

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

11(183)  
2011

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ

Издательство "Новые технологии"

## СОДЕРЖАНИЕ

### ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

- Журавлева Л. М. Закон Мура и развитие устройств обработки и передачи информации . . . . . 2  
Саак А. Э. Алгоритмы диспетчеризации в Grid-системах на основе квадратичной типизации массивов заявок . . . . . 9

### ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

- Леденева Т. М., Нгуен Нгок Хуи. Влияние функции подобия на результаты нечеткой классификации . . . . . 14  
Савченко А. В. Трехпороговая система для распознавания изображений на основе метода направленного перебора . . . . . 21  
Якимов В. Н., Мошков И. С. Методика анализа сложных составных терминов в тексте на естественном языке . . . . . 26

### МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Бояринов Ю. Г., Борисов В. В. Анализ систем и процессов на основе нечетких полумарковских моделей . . . . . 31  
Кузьмина И. А. Генетические методы синтеза сетей энергоснабжения . . . . . 36

### ГЕОИНФОРМАЦИОННЫЕ СИСТЕМЫ

- Ефанов Д. В., Ибрагимов А. М. Динамические условные знаки и способы их реализации в современных геоинформационных системах . . . . . 40  
Архипова О. Е., Бойко В. В., Ковалева Г. В., Москаленко В. А., Тарасова Т. Т. Методические подходы к оценке природных и антропогенных изменений Азово-Черноморского побережья . . . . . 44

### ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ СИСТЕМЫ

- Малыхина Г. Ф. Статистическая модель для группы параметров, контролируемых в интеллектуальной измерительной системе . . . . . 49

Журнал в журнале

### НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ

- Четырбоцкий А. Н. Параметрическая идентификация радиальных базисных функций нейронных сетей методами глобальной оптимизации . . . . . 54  
Потапов В. И. Модели и задачи оценки надежности нейронной системы при обучении и переобучении нейронной сети после отказов . . . . . 59  
Красовская Л. В. Имитация отработки действий на нечетких семантических сетях в алгоритмах самообучения интеллектуальных систем . . . . . 64  
Куравский Л. С., Мармалюк П. А., Панфилова А. С., Ушаков Д. В. Исследование факторных влияний на развитие психологических характеристик с применением нового подхода к оценке адекватности моделей наблюдениям . . . . . 67  
Contents . . . . . 78  
Приложение. Буянкин В. М., Гоменюк С. М., Карпенко А. П., Литун Т. О. Нейросетевая идентификация и управление некоторыми механизмами параллельной кинематики

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 621.38

**Л. М. Журавлева**, канд. техн. наук, доц.,  
Московский государственный университет  
путей сообщения (МИИТ),  
e-mail: zhlubov@mail.ru

## Закон Мура и развитие устройств обработки и передачи информации

*Анализируются результаты развития элементной базы электроники и оптоэлектроники, приводятся причины расхождений с прогнозами по "закону Мура", рассматриваются перспективы совершенствования устройств обработки и передачи информации.*

**Ключевые слова:** прогнозы, скорость, микропроцессоры, оптическая связь

### Значение эмпирического закона Мура

Постоянный рост количества информации, которую необходимо обрабатывать и передавать с помощью компьютерной техники и систем связи, заставляют задуматься о будущем устройств обработки и передачи информации. Открытые в 1965 г. Гордоном Муром эмпирические законы возрастания числа транзисторов на одном кристалле и повышения расходов на строительство заводов свидетельствуют о скорости количественных и качественных изменений в области микроэлектроники [1]. Первоначальный вид двух "законов Мура" представлен на рис 1. 1-й закон Мура касается размеров транзисторов, 2-й закон — стоимости производства. Как показала практика, экспоненциальный характер этих законов был обусловлен непрерывно увели-

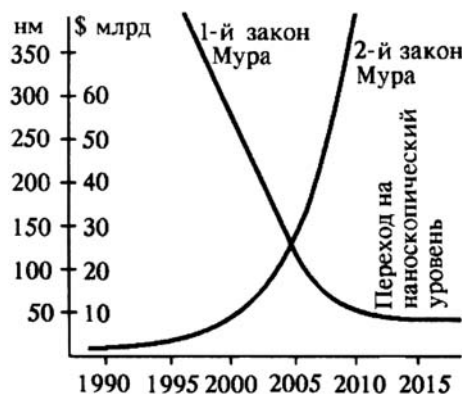


Рис. 1. Законы Мура [1]

чивающейся потребностью в повышении производительности вычислительных устройств и ростом затрат на новые технологии. Наиболее известен первый закон, согласно которому число транзисторов в одной микросхеме должно было удваиваться каждый год [2]. Хотя временные параметры удвоения "закона Мура" несколько раз корректировались, действие его продолжается. Интерес к нему не угасает, потому что его считают своеобразным "хронометром полупроводниковой технологии". Прогнозы, сделанные с помощью "закона Мура", становятся на определенном этапе руководящим принципом развития не только электронной промышленности, но и различных информационных технологий обработки и передачи информации.

Более точную математическую трактовку первого "закона Мура" можно представить следующим выражением [3]:

$$N(y) = N_0 2^{\frac{y}{\mu}}, \quad (1)$$

где  $N_0$  — число транзисторов на кристалле в нулевой год, условно принятый за начало отсчета;  $N(y)$  — число транзисторов на кристалле спустя  $y$  лет;  $\mu$  — срок (в годах и долях года), за который число транзисторов возрастает вдвое.

Под символом  $N(y)$  можно подразумевать различные характеристики, например, число ячеек в устройствах памяти, быстродействие микропроцессоров, число каналов и пропускную способность в системах связи, скорость передачи сигналов и т. д. Заданием значений  $N_0$  и  $\mu$  можно перемещать точку отсчета периода прогнозирования выбранного параметра, подчиняющегося "закону Мура". В этом заключается универсальность и значение этого закона для различных областей электроники и оптоэлектроники.

Для составления прогноза на основании соответствующей статистики необходимо решить математическую задачу нелинейной регрессии [3]. Так, анализ динамики роста числа транзисторов в микропроцессорах позволил уточнить в соответствии с формулой (1) первый "закон Мура" (рис. 2). Полученная зависимость характеризуется тремя отрезками прямых в логарифмическом масштабе с параметрами  $\mu = 2$  (с 1971 по 1993 г.),  $\mu = 4$  (с 1993 по 1999 г.),  $\mu = 0,6$  (с 1999 по 2002 г.) [3].

Временные этапы коррекции связаны с действием всеобщего закона перехода количества в качество. Это легко объяснить с позиций непрерывного роста потребностей в повышении производительности

вычислительной техники. Быстродействие микропроцессоров зависит от числа транзисторов в микросхемах и пропорционального уменьшения размеров элементной базы. Для изготовления таких устройств постоянно требуется внедрение новых технологий производства. Именно совершенствование технологического процесса промышленного производства интегральных схем гарантировало непрерывное увеличение числа транзисторов на кристалле. Технологии производства развивались в несколько этапов. Началом можно считать 1965 г. (через шесть лет после изобретения первой интегральной схемы). В то время технология производства микросхем позволяла интегрировать в одном чипе порядка 30 транзисторов [2]. В дальнейшем наметился быстрый рост числа транзисторов на кристалле. Предлагались все новые типы микросхем (доминировали три вида ТТЛ, КМОП, ЭСЛ), которые отличались друг от друга экономичностью, потреблением энергии и быстродействием. Этот период, когда удвоение числа транзисторов происходило за два года, продолжался примерно до 1993 г. Со временем сокращение размеров транзисторов привело к тому, что элементная база становилась не только высокоскоростной, но и более дешевой, мощной и доступной для населения [2]. Поэтому еще одним немаловажным следствием "законов Мура" стало стремительное падение цен в расчете на один транзистор и соответственно непрерывный рост покупательной способности. Уже в начале 2000 г. за 1 долл. можно было купить 1 млн транзисторов [2]. Несмотря на то, что второй "закон Мура" предсказывал резкое повышение стоимости постройки заводов по изготовлению чипов (см. рис. 1), производительность новых технологий привела к быстрому снижению себестоимости и экспоненциальному росту экономической эффективности. Чтобы оценить темпы роста эффективности производства вычислительной техники, в журнале "В мире науки" № 10 за 1983 год было приведено интересное сравнение: "Если бы авиационная промышленность в последние 25 лет развивалась столь стремительно, то сейчас самолет Boeing 767 стоил бы 500 долл. и совершал облет земного шара за 20 мин, затрачивая при этом пять галлонов ( $\approx 18,9$  л) топлива. Приведенные цифры весьма точно отражают снижение стоимости, рост быстродействия микропроцессоров".

Таким образом, первый период совершенствования элементной базы микроэлектроники длился чуть более 20 лет. Этот путь от процессоров 4004 до Pentium соответствовал эволюционному развитию технологии производства [3]. К концу этого периода (начало 90-х годов прошлого века) тактовая частота удвоилась за три года (с 25 до 50 МГц) [2].

