной таблице;
- заполнение каждого следующего пробела с использованием исходной информации и прогнозных значений ранее заполненных пробелов.

## Список литературы

1. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. Новосибирск: ИМ СО РАН, 1999. — 270 с.

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И УПРАВЛЕНИИ

УДК 004.94

**Г. В. Ващенко**, канд. физ.-мат. наук, доц.,  
e-mail: gavaki@yandex.ru

**Д. В. Капулин**, канд. техн. наук, доц.,  
e-mail: kapulin@gmail.com

Сибирский федеральный университет,  
г. Красноярск

## Возможности интеграции бизнес-процессных моделей и ERP-систем

*Представлено универсальное решение на базе платформы IC: Предприятие 8.1, позволяющее импортировать модели ARIS в ERP-систему. Применение предложенного подхода позволяет осуществлять поддержку и обновление бизнес-процессов, не изменяя конфигурации ERP-системы.*

**Ключевые слова:** процессный подход, проектирование информационных систем, моделирование бизнес-процессов, ARIS

### Введение

Наиболее перспективными корпоративными информационными системами (КИС) на сегодняшний день являются ERP-системы, позволяющие достичь согласованности в работе различных подразделений компании, что значительно снижает административные издержки и устраняет проблему интеграции данных для разных приложений, поскольку все предприятие работает в едином информационном пространстве. Кроме того, применение новых подходов к управлению организацией, таких как процессный подход, также дает ERP-системам следующие конкурентные преимущества:

- возможность оптимизации системы управления, способной гибко реагировать на трансформацию внешней среды;
- возможность получить и реализовывать систему показателей и критериев оценки результативности управления на каждом этапе управленческой цепочки;
- сосредоточенность управления на стратегических процессах;
- обеспечение качества выпускаемой продукции в соответствии с требованиями ИСО 9000;

- четко установленный порядок в разработке, согласовании, утверждении и ведении документации;
- результаты одного управленческого процесса способствуют улучшению результатов другого.

С введением процессного подхода появляется единый язык описания деятельности, доступный и понятный всем участникам процесса, а также возможность простой и наглядной графической трактовки деятельности. Также при процессном подходе происходит выделение рабочих зон, что облегчает формулирование требований к персоналу, привлеченному к конкретным работам, и составление более точного и прицельного штатного расписания организации.

Процессный подход к управлению — это простое и надежное определение точек контроля, что достигается разбиением основного бизнес-процесса на subprocesses. Процессный подход обеспечивает единую точку зрения на управление предприятием, что приводит к упрощению многоуровневых иерархических организационных структур, присущих функциональному подходу, и обеспечивает большую ориентацию организации на потребителя.

За счет сокращения уровней организационной структуры упрощается обмен информацией между подразделениями и устраняется обособленность подразделений и должностных лиц [1, 2].

Таким образом, процессный подход является эффективным обоснованным способом решения проблем, связанных с разработкой и внедрением КИС. Но эффективное применение такого подхода невозможно без интеграции результатов бизнес-процессного моделирования и ERP-систем. Такого рода интеграция является задачей многошаговой, занимающей значительное количество времени, трудовых и материальных затрат, поэтому для ее реализации принципиально важно выбирать и использовать инструментальные средства, позволяющие описывать и анализировать бизнес-процессы, делать их прозрачными и управляемыми.

### Постановка задачи

Для внедрения процессного подхода в управление организацией широко применяют разработки фирмы *IDS Scheer*, представляющие собой методологию и программный продукт *ARIS*. А для перепроектирования или реорганизации компании или отдельного ее подразделения, чтобы снизить затраты или быстро провести интеграцию после слияния, применяется такое решение, как *ARIS Value Engineering (AVE)* [3].

Методология проектирования *AVE* используется на всех этапах цикла работ по созданию и развитию бизнеса — от разработки стратегии предприятия до реорганизации основополагающих бизнес-процессов, от управления стоимостью процессов до внедрения информационных систем и последующей оптимизации деятельности предприятия [3, 4].

Для того чтобы получить возможность модификации и настройки модели описанных бизнес-процессов предприятия непосредственно с участием корпоративной информационной системы, необходимо, чтобы модель была определенным образом связана с ней. Эта задача решается интеграцией (объединением) этих двух сред "мостом", который позволит переносить новые способы ведения бизнеса из моделей процессов в пространство информационных систем.

Широко используемая для построения КИС платформа 1С: Предприятие 8.1 имеет свой базовый набор инструментов по работе с управлением процессами (объект конфигурации "Бизнес-процессы"), но не имеет достаточных возможностей для описания структуры всех процессов и подпроцессов предприятия и того функционала, который заложен в программных продуктах *ARIS*, специально для этого предназначенных. Также работа с бизнес-процессами в 1С: Предприятие заложена только на уровне программного кода, т. е. существует возможность запрограммировать действующие процессы на предприятии, но нет возможности модифицировать и редактировать схемы их функционирования в пользовательском режиме. Тем не менее гибкость рассматриваемой платформы, заложенные в нее возможности совместного функционирования с различным программным обеспечением позволяют решить актуальную задачу интеграции методологии *ARIS* и *ERP* для разработки универсального прикладного решения, которое, в свою очередь, позволит импортировать модели бизнес-процессов из *ARIS* в *ERP*-систему, что послужит основой для формирования инструмента модификации бизнес-процессов предприятия непосредственно в рабочей среде.

### Методика решения и анализ результатов

Поставленная цель может быть достигнута выполнением следующих задач:

1) описать конкретный экземпляр бизнес-процесса предприятия с помощью методологии *ARIS Value Engineering*;

2) разработать алгоритм выгрузки моделей бизнес-процессов, представленных в программных продуктах *ARIS*;

3) разработать алгоритм загрузки и конвертации выгруженных моделей в модели бизнес-процессов 1С: Предприятие.

В качестве примера рассмотрим типовой процесс оптовой продажи с постоплатой, который может использоваться для производственных предприятий различной направленности. Рассмотрим модель сце-

нария бизнес-процесса, представленного с помощью методологии *ARIS Value Engineering* (рис. 1).

Сценарная модель бизнес-процесса, которая отражает последовательность действий на предприятии в рамках данного процесса, позволяет назначить владельца процесса, распределить все операции по исполнителям, а также определить ключевые показатели эффективности, влияющие на достижение целей компании. Но представленная модель существует отдельно от информационной системы, в которой ведется учет торговых операций, что не позволяет использовать реальные учетные данные для отслеживания процедуры выполнения бизнес-процессов на предприятии, настраивать и модифицировать их функционирование, а также отсутствует возможность анализа статистических данных для формирования решений и прогнозов для дальнейших действий по реализации стратегической цели. Таким образом, требуется, чтобы модель процесса была связана с используемой *ERP*-системой.

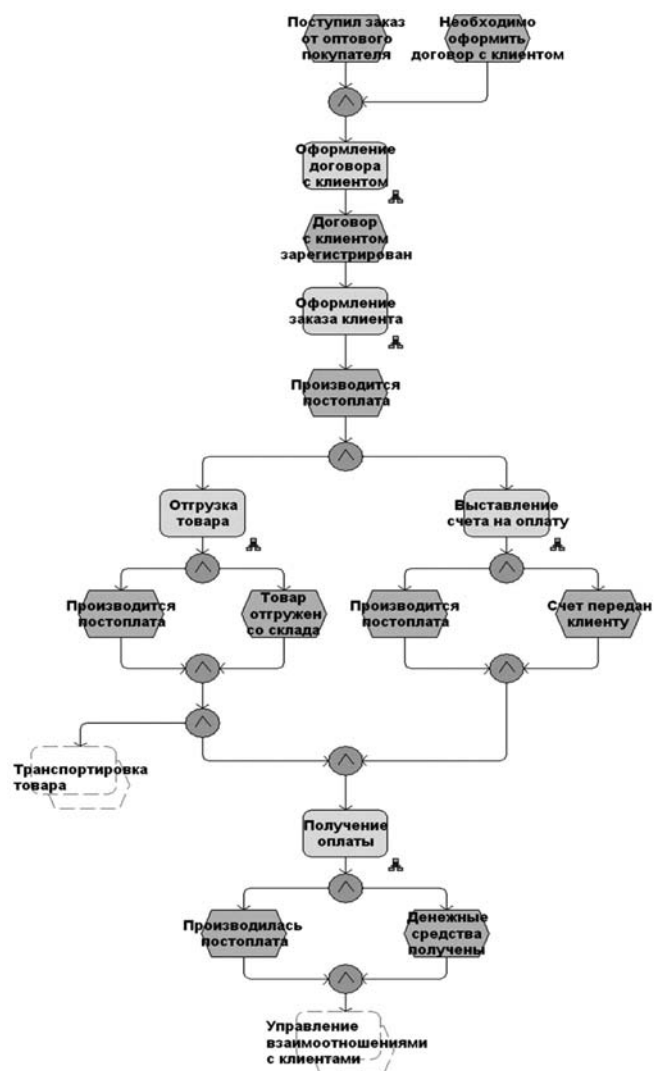


Рис. 1. Модель сценария процесса оптовой продажи с постоплатой

Работа с процессами в 1С: Предприятие реализована в виде объектов конфигурации "Бизнес-процессы" и "Задачи". Бизнес-процессы предназначены для управления последовательностью действий, направленных на достижение цели в контексте автоматизируемой предметной области. Задачи бизнес-процесса предназначены для отражения выдачи и исполнения заданий участниками бизнес-процессов

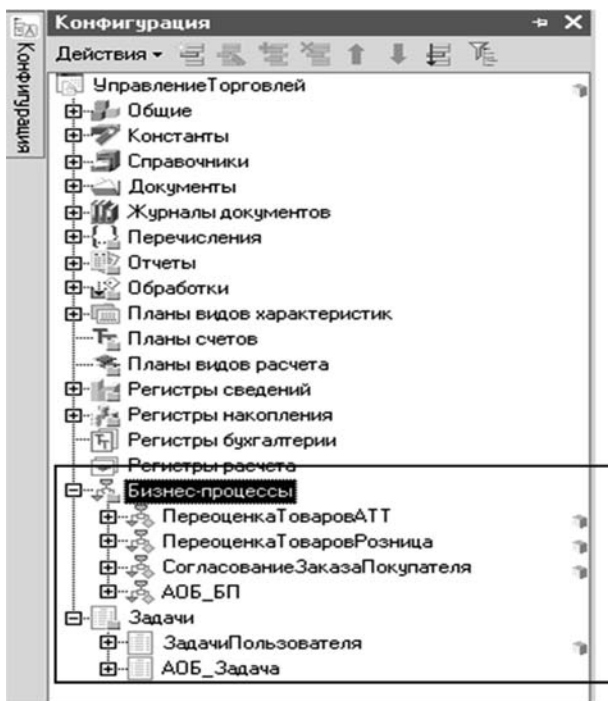


Рис. 2. Объекты конфигурации "Бизнес-процессы" и "Задачи"

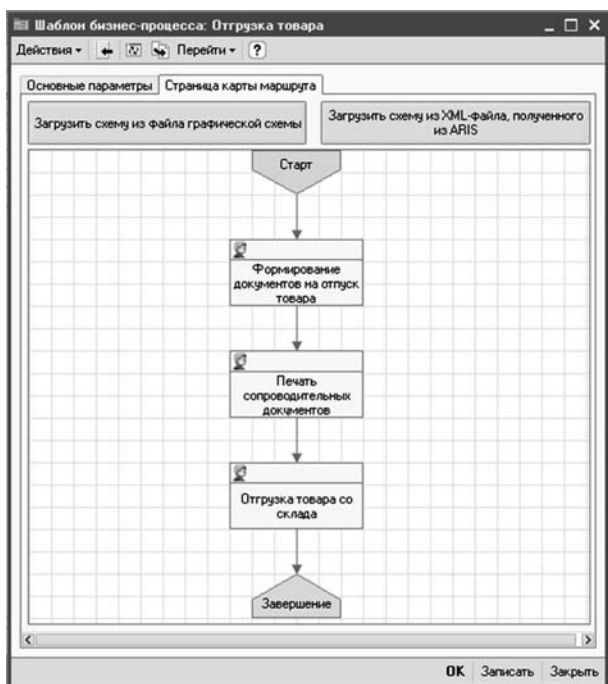


Рис. 3. Пример карты маршрута бизнес-процесса в 1С: Предприятие

или обычными пользователями системы. Задачи могут применяться самостоятельно или использоваться для обеспечения функционирования бизнес-процессов разного вида. На рис. 2 представлено дерево конфигурации и выделены описанные объекты.

У бизнес-процесса есть свойство "Карта маршрута". Оно является ключевым свойством бизнес-процесса, наглядно описывает жизненный цикл от старта до завершения, позволяет реализовать визуальное проектирование в терминах предметной области, а также является нотацией, понятной не только специалистам, но и владельцам бизнес-процессов. Пример карты маршрута приведен на рис. 3.

Реализованный функционал по управлению бизнес-процессами, заложенный в 1С: Предприятие для настройки и модификации схем функционирования процессов, не может быть использован ввиду того, что в существующих в настоящее время конфигурациях работа с процессами заложена в самой конфигурации (то есть запрограммировать бизнес-процесс можно, но при этом отсутствует возможность модифицировать и настраивать его в пользовательском режиме).

Для реализации механизма конвертации моделей были определены соответствия объектов сценарных моделей ARIS и карт маршрутов бизнес-процессов 1С: Предприятие (см. таблицу).

Для переноса модели в ERP-систему разработана программа (скрипт) на языке ARIS Sax Basic, позволяющая перенести информацию из графических моделей в файлы документов в соответствии с определенными правилами. Результатом выполнения

Соответствие основных объектов сценарных моделей в ARIS с объектами карт маршрутов бизнес-процессов в 1С: Предприятие

Название объекта ARIS	Описание преобразования ARIS—1С
Функция (Function)	Объект соответствует процедуре. Преобразуется в 1С в точку "Действие"
Интерфейс процесса (Process Interface)	Объект используется для ссылок на функцию внешнего бизнес-процесса, который не описывается в данной модели. В 1С не отображается
Событие (Event)	Объект описывает начальное и конечное событие каждой процедуры. В 1С используется как точка "Выбор варианта" или не отображается
AND rule, XOR rule, OR rule	Правила ветвления или соединения процесса: "И", "Исключающее ИЛИ", "ИЛИ". В 1С "И" преобразуется в точки "Слияние" и "Разделение", "Исключающее ИЛИ" — в точку "Выбор варианта", "ИЛИ" — в совокупность указанных точек
Организационная единица (Organization unit)	Элемент организационной структуры, структурное подразделение. В 1С используется для адресации действия ("Отдел продаж")
Вложенный бизнес-процесс	Объект указывает на наличие вложенного подпроцесса. В 1С преобразуется в точку "Вложенный процесс"

программы является xml-файл, который может быть далее использован для загрузки в 1С: Предприятие. В свою очередь, в 1С: Предприятие написана обработка (программа), которая позволяет загрузить полученный xml-файл и преобразовать сведения об объектах модели в ARIS в объекты карты маршрута бизнес-процесса.

Карта маршрута бизнес-процесса представляет собой совокупность точек различного типа и соединительных линий. Каждая точка карты маршрута в 1С: Предприятие может быть представлена в виде определенной последовательности символов. Для каждой точки карты маршрута бизнес-процесса прописывается шаблон, куда потом при записи модели бизнес-процесса в качестве параметров передаются соответствующие значения из загружаемого xml-файла. Таким образом, поэлементно создается схема в 1С: Предприятие.

### Заключение

В статье представлено универсальное интегрированное решение на базе платформы 1С: Предприятие 8.1, которое позволяет импортировать мо-

дели, созданные с помощью методологии *ARIS Value Engineering* в ERP-систему для настройки модели функционирования бизнес-процессов в пользовательском режиме ERP-системы.

Разработанное решение является универсальным с точки зрения интеграции с типовыми решениями фирмы 1С. То есть независимо от конфигурации, в которую будет встраиваться разработанное решение, конфигурация остается типовой и сохраняется возможность ее поддержки и обновления в стандартизованном виде.

### Список литературы

1. **Выборнова В. В.** Процессный подход как самый эффективный метод управления организациями. [Электронный ресурс]. URL: <http://conf.bstu.ru/conf/docs/0033/0757.doc>.
2. **Аскарв Е.** Процессный подход в системе менеджмента качества // Региональный еженедельник "Без проблем". 2007. № 45. С. 12–18.
3. **Ильин В. В.** Реинжиниринг бизнес-процессов с использованием ARIS. 2-е изд. М.: Вильямс, 2008. 256 с.
4. **ARIS Value Engineering.** Ускорение проектов и обеспечение их успешности [Электронный ресурс]. URL: [http://www.ids-scheer.ru/ru/ARIS/ARIS\\_Reference\\_Models/ARIS\\_Value\\_Engineering/106085.html?mod\\_dct\[caching\]=62965754](http://www.ids-scheer.ru/ru/ARIS/ARIS_Reference_Models/ARIS_Value_Engineering/106085.html?mod_dct[caching]=62965754).