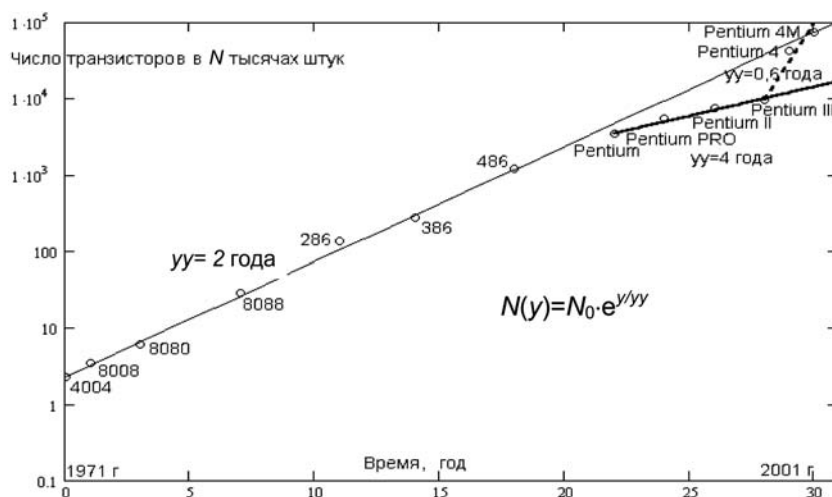


Рис. 2. Закон Мура в логарифмическом масштабе [3]

С появлением новых типов процессоров архитектура и алгоритмы работы сильно усложнились. Требовалось дальнейшее повышение числа транзисторов на кристалле. Поэтому когда число транзисторов превысило один из первых технологических пределов в 1 млн шт., появился устойчивый спад в темпах роста числа транзисторов на кристалле (см. рис. 2). Начался второй этап (с 1993 по 1999 г.), который был связан с преодолением технологических трудностей. Для производства высокоскоростных устройств необходимо было существенное снижение (на порядок) значений промышленных стандартов ввиду уменьшения геометрических размеров элементной базы [4]. Исторически сложилось так, что основные усилия исследователей и разработчиков были направлены на повышение разрешающей способности литографических методов. К этому периоду приближались физические возможности обычной оптической литографии [5]. Эта технология уже не удовлетворяла требованиям производства приборов с повышенным быстродействием по точности изготовления устройств. Поэтому в 80-е годы XX века были начаты разработки новых промышленных технологий (молекулярно-пучковой эпитаксии и газовой эпитаксии на основе металлоорганических соединений). Однако эти методы производства пока еще оставались дорогими и малопродуктивными [5]. Более перспективными с точки зрения массового производства оказались новые литографические технологии на основе сверхжесткого ультрафиолетового излучения [2].

Эти технологии были внедрены к 2000 г., что привело к значительному увеличению числа транзисторов в микросхеме в течение первых трех лет (см. рис. 2). В это время тактовая частота процессоров повышалась со скоростью 25 МГц в неделю [2]. Так, к 2004 г. тактовая частота новейших микросхем двухъядерных процессоров для серверов компаний Intel и Advanced Micro Devices достигла значения 2,5...3 ГГц [6].

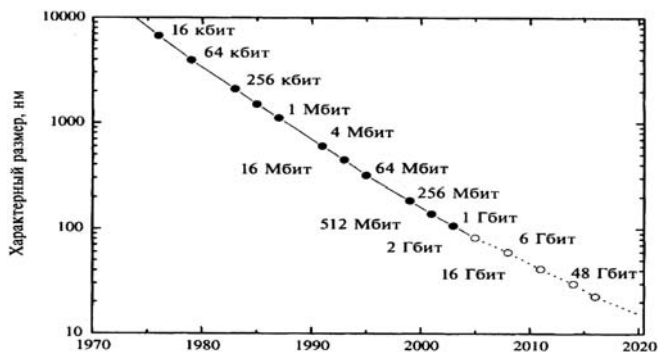


Рис. 3. Зависимость плотности записи (числа битов на чип) для кремниевых ОЗУ от размеров микросхемы [5]

Влияние технологий можно проследить также на примере динамики изменения геометрических размеров элементов кремниевых чипов ОЗУ (оперативных запоминающих устройств) типа *DRAM* и плотности записи информации (рис. 3) [5]. Собранный статистика за более чем двадцатилетний период позволила сделать по формуле (1) прогнозы до 2015 г. Однако уже к концу первого десятилетия XXI века началось замедление процесса уменьшения размеров элементной базы электроники, что потребовало пересмотра прогнозов. Начался этап спада роста быстродействия микропроцессоров. Сколько по времени продлится этот период, и когда начнется очередной резкий скачок производительности вычислительной техники, зависит от внедрения научных открытий в области новых материалов и успехов в нанотехнологиях. Можно предположить на основании приведенных выше цифр, что новый цикл (рост и спад увеличения быстродействия) будет примерно в 2 раза короче. Если предыдущий цикл (см. рис. 2) длился 28 лет (22 + 6), то наступивший цикл закончится приблизительно к 2014 г. Возможно, что уже к концу этого десятилетия мы станем свидетелями появления первых квантовых компьютеров.

### Влияние "закона Мура" на развитие оптической связи

Развитие элементной базы вычислительной техники определяет темпы роста скорости передачи информации в цифровых телекоммуникационных системах. Среди них наиболее перспективными являются волоконно-оптические системы передачи (ВОСП). Поэтому скачкообразность развития, прослеживаемая на примере анализа "закона Мура" для электроники, характерна и для оптоэлектроники (элементной базы ВОСП).

Это объясняется тем, что скорости компьютерной обработки информации и обмена информацией по компьютерным сетям должны соответствовать друг другу. Кроме того, многие информационные технологии в оптической связи и радиосвязи невозможны без использования высокоскоростных процессоров для реализации, например, сетей с переменной конфигурацией каналов и поддержания

различных протоколов обмена [2]. Поэтому история развития телекоммуникационных систем тесно связана с историей развития микропроцессоров.

Так, возникновение в конце 60-х и начале 70-х годов прошлого века вычислительных центров привело к необходимости строительства локальных, а затем региональных сетей передачи данных *Ethernet* в пределах одного ведомства. Затем на повестку дня была поставлена задача межведомственного обмена информацией и создания единой сети передачи данных. Создание в 1971 г. первого микропроцессора компании *Intel*, увеличение быстродействия и емкости памяти персональных компьютеров (ПК) привело к быстрому росту числа пользователей сетей передачи данных. В этот период актуальной проблемой стало повышение дальности, пропускной способности и строительство новых сетей. Еще в начале 80-х годов прошлого века широкое внедрение цифровых методов передачи информации (голоса и данных) позволило стереть различия между компьютерной и телефонной технологиями. Информационная сеть становится единой системой. В настоящее время подавляющая часть передаваемого трафика является данными компьютерных сетей (99 %) и только 1 % представляет речь [6].

В конце 80-х годов прошлого века для обмена информацией с помощью компьютерных и волоконно-оптических технологий была создана глобальная сеть *Internet*. Пропускная способность этой сети повышалась в соответствии с ростом быстродействия микропроцессорной техники. Так, например, с 1995 по 1999 год трафик *Internet* увеличился в 30 раз [7]. Такой осязаемый скачок был следствием внедрения новых процессоров *Pentium* и повсеместного перехода на технологию *SDH* (синхронного временного уплотнения). Это позволило в среднем трафику ВОСП удваиваться в каждые два года (такие темпы сохранились до сих пор). При этом существенно снижалась стоимость передачи бита информации. В развитии оптических систем разработчики всегда стремились реализовать два принципа: "быстрее" и "дальше". Поэтому постоянно улучшались эксплуатационные характеристики ВОСП. В качестве примера можно привести введенную в тот период первую трансатлантическую волоконно-оптическую линию связи ТАТ-8, обеспечившую резкое увеличение объема передаваемой информации. Стоимость телефонного канала при этом уменьшилась в 100 раз: с 1 млн до 10 тыс. долл. [7]. Это стало возможным благодаря совершенствованию технологий изготовления оптического волокна, оптических усилителей, полупроводниковых лазеров и фотодетекторов. Быстрыми темпами повышалась пропускная способность ВОСП. Довольно длительный период в истории развития систем связи господствовало положение, при котором потребность в увеличении пропускной способности опережала возможности, предоставляемые средствами связи. Такая ситуация была стимулом развития сетей передачи

информации. Затем наметился спад спроса на услуги систем передачи данных. Он объяснялся тем, что предложения стали опережать спрос. Скорость роста пропускной способности ВОСП, между тем, продолжала увеличиваться (особенно резко с внедрением технологии волнового уплотнения). Предоставляемые возможности стали значительно превосходить потребности. Поэтому с начала 2000 года начался третий период развития ВОСП, нацеленный на ориентацию рынка на более широкополосные услуги связи.

Отмеченные выше этапы развития ВОСП связаны с определенными вехами в совершенствовании волоконно-оптических линий [7]:

1) появление в середине 70-х годов прошлого века полупроводниковых лазеров и многомодовых волоконных светодиодов с небольшим затуханием, работающих в первом окне прозрачности;

2) увеличение дальности за счет освоения второго окна прозрачности и борьба с межмодовой дисперсией;

3) выпуск с 1983 г. одномодового волоконно-оптического кабеля, борьба с хроматической дисперсией, использование регенераторов;

4) внедрение с конца 90-х годов прошлого века методов волнового уплотнения (*WDM*), оптических усилителей, освоение третьего окна прозрачности.

В результате значительно увеличилась дальность связи, пропускная способность и скорость передачи информации.

В середине 1990-х годов электронные методы временного уплотнения достигли для того времени своего предела (скорости передачи информации в 40 Гбит/с). Причем рост скорости и пропускной способности ВОСП постоянно опережал первоначальные прогнозы, сделанные на период с 1978 по 2000 год (рис. 4) [7]. Такое опережение можно объяснить появлением новых технологий производства элементной базы ВОСП. Если разделить кривую (рис. 4) на десятилетние отрезки и определить для них параметр  $u$ , то в пределах каждого периода можно получить по формуле (1) значения, близкие к реальным цифрам.

Следует отметить, что устойчивый рост скорости и пропускной способности ВОСП наблюдался и после 2000 года. Об этом свидетельствуют разработки ведущих компаний. Так, в 2002 г. компания Алкатель организовала передачу 320 спектральных каналов в двух диапазонах с канальной скоростью 10,66 Гбит/с [8]. Однако достигнутое благодаря системам плотного волнового уплотнения резкое повышение пропускной способности оказалось невостребованным. Так, в США 80-канальная система плотного волнового уплотнения, работающая с канальной скоростью 2,5 Гбит/с, имеет загрузку в среднем равную шести каналам [8]. Такая ситуация заставила изменить приоритеты развития ВОСП и направить усилия на повышение экономической эффективности, поиск клиентов, формирование нового рынка услуг [7].

Как показала практика, объем рынка телекоммуникационных услуг и оборудования во многом определяется развитием сетей доступа [6]. Для этого была разработана концепция "последней мили" или "волокно в дом", которая включала и модели расчета эффективности таких сетей. Так, оценка эффективности инвестиций при создании локальных сетей, сделанная Робертом Меткалфом, показала следующее: при линейных инвестициях в сети интернет-бизнеса отдача будет расти в геометрической прогрессии [9]. Этот вывод долгое время стимулировал непрерывный количественный рост сетей доступа: повышалось число пользователей и число сайтов. Венчурные инвесторы, предприниматели, инженеры назвали это явление "законом Меткалфа". Они были уверены, что полезность сети увеличивается в геометрической прогрессии, даже если число пользователей растет линейно. Однако, как и в случае с "законом Мура", со временем пришлось пересмотреть прогнозы. На практике оказалось, что далеко не все узлы сети будут устанавливать связи друг с другом. На самом деле, по мнению математика Оддыжко, ценность сети размером  $n$  изменяется по формуле  $n \log(n)$ . Данная формула выведена из "закона Ципфа", согласно которому каждый новый элемент системы имеет пропорционально меньшую ценность [9]. Это свидетельствует о том, что и "закон Меткалфа", и "закон Мура" не являются фундаментальными законами природы. Скорее это эмпирические правила, по которым можно делать прогнозы, их продолжительность зависит от более общих закономерностей развития общества.

Таким образом, после 2000 г. в связи с заметным спадом спроса на услуги телекоммуникационных компаний возникли определенные проблемы, которые заставили пересмотреть рынок предложений. Это привело к тому, что акцент научных поисков сместился на способы повышения скорости передачи по ВОСП [6, 7, 8, 10]. Если раньше нужно было обеспечить рост пропускной способности ВОСП (т. е. произведения числа волновых каналов на канальную скорость), то теперь требуются системы для передачи широкополосных сигналов, в которых главное — не число каналов, а высокая скорость переда-



Рис. 4. Прогнозы повышения пропускной способности ВОСП [7]

чи информации. Потребность в таких высокоскоростных системах увеличивается с течением времени. Это связано с необходимостью передачи в цифровом виде широкополосных сигналов видеоизображения высокой четкости, например, для прямой трансляции оперных спектаклей ведущих театров мира в формате 3D [10]. Поэтому вопросы повышения быстродействия актуальны не только для микропроцессоров, но и для устройств оптической связи.

Таким образом, общей тенденцией развития оптических технологий в настоящее время является создание широкополосных каналов. Как считают специалисты, клиенты должны иметь возможность не только слышать, но и видеть с хорошим качеством. По прогнозам российских аналитиков, если сейчас среднестатистическое потребление информации в расчете на одного пользователя сети *Internet* составляет более 2 Гбайт в месяц, то в ближайшем будущем эта цифра увеличится до 200 Гбайт и более. В наступающей эре цифрового телевидения высокого разрешения *HDTV* потребление количества информации может возрасти до 1 Тбайт в месяц [6]. Этот прогресс не в последнюю очередь обязан быстрому развитию двух связанных информационных технологий: *VoIP* (передача видео поверх *IP*-сети) и *VoD* (доставка видео по требованию). Сверхбыстрые системы передачи (скорость в одном канале более 40 Гбит/с) — это следствие требований клиентов все большей полосы пропускания. Поэтому уже тестируются мультипроцессоры *SDH* с агрегатным потоком 160 Гбит/с и коммутаторы *Ethernet* со скоростями переключения 100 Гбит/с. Для их промышленной реализации приходится совершенствовать информационные технологии. К ним относятся специальные форматы оптической модуляции и кодирования на передающей стороне, специальные способы демодуляции/декодирования на приемной стороне, а также усовершенствованные методы компенсации дисперсии, в том числе электронные [6].

### **Причины замедления роста быстродействия элементной базы**

Отмеченные выше тенденции в развитии рынка услуг требуют дальнейшего повышения быстродействия в устройствах обработки и передачи информации. Предельная рабочая частота электронных и оптоэлектронных устройств зависит от размеров отдельных элементов. Она определяется временем пролета электрона, а следовательно, его скоростью и длиной затвора транзистора [5]. Внедряемые уникальные технологии позволяют создать уже сейчас тонкий оксидный слой затвора толщиной 1,2 нм (менее пяти атомных слоев) и достичь рабочей частоты транзистора 1 ТГц.

Хотя "закон Мура" оправдывает себя в течение почти 40 лет, многие скептически относятся к тому, что он будет действовать и в дальнейшем. Действительно, ведь не может происходить уменьшение

электронных приборов до бесконечности. Дальнейшее уменьшение размеров (увеличение скорости) должно замедлиться. Это объясняется ограничениями, обусловленными фундаментальными законами природы. Первый предел — релятивистский, который говорит о том, что скорость распространения сигнала не может превышать скорости света. Поэтому с учетом минимальной длины затвора транзистора (порядка значения, равного значению постоянной кристаллической решетки, т. е. меньше одного нанометра) легко можно подсчитать минимальное время пролета электрона (сотые доли фемтосекунды) и максимальную частоту. Второй предел связан с принципом неопределенности Гейзенберга, согласно которому в транзисторе с ростом частоты будет легко достигнут квантовый предел (значения произведения времени на энергию процесса ограничены постоянной Планка) [1, 2, 5].

Существуют также и технологические пределы, связанные с разбросом параметров, который невозможно контролировать при пропорциональном уменьшении размеров, а также энергопотреблением и тепловыделением. Например, чем больше транзисторов в микросхеме, тем больший ток он потребляет, а следовательно, увеличивается и тепловыделение. Если бы не наметился спад в "законе Мура", то уже после 2010 г. плотность выделяемой мощности, измеряемой в ваттах на квадратный сантиметр, достигла бы такого гигантского значения, которое можно было бы сравнить со значением тепловыделения на поверхности Солнца [2].

### **Новый этап развития элементной базы устройств обработки и передачи информации**

Для дальнейшего повышения быстродействия требуются новые архитектура и принципы работы устройств. Несмотря на последние достижения в области повышения тактовой частоты кремний-германиевых микросхем (в компании *IBM* совместно с учеными из Технологического университета шт. Джорджия уже создана микросхема с рекордной рабочей частотой более 500 ГГц при охлаждении до 4,5 К) [6], эра дешевых кремниевых приборов заканчивается. Существует мнение, что современная кремниевая технология не сможет преодолеть предел в 10 нм [5]. Поэтому для внедрения в производство электронных приборов типа одноэлектронных транзисторов, приборов с резонансным тунелированием [5], основанных на принципах квантовой физики, потребуются другие материалы и технологии. Так, большие надежды ученые связывают с новым материалом — графеном. Этот материал оказался стабильным, очень гибким, прочным и электропроводным. С его помощью надеются создать суперкомпьютер, где размеры вычислительных ячеек будут уменьшены в миллионы раз. Сейчас главное, найти технологии изготовления графена в промышленных

масштабах, т. е. научиться производить графеновую пленку размерами не менее  $10 \times 10$  дюймов [11]. Поэтому опять на первый план выходит развитие технологии производства элементной базы, которое во многом определяется новыми идеями и объемом вкладываемых материальных средств.