УДК 519.86

**Ю. И. Димитриенко**, д-р физ.-мат. наук, проф.,  
e-mail: [dimit@serv.bmstu.ru](mailto:dimit@serv.bmstu.ru),  
**О. Ю. Димитриенко**, аспирант  
e-mail: [olyad@mail.ru](mailto:olyad@mail.ru),  
МГТУ им. Н. Э. Баумана

## Континуальное моделирование динамики кластеров экономических данных в условиях внешних кризисных воздействий

### Введение

Прогноз воздействий, оказываемых внешними силами на динамику данных в экономике, является одной из основных задач математического моделирования экономических процессов. Статистические и эконометрические методы, обычно используемые на практике при прогнозировании динамики данных по продажам, применимы для эволюционных процессов и при отсутствии кризисных явлений в экономике. Для прогнозирования влияния резких экономических скачков нужны фундаментальные модели, описывающие характер воздействий, оказываемых на экономическое поведение покупателей. В ряде предыдущих работ [1, 2] был предложен принципиально новый подход к моделированию динамики рынка массовых продаж, основанный на введении многомерного пространства товаров, кластеризации покупателей в этом пространстве по данным об их покупках, переходе к континуальному описанию данных и анализу движения кластеров с помощью введенных законов сохранения, аналогичных законам обобщенной на  $n$ -мерный случай механики сплошных сред.

Целью настоящей статьи является построение модели влияния внешних воздействий на стабильный процесс покупок [2] и динамику движения экономических кластеров в пространстве товаров  $E_n$ .

*Для модели динамики рынка массовых продаж, основанной на кластеризации покупателей в многомерном пространстве, а также на построении континуальной модели движения кластеров, предложена модель воздействия внешних сил на покупательское поведение кластеров. Представлены два основных случая на рынке: стабильный и нестабильный процессы покупок. Приведены примеры численного расчета параметров влияния внешних воздействий, проведено сравнение результатов прогнозирования с экспериментальными данными и популярными методами прогноза.*

**Ключевые слова:** математическое моделирование, влияние рынка, внешние факторы, многомерное пространство, уравнение движения, кластеры покупателей

## Общие уравнения динамики кластеров экономических данных

Модель движения экономических кластеров [1] основывается на введении многомерного пространства товаров  $E_n$ , где  $n$  — число наименований различных товаров. В этом пространстве каждая точка  $\mathbf{x} = \{x^1, \dots, x^n\}$  представляет собой набор товаров  $i = 1, \dots, n$ ,  $x^i \in \mathbf{R}$ , купленный или планируемый к покупке одним покупателем за время  $t$ . Таким образом, закон движения покупателя во введенном пространстве описывается функцией  $\mathbf{x} = \mathbf{x}(t) = \{x^1(t), \dots, x^n(t)\}$ , дважды дифференцируемой на анализируемом интервале времени  $[0, t_{\max}]$ .

Далее, при построении модели предполагается, что существуют взаимосвязи между отдельными группами покупателей, которые заставляют покупателей двигаться в пространстве товаров согласованно, подчиняясь групповым закономерностям. Такие связи могут быть представлены схожими социальными, демографическими и психологическими качествами покупателей, а также информационными связями. Таким образом, применяя алгоритм кластеризации, предложенный в работе [1], можно выделить из общей массы покупателей  $N$  в фиксированный момент времени  $t$  более крупные группы покупателей, обладающих коллективным поведением, — кластеры. При переходе к континуальному описанию пространства  $E_n$  каждый кластер представляется областью  $V$  в пространстве.

Если предполагается, что связи, объединяющие покупателей, входящих в один кластер  $V$ , достаточно сильны и поэтому расстояния между покупателями внутри кластера не меняются в зависимости от времени, то имеет место модель жесткого кластера, согласно которой движение каждой точки кластера может быть представлено как

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{Q} \cdot \tilde{\mathbf{x}}', \quad (1)$$

где  $\mathbf{x}_0 = x_0^i(t)\mathbf{e}_i$  обозначает поступательное движение центра масс кластера, а также допускается вращательное движение кластера в пространстве, введенное через ортогональный тензор поворота  $\mathbf{Q} = Q^{ij}(t)\mathbf{e}_i \otimes \mathbf{e}_j$  и радиус-вектор  $\tilde{\mathbf{x}}'$  рассматриваемой точки в лагранжевом описании  $\mathbf{x}' = X^i\mathbf{e}_i$ ,  $\mathbf{e}_i$  — ортонормированный декартов базис в пространстве  $E_n$ ,  $X^i$  — Лагранжевы координаты точки.

Для движения жесткого кластера в  $n$ -мерном пространстве товаров формулируются следующие законы сохранения:

1) закон сохранения числа индивидуумов (покупателей) в кластере (2);

2) закон изменения количества движения (частоты покупок) кластера (3) — отражает изменение суммарного для кластера числа покупок каждого вида в единицу времени;

3) закон изменения момента количества движения (момента частоты покупок) (4):

$$\frac{d}{dt} \int_V \rho dV = 0; \quad (2)$$

$$\frac{d}{dt} \int_V \rho \mathbf{v} dV = \int_V \rho \mathbf{f}_m dV + \int_{\Sigma} \mathbf{t}_{\Sigma} d\Sigma; \quad (3)$$

$$\frac{d}{dt} \int_V \rho \mathbf{x} \times \mathbf{v} dV = \int_V \rho \mathbf{x} \times \mathbf{f}_m dV + \int_{\Sigma} \mathbf{x} \times \mathbf{t}_{\Sigma} d\Sigma, \quad (4)$$

где  $\rho$  — плотность покупателей в кластере;  $V$  — объем кластера;  $\mathbf{f}_m$  — вектор плотности внешних массовых сил, действующих на кластер;  $\mathbf{t}_{\Sigma}$  — вектор поверхностных сил. Радиус-вектор точек  $\mathbf{x} = x^i\mathbf{e}_i$  в пространстве  $E_n$  подчиняется закону движения (за-

кону накопления покупок)  $\mathbf{x} = \mathbf{x}(X^i, t)$ ;  $\mathbf{v} = \frac{\partial \mathbf{x}(X^i, t)}{\partial t}$  — вектор скорости (частоты покупок). Все основные элементы пространства  $E_n$ , в том числе векторное произведение  $\mathbf{a}_1 \times \mathbf{a}_2 = (\mathbf{a}_2 \otimes \mathbf{a}_1) \cdot \overset{n}{\mathbf{e}}$ , введены в работе [1], здесь  $\overset{n}{\mathbf{e}} = e_{i_1 \dots i_n} \mathbf{e}^{i_1} \otimes \dots \otimes \mathbf{e}^{i_n}$  — тензор Леви—Чивиты,  $e_{i_1 \dots i_n}$  — символы Леви—Чивиты,  $(\cdot)$  — оператор двойного скалярного умножения [3].

### Модель внешних экономических воздействий на кластеры

Плотность массовой силы задается следующей формулой:

$$\mathbf{f}_m = \frac{d\bar{\mathbf{f}}}{dM} \quad (5)$$

как отношение силы  $d\bar{\mathbf{f}}$ , изменяющей частоту покупок локальной группы покупателей  $dV$ , к числу покупателей  $dM$  в группе. Эта сила бывает только внешней, т. е. вызванной "внешними" причинами по отношению к кластеру покупателей.

Введем плотность поверхностной силы:

$$\mathbf{t}_{\Sigma} = \frac{d\bar{\mathbf{f}}}{d\Sigma} \quad (6)$$

как отношение силы  $d\bar{\mathbf{f}}$ , изменяющей частоту покупок локальной группы покупателей  $d\Sigma$ , к числу контактных покупателей  $d\Sigma$  в этой группе. Элементарную площадку  $d\Sigma$  следует трактовать как совокупность покупателей в локальной группе  $dV$ , которая контактирует с соседней локальной группой, т. е.  $d\Sigma \in dV$ .

Плотность внешней силы  $\mathbf{f}_m$  определяет воздействие продавца на покупателя, т. е. с учетом того, что влияние выражается через изменение частоты покупок, а любая деятельность продавца, направлен-



ная на изменение покупок, называется маркетингом, плотность внешней силы отражает влияние маркетинга на частоту покупок. Задание выражения для  $\mathbf{f}_m$  или  $\mathbf{t}_\Sigma$  не лимитируется никакими законами сохранения (по крайней мере, известными в настоящее время), а определяется только "смысловыми" ограничениями. Таким образом, вид функций плотности внешних сил  $\mathbf{f}_m$  и поверхностных сил  $\mathbf{t}_\Sigma$  необходимо задавать с помощью построения дополнительных моделей.

Будем говорить, что рассматривается *потокковая модель* покупок товаров покупателями кластера, если плотность внешней массовой силы  $\mathbf{f}_m$ , вызывающей изменение частоты покупок, определяется суммарной скоростью изменения частоты приобретения товара покупателем "в запас" (или "на склад") и для непосредственного потребления (или "производства"):

$$\mathbf{f}_m = \frac{d\mathbf{h}^s}{dt}, \quad (7)$$

где  $\mathbf{h}^s$  — вектор, называемый *вектором частот приобретения товара покупателем* на склад и для непосредственного потребления (производства). Выражения для  $\mathbf{h}^s$  будут заданы ниже.

В компонентах в неподвижном базисе формула (7) имеет следующий вид:

$$f_{mi} = \frac{dh_i^s}{dt}. \quad (8)$$

Плотность поверхностных сил в потокковой модели считается равной нулю, поскольку полагается, что кластер приводится в движение только естественной частотой покупок и частотой покупок, вызванной внешней силой:

$$\mathbf{t}_\Sigma = 0.$$

Принципиальным моментом в потокковой модели является факт того, что в ней полагаются заданными компоненты вектора плотности массовых сил  $f_{mi} = \frac{dh_i^s}{dt}$  в неподвижном базисе, это так называемые замороженные силы [4], которые не зависят от движения самого кластера. В то же время известно, что могут существовать "следающие силы" [4], для которых заданы компоненты  $f'_{mi}$  в подвижном базисе, причем

$$f'_{mi} = f_{mj} Q_i^j.$$

Для рассматриваемых здесь экономических объектов модель замороженных сил является более адекватной, поскольку в ней полагаются заданными компоненты  $f_{mi}$  в координатах  $x^i$ , соответствующих количеству покупаемых товаров, — именно по изменению этих величин обычно и устанавливается

наличие кризисных явлений и эффективность маркетинговых мероприятий.

Если подставить выражение (7) в формулу (5), то с учетом правила дифференцирования интеграла по подвижному объему получим:

$$\bar{\mathbf{f}} = \int_V \rho \mathbf{f}_m dV = \int_V \rho \frac{d\mathbf{h}^s}{dt} dV = \frac{d}{dt} \int_V \rho \mathbf{h}^s dV = \frac{d\bar{\mathbf{h}}^s}{dt}; \quad (9)$$

$$\bar{\mathbf{h}}^s = \int_V \rho \mathbf{h}^s dV. \quad (10)$$

Если теперь подставить это выражение в уравнение движения центра масс (3), то придем к следующему дифференциальному уравнению:

$$M \frac{d\mathbf{v}_0}{dt} = \frac{d\bar{\mathbf{h}}^s}{dt},$$

которое легко интерпретируется:

$$\mathbf{v}_0 = \mathbf{v}_0^0 + \frac{1}{M} (\bar{\mathbf{h}}^s - \bar{\mathbf{h}}_0^s); \quad \bar{\mathbf{h}}_0^s = \bar{\mathbf{h}}^s(0). \quad (11)$$

Таким образом, в потокковой модели уравнение скорости движения центра масс кластера всегда может быть проинтегрировано и приведено к виду (11).

С помощью формулы (11) приходим к уравнению движения центра масс кластера:

$$\frac{d\mathbf{x}_0}{dt} = \mathbf{v}_0^0 + \frac{1}{M} (\bar{\mathbf{h}}^s - \bar{\mathbf{h}}_0^s). \quad (12)$$

Если подставить выражение (12) в формулу (4), то с учетом правила дифференцирования интеграла по подвижному объему получим следующее выражение для компонент тензора суммарных моментов внешних сил  $\tilde{\mu}^{ij}$ :

$$\begin{aligned} \tilde{\mu}^{ij} &= \int_V \rho (\tilde{x}^j f_m^i - f_m^j \tilde{x}^i) dV = \\ &= Q^{lk} \int_V \rho (\tilde{x}^j \delta_k^i - \delta_k^j \tilde{x}^i) \frac{dh_l^s}{dt} dV = \\ &= Q^{lk} \frac{d}{dt} \int_V \rho (\tilde{x}^j \delta_k^i - \delta_k^j \tilde{x}^i) h_l^s dV = Q^{lk} \frac{d}{dt} H_{kl}^{ij}; \\ H_{kl}^{ij} &= \int_V \rho (\tilde{x}^i \delta_k^j - \delta_k^j \tilde{x}^i) h_l^s dV, \end{aligned}$$

где  $\delta_k^i$  — символ Кронекера.

Из этих формул видно, что функции  $\tilde{\mu}^{ij}$  не интегрируемы по времени, при численном решении уравнений (4) необходимы алгоритмы вычисления этих функций непосредственно по формулам (12).

### Модель стабильного процесса покупок

Будем говорить, что покупатели осуществляют *стабильный процесс покупок товаров*, если компоненты  $h_i^s$  вектора  $\mathbf{h}^s$  в неподвижном базисе являются постоянными, не изменяющимися во времени

(но они могут различаться для разных покупателей, т. е. зависеть от координат):

$$\mathbf{h}^s = \mathbf{h}_0^s = \text{const}, \quad \bar{\mathbf{h}}^s = \int_V \rho \mathbf{h}_0^s dV = \bar{\mathbf{h}}_0^s = \text{const}. \quad (13)$$

Для стабильного процесса покупок внешние силы  $\mathbf{f}_m$  отсутствуют:

$$\mathbf{f}_m = 0,$$

и покупки осуществляются покупателями "по инерции", т. е. под действием начальных данных, обусловленных "стабильными потребностями людей в потреблении и запасах".

Из уравнения движения жесткого кластера (11) следует, что

$$\mathbf{v}_0 = \mathbf{v}_0^0 = \frac{1}{M} \bar{\mathbf{h}}_0^s = \text{const} \quad (14)$$

— средний вектор частот покупок товара в кластере при стабильном процессе потребления и накопления — является постоянным (по величине и по направлению), он определяется только начальной частотой покупок  $\mathbf{v}_0^0$ , которая совпадает с  $\frac{1}{M} \bar{\mathbf{h}}_0^s$ .

Этот вывод вполне согласуется с обычной "экономической логикой покупок": в стабильном состоянии (в отсутствие внешних воздействий) каждый раз покупатели покупают одинаковые товары в количестве, необходимом им для существования.

Подставляя в уравнение (12) формулу (14), находим изменение количества покупаемого товара "типовым" покупателем кластера в стабильном процессе:

$$\mathbf{x}_0 = \mathbf{x}_0^0 + \frac{t}{M} \bar{\mathbf{h}}_0^s, \quad (15)$$

оно представляет собой равномерное прямолинейное движение в пространстве товаров  $E_n$ . Таким образом, мы получили, что в отсутствие внешних сил типовой покупатель кластера (центр масс) движется по пространству товаров равномерно и прямолинейно, покупая постоянно одно и то же количество необходимых ему товаров.

Для расширения возможностей рассматриваемых моделей необходимо принять во внимание и более сложные экономические явления, например кризисные, в которых возникает эффект нарушения стабильности процесса покупок покупателями кластера. В такие моменты общие изменения экономической среды влияют на покупательское поведение людей. В этом случае внешние силы отличны от нуля и, вообще говоря, распределены неравномерно по кластеру. Будем говорить, что задана модель *нестабильного процесса покупок*, если  $\mathbf{f}_m$  определяется изменением только экономического положе-

ния рынка, представленного колебанием финансового индикатора рынка  $\hat{e}(t)$ :

$$\mathbf{f}_m = \frac{d\mathbf{h}^s}{dt}; \quad (16)$$

$$\mathbf{h}^s = \mathbf{h}_0^s \cdot (\mathbf{E} + \mathbf{B}\Delta\hat{e}); \quad \Delta\hat{e} = \hat{e}(t) - \hat{e}(0),$$

где  $\mathbf{B}$  — тензор коэффициентов сопротивления кластера внешним воздействиям, в базисе  $\bar{\mathbf{e}}_i$  он имеет диагональный вид:

$$\mathbf{B} = \sum_{\gamma=1}^n B^\gamma \bar{\mathbf{e}}_\gamma \otimes \bar{\mathbf{e}}_\gamma.$$

Коэффициенты сопротивления  $B^\gamma$  являются характеристиками кластера и должны быть заданы (например, установлены по результатам предварительной обработки экспериментальных данных по прошлым реакциям покупателей на изменение финансового индикатора рынка).

Функция  $\hat{e}$  вводится в модели аксиоматически — она характеризует *общее финансовое благосостояние покупателей* на рынке, предполагается, что эта функция известна. В отличие от функции финансового запаса  $e$ , который был причиной изменения объема деформируемого кластера, введенная здесь величина является общей характеристикой рынка, а не конкретной группы покупателей.

Если функция финансового благосостояния покупателей остается стабильной  $\hat{e} = \text{const}$ , то  $\Delta\hat{e} = 0$ , вектор  $\mathbf{h}^s$  остается постоянным и мы приходим снова к модели стабильного процесса покупок.