На сегодняшний день прогресс в области нанотехнологии оказался возможным, прежде всего, вследствие развития техники осаждения очень тонких пленок, образующих гетеропереходы. Существовавшие ранее стандартные методы получения тонких пленок (осаждение, напыление) не позволяли формировать гетеропереходы нужного качества [4, 5, 11]. Переход на массовое производство низкоразмерных (квантовых) устройств для высокоскоростной электронной элементной базы потребовал не только снижения себестоимости, но и снижения значений промышленных стандартов. Поэтому в промышленности сейчас используется новая нанолитографическая технология высокого разрешения, основанная на электронных или ионных пучках, рентгеновском излучении и др. [5]. На рис. 5 изображен график, иллюстрирующий этапы совершенствования современной литографии. На рисунке представлены значения прогнозируемого повышения разрешающей способности литографических методов и времени технологической готовности их внедрения по годам [4]. Следует отметить, что уже достигнутая точность изготовления размеров квантовых структур опережает прогнозы [5]. Однако существующие сейчас пределы нанотехнологий (порядка 10 нм) не позволяют пока перейти на качественно новый уровень работы вычислительной техники.

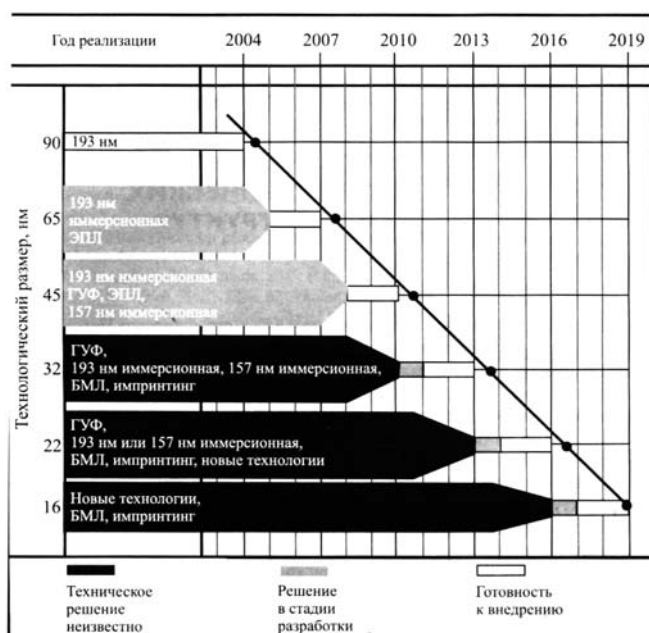


Рис. 5. Прогноз изменений разрешающей способности литографической нанотехнологии: ЭПЛ — электронная проекционная литография, БМЛ — безмасочные методы литографии, ГУФ — литография глубокого ультрафиолетового диапазона [4]

Одним из путей перехода на новый уровень является дальнейшее развитие устройств обработки информации с помощью молекулярной электроники [5]. В качестве элементной базы предполагается использовать отдельные молекулы (азобензола, гидрабензола и других органических соединений). Для кодирования различных состояний элементов можно выбирать разные конфигурации молекул (например, использование *цис*- и *транс*-форм, параллельное либо антипараллельное упорядочение спинов неспаренных электронов и др.). Изменение состояний может быть быстрым и не требовать значительных расходов энергии, им легко управлять внешними сигналами и регистрировать с помощью зондовых методов. Это приведет к удивительному прогрессу в миниатюризации, поскольку характерные размеры молекул на порядки меньше, чем размеры элементов существующих электронных схем. Последние достижения в области сканирующей атомно-силовой микроскопии позволяют манипулировать отдельными молекулами. Однако пока еще существуют проблемы в реализации этих проектов. Одной из них является задача соединения "вычислительных молекул" с внешним миром, т. е. крупными системами [5].

Разновидностью молекулярной электроники является изотоптроника. В качестве кодирующих признаков отдельных элементов предлагается использовать различия в оптоэлектронных характеристиках изотопов одного и того же вещества. Изотоптроника изучает влияние свойств ядер изотопов химического элемента на электронные и оптические характеристики низкоразмерных структур, которые могут иметь размеры меньше 1 нм [10, 12]. С помощью изотоптроники можно получать новые материалы, используя особенности изотопического строения вещества. Эти материалы могут быть востребованы для создания более совершенных устройств передачи, хранения и обработки информации в различных отраслях науки и техники. Основолагающей технологией изотоптроники является ядерная нанотехнология. С ее помощью можно влиять на изотопический состав химических элементов, и тем самым получать материалы с новыми электронными и оптическими свойствами. Так, с помощью нейтронного облучения можно изменять внутреннюю структуру кристаллической решетки вещества. Тем самым можно моделировать свойства материала для достижения требуемых параметров элементной базы.

### Новые принципы работы устройств обработки и передачи информации

Анализ динамики временных коррекций "закона Мура" позволяет сделать вывод о том, что мы на пороге очередного этапа "перехода количества в качество". Можно говорить, что в ближайшее десятилетие нас ждут кардинальные перемены: на смену

замедления роста быстродействия устройств электроники придет очередной скачок производительности вычислительной техники и скорости передачи информации. Технологический прогресс, который позволит уменьшить размеры элементной базы до размеров отдельных молекул или даже атомов, приведет к принципиальному изменению не только алгоритмов обработки и передачи информации, но и самого понятия информации. Постепенно на смену классическому представлению об устройствах обработки и передачи информации придут квантовые системы и квантовая теория информации.

Уже сейчас по всему миру развернулась настоящая научная гонка по созданию квантовых вычислительных устройств (компьютеров) и систем передачи (квантовой телепортации и квантовой оптической связи) [13, 14]. Следует отметить, что квантовую телепортацию и квантовую оптическую связь можно условно считать аналогами беспроводных и волоконно-оптических систем передачи классической информации.

Активный поиск способов реализации квантового процессора объясняется тем, что он по быстродействию несоизмеримо выше, чем самый скоростной современный процессор [14]. Какими же преимуществами обладают квантовые вычисления, можно пояснить на следующем примере. Несмотря на успехи вычислительной техники сегодняшнего дня, которая способна развивать производительность более  $10^{15}$  операций в секунду, требуются еще большие мощности. Так, для отыскания простых сомножителей 300-значного числа необходимо сделать  $5 \cdot 10^{24}$  шагов в секунду. При производительности современной вычислительной техники на это потребуется 150 тысяч лет. Для сравнения следует отметить, что квантовый компьютер должен будет сделать  $5 \cdot 10^{10}$  шагов и затратить на все операции меньше секунды [1].

Для обмена квантовой информацией большое значение имеет квантовая телепортация.

Квантовая телепортация — это удивительный процесс передачи квантовой информации на любые расстояния, который может оказать огромное влияние на алгоритмы управления различными объектами, передачи секретных кодов, квантовых вычислений и т. д.

Телепортация — это исчезновение объекта в одном месте и появление его в другом, пространственно удаленном от первого места. Ключевую роль в квантовой телепортации играют фотонные пары, находящиеся в так называемом сцепленном состоянии. Сцепленные фотонные пары имеют парадоксальные свойства, описанные в 1964 г. Дж. Беллом [13]. Так, измерение, проведенное над одной частицей (фотоном), мгновенно влияет на состояние другой, которая может быть удалена на произвольное расстояние от первой. Такие сцепленные состояния получают на основе спонтанного параметрического распада в кристаллах с квадратичной нели-

нейностью. Таким образом, с помощью фотонных пар реализуется квантовый канал связи. В конце XX века удалось осуществить квантовую телепортацию с помощью поляризационных сцепленных фотонов на расстояние более 10 км [13].

Не менее важна реализация идеи передачи квантовых состояний (отдельных фотонов света) с помощью оптической связи. Такая система необходима для реализации идей квантовой криптографии. Задача криптографии состоит в обеспечении секретности передачи информации между двумя объектами. Для этого используется секретный код, который должен быть известен объектам. Согласно К. Шеннону абсолютно секретная передача может быть тогда, когда код представляет собой большую последовательность "истинно случайных чисел" и используется только один раз [13]. На практике реализация такой системы, когда каждый раз при посылке сообщения, нужно создавать и передавать новую последовательность чисел (код), наталкивается на серьезные трудности. Изменить ситуацию может передача квантовых состояний, которая позволяет на приемной стороне обнаружить факт попытки "перехвата" секретного кода. Так, любое проникновение в систему передачи квантовой информации сопровождается воздействием на квантовый объект, который необратимо изменяет свое состояние. Это фиксируется на приемной стороне с помощью специального алгоритма "сравнения уровня ошибок" после измерения состояния квантового объекта. В конце 1990-х годов учеными была успешно проведена операция передачи квантового кода по дну Женевского озера по стандартному оптическому волокну на расстояние 67 км [13].

Таким образом, развитие квантовых устройств обработки и передачи информации является приоритетной задачей уже сегодняшнего дня. Решить ее можно с помощью внедрения новых нанотехнологий, позволяющих манипулировать отдельными молекулами и атомами. Следовательно, начинается новый этап развития устройств обработки и передачи информации на основе молекулярной электроники, где будут действовать свои эмпирические законы.