Такой подход к моделированию внешних воздействий позволяет учитывать одновременное влияние нескольких видов внешних сил. При наличии нескольких видов воздействий суммарная внешняя массовая сила определяется выражением

$$\mathbf{f}_m = \frac{d\mathbf{h}^{s1}}{dt} + \frac{d\mathbf{h}^{s2}}{dt} + \dots + \frac{d\mathbf{h}^{sp}}{dt},$$

где  $p$  — число одновременно действующих внешних воздействий. При расчете изменения потоков покупок для всех внешних воздействий, кроме  $\mathbf{h}^{s1}$ , необходимо исключить естественный (стабильный) уровень покупок  $\mathbf{h}_0^s \cdot \mathbf{E}$ , который учтен в первом слагаемом формулы.

Вектор частот приобретения товара покупателем  $\mathbf{h}_0^s$ , вообще говоря, различен для разных покупателей, т. е.  $\mathbf{h}_0^s$  зависит от координат  $X^i$  (эта зависимость приводит к тому, что суммарные моменты внешних сил  $\tilde{\mu}^{ij}$  отличны от нуля, т. е. приводят к вращению). Для проведения расчетов можно заранее задавать вид зависимости  $\mathbf{h}_0^s(X^i)$  исходя из информации о распределении покупателей в кластере по отношению к типовому покупателю — центру масс кластера  $\mathbf{x}_0$ :

$$\mathbf{h}_0^s(X^i) = \mathbf{h}^s \varphi(X^i),$$

где  $\mathbf{h}^s$  — вектор максимальных для данного кластера значений частот приобретения товаров, а  $\varphi(X^i)$  — функциональная зависимость, характеризующая распределение покупателей кластера относительно типового покупателя.

### Результаты численного моделирования

Разработанные положения модели были применены для анализа данных по продажам он-лайн магазина автомобилей в течение 2,5 лет. В качестве исходных данных при этом использованы данные по индивидуальным покупкам группы покупателей. Для изучения группового поведения покупатели были разделены на группы исходя из данных по их покупательскому поведению. Перед применением алгоритма обобщающей кластеризации данные, обладающие первоначально большой степенью дискретности, были континуализованы по дополнительной информации о покупателях.

В качестве осей координат  $x^i$  пространства  $E_n$  были введены суммарные значения купленных или планируемых к покупке автомобилей пяти ценовых классов, поэтому размерность  $n$  пространства товаров  $E_n$  была равна 5. Данные анализировались через 3 месяца в моменты времени  $t_m$ ,  $m = 1, \dots, 11$ . Моменты времени  $t_4$  и  $t_6$  были идентифицированы как первая и вторая волна финансового кризиса, который выражался в изменении общего экономического благосостояния покупателей на рынке  $\hat{e}$  покупателей согласно модели (16).

Для аппроксимации данных применяется модель

$$\sum_{\alpha=1}^n \left( \frac{X^\alpha}{R^\alpha} \right)^2 = 1 \text{ в Лагран-}$$

жевых координатах, где  $R^\alpha$  — полуоси эллипсоида, которые находятся путем решения задачи минимизации объема кластера с ограничениями на вхождение в эллипсоид не менее 95 % точек, относящихся к кластеру. Векторы базиса  $\bar{\mathbf{e}}_j^i(0) = Q_j^i(0)\mathbf{e}_i$  в начальный момент определяются путем нахождения элементов матрицы инерции  $I^{ij} = \int_V \rho \tilde{x}^i \tilde{x}^j dV$  по экспериментальным точкам и нахождения ее собственных векторов. Для вычисления многомерных интегралов применялись четыре метода, описанные в [1]:

- метод  $n$ -мерных шаров;
- метод равного веса точек;
- метод нормального распределения;
- метод локальных групп.

Результат разделения покупателей на кластеры и аппроксимация полученных кластеров эллипсоидами показаны на рис. 1, где эллипсы представляют сечения эллипсоидов плоскостью  $(x^3, x^4)$ , тем-

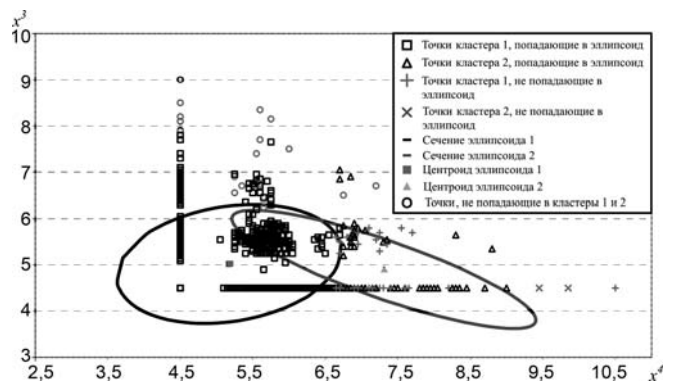


Рис. 1. Деление точек на кластеры и аппроксимирующие эллипсоиды для 8-го периода, рассчитанные методом  $n$ -мерных шаров

ные фигуры — точки, попадающие в эллипсоид, а светлые — не попадающие.

Исследование движения центров масс кластеров показало, что оно носит линейный характер по всем осям товаров  $x^i$  в пространстве  $E_5$  (рис. 2) при отсутствии внешних сил. Наличие внешних воздействий (кризисных явлений) приводит к изменению углов наклона линейной траектории движения кластеров. Таким образом, экспериментальные данные показали, что в данном случае в стабильном процессе в отсутствие внешних сил типовой покупатель движется по пространству товаров равномерно и прямолинейно, как это и прогнозирует формула (15). Изменение угла наклона линейной

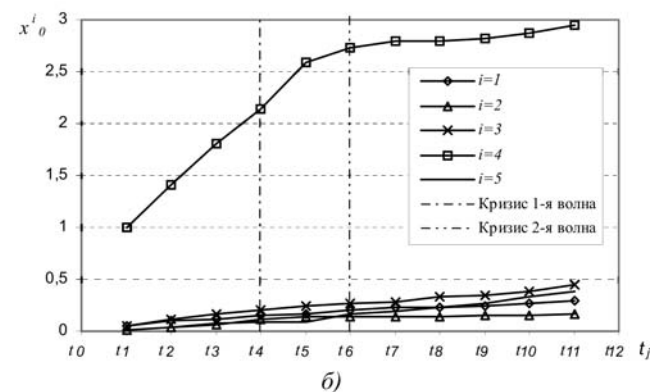
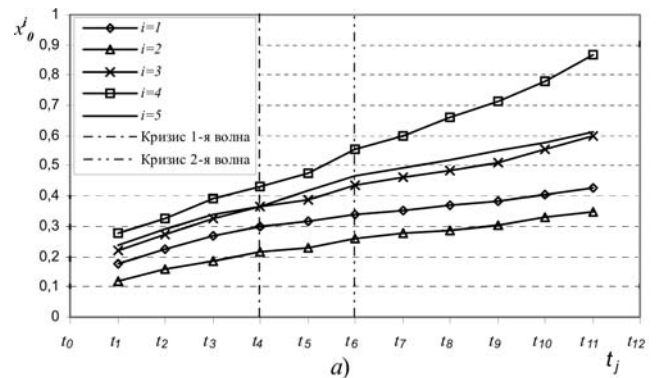


Рис. 2. Динамика центра масс кластеров 1 (а) и 2 (б)

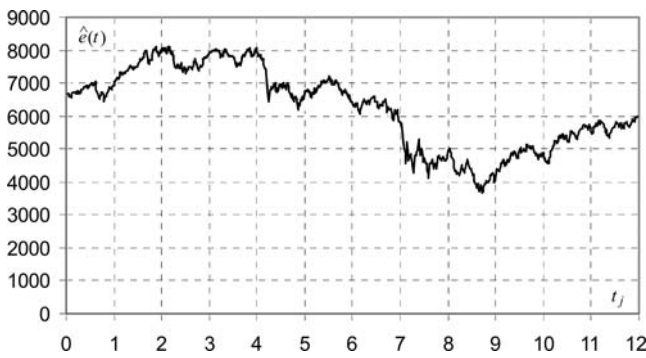


Рис. 3. Динамика биржевого индекса  $\hat{e}(t)$

траектории вследствие действия внешней силы также объясняется вышеизложенной моделью (11).

В соотношения (16) входит функция изменения общего благосостояния покупателей на рынке  $\hat{e}(t)$ . Для моделирования кризисных явлений с помощью функции  $\hat{e}(t)$  необходимо использовать некоторый репрезентативный показатель общего благосостояния рынка. В данной работе в качестве такого показателя был использован финансовый биржевой индекс, его динамика показана на рис. 3.

В формулу для вычисления внешней силы (16) входит также тензор коэффициентов сопротивления

кластера  $\mathbf{B} = \sum_{\gamma=1}^n B^\gamma \bar{\mathbf{e}}_\gamma \otimes \bar{\mathbf{e}}_\gamma$ . Этот тензор предпо-

лагается ортотропным, и в неподвижном базисе он является диагональным [3]. Его компоненты  $B^\gamma$  — коэффициенты сопротивления кластера — могут быть установлены путем соотнесения части экспериментальных данных по динамике движения центра масс кластера для фиксированного момента времени  $t_c$  с динамикой биржевого индекса за тот же период. Применяя формулу (11) для скорости движения центра масс кластера, получаем

$$\mathbf{v}_0 = \mathbf{v}_0^0 + \frac{\bar{\Phi}}{M} \mathbf{h}_0^s \cdot \mathbf{B} \Delta \hat{e};$$

$$v_0^j = v_0^{j0} + \frac{\bar{\Phi}}{M} h_{0q}^s B^{iq} \Delta \hat{e},$$

где  $\bar{\Phi} = \int_V^0 \rho \varphi dV$  — интегральный показатель рас-

пределения покупателей в кластере относительно типового покупателя.

Отсюда для момента времени  $t_c$  находим

$$B^\gamma = B^{\gamma\gamma} = M \frac{(v_0^\gamma(t_c) - v_0^{\gamma0})}{\bar{\Phi} h_{0\gamma}^s \Delta \hat{e}(t_c)}.$$

Значения рассчитанных по этой формуле элементов матрицы  $B^\gamma$  приведены в таблице. Значение

изменения благосостояния  $\Delta \hat{e}(t_c)$  взято по изменению биржевого индекса для первой и второй волны кризиса. Моменты времени  $t_c$  выбраны совпадающими с проведением очередных волн опроса. В качестве  $v_0^0$  была использована средняя скорость для контрольных точек  $t_3$  и  $t_4$  для первой волны кризиса и для точки  $t_6$  для второй волны кризиса.

Результаты расчетов коэффициентов влияния  $B^\gamma$  показывают, что кластер 1 склонен сильнее реагировать на уменьшение благосостояния (коэффициенты влияния на кластер 1 больше для обеих волн кризиса по товарам  $i = 1, 2, 3$ ). Также можно сказать, что товары  $x^4, x^5$  значительно меньше подвержены влиянию уменьшения благосостояния  $\Delta \hat{e} < 0$ . Это может быть объяснено тем, что их покупают люди с большими финансовыми запасами, поэтому эти покупатели могут себе позволить покупать дорогие товары и в кризис. Также следует заметить, что коэффициенты сопротивления  $B^\gamma$  больше для второй волны кризиса практически для всех товаров. Это может быть объяснено тем, что одна за другой две волны кризиса заставили людей понять серьезность ситуации и, возможно, больше экономить.

В первую очередь, модели экономических процессов разрабатываются для решения задач прогнозирования. Поэтому для экспериментальных данных была рассмотрена тестовая задача. По известным данным о продажах в течение 11 контрольных точек для двух групп покупателей необходимо, используя для идентификации (вычисления необходимых параметров) методов только часть данных по  $m < 11$  периодам, осуществить прогноз на оставшиеся  $11 - m$  периодов.

Для исследования были использованы четыре метода прогнозирования:

- метод построения линейного тренда по методу наименьших квадратов;
- метод прогнозирования с помощью кубического сплайна;
- метод прогнозирования на основе тригонометрического ряда Фурье с применением методов спектрального анализа;

Значения коэффициентов сопротивления  $B^\gamma$  для двух кластеров, рассчитанных для двух волн кризиса

№ кластера	Кластер 1		Кластер 2	
	1-я волна	2-я волна	1-я волна	2-я волна
$\gamma = 1$	0,000465	0,000956	0,000397	0,00051
$\gamma = 2$	0,00049	0,000404	0,000308	0,000998
$\gamma = 3$	0,000438	0,000428	0,00008	0,000667
$\gamma = 4$	0,000164	0,000471	-0,00017	0,000596
$\gamma = 5$	-0,00032	0,000505	0,000849	0,000537

- метод прогнозирования с помощью разработанной модели жесткого кластера с учетом воздействия внешних сил.

Прогнозирование с помощью линейного тренда, вычисленного по методу наименьших квадратов (МНК), подразумевает нахождение на основе анализа экспериментальных данных коэффициентов в уравнении прямой, минимизирующей отклонение значений реальных данных от тренда.

Кубический сплайн, построенный для целей прогнозирования в данной задаче, представлял собой на каждом анализируемом отрезке полином третьей степени, коэффициенты которого определялись из анализируемых данных методом прогонки для трехдиагональной матрицы. Использование метода тригонометрического ряда Фурье для целей прогнозирования данных подразумевает вычисление коэффициентов ряда на основе многомерной регрессии, построенной по имеющимся данным. Число гармоник тригонометрического ряда определялось в зависимости от количества данных, заданных в варианте тестовой задачи. Использование метода спектрального анализа позволило определить периодичность колебаний значений продаж.

Рис. 4 демонстрирует *среднесрочный прогноз* ( $m = 6$ , шесть контрольных точек использованы для идентификации методов, для пяти точек осуществлено

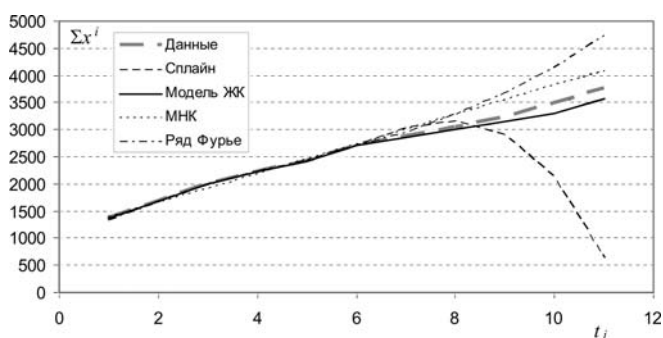


Рис. 4. Прогноз на пять временных периодов общей суммы накопленных покупок по данным за шесть временных периодов различными методами для кластера 1

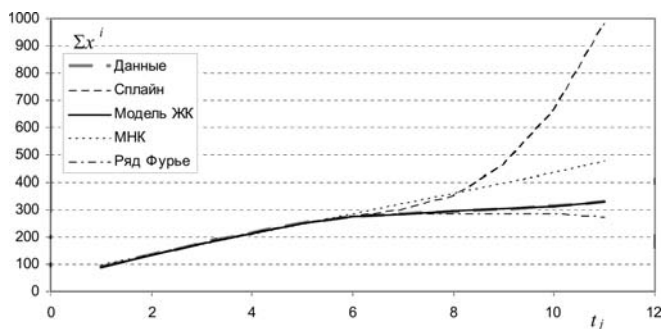


Рис. 5. Прогноз на пять временных периодов общей суммы накопленных покупок по данным за шесть временных периодов различными методами для кластера 2

прогнозирование) общей суммы накопленных покупок, рассчитанный с помощью всех используемых методов для кластера 1. Заметно расхождение точности методов прогнозирования: метод сплайна дает большую погрешность прогноза — более 80 %, ряд Фурье также дает большую погрешность — более 25 %, в то время как для метода МНК и модели жесткого кластера погрешность не превышает 10 %.

На рис. 5 показан среднесрочный прогноз общей суммы накопленных покупок для кластера 2. Из-за нелинейного характера покупок кластера метод МНК здесь дает большую погрешность, чем для кластера 1 — более 40 %, метод сплайна не может быть рекомендован для использования в данном случае, поскольку ошибка прогноза превышает 150 %, ряд Фурье заметно меньшую погрешность — менее 18 %. Наиболее точным методом для данного случая является модель жесткого кластера — погрешность не превышает 2 %.

Учет воздействия внешних сил изменяет линейный характер движения центра масс кластера в модели, что увеличивает точность прогноза и помогает более детально и фундаментально понять причины покупательского поведения и его изменения.

## Выводы

Разработана модель воздействия внешних сил на покупательскую динамику экономических кластеров, основанная на применении методов механики многомерных сплошных сред для решения задач анализа и прогнозирования динамики больших массивов плохо структурированных экономических данных. Предложены численные алгоритмы решения систем нелинейных дифференциальных уравнений, возникающих в этой модели. Представлен пример анализа экспериментальных данных рынка продаж, показано, что разработанная модель воздействия внешних сил на динамику кластеров представляет большие возможности для прогнозирования поведения кластеров покупателей при наличии изменений внешней экономической среды. Также модель может быть использована для прогнозирования влияния, оказываемого несколькими видами внешних экономических сил. Учет воздействия сил увеличивает точность модели динамики экономических кластеров.