#### Список литературы

1. **Щука А. А.** Нанoeлектроника. М.: Физматкнига, 2007. 463 с.
2. **Пахомов С. О.** Экспансия закона Мура // КомпьютерПресс. 2003. № 1. URL: <http://www.compress.ru/article.aspx?id=9558&iid=404#begin>
3. **Дьяконов В. П.** Mathcad в математике: справочник. М.: Горячая линия—Телеком, 2007. 958 с.
4. **Нанотехнология** в электронике / под ред. Чаплыгина Ю. А. М.: Техносфера, 2005. 446 с.
5. **Маргинес-Дуарт Дж. М., Мартин-Палма Р. Дж., Агулло-Рueda Ф.** Нанотехнологии для микро- и оптоэлектроники. М.: Техносфера, 2007. 368 с.
6. **Слепов Н. Н.** Развитие оптической связи и волокон // Электроника: Наука. Технология. Бизнес. 2006. № 4. С. 62—65.
7. **Кунегин С. В.** Линии связи на основе волоконно-оптических кабелей. URL: <http://kunegin.narod.ru>. ноябрь, 2004.
8. **Скляр О. К.** Волоконно-оптические сети и системы связи. М.: Солон-Пресс, 2004. 261 с.



9. Закон Меткалфа (Metcalfе's Law). URL: <http://ecsocman.edu.ru/text/19281015/> Опул. на портале 24.10.2003.

10. Журавлева Л. М., Плеханов В. Г. Изотопическая нанотехнология низкоразмерных структур // Нано- и микросистемная техника. 2010. № 9. С. 8—13.

11. Нанотехнологии. Наноматериалы. Наносистемная техника / под ред. П. П. Мальцева. М.: Техносфера, 2008. 432 с.

12. Журавлева Л. М., Плеханов В. Г. Ядерная нанотехнология низкоразмерных изотопически-смешанных структур // Наноиндустрия. 2009. № 4. С. 28—30.

13. Килин С. Я. Квантовая информация // Успехи физических наук. 1999. Т. 169, № 5.

14. Журавлева Л. М., Плеханов В. Г. Изотопический квантовый процессор // Нано- и микросистемная техника. 2011. № 3. С. 46—54.

УДК 004.272.43

А. Э. Саак, канд. техн. наук, доц.,  
Технологический институт Южного федерального  
университета в г. Таганроге,  
e-mail: saak@tti.sfedu.ru

## Алгоритмы диспетчеризации в Grid-системах на основе квадратичной типизации массивов заявок

*Линейные и триодные полиэдры координатных ресурсных прямоугольников — заявок пользователей — предлагаются в качестве модели множественного обслуживания для задач Grid-технологии. Вводятся понятия быстрой и медленной монотонности изменения параметров полиэдральной модели. С помощью данного различия и других свойств массива определяется квадратичная типизация полиэдров планарных элементов. Вводятся целевые меры и многоцелевой критерий эвристики упорядочения указанных массивов в качестве технологии диспетчеризации вычислительных ресурсов Grid-технологий. Предлагаются алгоритмы рамочного перераспределения массивов на основе кольцевой, угловой и инверсной локализации в зависимости от квадратичной типизации.*

**Ключевые слова:** Grid-система, многопроцессорная вычислительная система, диспетчеризование, квадратичный тип массива требований пользователей, эвристика критериев симметризации ресурсной оболочки, ресурсной меры, заполненности оболочки заявками пользователей

### Введение

В работе [1] для множества заявок пользователей на компьютерное обслуживание в Grid-системах, многопроцессорных вычислительных системах (МВС) [3—7] намечена квадратичная классификация. В настоящей статье квадратичная типизация строится более полно на существенно новой методической основе. Квадратичная типизация дополняется понятием эвристики контролируемых упорядочений указанных массивов. Последняя заключается в согласовании производимых упорядочений, перераспределений массивов координатных ресурс-

ных прямоугольников с направленным изменением значений целевых мер — меры симметризации измерений оболочки, ресурсной меры оболочки, меры заполнения оболочки гранями массива [2], в сторону повышения эффективности функционирования компьютерного сервиса.

Намеченный аппарат используется в качестве базы технологии диспетчеризации множественным компьютерным обслуживанием, приводящим к перераспределению массивов координатных ресурсных прямоугольников с большими значениями протяженности и уровня по ресурсным рамкам МВС.

Указанное перераспределение имеет вид кольцевой локализации для массивов кругового типа, угловой локализации для гиперболических массивов и локализации на основе инверсного синтеза для массивов параболической типизации.

### 1. Постановка задачи

По отношению к триодным полиэдрам, имеющим существенные значения протяженности и уровней, ставится задача динамического упорядочения в целях локализации в рамках  $M \times M$  поля ресурсных распределений МВС согласно эвристике критериев симметризации ресурсной оболочки, ресурсной меры, заполненности оболочки гранями массива и некоторыми другими [1, 2]. Особенно отметим важный критерий минимизации числа рамок при многостадийном обслуживании.

При этом под эвристикой многоцелевого критерия динамического упорядочения планарного массива координатных ресурсных прямоугольников понимается направленное изменение массива граней по правилам аддитивности (неналожения), ориентированности (сохранения координатности) и целостности, с одной стороны, и надлежащей монотонности поведения целевых мер с некоторого индекса целочисленного алгоритма локализации, с другой стороны. При представлении заявки пользователя для обслуживания диспетчером центра Grid-технологий или операционной системы МВС координатным ресурсным прямоугольником горизонтальное и вертикальное измерения, соответственно, принимаются равными числу единиц ресурса процессоров и времени, требуемому для обработки.

Указанная эвристика динамического упорядочивания основывается на метрировании полиэдральных массивов и связанной с метрикой граней квадратичной типизацией полиэдралей на массивы кругового, гиперболического и параболического типов.

## 2. Классификация триодных полиэдралей

В основу классификации массивов линейных (рис. 1) и триодных полиэдралей заявок пользователей (рис. 2), соответственно, полагаем различие медленного и быстрого убывания уровней вертикальных слоев (рис. 3).

$$\text{Здесь } A(j_1) = \sum_{j_1'=0}^{j_1-1} a(j_1') \text{ — интегральная протя-$$

женность. Символом  $a(j_1) \times b_{j_1}(j_2)$  или  $\lfloor (a(j_1), b_{j_1}(j_2)) \rfloor$  обозначается заявка с индексацией  $(j_1, j_2)$ , требующая  $a(j_1)$  единиц процессоров и  $b_{j_1}(j_2)$  единиц времени.

Пусть вертикальные слои триодной полиэдралей упорядочены по убыванию суммарных уровней  $\sum_{j_2} b_{j_1}(j_2) = H(A(j_1)) \downarrow, j_1 \uparrow$ . Проведем хорду, соединяющую точку левого максимального и правого минимального уровней. При этом, если все вертикальные слои находятся под хордой, то убывание считаем быстрым, иначе — медленным. В дальнейшем медленное убывание уровней полагается в основу круговой и гиперболической квадратичной типизации, тогда как быстрое относится к параболическому квадратичному типу.

Линейную полиэдраль последовательных пар координатных ресурсных сравнимых прямоугольников  $a(j_1) \times b(j_1), \Delta a(j_1) \cdot \Delta b(j_1) \geq 0$  (рис. 4) относим к массивам кругового квадратичного типа при медленном убывании высот после соответствующего упорядочения  $b(j_1) \downarrow, j_1 \uparrow$ .

Линейную полиэдраль последовательных пар координатных ресурсных несравнимых прямоугольников  $a(j_1) \times b(j_1), \Delta a(j_1) \cdot \Delta b(j_1) < 0$  (рис. 5) относим к массивам гиперболического квадратичного типа при медленном убывании высот и медленном росте оснований граней  $b(j_1) \downarrow, a(j_1) \uparrow, j_1 \uparrow$ .

Линейную полиэдраль координатных ресурсных прямоугольников с быстрым убыванием высот  $a(j_1) \times b(j_1), b(j_1) \downarrow \downarrow, j_1 \uparrow$  относим к массивам параболического квадратичного типа при не более, чем медленном росте оснований  $a(j_1) \leq \tilde{a}(j_1), \tilde{a}(j_1) \uparrow, j_1 \uparrow$ .

Квадратичный тип триодных полиэдралей определяем переходом к линейным полиэдралам граней-оболочек вертикальных слоев (рис. 6).