## Список литературы

1. Димитриенко Ю. И., Димитриенко О. Ю. Кластерно-континуальное моделирование в экономике на основе методов механики многомерных сплошных сред // Информационные технологии. 2010. № 8. С. 54–62.
2. Димитриенко Ю. И., Димитриенко О. Ю. Модель деформируемых кластеров для анализа динамических данных в экономике // Информационные технологии. 2010. № 9. С. 43–50.
3. Димитриенко Ю. И. Тензорное исчисление. М.: Высшая школа. 2001. 575 с.
4. Димитриенко Ю. И. Нелинейная механика сплошной среды. М.: Физматлит. 2009. 624 с.

**ЖУРНАЛ В ЖУРНАЛЕ**



**НЕЙРОСЕТЕВЫЕ  
ТЕХНОЛОГИИ**

**№ 1**

**ЯНВАРЬ**

**2012**

**Главный редактор:**

ГАЛУШКИН А. И.

**Редакционная коллегия:**

АВЕДЬЯН Э. Д.  
БАЗИАН Б. Х.  
БЕНЕВОЛЕНСКИЙ С. Б.  
БОРИСОВ В. В.  
ГОРБАЧЕНКО В. И.  
ЖДАНОВ А. А.  
ЗЕФИРОВ Н. С.  
ЗОЗУЛЯ Ю. И.  
КРИЖИЖАНОВСКИЙ Б. В.  
КУДРЯВЦЕВ В. Б.  
КУЛИК С. Д.  
КУРАВСКИЙ Л. С.  
РЕДЬКО В. Г.  
РУДИНСКИЙ А. В.  
СИМОРОВ С. Н.  
ФЕДУЛОВ А. С.  
ЧЕРВЯКОВ Н. И.

**Иностранные  
члены редколлегии:**

БОЯНОВ К.  
ВЕЛИЧКОВСКИЙ Б. М.  
ГРАБАРЧУК В.  
РУТКОВСКИЙ Л.

**Редакция:**

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

**Мышев А. В.**

Информационная модель нейросети в технологиях вычислительного интеллекта и формах реализации компьютеринга . . . 63

**Кулик С. Д., Никонец Д. А.**

Автоматизация классификационно-диагностических почерковедческих исследований с помощью нейронных сетей . . . 70

**Нгуен Виет Хунг**

Сжатие изображения с использованием нейронной сети в преобразованном пространстве . . . . . 76

**А. В. Мышев**, канд. физ.-мат. наук,  
Национальный исследовательский ядерный  
университет МИФИ — Обнинский институт  
атомной энергетики  
e-mail: mishev@iate.obninsk.ru

## **Информационная модель нейросети в технологиях вычислительного интеллекта и формах реализации компьютинга**

*Рассматривается информационная модель нейросети для реализации новых форм компьютеринга и разработки технологий вычислительного интеллекта на основе виртуальной модели активной памяти вычислительной системы. Формализм модели описывает такую нейросеть в виде глобальной и локальной сети динамически взаимодействующих информационных элементов с набором дуплексных входных—выходных информационных связей. Информационная динамика взаимодействия между элементами таких сетей определяется как процессы с локальным взаимодействием.*

**Ключевые слова:** информационная модель, нейросетевой процессор образа, модель активной памяти, вычислительный интеллект, технологии виртуализации, информационная динамика

### **Введение**

Вычислительный интеллект в информационных технологиях AI-компьютеров и интеллектуальных вычислительных систем является не столько вычислительным и транспортно-информационным устройством или памятью, сколько когнитивным инструментом интеллектуальных технологий среды вычислений и осуществления новых форм компьютеринга. Феномен когнитивных функций технологий вычислительного интеллекта, который является одним из важнейших факторов и проявлений интеллектуальной эволюции таких систем, в логическом базисе образного представления информации может быть определен следующим образом: синтез символической цепочки, образа и алгоритма. Этот феномен позволяет, с одной стороны, в концентрированной, более содержательной и образной форме отражать, анализировать и наглядно интерпретировать информацию, а с другой стороны, — синтезировать новые знания. Когнитивные функции компьютерных и информационных технологий таких систем являются основой некой интеллектуальной среды, погружаясь в которую осуществляется переход от виртуальных образов вербально-логического представления информации к ее образному

представлению средствами вычислительного интеллекта в информационной среде когнитивной реальности в целях получения новых знаний.

Нейросетевые процессоры образов как базовые элементы вычислительного интеллекта, с одной стороны, выполняют функции синтеза и образного представления информации в системах виртуальной и когнитивной реальности. А с другой стороны, образное представление и анализ информации с помощью таких процессоров позволяет не только реализовать технологии визуализации и в иллюстративной образной форме выразить суть и смысл сложного абстрактного виртуального моделируемого объекта, но и увидеть в его образной форме новые скрытые, неожиданные свойства и неординарные решения, а в информационной динамике — новое соображение, мысль, идею. Разработка и реализация процессоров такого типа предполагает создание информационных моделей искусственной нейросети и формального нейрона другого типа или класса, нежели те, которые предлагаются в рамках подходов и моделей традиционной теории нейронных сетей [1, 2].

Здесь следует отметить важный и принципиальный вопрос, который в традиционной теории нейронных сетей не отражается (в доступных источниках автор не встречал), а именно: в чем и как проявляются закономерности взаимосвязи и взаимодействия сигнала и символа (образа)? С одной стороны, как они проявляются в информационной динамике обозначенных сущностей в нейросетевых технологиях как средствах искусственного интеллекта и атрибутов различной природы и среды в динамическом, информационном, математическом и метрологическом аспектах? А с другой стороны, в чем же суть, смысл и содержание этих категорий в восприятии человеком информации, генерируемой посредством обозначенных нейросетевых технологий? Этот момент тесно связан с тем, что в информационной среде вычислительных систем, включая системы компьютерного и человеческого восприятия, развитую сетевую и системную периферию, память и процессорные элементы, информация на логическом и аналитическом уровнях имеет символическое представление и описание. В свою очередь, практический аспект математического описания и реализации технологий вычислительного интеллекта в виртуальной информационной среде вычислительных систем базируется на символическом представлении информационных объектов в виде логических структур, образующими элементами которых являются ограниченные или бесконечные цепочки взаимодействующих символов. Цепочки символов являются информационными атрибутами, посредством которых абстрактные математические объекты идентифицируются и инициализируются в виртуаль-

ной информационной среде вычислительного интеллекта [3]. Информационная динамика символов в цепочках и динамика цепочек в операциях и алгоритмах технологий вычислительного интеллекта определяет метамеханизм функционирования обозначенного нейросетевого процессора образов. Здесь мы сталкиваемся с информационной динамикой взаимодействующих цепочек символов в виртуальной среде адресного пространства ячеек реальной физической памяти на базовом логическом уровне.

В рамках такого подхода построения информационных моделей процессора образов нейросеть логически и структурно определяется в виде локальной виртуальной информационной динамической сети. А нейрон, как элемент такой сети, определяется в виде локальной виртуальной вычислительной системы — виртуального процессорного элемента и локализованного сегмента виртуальной активной памяти, базовыми элементами которой являются виртуальные ячейки различных типов. Виртуальные ячейки активной памяти локализованы на сегментах адресного пространства физической памяти различных уровней.

В сетевых алгоритмах и процедурах вычислительного интеллекта такая модель определяет как информационные и адресные системы координат для образующих и конфигураций, так и сетевой шаблон нейросетевого процессора образов. Модель формального нейрона в логической структуре и функциональной схеме обозначенной искусственной нейросети определяет и описывает его как функциональную единицу глобальной нейросети.

Нейрон можно определить двояким образом: 1) как модель активной виртуальной памяти, в которой динамически изменяются структура ее элементов, топология связей, логическая и информационная организация; 2) как виртуальный процессорный элемент. Процессы информационного взаимодействия нейронов в сети определяются как параллельные потоковые процессы обработки, хранения,

приема и генерации информации на "взаимодействующих" сегментах процессорных элементов и адресного пространства активной памяти.

Секреты информационной основы нейросетевых технологий вычислительного интеллекта, разрабатываемых на основе обозначенных моделей нейросети, во многом сокрыты в тайне гиппокампа, который привлекает внимание как возможная модель гиппокампальной памяти с доминантной архитектурой. Такая модель памяти позволяет с определенной степенью приближения выявить и определить связь внимания с восприятием, памятью и другими когнитивными функциями человека и когнитивными возможностями виртуальных вычислительных систем. В этом случае информационные процессы в схемах и алгоритмах вычислительного интеллекта развиваются не под централизованными воздействиями, а за счет коллективного воздействия его компонентов в нейросети. Кооперация компонентов нейросети как прототипа схемы вычислительного интеллекта позволяет использовать ее резервные возможности и существенно повышать степень эмерджентности (системного эффекта).

### Геометрия и топология информационной модели нейросетевого процессора образов

Геометрия информационной модели нейросети процессора образов определяется в виде многомерного "потокowego" ориентированного графа с набором входных и выходных информационных связей. Вершины графа соответствуют нейронам сети, а ребра (биориентированные) обозначают логические и физические информационные шины для процессов хранения, "транспортировки" и управления в сети. Определены информационные потоки трех типов: управляющие, информационные, адресные. Организация связей между вершинами удовлетворяет условию, что каждая вершина любого слоя графа имеет связь со всеми вершинами только соседних слоев. Геометрия подобных сетей и топология связей в них определяет тип информационного взаимодействия между элементами сети как процессы с локальным информационным взаимодействием [3, 4]. Геометрическая иллюстрация двумерной схемы такой нейросети приведена на рис. 1.

Схема сети, с одной стороны, является графическим образом (ГО) искусственной нейросети и иллюстрирует геометрию ее логической и функциональной организации в информационном пространстве вычислительной системы, а с другой стороны — это сетевая "модель", которая имеет свой прообраз и строится с определенной целью. Информационной средой ГО нейросети в адресном пространстве виртуальной активной памяти будет множество виртуальных ячеек, на котором определена конечная топологическая струк-

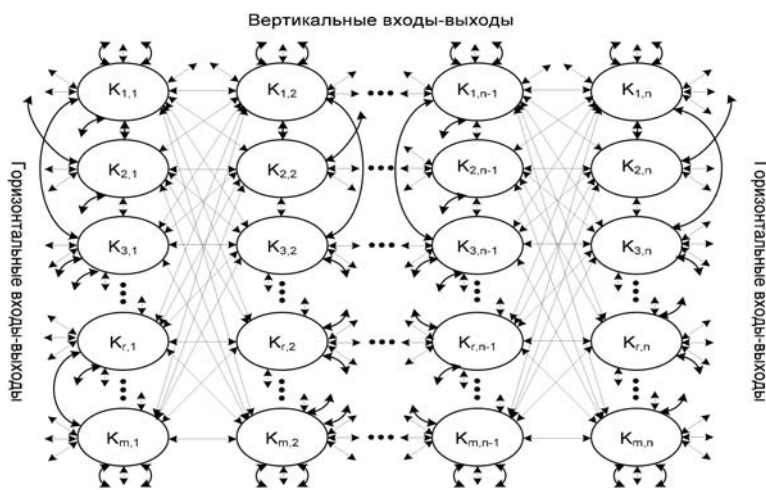


Рис. 1. Геометрическая иллюстрация нейросетевого процессора образов в виде двумерного "потокowego" ориентированного графа



тура, удовлетворяющая условиям частичной упорядоченности.

Все связи элементов в сети являются двунаправленными: входными—выходными.

Геометрия и топология связей между элементами нейросети иллюстрирует характер информационно-пространственного взаимодействия элемента  $K_{ij}$  с его окружением в сети, которое определено в пределах границ соседних столбцов (слоев, которые в многомерном случае определяются как подграфы). Любой элемент  $j$ -го столбца взаимодействует только с элементами  $(j + 1)$ -го или  $(j - 1)$ -го столбцов. Для механизма информационного взаимодействия нейрона с его окружением характерны диффузия информации по вертикали и локализация процессов взаимодействия по горизонтали в ограничениях справа и слева.

На рис. 2, а показана геометрическая иллюстрация схемы информационных связей элемента сети

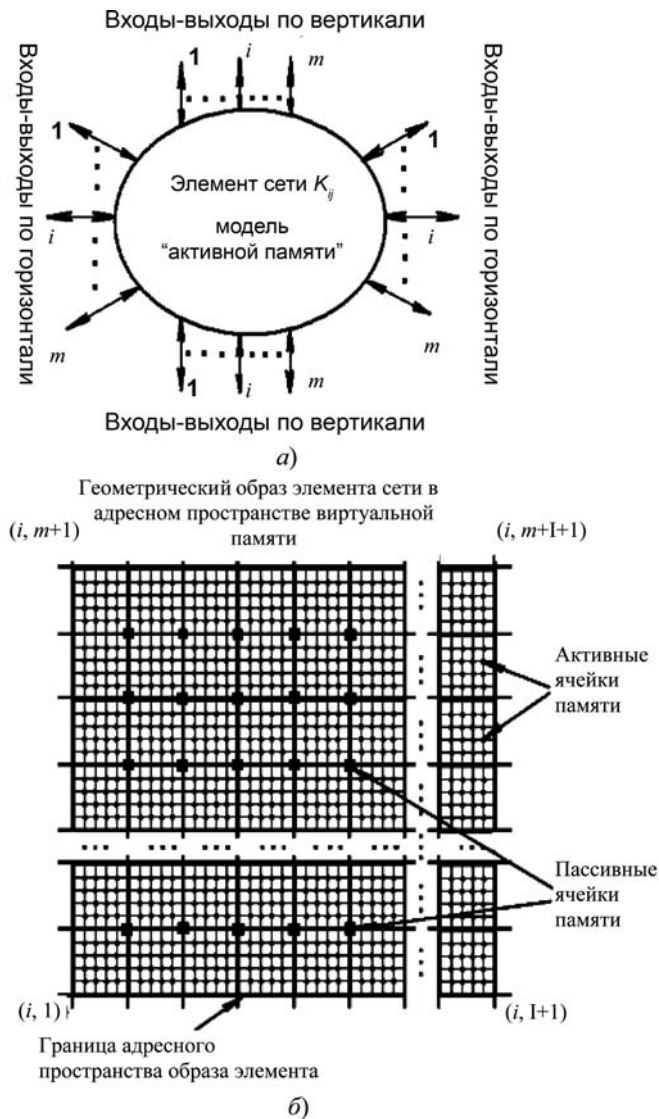


Рис. 2. Геометрическая иллюстрация схемы: а — информационных связей элемента сети с его окружением; б — образа элемента сети, как сегмента активной памяти

с его окружением, которая включает связи по вертикали и по горизонтали. Связи определяют тип информационного взаимодействия [3, 4, 5]: по вертикали — это информационная диффузия, а по горизонтали — локальная динамика со смещением.

Функциональные, динамические и топологические свойства информационного взаимодействия ее элементов описываются на двух уровнях: 1) детализации топологии и метрики нейросети; 2) управления и контроля информационной динамикой на элементах нейросети. Область памяти вычислительного интеллекта (домен—прототип нейрона) логически и функционально структурирована и определяется как виртуальная локальная динамическая информационная система или сеть, в которой взаимодействие или функционирование ее элементов описывается не на уровне передачи—приема сигналов, а на информационном уровне: обмен, хранение и генерация информации в каналах хранения и передачи.

Нейрон в такой динамической информационной среде определяется как локальная виртуальная вычислительная система или виртуальный процессорный элемент. С математической точки зрения смысл введенной таким образом сущности заключается в следующем. Во-первых, информационным прототипом нейрона в адресном пространстве активной памяти будет сегмент, который задает область его информационного определения в пространстве виртуальных адресов — его домен — и область значений — его диапазон. Во-вторых, элементами памяти таких сегментов являются виртуальные ячейки разных типов [6], которые различаются по функциональному признаку, логической организации и информационному наполнению.

Ячейки, которые относятся к классу пассивных, являются функциональными элементами, т. е., кроме операций хранения и передачи информации, на них определены виртуальные операторы информационного взаимодействия, а именно: между ячейками различных типов; между ячейками и информационной средой.

Ячейки активного типа являются элементами виртуальной памяти в привычном смысле и определении, т. е. выполняют функции хранения и передачи информации, но логическая структура таких ячеек и организация информации в них динамически изменяется. Геометрическая иллюстрация схемы образа элемента сети как сегмента активной памяти приведена на рис. 2, б.

### Механизм информационного взаимодействия цепочек символов в виртуальных ячейках активной памяти

Модели алгоритмов и процедур, описывающих информационные процессы на взаимодействующих цепочках символов, в технологиях вычислительного интеллекта строят на основе механизма процессов

с локальным взаимодействием и метода виртуальной перспективы [3, 4].

Уровень информационного состояния нейронов определяется оценкой вероятностной меры, заданной на размытом подмножестве информационных признаков пассивных виртуальных ячеек соответствующих сегментов активной памяти. Нейрон определяется в виде подмножества соседних пассивных виртуальных ячеек памяти, которые логически организованы и функционально связаны. Вычислительная и информационная среда нейрона задается подмножеством пассивных и активных виртуальных ячеек в локальной области активной памяти. Если мера на пассивных ячейках сегмента отлична от нуля, то соответствующий элемент нейросети находится в активном состоянии, а в противном случае — в пассивном состоянии.

В рассматриваемых схемах построения нейросетевых технологий вычислительного интеллекта память разбивается на виртуальные ячейки следующих типов [6]. Первый тип — это пассивные ячейки, размерность которых фиксирована, и информация в них не изменяется. Они также являются элементами системы информационной привязки. Второй тип — это пассивно-активные ячейки, размерность которых фиксирована, и информация в них динамически изменяется. Третий тип — это активные ячейки, размерность которых и информация в них динамически изменяются. Четвертый тип — это активно-пассивные ячейки, размерность которых динамически изменяется, но смысловая информация не изменяется. Характер информационного взаимодействия между символьными цепочками в активных ячейках определяется как процессы с локальным взаимодействием и взаимодействие с виртуальной информационной средой [3, 7], а тип информационного взаимодействия между цепочками символов задается эволюционным оператором, определенным на информационном пространстве цепочек взаимодействующих символов. Закон информационного взаимодействия между цепочками символов в пассивных и активных ячейках задается в виде двух операторов: оператора взаимодействия и оператора проектирования.

Оператор взаимодействия  $f$  определяется между символьными цепочками  $x_i$  и  $x_{i+1}$ , хранящимися в активных ячейках, и для простой реализации имеет следующий вид:

$$x_{i+1} = f(x_i), \quad (1)$$

где  $x_i$  — это адрес активной виртуальной ячейки, в которой хранится первый операнд  $x_i$  оператора  $f$ , а  $x_{i+1}$  — это адрес активной виртуальной ячейки, где хранится второй операнд  $x_{i+1}$  оператора  $f$  и куда будет записан результат. Следует отметить, что уравнение (1) будет справедливо и для векторной формы.

Оператор проектирования  $\psi$  задает механизм информационного взаимодействия между сим-

вольными цепочками, хранящимися в активных и пассивных виртуальных ячейках (рис. 3). На рис. 3, б  $z_i$  — пассивная ячейка;  $\{m_k\}$  — подмножество активных ячеек;  $d_{ik}$  — информационное расстояние между символьными цепочками в ячейках  $z_i$  и  $m_k$ ;  $d_{\text{доп}}$  — информационная сфера влияния пассивной ячейки  $z_i$ . Для информационных объектов подмножества  $\{m_k\}$  активных ячеек, взаимодействующих с информационным объектом в пассивной ячейке  $z_i$ , оператор  $\psi$  задается в виде следующего обобщенного выражения:

$$\psi: \{m_k\} \rightarrow z_i, \quad k = 1, 2, \dots$$

А весовая функция  $\chi(z_i)$  ячейки  $z_i$  определяется на основе следующих выражений:

$$\chi(z_i) = n_i / N,$$

$$n_i = \sum_{k=1}^M m_k,$$

где  $n_i$  — мощность подмножества  $\{m_k\}$  из множества активных ячеек, попавших в  $d_{\text{доп}}$  — окрестность

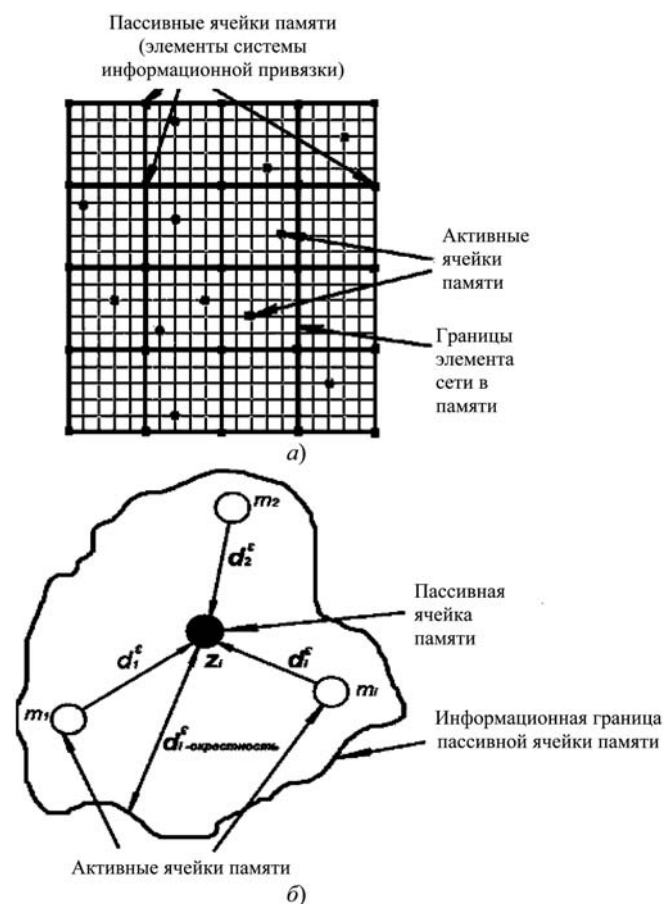


Рис. 3. Геометрия относительного расположения взаимодействующих цепочек символов, хранящихся в активных и пассивных виртуальных ячейках (а), и геометрическая иллюстрация действия оператора проектирования (б)

ячейки  $z_j$ , а  $N = \sum_i n_i$ . Механизм действия оператора  $\psi$  заключается в следующем. Так как символьные цепочки в пассивных виртуальных ячейках имеют фиксированную длину всегда меньшую, чем переменная длина цепочек в активных виртуальных ячейках, то вначале реализуется операция информационного "выравнивания" символьных цепочек разной длины. Логическая схема и механизм такой операции определяются как взаимодействие с информационной средой [3, 4, 5, 7]. Численная оценка информационного расстояния между символьными цепочками, хранящимися в активных и пассивных виртуальных ячейках, вычисляется по следующей формуле:

$$d(x_i, x_{i+1}) = \sum_{i=1}^M w_i |a_i - b_i|,$$

где  $M$  — длина символьных цепочек  $x_i$  и  $x_{i+1}$ ;  $w_i$  — вес  $i$ -й позиции символьной цепочки;  $a_i$  и  $b_i$  —  $i$ -е символы цепочек, соответствующих  $x_i$  и  $x_{i+1}$ . Если это расстояние не превышает допустимый критерий  $d_{\text{доп}}$ , который является заданной информационной окрестностью символьной цепочки в пассивной виртуальной ячейке, т. е. справедливо неравенство

$$d(x_i, x_{i+1}) \leq d_{\text{доп}}, \quad (2)$$

то считается, что результат действия оператора взаимодействия находится в пассивной виртуальной ячейке, для которого выполняется условие (2).

Геометрически информационная окрестность такой пассивной виртуальной ячейки может иметь произвольную морфометрию или топологическую мозаику (рис. 3, б). Действие механизмов информационного взаимодействия символьных цепочек, хранящихся в активных и пассивных виртуальных ячейках активной памяти, состоит в том, что информация из активных ячеек "проектируется" в пассивные. Это означает следующее: если значение символьной цепочки активной ячейки попадает в информационную окрестность символьной цепочки, хранящейся в пассивной ячейке, то ее информационный вес увеличивается. На рис. 3, а показана геометрия относительного расположения взаимодействующих цепочек символов, хранящихся в активных и пассивных виртуальных ячейках. В такой интерпретации информационное взаимодействие символьных цепочек в активных и пассивных виртуальных ячейках активной памяти можно рассматривать как филогенетическое построение. Причины такого взаимодействия между символьными цепочками обусловлены филогенетическими взаимоотношениями и эволюционными силами, а характер таких взаимодействий можно отнести к типу кладических отношений на множестве генетических данных [8].

## Виртуальные информационные изображения нейросетевого процессора образов

Среда взаимодействия объектов информационных процессов вычислительного интеллекта в адресном (ассоциативном) и информационном пространстве виртуальной активной памяти включает следующие основные атрибуты-посредники: виртуальная алгоритмическая переменная; операторы взаимодействия; логическая структура в адресном пространстве ограниченной памяти вычислительной системы и другие, которые являются компонентами среды вычислений, включающей все или почти все, что относится к получению результата. В ней имеются переменные и их фактические значения, а сами они разнесены не только позиционно, но и контекстно.

Все информационные процессы среды вычислений можно свести к некоторой первооснове: берется то, что считается идентификатором, и для него относительно среды строится то, что будет считаться его значением.

Вводим следующие предположения: во-первых, в виртуальном информационном пространстве среды вычислений "обитает" некоторая сущность, обозначим ее через  $\Omega$ , откуда она получает входную информацию. Во-вторых, структура образов, с которой имеет дело  $\Omega$ , априори неизвестна, и в течение фиксированного интервала времени  $\Omega$  будет пытаться вывести ее.

Для технологий вычислительного интеллекта будем рассматривать такую разновидность информационных моделей нейросетевых процессоров образов, которые представляют собой информационное отображение сети в адресное и информационное пространства виртуальной активной памяти вычислительной системы. Тогда "целевая" установка обозначенных моделей сетевого процессора состоит в том, чтобы алгоритмически, процедурно и технологически описать, как такой процессор синтезирует информационные образы и при каких условиях.

В логическом базисе образного представления информации процессор будет синтезировать информационные образы в условиях следующего характера: многофакторной неопределенности, ограничений вычислительной системы, алгоритмической замкнутости моделей алгоритмов и процедур среды вычислений, а также в условиях, обусловленных другими причинами деформаций информационных образов. Поэтому точный результат (промежуточный и конечный) или информационный образ с высокой степенью детализации, достоверности и надежности в явной форме получить нельзя. В этом случае получаем деформированное изображение информационного образа точного результата на нейросети, которое порождает задачу (в общеизвестном смысле — это задача фильтрации), которая связана с выделением из такого изображения с заданной степенью детализации (в смысле

информационной метрики) точного образа. Изображение определяется в виде клеточной топологии на нейросети, элементами которой являются возможные конфигурации точного образа.

Смысл и содержание понятий "образ" и "изображение" определяется и задается в геометрическом, топологическом и информационном аспектах. В геометрическом аспекте это позволяет определить морфометрию изображения образа на нейросети и получить интегральные оценки: пространственные, временные и информационные. В топологическом — определить топологию связей (пространственных, временных и информационных) между элементами нейросети, образующими изображения образа, и их комбинациями. В информационном аспекте образ и изображение определяются как логически и функционально структурированные области виртуальной активной памяти, которые, в свою очередь, являются информационными атрибутами и прототипами образов для систем вычислительного восприятия в технологиях вычислительного интеллекта и для средств восприятия информации непосредственно человеком.

С математической точки зрения информационная модель нейросетевого процессора образов является его отображением в логически и функционально структурированную динамическую область активной виртуальной памяти. Для этого отображения известны область определения в виде множества пассивных виртуальных ячеек  $N$  — его домен — и область значений — его диапазон. В таком аспекте создание информационных моделей нейросетевых процессоров образа является развитием и разработкой нового направления теории отображений для технологий виртуализации и информационной динамики объектов вычислительного интеллекта с переменными доменами-диапазонами. Домен в этом случае однозначно задается организацией виртуальной активной памяти на множестве виртуальных ячеек различных типов и топологий связей (логических и функциональных) между ними, а диапазон задается типом и смысловым значением информации, которая хранится в виртуальных ячейках и "протекает" через них. Описание расположения конфигураций в адресном пространстве  $N$  задается в виде логической или любой другой структуры данных (СД), элементы которой являются адресами соответствующих виртуальных ячеек. Например, для нейросети произвольной пространственной размерности самый простой вариант — это односвязанный список. Информационное наполнение конфигурации определяется логической связью и смысловым содержанием потока данных, хранимых по адресам, указанным элементами СД. Множество возможных значений элементов такого потока данных задает систему информационных образующих  $G = \{g_i\}_1^L$  для конфигураций  $c = \{g_1, g_2, \dots, g_k\}$ , где

$k$  — любое натуральное число. Графическим образом (ГО) такой информационной конфигурации  $c$  на нейросети будет подмножество нейронов, которые образуют односвязанную ориентированную сеть или подграф. Информационным отображением нейрона как локальной информационной сети будет сегмент  $M \subset N$  виртуальной активной памяти. Множество пассивных ячеек сегмента  $M$  из  $m$  элементов активной виртуальной памяти образуют локальную информационную статистическую среду  $\text{env}(\Omega)$  для  $\Omega$  вычислительного интеллекта, т. е. они являются элементами кратковременной и долговременной памяти.

Топологическая структура обозначенного типа конфигураций  $c$  удовлетворяет комбинаторным правилам следующего вида:

$$\mathfrak{R} = \langle \rho, \Sigma \rangle,$$

где  $\Sigma$  — тип соединений между образующими, представляющими собой некоторое множество графов, описывающих способ соединения связей друг с другом;  $\rho$  — отношение связей.

Правила  $\mathfrak{R}$  представляют собой объединение отношения связей с типом соединений. Поэтому при разработке и реализации технологий вычислительного интеллекта на основе нейросетевого процессора образов можно использовать как одиночные правила, так и классы правил. Рассматриваем два класса конфигураций относительно правил  $\mathfrak{R}$ : регулярные и нерегулярные. К регулярным конфигурациям относительно  $\mathfrak{R}$  будем относить те, для которых  $\Sigma$  — это линейный тип связей (односвязанный ориентированный линейный граф), а  $\rho$  — попарное отношение связей.

Потенциальную среду для  $\Omega$  зададим как объединение регулярных конфигураций

$$\text{env}_{\text{pot}}(\Omega) = G(\mathfrak{R}) = \bigcup_{n=1}^{\infty} G_n(\mathfrak{R}).$$

В реальности среда для  $\Omega$  не определяется регулярностью  $\mathfrak{R}$  — эти правила лишь ограничивают возможные конфигурации множеством  $G(\mathfrak{R})$ , но не определяют меру неопределенности или вероятности "встречи"  $\Omega$  с тем или другим расположением объектов, той или иной конфигурацией. Для исключения этого момента введем меру неопределенности  $Q$  в пространстве конфигураций и зададим информационную статистическую среду.

Виртуальная информационная среда для  $\Omega$ , обозначим ее как  $\text{env}(\Omega)$ , логически и функционально структурирована в адресном и информационном пространстве виртуальной активной памяти.

Множество пассивных ячеек сегмента активной виртуальной памяти образуют локальную информационную статистическую среду  $\text{env}(\Omega)$  для объектов  $\Omega$  вычислительного интеллекта, т. е. являются информационными атрибутами элементов кратко-

временной и долговременной памяти. Формально эту среду (по аналогии с [9]) можно задать как

$$\text{env}(\Omega) = (G(\mathfrak{R}), Q),$$

где  $Q$  — некоторая вероятностная мера или другая мера неопределенности на множестве  $G(\mathfrak{R})$  допустимых конфигураций, которые представляют подмножество пассивных ячеек сегмента, объединенных логической и смысловой зависимостью, т. е.  $G(\mathfrak{R})$  является элементом топологической структуры  $T(M)$ , которая определена на множестве  $M$ . Все элементы  $T(M)$  частично упорядочены, а с каждой частичной упорядоченностью связана единственная диаграмма [10], что позволяет в рамках единого математического формализма на уровне сетевых моделей и алгоритмов построить информационную модель нейросетевого процессора, в котором среда вычислений и информационные процессы логически и функционально организованы, исходя из следующих сформулированных принципов компьютеринга.

1. Принцип взаимодействия объектов (символьные цепочки) среды вычислений с виртуальной информационной средой. Основная посылка состоит в том, что при выполнении логических или арифметических операций над операндами в различных типах виртуальных ячеек вначале реализуется "процедура" взаимодействия с виртуальной информационной средой.

2. Принцип виртуальной перспективы в операторе проектирования, смысл и содержание которого определены выше.

3. Принцип связанности результатов вычислений. Основная посылка состоит в том, что в  $\text{env}(\Omega)$  из  $G(\mathfrak{R})$  необходимо выделить одну единственную конфигурацию, которая удовлетворяет критериям фрактальной, топологической и информационной связанности.

### О практической реализации

Предложенный подход создания информационной модели нейросетевого процессора образов был положен в основу разработки технологий вычислительного интеллекта исследования поведения решений размытых динамических систем в рамках ограниченной задачи  $N$  тел на основе организации и реализации форм компьютеринга для соответствующих задач [11]. Фрактальный анализ большого объема результатов вычислительных экспериментов позволил получить ряд новых и уникальных результатов и выводов.

Во-первых, образы решений обозначенных размытых систем в условиях ограничений и неопределенности отражают закономерности самосоглазованного стремления к критическим режимам в их поведении. Эти образы имеют топологию перколирующего фрактального объекта и фрактального агрегата, образуя мультифрактальный объект в виде топологического комплекса.

Во-вторых, компьютеринг задач и информационная динамика объектов среды вычислений, обусловленных ограничениями и неопределенностью, обладают свойством самоорганизации (синергетические особенности), т. е. определяют те пространственно-временные и информационные условия, при которых процесс моделирования контролируем.

В-третьих, определен механизм перехода системы от регулярной динамики со сложной структурой к нерегулярной (размытой) — это перемежаемый коллапс неопределенностей, а также ряд других свойств как в поведении моделируемых систем, так и в технологиях компьютеринга.

В-четвертых, была показана и подтверждена одна из важных закономерностей поведения обратимых динамических систем, которая в рамках теорий традиционной теоретической и вычислительной математики считалась невозможной, а именно: моделирование обратимых динамических систем и формы реализации компьютеринга в условиях ограничений, обмена и неопределенности порождает необратимые процессы в информационной динамике объектов среды вычислений.