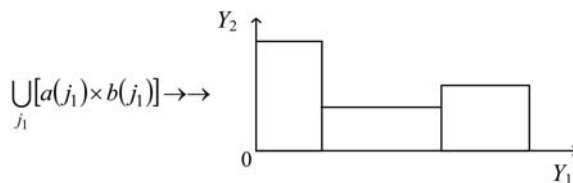


Рис. 1. Линейная полиэдраль заявок пользователей

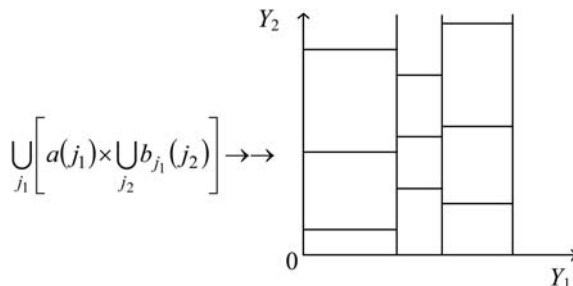


Рис. 2. Триодная полиэдраль заявок пользователей

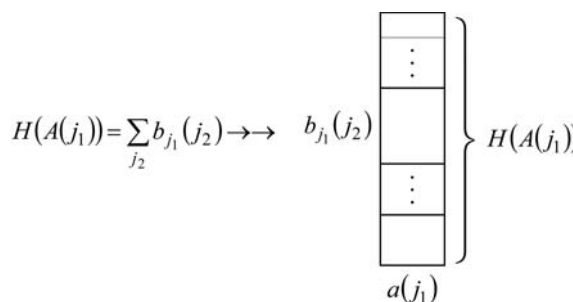


Рис. 3. Вертикальный слой заявок пользователей триодной полиэдралей

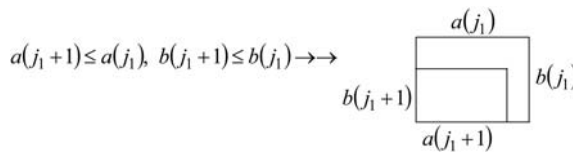


Рис. 4. Сравнимая пара ресурсных прямоугольников

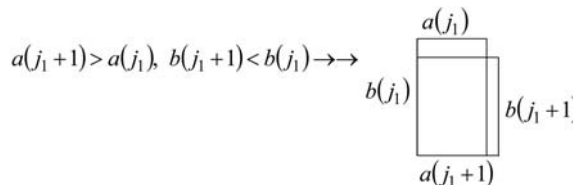


Рис. 5. Несравнимая пара ресурсных прямоугольников

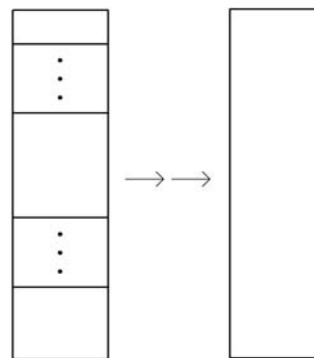


Рис. 6. Оболочка вертикального слоя триодной полиэдралей

### 3. Круговой тип триодных полиэдров

При медленном убывании интегральных уровней  $\sum_{j_2} b_{j_1}(j_2) = H(A(j_1)) \downarrow, j_1 \uparrow$ , ограничении оснований условием сравнимости граней-оболочек вертикальных слоев и медленном росте интегральных протяженностей  $\sum_{j_1'=0}^{j_1-1} a(j_1') = A(j_1) \uparrow, j_1 \uparrow$ , триодная полиэдраль имеет круговой квадратичный тип.

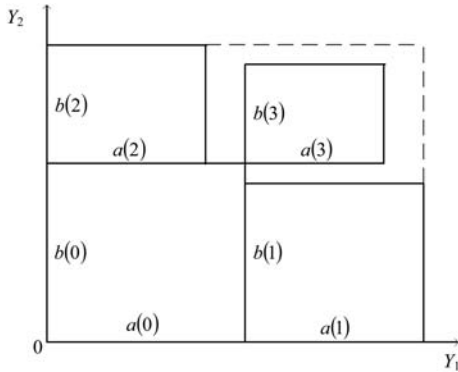


Рис. 7. Начальный шаг алгоритма кольцевой локализации

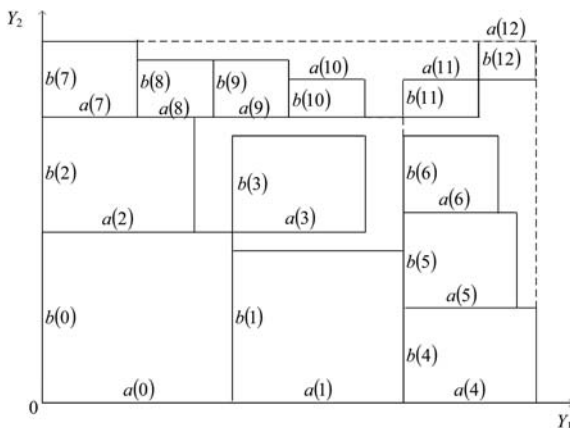


Рис. 8. Алгоритм кольцевой локализации

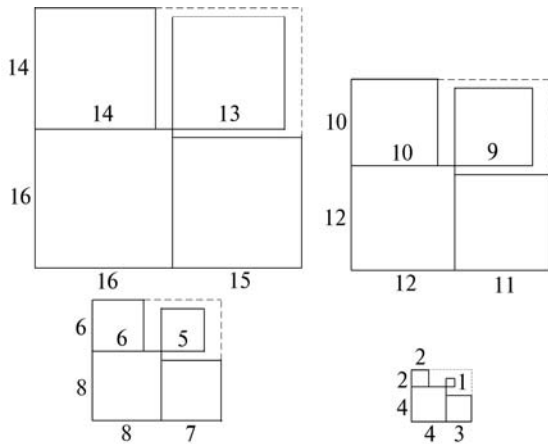


Рис. 9. Ресурсные блоки первого порядка кругового массива ресурсных квадратов

Триодную полиэдраль превращаем в линейную полиэдраль, принимая вертикальные слои  $\bigcup_{j_2} a(j_1) \times b_{j_1}(j_2) = a(j_1) \times \bigcup_{j_2} b_{j_1}(j_2)$  в качестве граней-блоков.

Укажем кольцевую локализацию линейной круговой полиэдраль координатных ресурсных прямоугольников, упорядоченных по вложению  $[(a(j_1 + 1), b(j_1 + 1))] \subseteq [(a(j_1), b(j_1))]$ , горизонтальной формы  $a(j_1) \geq b(j_1)$  и медленного убывания оснований  $a(j_1) \downarrow, j_1 \uparrow$ .

На первом шаге кольцевой локализации строим аддитивную графику с ресурсной оболочкой  $(a(0) + a(1)) \times (b(0) + b(2))$  (рис. 7). Здесь последующие за максимальным элементы суперпозируются справа, сверху и вершинно-диагонально к основному элементу, соответственно.

На втором шаге в качестве максимального элемента принимается указанная оболочка. Вдоль правой стороны оболочки вертикально суперпозируются последующие грани до наилучшего приближения уровня оболочки с недостатком. Затем над верхней стороной оболочки вдоль горизонтали на упомянутом уровне суперпозируются последовательно оставшиеся элементы до наилучшего приближения с недостатком значения протяженности оболочки. Наконец, с помощью вершинно-диагонального синтеза заполняем полученную новую ресурсную оболочку последующими гранями полиэдраль (рис. 8).

Введенный таким образом алгоритм кольцевой локализации повторяем до полного исчерпания массива.

Так, для модельного кругового массива ресурсных квадратов

$$(k - j_1) \times (k - j_1), j_1 = 0, 1, \dots, k - 1,$$

при  $k = 16$  имеем четыре блока первого порядка (рис. 9) в качестве ресурсных оболочек кольцевой локализации четверок последовательных квадратов с согласованными параметрами.

В итоге получаем локализацию первого порядка с последующим кольцевым синтезом второго по-

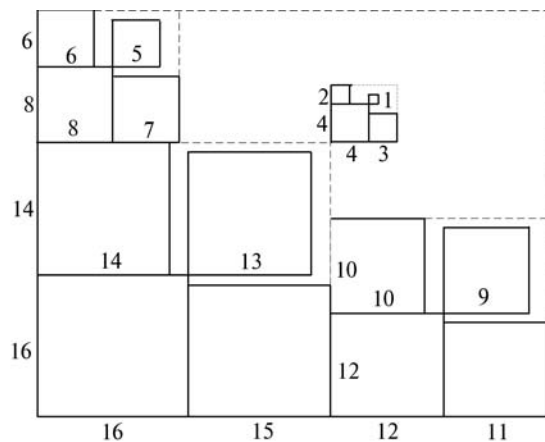


Рис. 10. Ресурсная оболочка модельной круговой линейной полиэдраль

рядка в ресурсную оболочку (рис. 10) с коэффициентом заполнения

$$\sum_{j_1=1}^k j_1^2 = \frac{k(k+1)(2k+1)}{6} = \frac{1496}{2376} = 0,63,$$

что при разбросе измерений от 1 до 16 можно считать удовлетворительным результатом.

#### 4. Гиперболический тип триодных полиэдралей

При медленном убывании интегральных уровней  $\sum_{j_2} b_{j_1}(j_2) = H(A(j_1)) \downarrow, j_1 \uparrow$ , медленно растущих оснований  $a(j_1) \uparrow, j_1 \uparrow$  и медленном росте интегральных протяженностей в форме оценки  $A(k) \leq H(0)$  протяженности через уровень триодная полиэдраль имеет гиперболический квадратичный тип.

Пусть имеется модельная гиперболическая линейная полиэдраль

$$j_1 \times (k - j_1), j_1 \in [1, k - 1] \subset Z^1,$$

из ресурсных прямоугольников с убывающими высотами  $(k - j_1) \downarrow, j_1 \uparrow$  и растущими основаниями  $j_1$ .