### Заключение

Рассмотренная методология разработки информационных моделей нейросетевого процессора образов технологий вычислительного интеллекта отражает и определяет новые формы реализации компьютеринга. В ней заложены новые основы осуществления компьютеринга в плане разработки и реализации проработанных и осмысленных моделей алгоритмов и процедур интеллектуальных вычислений, в которых не предполагается традиционный арифметический стиль работы с числовыми и другими типами данных, характерный для абсолютного большинства существующих вычислительных систем и систем программирования. В этом случае предлагается перейти к иному стилю рассуждений в терминах информационной динамики и образного представления объектов компьютерных процессов среды вычислений интеллектуальных технологий моделирования.

### Список литературы

1. **Осовский С.** Нейронные сети для обработки информации. М.: Финансы и статистика, 2002. 344 с.
2. **Борисюк Г. Н., Борисюк Р. М., Казанович Я. Б., Иваницкий Г. Р.** Модели динамики нейронной активности при обработке информации мозгом — итоги "десятилетия" // УФН. 2002. Т. 172, № 10. С. 1189—1214.
3. **Мышев А. В.** Теория компьютерного восприятия и технологии взаимодействия вычислительного интеллекта с виртуальной средой моделирования // Труды 7-й Международной научно-технической конференции "Кибернетика и высокие технологии 21 века", 16—18 мая 2006 г., Воронеж, Россия. Воронеж: ВГУ. 2006. С. 497—508.
4. **Мышев А. В., Иванов П. Г.** Системы компьютерного восприятия и когнитивные технологии взаимодействия объектов имитационного моделирования с виртуальной средой в условиях

информационной неопределенности // Труды Регионального конкурса научных проектов в области естественных наук. Вып. 14. Калуга, АНО "Калужский научный центр". 2009. С. 29—37.

5. **Мышев А. В.** Виртуальные модели активной памяти для построения нейросетевых процессоров синтеза образов решений размытых задач в технологиях вычислительного интеллекта // Труды 3-й международной научной конференции "САИТ-2009", 14—18 сентября 2009 г., Звенигород, Россия. М.: ИСА РАН. 2009. С. 873—882.

6. **Мышев А. В.** Модели активной памяти на бинарных полях в технологиях виртуализации каналов передачи и хранения информации // Программные продукты и системы. 2010. № 1. С. 54—58.

7. **Мышев А. В., Игнатенко П. И., Лебедева Н. С.** Динамика информационных процессов в системах компьютерного вос-

приятия // Труды Регионального конкурса научных проектов в области естественных наук. Вып. 13. Калуга: АНО "Калужский научный центр". 2008. С. 40—47.

8. **Вейр Б.** Анализ генетических данных. М.: Мир, 1995. 400 с.

9. **Гренандер У.** Лекции по теории образов: анализ образов. М.: Мир, 1981. 448 с.

10. **Kleitman D., Rothschild B.** The number of finite topologies // Proc. of American Mathematical Society. 1970. V. 25, N 2. P. 276—282.

11. **Мышев А. В., Куликова Н. В., Игнатенко П. И.** Информационные технологии системного анализа динамики объектов задачи N-тел в условиях неопределенности // Труды Регионального конкурса научных проектов в области естественных наук. Вып. 5. Калуга: Эйдос. 2003. С. 9—22.

УДК 004.032.26:343.98

**С. Д. Кулик**, д-р техн. наук, проф.,  
**Д. А. Никонен**, соискатель, вед. программист,  
e-mail: denis@informcontact.ru  
НИЯУ МИФИ

## **Автоматизация классификационно- диагностических почерковедческих исследований с помощью нейронных сетей**

*Рассматривается применение нейронных сетей для решения задач почерковедческой экспертизы, в частности, задачи определения пола и возраста по почерку. Приведены современные почерковедческие методики и программное обеспечение для определения пола и возраста исполнителя рукописи. Рассматривается формирование базы данных рукописных образцов и используемые признаки почерка. Описывается разработка нейронных сетей, используемых для принятия решения о поле и возрасте исполнителя рукописного текста.*

**Ключевые слова:** нейронные сети, почерк, почерковедческая экспертиза

### **Введение**

Число почерковедческих экспертиз, проведенных в РФ, по данным ЭКЦ МВД, составило в 2006 г. — 73 тыс., в 2007 г. — 76 тыс., в 2008 г. — 72 тыс., в 2009 г. — 78 тыс., а сама почерковедческая экспертиза занимает 3-е место среди других видов экспертиз по числу выполненных [1]. Если принять во внимание, что криминалистическими подразделениями ОВД в 1999 г. всего было выполнено 47 тыс. почерковедческих экспертиз [1], то можно сделать вывод, что за 10 лет их число выросло в 1,6 раза.

Согласно [2] выделяют следующие виды задач, встречающихся в практике экспертов-почерковедов:

- идентификационные задачи (идентификация почерка определенного лица);
- диагностические задачи (установлением факта наличия или отсутствия необычных условий выполнения рукописных текстов);
- классификационно-диагностические задачи (например, установление групповой принадлежности исполнителя рукописи по полу, возрасту и т. п.).

Решение классификационно-диагностических задач, в частности, определение пола и возраста по почерку характерно для почерковедческих исследований, назначаемых по уголовным делам, когда исследуются рукописные документы, исполнители которых неизвестны, а также для розыска лиц, причастных к совершению преступления.

При разработке почерковедческих методик специалистами редко используются нейросетевые технологии, что, по мнению авторов, не всегда оправдано. Данная работа нацелена на частичное устранение этого недостатка.

### **Методики и программы для определения пола и возраста по почерку**

При выполнении исследований эксперты руководствуются разного рода методическими документами (методиками). Следует заметить, что в нашей стране разработаны и апробированы на практике методики определения пола и возраста исполнителя рукописи [1—7].

Хотя зарубежными специалистами официально считается, что по почерку нельзя определить пол и возраст исполнителя рукописи (например, см. рекомендации ФБР [8]), однако за рубежом проводятся различные исследования, изучающие такие возможности [9, 10].

Специалистами экспертно-криминалистического центра (ЭКЦ) МВД РФ проведена паспортизация

используемых почерковедческих методик, в этот список вошла 31 действующая почерковедческая методика [1]. Среди них такие известные методики, разработанные совместно специалистами МИФИ и ЭКЦ МВД РФ, как методика определения возраста исполнителя рукописных текстов [5].

Разработаны программы для определения пола и возраста исполнителя рукописного текста (табл. 1) [1–3, 6, 11, 12]. При разработке программ ПАРМ, man-WO-man, Мастер, Почерк для определения пола исполнителя была использована методика определения пола исполнителя кратких рукописных текстов [4]. Методика определения возраста исполнителя рукописных текстов [5] была автоматизирована в программе ПАРМ.

Состояние дел в области почерковедческих экспертиз убедительно указывает на необходимость разработки новых почерковедческих методик, применения новых показателей эффективности, связанных с методиками, и создания комплексных автоматизированных средств поддержки почерковедческих экспертиз, например, в виде автоматизированного рабочего места (АРМ) эксперта-почерковеда.

Анализ уже созданных автоматизированных систем (АРМ) (табл. 1) показывает, что полностью автоматизировать проведение почерковедческой экспертизы для решения классификационно-диагностических задач пока не удается, а также то, что от-

сутствует программное обеспечение, решающее все основные задачи почерковедческих исследований.

Поэтому разработка АРМ для эксперта-почерковеда является крайне актуальной и необходимой. Было принято решение о разработке АРМ "FHWE v. 1.0" [1, 3], в рамках которого будут не только автоматизированы существующие методики определения пола и возраста исполнителя рукописного текста [4, 5], но и разработаны новые, модифицированные методики.

### Разработка базы данных образцов рукописных объектов

Создание базы данных (БД) рукописных объектов — первый этап разработки методики и создания программного обеспечения для почерковедческих исследований.

БД, содержащие рукописные объекты, обычно применяют для получения обучающей и контрольной выборки, используемых при обучении нейронной сети, для разработки и проверки алгоритмов идентификации, для определения параметров (признаков) рукописного текста и их информативности и т. д. Однако необходимо отметить, что существующие БД в основном разработаны для решения задачи идентификации и не предназначены для определения пола и возраста исполнителя рукописи. Также эти БД обычно содержат рукописные тексты на английском языке и не подходят для криминалистического исследования рукописных текстов на русском языке. Поэтому было принято решение о разработке необходимой отечественной БД. Для экспериментальных исследований и разработки новых почерковедческих методик в АРМ "FHWE v. 1.0" эта БД была заполнена образцами рукописных объектов (691 образец), для исполнителей которых были известны из пол, возраст, рост, национальность и профессия [1, 3]. Каждый образец представляет собой 1–2 страницы формата А4 рукописного текста. Известен пол исполнителей документов: 321 образец выполнен мужчинами, а 370 — женщинами. Возраст исполнителей — от 16 до 76 лет. Известна профессия исполнителей рукописей, например, число документов, выполненных: военными служащими — 14, инженерами и научными сотрудниками — 134, медицинскими работниками — 24, сотрудниками правоохранительных органов — 66, рабочими — 59, служащими — 78, учащимися и студентами — 179, учителями — 44, представителями других профессий — 93. Также известна национальность исполнителей, среди них (число): русские — 523, украинцы — 26, литовцы — 19, татары — 10, евреи — 6, белорусы — 5, армяне — 5, прочие — 16, не указано — 81.

Таким образом, можно сделать вывод, что в полученной БД образцов рукописных объектов широко представлены исполнители рукописей разных профессий, национальностей, возрастов и т. п., что повышает практическую достоверность решений,

Таблица 1

Программы для определения пола и возраста по почерку [1–3, 6, 11, 12]

Название	Описание	Разработчик
Тюльпан	Дифференциация рукописей на мужские и женские, установление возраста	Харьковский НИИСЭ
ПАРМ	Элементы автоматизированного рабочего места эксперта-почерковеда	ЭКЦ МВД РФ, А.Б. Левицкий, Э. П. Молоков
man-WO-man	Фактографическая система для определения пола исполнителя рукописи	С. Д. Кулик
Мастер	Определение пола исполнителя кратких рукописных текстов	ВНКЦ МВД СССР
POL	Дифференциация высоковыработанных почерков на мужские и женские	Ташкентский НИИСЭ
Prognoz, Priznak	Программная реализация комплексной методики установления пола, возраста и психологических характеристик исполнителя рукописи	РФЦСЭ, Л. В. Сидельникова, А. Н. Герасимов
Почерк	Автоматизированная система для определения пола исполнителя рукописи	ВА МВД России
CEDAR-FOX	Программа для почерковедческих исследований	США, 2003

принимаемых с помощью разрабатываемых нейронных сетей, использующих в качестве обучающей и контрольной выборки информацию этой базы данных.

При разработке АРМ "FHWE v. 1.0" совместно с экспертами-почерковедами был сформирован справочник частных признаков букв русского языка, используемых в качестве характеристик почерковых объектов. Для этого было использовано разбиение 28 прописных букв (не рассматривались признаки для букв "Е", "Й", "Б", "Ы", "Ь") на элементы.

Всего для 28 прописных букв русского языка было выделено 1865 частных признаков. Также в справочник были добавлены некоторые частные признаки строчных букв. В результате сформированный справочник частных признаков букв русского языка, используемых в качестве характеристик почерковых объектов, содержит 1963 признака (соответственно, 1865 для прописных и 98 для строчных букв).

Для всех рукописных документов из БД образцов рукописных объектов экспертами-почерковедами был составлен список встречающихся в этих документах частных признаков. Всего экспертами было выделено около 120 000 соответствий признаков образцам (соответственно, около 114 000 для прописных и 6 000 для строчных букв). Минимальное число признаков, выделенных в одном документе, — 89, максимальное — 278, среднее — 175. БД образцов и справочник частных признаков были далее использованы для проведения исследований и разработки почерковедческих методик, позволяющих определить пол, возраст и идентифицировать исполнителя рукописного текста.

Рассмотрим математическое описание для частных признаков букв рукописного текста. Пусть  $\Theta = \{\theta_1, \dots, \theta_m\}$  — множество (набор) частных признаков букв рукописного текста, используемых для описания рукописей в разрабатываемой БД рукописных объектов, где  $m$  — число используемых признаков букв. Мощность множества  $\Theta$  (число используемых признаков букв  $m$ ) равна  $|\Theta| = m = 1963$ .

Пусть дано множество исполнителей рукописных образцов (рукописей)  $D$ , где для каждого исполнителя рукописи экспертами выделено множество  $T$  частных признаков букв, характеризующих его почерк (встречающихся в почерке). То есть множество  $T$  представлено в виде набора элементов (частных признаков букв) из множества  $\Theta$ ,  $T \subset \Theta$ . Мощность множества  $D$  (число исполнителей рукописных образцов) равна  $|D| = 691$ . Каждому исполнителю рукописи из множества  $D$  соответствует рукописный образец в разрабатываемой БД. Известно, что все используемые образцы выполнены различными исполнителями.

Ошибки экспертов при выделении признаков не учитываются, т. е. предполагаем, что работают эксперты высокой квалификации. Тогда для двух

различных рукописей достаточной длины (обычно достаточно текста размером 1 страница и более), выполненных одним исполнителем, эксперты выберут одни и те же признаки, характеризующие почерк (множество  $T$ ).

Более удобно представлять почерк исполнителя рукописи как бинарный вектор:

$$\mathbf{T} = \{t_1, t_2, t_3, \dots, t_m\}, \quad (1)$$

где  $t_k = 1$ , если признак  $\theta_k$  характерен для почерка  $T$  (т. е.  $\theta_k \in T$ ), а иначе  $t_k = 0$ . Вектор  $\mathbf{T}$  представляет собой значения для входов нейронной сети, которую необходимо построить и обучить.

### Использование нейронных сетей для разработки методик

Для принятия решения в программных средствах для автоматизации почерковедческой экспертизы часто используют статистический дискриминатор (например, байесовский дискриминатор) или метод  $k$ -ближайших соседей [1, 12]. Также анализ показывает, что марковские модели (марковский процесс, сети Маркова) [13], вейвлет-преобразования [14] лонгитюдных данных [14], обучаемые многофакторные сети Маркова [15] имеют важное значение для различных практических приложений, в том числе и для криминалистики.

Специалисты считают, что использование статистических методов, байесовских дискриминаторов и байесовских сетей — одно из наиболее важных направлений создания решающих правил и количественных методик в криминалистике [16—18]. В отечественном почерковедении для реализации решающих правил рекомендуется специалистами [2] и часто применяется на практике [1—5] примитивный (наивный) байесовский дискриминатор. Однако, несмотря на очевидное удобство обучения и простоту реализации, примитивный байесовский дискриминатор обладает рядом недостатков. Наиболее существенным недостатком считают [2] требование статистической независимости признаков почерка (входов дискриминатора), что редко встречается на практике.

Следует отметить, что в настоящее время полный справочник взаимозависимостей частных признаков букв русского языка не существует. Специалистами представлены [2] только некоторые примеры (несколько десятков) наиболее типичных зависимостей для признаков почерка, хотя эксперименты показывают, что зависимостей значительно (в десятки раз) больше [1]. Зависимые признаки эксперты выделяют только на основании своего опыта. Поэтому корректное выделение зависимостей возможно, только если исследование выполняет эксперт достаточно высокой квалификации.



Специалисты предлагают следующие методы выявления и уменьшения числа зависимых признаков [2]:

- удаление некоторых высокоразвитых признаков;
- выявление групп взаимозависимых признаков и использование в дальнейшем вместо них только одного, группового признака.

Проблема и экспериментальное исследование влияния зависимости частных признаков на качество работы существующих почерковедческих методик рассмотрены в работе [1].

Учет зависимостей частных признаков букв при использовании существующих почерковедческих методик повышает число правильно принятых решений и снижает число ошибок. Поэтому одним из направлений автоматизации почерковедческих исследований является автоматизация существующих почерковедческих методик с разработкой и последующим применением справочника зависимостей признаков (справочник групп зависимых признаков). Более подробно формирование и использование такого справочника рассмотрено в работе [1].

Еще одним направлением автоматизации почерковедческих исследований является разработка новых почерковедческих методик, устойчивых к зависимости признаков. В качестве примера таких методик можно привести методики определения пола и возраста, реализованные с помощью нейронных сетей.

Следует заметить, что применение нейронных сетей для почерковедческих исследований не всегда удобно. Это связано с тем, что для работы нейронной сети требуется фиксированное число признаков (по числу входов), и, соответственно, затруднительно использовать нейронную сеть, если часть признаков отсутствует. А для реальных рукописных документов часто невозможно получить весь набор признаков. Связано это, прежде всего, с тем, что часто необходимо выполнять исследования рукописных текстов недостаточной длины, когда не все буквы присутствуют в тексте. И следовательно, не для всех признаков можно принять решение, встречаются ли они в почерке исполнителя рукописи или нет. То есть невозможно полностью сформировать вектор  $\mathbf{T}$  (1) значений для входов нейронной сети.

Для решения этой проблемы некоторые исследователи (см., например, [12]) предлагают вместо признаков, основанных на разбиении букв на элементы, использовать признаки, вычисление которых не зависит от текстового содержания и длины исследуемого рукописного документа. Так, в программе "CEDAR-FOX", где для принятия решений используются нейронные сети [12], признаками служат вычисляемые в автоматическом режиме некоторые характеристики сканированного изображения рукописного текста [12].

Однако в отечественном почерковедении использование автоматически вычисляемых признаков не

практикуется. Согласно действующему законодательству эксперт в ходе исследований принимает решения, сообразуясь со своим *внутренним убеждением*. Поскольку интерпретация значений автоматически вычисляемых признаков не всегда очевидна для эксперта, затруднительно использовать эти признаки для доказательства правильности принятого решения при оформлении экспертного заключения, а также в суде. Поэтому для разработки нейронных сетей было принято решение использовать традиционные частные признаки рукописных букв (описаны выше).

Следует заметить, что все рукописные объекты в используемой базе данных образцов имеют достаточную длину (1 страница и более). Причем обязательно в тексте встречаются (обычно несколько раз) все буквы. Поэтому для каждого образца эксперт может для всех используемых признаков принять решение, встречаются ли они в почерке исполнителя или нет. То есть возможно полностью сформировать вектор значений для входов нейронной сети  $\mathbf{T}$  (1).

### Порядок разработки нейронных сетей

Для разработки нейронных сетей было решено использовать нейропакет *NeuroSolutions* фирмы *NeuroDimension, Inc.* Данная программа была выбрана потому, что она показала приемлемую скорость обучения нейронной сети с 1963 входами, а также из-за того, что программа позволяет разрабатывать нейронные сети, на вход которых подаются и с выхода получают нечисловые данные.

Было решено сформировать несколько вариантов нейронных сетей, используя различные парадигмы и меняя число слоев, нейронов в слое и функции активации. В качестве показателя эффективности нейронной сети примем  $G$  — процент правильно принятых решений на контрольной выборке. Введем критерий эффективности нейронной сети, в качестве него примем —  $\max(G)$ , т. е. та нейронная сеть, у которой показатель  $G$  будет максимален, и будет эффективной. Исследование в целях сокращения времени расчетов было решено проводить в два этапа:

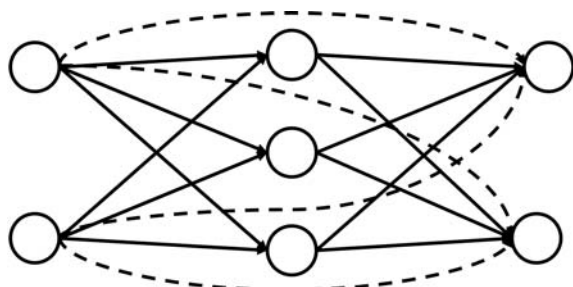
- для формирования нейронных сетей в ходе первого этапа вместо 1963 признаков для 28 букв используются 697 признаков первых 10 букв. В результате отбираются варианты нейронных сетей, показавшие лучший результат. Эти варианты сетей используются на втором этапе;
- в ходе второго этапа исследования, используя топологии, полученные на первом этапе, формируются и обучаются нейронные сети для всех 1963 признаков. В результате отбирается нейронная сеть, показавшая лучший результат.

Полученная нейронная сеть в дальнейшем используется для принятия решения в разрабатываемой почерковедческой методике.

## Разработка нейронной сети для определения пола по почерку

Для разработки методики определения пола исполнителя рукописи с помощью нейронной сети из 691 образца рукописного текста были случайным образом выделены обучающая выборка из 601 и контрольная выборка из 90 рукописных объектов (по 45 образцов для каждого пола).

В качестве исследуемых топологий нейронных сетей были выбраны многослойный перцептрон (MLP-сеть) и обобщенная сеть прямого направления (GFFN-сеть, см. рисунок).



Пример сети GFFN с одним скрытым слоем (штриховыми линиями показаны связи через слой)

Таблица 2

Параметры нейронных сетей для определения пола

Номер нейронной сети	Параметры скрытых слоев			Параметры выходного слоя
	Число слоев	Число нейронов	Функция активации	Функция активации
1 (MLP)	1	15	SigmoidAxon	LinearSigmoid-Axon
2 (MLP)	1	15	Linear-SigmoidAxon	TahnAxon
3 (MLP)	1	15	LinearSigmoidAxon	SigmoidAxon
4 (GFNN)	1	15	TahnAxon	TahnAxon
<b>5 (GFNN)</b>	<b>1</b>	<b>15</b>	<b>Linear-SigmoidAxon</b>	<b>TahnAxon</b>
6 (GFNN)	2	10	TahnAxon	TahnAxon

Таблица 3

Результаты проверки нейронных сетей для определения пола

Номер нейронной сети	Мужской пол, правильно		Женский пол, правильно		Всего правильно, %
	Число	%	Число	%	
1	44	97,8	39	86,7	92,2
2	44	97,8	38	84,4	91,1
3	42	93,3	39	86,7	90,0
4	45	100,0	41	91,1	95,6
<b>5</b>	<b>44</b>	<b>97,8</b>	<b>43</b>	<b>95,6</b>	<b>96,7</b>
6	45	100,0	41	91,1	95,6

В ходе первого этапа с помощью программы NeuroSolutions были сформированы и обучены несколько нейронных сетей с числом входов 697, выходов — 2, в которых варьировались следующие параметры: число слоев (1 или 2), число нейронов в слое (10 или 15), функции активации в скрытых и выходных слоях (сигмоидальная функция (*SigmoidAxon*), гиперболический тангенс (*TahnAxon*), линейная сигмоидальная функция (*LinearSigmoidAxon*)).

Всего было сформировано и обучено 72 нейронные сети (по 36 для каждой используемой топологии нейронной сети). Из полученных 72 нейронных сетей было выбрано шесть, показавших лучшие результаты (меньшее число ошибок), по три для каждой топологии нейронной сети.

В ходе второго этапа исследований на основании результатов, полученных на предыдущем этапе, были сформированы шесть нейронных сетей (табл. 2) с числом входов 1963 (для признаков всех 28 букв), выходов — 2. Результаты проверки с помощью контрольной выборки полученных нейронных сетей приведены в табл. 3.

Лучший результат показала обобщенная сеть прямого распространения (выделена в табл. 2, 3 жирным) с одним скрытым слоем в 15 нейронов. Функция активации скрытого слоя — линейная сигмоидальная, выходного слоя — гиперболический тангенс. В результате проверки данной нейронной сети с помощью контрольной выборки из 90 рукописей было получено 87 правильных результатов (96,7 %).

## Разработка нейронной сети для определения возраста по почерку

Результатом работы нейронной сети должно быть отнесение исполнителя рукописи к одному из трех возрастных диапазонов: не более 27 лет; 28—40 лет; не менее 41 года. Для разработки методики определения возраста исполнителя рукописи с помощью нейронной сети из 691 образца рукописного текста были случайным образом выделены обучающая выборка из 601 и контрольная выборка из 90 объектов (по 30 образцов для каждого возрастного диапазона). В качестве топологии нейронных сетей был выбран многослойный перцептрон (MLP-сеть).

В ходе первого этапа были сформированы и обучены несколько нейронных сетей с числом входов 697, выходов — 3 (по числу возрастных диапазонов), в которых варьировались следующие параметры: число слоев — 1, 2, 3, число нейронов в слое — 10, 15, 20, функции активации в скрытых и выходных слоях (сигмоидальная функция или гиперболический тангенс). Всего было сформировано и обучено 27 нейронных сетей. Из полученных 27 нейронных сетей были выбраны три, показавших лучшие результаты (меньшее число ошибок).

Таблица 4

## Параметры нейронных сетей для определения возраста

Номер нейронной сети	Параметры скрытых слоев			Параметры выходного слоя
	Число слоев	Число нейронов	Функция активации	Функция активации
1 (MLP)	1	20	TahnAxon	SigmoidAxon
2 (MLP)	2	15	TahnAxon	SigmoidAxon
3 (MLP)	2	20	SigmoidAxon	TahnAxon

Таблица 5

## Результаты проверки нейронных сетей для определения возраста

Номер нейронной сети	1-й диапазон (не более 27 лет) правильно		2-й диапазон (28—40 лет) правильно		3-й диапазон (не менее 41 года) правильно		Всего правильно, %
	Число	%	Число	%	Число	%	
1	17	56,7	22	73,3	27	90,0	73,3
2	9	30,0	19	63,4	29	96,7	63,3
3	21	70,0	23	76,7	29	96,7	81,1

В ходе второго этапа исследований на основании результатов, полученных на предыдущем этапе, были сформированы три нейронные сети с числом входов 1983, выходов — 3 (табл. 4).

Результаты проверки с помощью контрольной выборки полученных нейронных сетей приведены в табл. 5.

Лучший результат показала сеть (выделена в табл. 4, 5 жирным) с двумя скрытыми слоями в 20 нейронов. Функция активации скрытых слоев — сигмоидальная, выходного слоя — гиперболический тангенс. В результате проверки данной сети с помощью контрольной выборки из 90 рукописей было получено 73 правильных результата (81,1 %).

### Выводы

Были сформированы и исследованы 72 нейронные сети для решения задачи определения пола и 27 нейронных сетей для решения задачи определения возраста исполнителя рукописи. Из этих сетей были выбраны лучшие (с меньшим числом ошибок), которые в дальнейшем были использованы в разрабатываемых почерковедческих методиках.

С помощью нейронных сетей разработан ряд экспериментальных почерковедческих методик, повышающих эффективность работы эксперта-криминалиста (почерковеда) и позволяющих определить пол, возраст исполнителей рукописных текстов.

### Список литературы

- Кулик С. Д., Никоненко Д. А., Ткаченко К. И., Лукьянов И. А. Методы и средства повышения эффективности информационных систем (нейронные сети, криминалистика, формирование фактографических данных, морфологический анализ). Т. 1. Криминалистика. М.: Радиотехника, 2011. 300 с. Деп. в ВИНТИ 05.05.2011. № 206-В2011.
- Судебно-почерковедческая экспертиза: общая часть: теор. и метод. основы / Под науч. ред. В. Ф. Орловой. Государственное учреждение РФЦЭС при Минюсте России. 2-е изд., перераб. и доп. М.: Наука, 2006. 544 с. (Библиотека судебного эксперта).
- Никоненко Д. А. Нейросетевые технологии для решения задач криминалистики // Нейрокомпьютеры: разработка, применение. 2009. № 4. С. 26—45.
- Методика определения поля исполнителя кратких рукописных текстов: учеб. пособие / М. М. Чельшев, С. Д. Кулик, С. Ю. Мирошникова, Г. А. Бажакин, О. Д. Белоусова, Е. Ю. Колесова, А. Б. Левицкий, М. Т. Марушкин, Э. П. Молоков, О. С. Мурашова, В. В. Серегин. М.: ВНКЦ МВД СССР, 1990. 185 с.
- Методика определения возраста исполнителя рукописных текстов: учеб. пособие / М. М. Чельшев, С. Д. Кулик, Г. А. Бажакин, Е. Ю. Колесова, А. Б. Левицкий, Э. П. Молоков, О. С. Мурашова, В. В. Серегин, А. Г. Скоморохова, М. Н. Сосенушкин, А. М. Черенков, Т. Г. Шаова. М.: ЭКЦ МВД России, 1995. 255 с.
- Сидельникова Л. В., Герасимов А. Н. Комплексная методика установления пола, возраста и психологических свойств исполнителя текста, выполненного почерком высокой и выше средней степени выработанности: Методическое письмо. М.: РФЦЭС, 2005.
- Комаров А. С. Логические и программные средства интеллектуального анализа криминалистических данных: Дис. ... канд. техн. наук: 05.13.17. М., 2010.
- Handbook of Forensic Services (revised 2007) / K. Waggoner (ed.). Federal Bureau of Investigation. Quantico, Virginia: An FBI Laboratory Publication, 2007.
- Huber R. A., Headrick A. M. Handwriting Identification: Facts and Fundamentals. Boca Raton. New York: CRC Press, 1999.
- Tomai C., Kshirsagar D. M., Srihari S. N. Group Discriminatory Power of Handwritten Characters // Document Recognition and Retrieval XI: Proceedings of Electronic Imaging (E. H. Barney Smith, Jianying Hu and J. Allan, ed.), January 2004. San Jose, CA. P. 116—123.
- Кошманов П. М., Кошманов М. П. Этапы и основные направления внедрения компьютерных технологий в судебное почерковедение и почерковедческую экспертизу // Эксперт-криминалист. 2008. № 3. С. 35—40.
- Srihari S. N., S.-H. Cha, Arora H., Lee S. Individuality of Handwriting // Journal of Forensic Sciences. 2002. 47(4), July. P. 856—872.
- Куравский Л. С., Юрьев Г. А. Адаптивное тестирование как марковский процесс: модели и их идентификация // Нейрокомпьютеры: разработка и применение. 2011. № 2. С. 21—29.
- Куравский Л. С., Мармалок П. А., Баранов С. Н., Абрамочкина В. И., Петрова Е. А. Факторный анализ результатов вейвлет-преобразований лонгитюдных данных как новый метод исследования динамических характеристик сложных систем // Нейрокомпьютеры: разработка и применение. 2009. № 9. С. 5—19.
- Куравский Л. С., Баранов С. Н., Корниенко П. А. Обучаемые многофакторные сети Маркова и их применение для исследования психологических характеристик // Нейрокомпьютеры: разработка и применение. 2005. № 12. С. 65—76.
- Кузьмин С. А. Перспективные направления использования современных методов обработки судебно-экспертных данных // Теория и практика судебной экспертизы. 2009. № 3. С. 135—137.
- Bayesian Networks and Probabilistic Inference in Forensic Science / F. Taroni, C. Aitken, P. Garbolino, A. Biedermann. John Wiley & Sons, 2006.
- Köller N., Nissen K., Rieß M., Sadorf E. Probabilistische Schlussfolgerungen in Schriftgutachten = Probability Conclusions in Expert Opinions on Handwriting. München: Luchterhand, 2004.

**Нгуен Виет Хунг**, аспирант,  
Московский физико-технический институт,  
viethung81@yahoo.com

## Сжатие изображения с использованием нейронной сети в преобразованном пространстве

*Рассматривается новый подход к задаче сжатия изображения — нейросетевой алгоритм сжатия изображений на основе самоорганизующейся карты Кохонена в преобразованном пространстве. Приводятся результаты экспериментальных исследований.*

**Ключевые слова:** сжатие изображения, нейронные сети, сеть Кохонена, вейвлетное преобразование

### Введение

Проблема сжатия цифровых видео и изображений является актуальной во многих прикладных и научных задачах. Эта проблема особенно ощущается в условиях возрастания объемов передаваемой информации из-за ограничений пропускной способности каналов связи. Сжатие изображений имеет также важное значение при создании центров хранения, архивов и каталогов (баз данных) изображений в цифровом виде (медицинские изображения, космические изображения, получаемые с помощью датчиков дистанционного зондирования, фотоизображения и др.) [2]. Решение проблемы сжатия позволит уменьшить объем информации, хранимой на носителях.

Существует три основных метода сжатия (кодирования) статических изображений [6]:

- метод с предсказанием, использующий экстраполятор для предсказания уровня яркости данного элемента изображения исходя из значений предшествующих элементов. Кодированию подвергается разница между предсказанным и фактическим значениями;
- метод векторной квантизации, использующий для кодирования блоков кадра изображения кодовый словарь;
- метод кодирования с преобразованием, состоящий в разложении кодируемых блоков изображения по элементам некоторого ортогонального базиса. В результате преобразования удается представить более компактно большую часть наиболее значимой информации за счет минимизации избыточности исходного изображения без заметного ухудшения его качества. К таким преобразованиям относятся преобразование синусно-косинусное, вейвлетное, Адамара и др.

Методы с предсказанием недостаточно робастны и не обеспечивают высокой степени сжатия из-за небольших выигрышей в избыточности, причем только между соседними пикселями.

Методы векторной квантизации позволяют достичь высокой степени сжатия изображения, однако их эффективность обычно существенно зависит от характера изображения.

Методы кодирования с преобразованием дают хорошие результаты и по степени сжатия, и по качеству. Однако эти методы нелегко реализовать, и они требуют значительного времени для сжатия изображения.

Мы предлагаем новый подход — векторную квантизацию на основе нейронной сети (сеть Кохонена) с вейвлетным преобразованием.

### Векторная квантизация (VQ) на основе самоорганизующейся карты Кохонена

Основная идея векторной квантизации заключается в разбиении изображения на блоки [2]. Как правило, некоторые блоки оказываются похожими друг на друга. В этом случае компрессор идентифицирует класс похожих блоков и заменяет их одним общим блоком. Кроме того, генерируется двоичная таблица (карта) таких общих блоков из самых коротких кодовых слов. VQ-декодер затем, используя таблицу, собирает изображение поблочно из общих блоков.

Сеть Кохонена самообучается на основе информации входных векторов и группирует "подобные" входные векторы в одну группу. Это не что иное, как один из подходов метода векторной квантизации. В последние годы в связи с широким распространением нейросетевых методов [1] были разработаны нейросетевые алгоритмы сжатия изображений [4, 5]. И наибольшее число исследований и разработок нейросетевых методов сжатия проведено на основе самоорганизующейся карты Кохонена, относящейся к методу векторной квантизации.

Обычно изображение сжимается сетью Кохонена следующим образом. Входной вектор, подаваемый на входной слой нейросети (НС) — ретину, соответствует блоку изображения (рис. 1), а элементы вектора — пикселям этого блока. Первый входной слой НС не содержит нейронов, он выполняет коммутационные функции, распространяя каждый входной вектор по всем нейронам второго слоя. Число узлов входного слоя равно числу элементов/пикселей входного вектора.