При этом падению уровня на единицу отвечает приращение основания на то же значение. Поэтому убывание высот следует считать медленным. Для локализации данного массива по критерию симметрии измерений ресурсной оболочки выделяем центральный элемент

$$j_1 = \left[ \frac{k}{2} \right] \rightarrow \left[ \frac{k}{2} \right] \times \left[ \frac{k}{2} \right]$$

квадратной формы и строим вертикаль

$$Y_1 = \max \left\{ \left[ \frac{k}{2} \right], \sum_{j_1=1}^{\left[ \frac{k}{2} \right]-1} j_1 \right\} = \max \left\{ \left[ \frac{k}{2} \right], \frac{\left[ \frac{k}{2} \right] \left( \left[ \frac{k}{2} \right] - 1 \right)}{2} \right\} = \frac{\left[ \frac{k}{2} \right] \left( \left[ \frac{k}{2} \right] - 1 \right)}{2}.$$

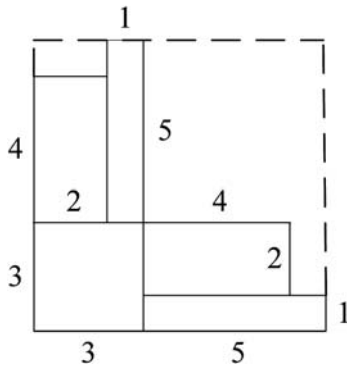


Рис. 11. Ресурсная оболочка модельной гиперболической линейной полиэдраль

На верхнем основании центрального элемента слева от вертикали суперпозируем горизонтально предшествующие элементы  $j_1 < \left[ \frac{k}{2} \right]$ . Вдоль вертикальной линии справа суперпозируем вертикально последующие элементы  $j_1 \in \left( \left[ \frac{k}{2} \right], k \right] \subset Z^1$ . Получаем аддитивную графику заданных ресурсных прямоугольников с ресурсной оболочкой

$$\left( \frac{\left[ \frac{k}{2} \right] \left( \left[ \frac{k}{2} \right] - 1 \right)}{2} + (k - 1) \right) \times \max \left\{ \left[ \frac{k}{2} \right] + (k - 1), \sum_{j_1=\left[ \frac{k}{2} \right]+1}^{k-1} (k - j_1) \right\}.$$

Так, при  $k = 6$  имеем ресурсные прямоугольники  $1 \times 5, 2 \times 4, 3 \times 3, 4 \times 2, 5 \times 1$  внутри ресурсной оболочки  $8 \times 8$  (рис. 11) с коэффициентом заполнения

$$\frac{5 + 8 + 9 + 8 + 5}{8 \times 8} = \frac{35}{64} = 0,55,$$

что при разбросе измерений от 1 до 5 можно считать удовлетворительным результатом.

#### 5. Параболический тип триодных полиэдралей

При быстром убывании интегральных уровней  $\sum_{j_2} b_{j_1}(j_2) = H(A(j_1)) \downarrow \downarrow, j_1 \uparrow$  и не более, чем медленном

росте оснований  $a(j_1) \leq \tilde{a}(j_1), \tilde{a}(j_1) \uparrow, j_1 \uparrow$ , триодная полиэдраль имеет параболический квадратичный тип.

Модельной параболической линейной полиэдралью служит массив координатных ресурсных прямоугольников  $j_1 \times (k - j_1)^2, j_1 \in [1, k - 1] \subset Z^1$  с быстрым убыванием высот, изображенный при  $k = 6$  на рис. 12.

Локализацию строим в виде координатной трапеции с протяженностью  $\sum_{j_1=1}^{k-1} j_1 = \frac{k(k-1)}{2}$ , меньшим

основанием 1 и большим основанием  $(k - 1)^2 + \frac{(k - 1)^2 - 1}{2} \approx (k - 1)^2$  для  $k \gg 1$ . В нашем случае

ресурсная оболочка имеет вид, приведенный на рис. 13.

Пусть имеется линейная полиэдраль из  $k = 2m$  модулей координатных ресурсных прямоугольников с быстро убывающими высотами в каждом модуле и идентичными координатными трапециями

ресурсных оболочек транспонировано-симметричных модулей индексации  $i'$ ,  $k - i'$  (рис. 14).

По совокупности указанных свойств  $k$ -модульная линейная полиэдраль координатных ресурсных прямоугольников относится к однородно-параболическому квадратичному типу. Для линейной  $k$ -модульной полиэдральной со свойством однородной параболичности применяем транспонированный синтез парных идентичных координатных трапеций — ресурсных оболочек параболических модулей (рис. 15).

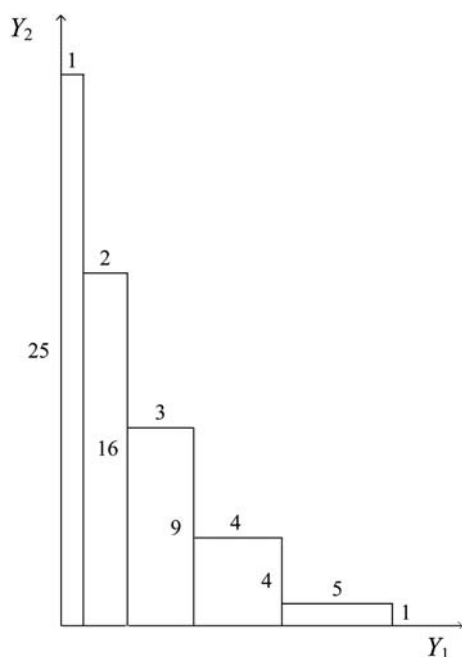


Рис. 12. Модельная параболическая линейная полиэдраль

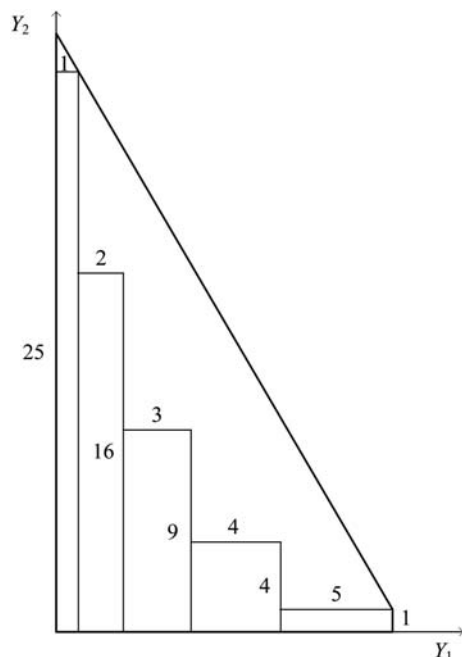


Рис. 13. Ресурсная оболочка модельной параболической линейной полиэдральной

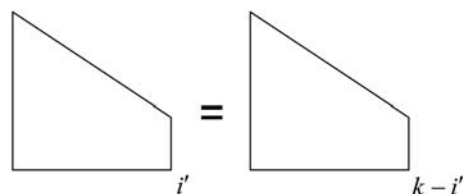


Рис. 14. Ресурсные оболочки транспонировано-симметричных модулей

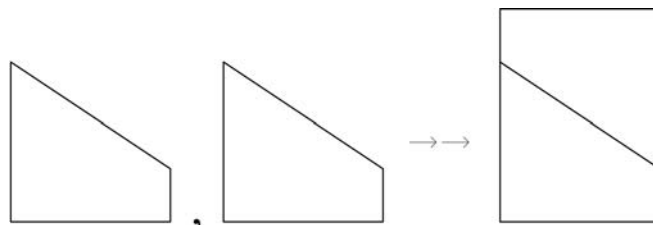


Рис. 15. Транспонированный синтез ресурсных оболочек параболических модулей

Полученные координатные блоки упорядочим в рамочную оболочку с помощью одного из указанных выше видов локализации.

### Заключение

Линейные и триодные полиэдры координатных ресурсных прямоугольников — заявок пользователей — могут служить адекватными моделями множественного компьютерного обслуживания и задач Grid-технологии. Анализ быстрой и медленной монотонности изменения параметров полиэдральной модели, наряду с квадратичной типизацией и эвристикой контролируемого изменения целевых мер многоцелевого критерия упорядочения протяженного массива в локальную часть координатного квадранта, позволяет осуществлять диспетчеризацию вычислительных ресурсов Grid-технологий.

Предложены алгоритмы рамочного перераспределения массивов линейных и триодных полиэдров большой протяженности на основе кольцевой, угловой и инверсной локализации в зависимости от квадратичной типизации массива. Указанные алгоритмы могут использоваться в диспетчере как МВС, так и центра Grid-технологий.

### Список литературы

1. Саак А. Э. Локально-оптимальные ресурсные распределения // Информационные технологии. 2011. № 2. С. 28—34.
2. Саак А. Э. Локально-оптимальный синтез расписаний для Grid-технологий // Информационные технологии. 2010. № 12. С. 16—20.
3. Барский А. Б. Параллельные информационные технологии. М.: ИНТУИТ; БИНОМ. Лаборатория знаний, 2007. 503 с.
4. Барский А. Б. Параллельные информационные технологии в основе Grid-системы // Информационные технологии. 2006. № 12. С. 54—60.
5. Хорошевский В. Г. Архитектура вычислительных систем. М.: Изд-во МГТУ им. Н. Э. Баумана, 2005. 512 с.
6. Воеводин В. В., Воеводин Вл. В. Параллельные вычисления. СПб.: БХВ-Петербург, 2002. 608 с.
7. Калиев И. А., Левин И. И., Семерников Е. А., Шмойлов В. И. Реконфигурируемые мультиконвейерные вычислительные структуры / Изд. 2-е, перераб. и доп. / Под общ. ред. И. А. Каляева. Ростов н/Д.: Изд-во ЮНЦ РАН, 2009. 344 с.