Каждый нейрон второго слоя (слой Кохонена) имеет обучаемые весовые коэффициенты своих входов, число входов соответствует числу узлов первого слоя. Внутри второго слоя все нейроны дополнительно могут быть связаны друг с другом латеральными стимулирующими и тормозящими связями.

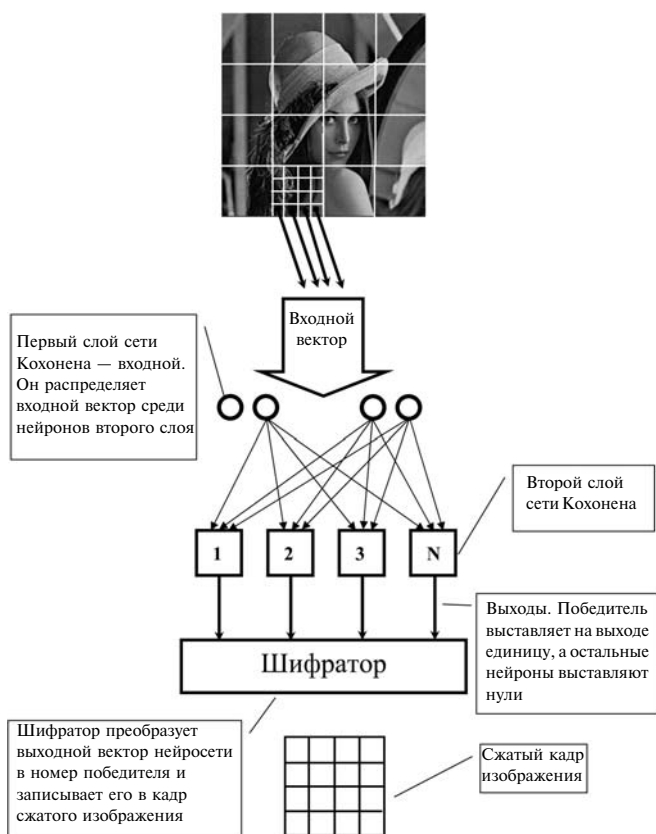


Рис. 1. Структура сжатия изображения с помощью нейронной сети Кохонена

Однако первоначально предложенная Кохоненом НС обладает при обучении плохой сходимостью и некоторыми недостатками в качестве восстанавливаемого изображения [4, 6]. Классические недостатки, которые встречаются при простых алгоритмах блочного сжатия, следующие:

- блочный эффект в низкочастотных (однотонных) областях — видны границы блоков;
- размывание (сглаживание) высококачественных (мелких) компонентов;
- искажения в текстуре и др.

Для их уменьшения предлагаем обучать сеть не по значениям пикселей изображения, а по их значениям коэффициентов вейвлетного преобразования, благодаря чему нарушается корреляция между соседними пикселями.

### Сжатие изображения в преобразованном пространстве на основе сети Кохонена

Для получения коэффициентов вейвлетного преобразования сначала все изображение делится на части одинакового размера, над каждой из которых независимо от других и будут происходить дальнейшие преобразования. Далее каждый канал

проходит фильтрацию низкочастотным и высокочастотным фильтрами отдельно по строкам и по рядам, в результате чего после первого прохода в каждой части формируются четыре более мелких изображения. Все они несут информацию об исходном изображении, но их информативность значительно различается [7].

Например, изображение, полученное после низкочастотной фильтрации по строкам и рядам (вверху слева), несет наибольшее количество информации, а полученное после высокочастотной — минимальное. Информативность у изображений, полученных после НЧ фильтрации строк и ВЧ для столбцов (и наоборот), — средняя. Наиболее информативное изображение опять подвергается фильтрации и т. д. И в результате получается одно маленькое изображение и набор картинок с мелкими деталями, последовательно и с определенной точностью восстанавливающих его до нормального размера.

Для сжатия этих коэффициентов используем такую же структуру, что показана на рис. 1, только входные значения будут не пиксели изображения, а коэффициенты, получающиеся по вейвлетному преобразованию. Для процесса обратного преобразования чем больше коэффициенты, тем важнее роль они играют. Поэтому используются разные сети Кохонена с разными размерами входов (т. е. разные коэффициенты сжатий при фиксированных размерах слоя Кохонена  $16 \times 16$ ) для разных каналов коэффициентов. Чем меньше значения коэффициентов канала, тем больше коэффициент сжатий.

### Эксперименты и результаты

Первые версии алгоритмов были упрощенными и обеспечивали сжатие монохромных изображений формата  $512 \times 512 \times 8$  бит. Было выбрано вейвлет-преобразование Добеши 9/7 (*Daubechies*), и по этому вейвлетному преобразованию изображение разложили до 3-й степени (рис. 2, см. третью сторону обложки).

Чем выше степени вейвлетного разложения изображения, тем меньше коэффициенты разложения. Для 3-й степени используем сеть Кохонена с размером входа, равным  $4 = 2 \times 2$ ; для 2-й степени — сеть Кохонена с размером входа, равным  $16 = 4 \times 4$ , и для 1-й степени — сеть Кохонена с размером входа, равным  $64 = 8 \times 8$ .

При использовании сети Кохонена выходом будет только позиция победительного нейрона, поэтому для каждого входного вектора с размером  $P$  (байт) необходимо сохранить только 1 байт при кодировании. Заметим, что у 3-й степени есть четыре части (ННЗ, НВЗ, ВНЗ, ВВЗ) с размером  $64 \times 64$ , у 2-й степени — три части с размером  $128 \times 128$ , у 1-й сте-

пени — три части с размером  $256 \times 256$ . Таким образом, коэффициент сжатия

$$CR = \frac{\text{Вход(байт)}}{\text{Выход(байт)}} = \frac{512 \times 512}{4 \times 64 \times 64 \times \frac{1}{4} + 3 \times 128 \times 128 \times \frac{1}{16} + 3 \times 256 \times 256 \times \frac{1}{64}} = \left[ \frac{1 + 3 \times 4 \times \frac{1}{16} + 3 \times 16 \times \frac{1}{64}}{64} \right]^{-1} = 25,6.$$

В качестве меры искажения используется отношение сигнал/помеха декодируемого изображения (дБ)

$$PSNR = 10 \lg \left( \frac{D^2 N_x N_y}{E(A, \tilde{A})} \right).$$

Здесь  $N_x \times N_y$  — размер изображения;  $D$  — диапазон амплитуд отсчетов изображения  $A(x, y)$ ;  $E(A, \tilde{A})$  — евклидово расстояние между исходным  $A(x, y)$  и декодируемым изображениями  $\tilde{A}(x, y)$ :

$$E(A, \tilde{A}) = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} (A(x, y) - \tilde{A}(x, y))^2.$$

Отношение сигнал/помеха при сжатии монохромных изображений разными алгоритмами (CR = 25,6)

Изображение	Сеть Кохонена	JPEG	Новый алгоритм
Elaine	25,3454	27,7589	28,8698
Lena	25,6785	27,4508	28,5714
Camera man	23,7325	27,0923	27,1945
Woman blonde	22,2954	27,2954	27,0256
Pepper	25,6115	27,4115	27,5327

Выбирали несколько изображений как образцы для обучения сети. Рис. 3 (см. третью сторону обложки) показывает исходное изображение  $512 \times 512$  elaine.bmp (не входит в группу обучающих изображений) и его сжатое изображение. Коэффициент PSNR получается 28,8698 дБ.

В таблице показываются результаты PSNR сжатия разных изображений данным алгоритмом, стандартной сетью Кохонена и JPEG при коэффициенте сжатия = 25,6.

## Заключение

Представленный алгоритм разрешает сжимать изображения с большим коэффициентом сжатия и достаточным хорошим качеством. Нейронные сети могут быть легко распараллелены, что позволяет сократить время вычисления и дает возможность сжимать изображения в реальном времени. В дальнейшем планируется реализовать структуру предполагаемой сети на базе графической карты NVIDIA.

## Список литературы

1. Галушкин А. И. Теория нейронных сетей. М.: Радиотехника, 2000.
2. Балухто А. Н., Галушкин А. И., Ковальчук Д. В., Назаров Л. Е., Томашевич Н. С. Нейрокомпьютеры в прикладных задачах обработки изображений. М.: Радиотехника, 2003.
3. Hadi Veisi, Mansour Jamzad. A Complexity-Based Approach in Image Compression using Neural Networks // International Journal of Signal Processing, 2009. Vol. 5.2.
4. Durai S. A., Saro A., Phil M. A Improved Image Compression approach with Self-Organizing Feature Maps using Cumulative Distribution Function // GVIP Journal. 2006. Vol. 6, N 2.
5. Luttrell S. P. Image compression on using multilayer neural network // Pattern recognition letters. 1989. Vol. 10. N 1.
6. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиватора, сжатие изображений и видео. М.: ДИАЛОГ-МИФИ, 2003.
7. Jung-Hua Wang, Ker-Jiang Gou. Image compression using wavelet transform and Self-development Neural Network // IEEE International Conference on Systems, Man, and Cybernetics. 1998.

# CONTENTS

## Vasenin V. A., Shundeev A. S. Evolution of the Grid Technology . . . . . 2

Architecture and concept of GRID was developed in the period 2001—2006. This period, associated with rapid technological development of Big Web services, laid the foundation architecture of the GRID. After 2006, a number of alternative approaches to the construction of distributed software systems (such as RESTful Web services) and to solve problems that traditionally belong to the field of GRID (cloud computing) emerged. In this regard, current studies are aimed at analysis of the history and prospects of the GRID technology. The results of such a work are presented in this article.

**Keywords:** grid, cloud computing, middleware

## Bogatyrev V. A. Failure-Safe Clusters the Duplicated Computer Complexes . . . . . 9

Variants of association of the duplicated computer complexes in cluster systems with the organisation of the duplicated calculations taking into account structure influence cluster and the organisations of an inter-complex exchange for reliability, fault tolerance and average time of stay of inquiries to system are considered.

**Keywords:** the fault tolerance, the duplicated complex, reliability, cluster

- Vladimirsky E. I., Ismailov B. I.** *Synchronization in Control of Chaotic Systems* . . . . . 16
- In work the approach to a choice of structure any spatially — is distributed of chaotic system is offered. The control of such system and identification of the latent transfer of the information is realized by a method of synchronization in the terms of nonlinear recurrence analysis.
- Keywords:** spatially-distributed system, percolation, synchronization, nonlinear recurrence analysis, Henon, Chirikov and Ikeda maps, latent transfer of the information
- Zack Yu. A.** *Adaptive Methods of Global Random Search in Solving a Class of Problems in Multiple Limited* . . 20
- Studied the properties of admissible plans for a sufficiently broad class of quasi-monotone objective function and constraints of the problem and put the lower limit of estimates of optimal solutions. Proposal 2 pozvoliyayut algorithm based on exclusion of areas that do not contain valid and optimal solutions, to reduce search costs, increase efficiency and improve quality assessment of the result.
- Keywords:** global random search, multiextremal problems with constraints, quasi-monotone functions
- Malykhin V. M., Malychina G. F., Merkusheva A. V.** *Regularization Method for Increasing the Quality of Image Restoration* . . . . . 28
- The registered image has imperfections in recording and transmission medium. They result in the defects associated with some forms of blur and noise. Reduction of these defects is termed as image reconstruction (IR) and is realized on the minimization for functional that represent integral square error of registered image relatively to original image. As IR procedure is the inverse problem that is unstable to imprecision of data (i. e. ill-conditioned problem) for these types of cases there are applied methods of regularization. For IR the using of such method results in the smoothing if image and (that especially is non-permissible) diminish the sharpness of local edges of image. The special form of regularization is considered that is selective to orientation of image intensity gradient so that it significantly diminishes the smoothing for the areas of contrast edges in image.
- Keywords:** image, degradation model, restoration criteria, regularization functional, intensity, orientation, adaptation
- Mandrikova O. V., Polozov Yu. A.** *Criteria of a Choice Wavelet-Function in Problems of Approximation of Natural Time Numbers Difficult Structure* . . . . . 31
- In the given work criteria of a choice of wavelet-function are defined and the scheme of transformation of the time number is constructed, allowing to minimize number approximating composed and an approximation error. Using the given technology, the data of critical frequency of an ionosphere is processed and communication between the processes proceeding in an ionosphere and in a lithosphere is revealed.
- Keywords:** wavelet-transformation, time numbers of difficult structure, ionosphere parameters, abnormal features
- Kukhareenko B. G., Ponomarev D. I.** *Use of the Prony Method for Time-Scale Estimation under Discovering Time-Series Patterns* . . . . . 37
- Single-frequency approximation of time-series by the Prony method estimates its characteristic period — a time-scale for finding its approximately matching subsequences (patterns). Under the time-scale, the approximately matching subsequences are identified with pattern discovering algorithms in use. As example, records of a telerobotic manipulator with precise accelerometer are under study. Patterns of the manipulator control signal represent voluntary movements — gestures of operator hand.
- Keywords:** time-series, pattern, motif, approximation, the Prony method, motif discovering algorithm, telerobotic manipulator
- Brailovskiy I. V., Solomeshch N. A.** *Videocodec Quality Modeling* . . . . . 42
- This work is dedicated to Video Encoding Quality estimation and modeling. A new two dimensional non-reference Video Encoding Quality metric is proposed in this paper. Based on proposed model classification of source video streams is discussed.
- Keywords:** visual quality metrics, video encoding, rate distortion, bitrate
- Senotrusov I. A., Neshkov P. A.** *Analysis and Filling Data Gaps in the Databases of Information-Analytical Systems* . . . . . 49
- The authors considers the problem of improving the monitoring of socio-economic processes in regions of the Russian Federation. As a methodology supports the process of filling in missing data in the information fund of information-analytical systems. In the article is considered using the basic algorithm ZET fill the gaps.
- Keywords:** information-analytical system, prediction, ZET algorithm to fill the gaps, competent rows and columns

**Vaschenko G. V., Kapulin D. V.** *Integration Capabilities of Business Process Models and ERP-Systems* . . . . .52

The universal solution based on 1C: Enterprise 8.1 platform for model import from ARIS to ERP-system is presented. The use of the approach provides to support and update business processes without changing configuration of the ERP-system.

**Keywords:** process approach, design of information systems, business process modeling, ARIS

**Dimitrienko Yu. I., Dimitrienko O. Yu.** *Continual Modeling of Economic Data Cluster Dynamics in the Presence of External Crisis Influences* . . . . .55

The article suggests the model of external forces influence on the buying behaviour of buyers' clusters for the model of mass sales market dynamics, which is based on the buyers clustering in the multidimensional space and on the introduction of the continuum model of clusters motion. The two main cases for the market are presented: stable and nonstable buying processes. The example of the numerical calculation of the external forces influence parameters is given, and the comparison of forecast results with the experimental data and the results obtained with the popular forecasting methods is executed.

**Keywords:** mathematical modeling, market influence, external factors, multidimensional space, motion equation, buyers clusters

**Myshev A. V.** *Information Model of Neural Network in Technologies of Computing Intelligence and Forms Realization of Computing* . . . . .63

The information model of neural networks for developing technologies of computational intelligence in the form of virtual model of the active memory of a computer system are considered. Formalism of the model describes the neural network processor patterns neuron as global and local networks of dynamically interacting of information elements with a set of duplex input and output information links. Information dynamics of interaction between the elements of such networks is defined as processes with local interaction.

**Keywords:** information model, neural networks processor patterns, model of the active memory, computing intelligence, technologies virtualization, information dynamics

**Kulik S. D., Nikonets D. A.** *Using Neural Networks for Writer Sex and Age Identification* . . . . .70

The paper deals with the application of the neural networks for the solution of the problems of the forensic handwriting expertise, especially writer sex and age recognition. It is short described modern forensic techniques and the software for writer sex and age identification. Formation of reference database of handwritten samples and of handwritten letters features, used as characteristics of handwriting is considered. Authors describe working out of the neural networks used for decision-making for sex and age identification of the executor of the handwritten text.

**Keywords:** neural networks, handwriting, forensic handwriting expertise

**Nguyen Viet Hung.** *Image Compression Using Neural Network in Transform Domain* . . . . .76

The neural approach for image compression issue is considered in this paper — neural algorithm on the basis of Kohonen's network in wavelet's transform domain. The computing experiment results have been produced.

**Keywords:** image compression, neural network, Kohonen's network, wavelet transform

---

---

**Адрес редакции:**

107076, Москва, Стромьинский пер., 4

Телефон редакции журнала **(499) 269-5510**

E-mail: [it@novtex.ru](mailto:it@novtex.ru)

Дизайнер *Т.Н. Погорелова*. Технический редактор *Е.В. Конова*.

Корректор *М.Г. Джавадян*.

Сдано в набор 05.10.2011. Подписано в печать 20.12.2011. Формат 60×88 1/8. Бумага офсетная. Печать офсетная.

Усл. печ. л. 9,8. Уч.-изд. л. 10,87. Заказ ИТ112. 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.

Оригинал-макет ООО "Авансед солюшнз". Отпечатано в ООО "Авансед солюшнз".

105120, г. Москва, ул. Нижняя Сыромятническая, д. 5/7, стр. 2, офис 2.