УДК 519.72

**Т. М. Леденева**, д-р техн. наук, зав. каф.,  
**Нгуен Нгок Хуи**, аспирант  
 Воронежский государственный университет,  
 e-mail: shashkin@amm.vsu.ru

## Влияние функции подобия на результаты нечеткой классификации

*Рассматриваются результаты исследования алгоритма нечеткой классификации [1], в котором для формирования отношения сходства используется функция подобия.*

**Ключевые слова:** нечеткие числа, функция подобия, нечеткое отношение эквивалентности, транзитивность, декомпозиционное дерево, нечеткая классификация

### Введение

Рассмотрим задачу нечеткой классификации в следующей постановке: пусть задано конечное множество объектов  $X = \{x_1, \dots, x_N\}$ , каждый из которых  $x_i$  может быть описан нечетким подмножеством  $A_i$  некоторого универсального множества  $U$ . Требуется определить транзитивно ближайшие подмножества, т. е. такие, которые попадут в один класс, если в качестве расстояния между ними использовать транзитивные расстояния, определенные с помощью различных композиций нечетких отношений.

Решение данной задачи осуществляется с помощью одного из двух вариантов алгоритма нечеткой классификации (рис. 1).

Каждый вариант обусловлен тем, какое отношение — сходства или несходства — формируется вначале. В работе [1] предлагается построить матрицу отношения несходства, определяя расстояние

(Хемминга или Евклида) между соответствующими нечеткими множествами. Переход к отношению различия осуществляется на основе  $(\min - \max)$ -транзитивного замыкания, что позволяет отношение несходства, обладающее свойствами антирефлексивности и симметричности, наделять еще и  $(\min - \max)$ -транзитивностью. Получаемое при этом отношение различия, по сути, задает матрицу транзитивных расстояний, а его декомпозиция позволяет построить классы объектов, расстояние между которыми не превосходит задаваемой в качестве порога величины уровня. В работе [2] рассматривались обе схемы и другие варианты транзитивности. Оценивалось влияние типа транзитивности на структуру декомпозиционного дерева.

Целью данной статьи является исследование результатов нечеткой классификации при условии, что на начальном этапе алгоритма строится отношение сходства с помощью функции подобия, а оценки объектов заданы в виде нечетких треугольных или трапециевидных чисел. Данная ситуация в наибольшей степени соответствует практическим задачам, в которых в качестве оценки объекта может использоваться нечеткое число, выражающее приближенную комплексную (обобщенную) оценку, возможно, по нескольким показателям, которые характеризуют свойства объектов рассматриваемого множества.

### Функция подобия

#### как инструмент сравнения нечетких чисел

Современные модели обработки информации в задачах принятия решений ориентированы на использование нечетких чисел, причем одна из базовых процедур — это их сравнение. В настоящее время существуют различные подходы к сравнению нечетких чисел. Нами использовался подход, основанный на функции подобия [3], которая, в свою очередь, определяется через мощность нечеткого множества

$$|A| = \int_{x \in \text{Supp}(A)} \mu_A(x) dx,$$

где  $\text{Supp}(A) = \{x: \mu_A(x) > 0\}$ ;  $\mu_A(x)$  — функция принадлежности нечеткого множества  $A$ .

Пусть  $F(U)$  — семейство нечетких подмножеств универсального множества  $U$ ,  $A_1, A_2 \in F(U)$  — не-

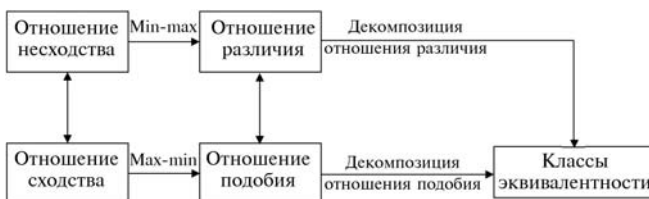


Рис. 1. Схема алгоритма нечеткой классификации

четкие множества, тогда функция подобия этих нечетких множеств определяется по формуле

$$S(A_1, A_2) = \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|} = \frac{|A_1 \cap A_2|}{|A_1| + |A_2| - |A_1 \cap A_2|}.$$

Среди основных свойств функции подобия отметим следующие:

- 1) если  $A_1 = A_2$ , то  $S(A_1, A_2) = 1$ ;
- 2) если  $A_1 \cap A_2 = \emptyset$ , то  $S(A_1, A_2) = 0$ ;
- 3) если  $A_1 \subset A_2$ , то  $A_1 \cap A_2 = A_1$  и тогда  $S(A_1, A_2) =$

$$= \frac{|A_1|}{|A_1| + |A_2| - |A_1|} = \frac{|A_1|}{|A_2|}.$$

Рассмотрим нечеткие трапециевидные числа  $T_1 = (a_1, d_1, e_1, b_1)$  и  $T_2 = (a_2, d_2, e_2, b_2)$ , считая при этом, что функция принадлежности нечеткого числа  $T = (a, d, e, b)$ , заданного параметрами  $a, d, e, b$ , имеет вид

$$T = T(x) = \begin{cases} \frac{x-a}{d-a}, & \text{если } a \leq x \leq d; \\ \frac{b-x}{b-e}, & \text{если } e \leq x \leq b; \\ 1, & \text{если } d \leq x \leq e; \\ 0, & \text{иначе.} \end{cases}$$

Для нечеткого числа  $T(x)$  с помощью простых вычислений получим

$$|T| = \int_a^b T(x) dx = \frac{b-a}{2} + \frac{e-d}{2} = \frac{w_0(T) + w_1(T)}{2},$$

т. е. мощность представляет собой среднее арифметическое между длинами ядра ( $w_0$ ) и носителя ( $w_1$ ).

В качестве примера рассмотрим частный случай взаимного расположения двух трапециевидных чисел  $T_1$  и  $T_2$ , когда  $a_1 \leq a_2 \leq b_2 \leq b_1$ ,  $d_1 \leq e_1 \leq d_2 \leq e_2$ . Графики функций принадлежности нечетких чисел пересекаются, как показано на рис. 2.

Заметим, что  $|T_1| + |T_2| = \frac{w_0 + w_1}{2}$ , где  $w_0 = w_0(T_1) + w_0(T_2)$ ,  $w_1 = w_1(T_1) + w_1(T_2)$ . Решая уравнение  $\frac{b_1 - x}{b_1 - e_1} = \frac{x - a_2}{d_2 - a_2}$ , получим  $h_1 = \frac{b_1 - a_2}{(b_1 - e_1) + (b_2 - e_2)}$ .

Аналогично, решая уравнение  $\frac{b_1 - x}{b_1 - e_1} = \frac{b_2 - x}{b_2 - e_2}$  для нахождения  $h_2$ , получим  $h_2 = \frac{b_2 - b_1}{(b_2 - e_2) - (b_1 - e_1)}$ .

Тогда  $|T_1 \cap T_2| = \frac{(b_1 - a_2)h_1}{2} - \frac{(b_1 - b_2)h_2}{2}$  или с учетом замены  $l_1 = (b_1 - a_2)h_1$ ,  $l_2 = (b_2 - b_1)h_2$  имеем  $|T_1 \cap T_2| = \frac{1}{2}(l_1 + l_2)$ .

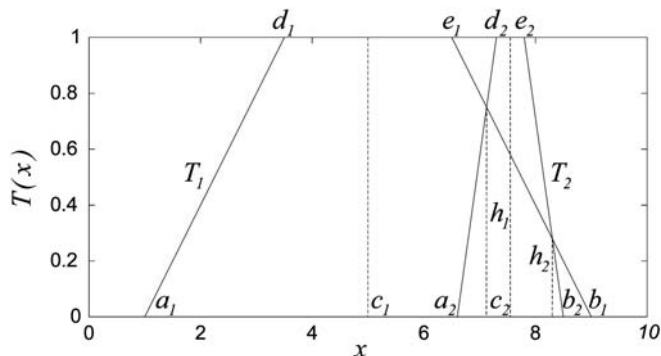


Рис. 2. Графики функций принадлежности нечетких чисел

Таблица 1

№	Условия параметров	Функция подобия
I	$a_1 \leq a_2 \leq b_2 \leq b_1$ , $d_1 \leq d_2 \leq e_2 \leq e_1$	$S(T_1, T_2) = \frac{b_2 - a_2 + e_2 - d_2}{b_1 - a_1 + e_1 - d_1}$
II	$a_1 \leq a_2 \leq b_2 \leq b_1$ , $d_1 \leq d_2 \leq e_1 \leq e_2$	$h = \frac{b_2 - b_1}{(b_2 - e_2) - (b_1 - e_1)}$ ; $l_1 = e_1 - d_2$ ; $l_2 = b_1 - a_2$ ; $l_3 = (b_2 - b_1)h$ ; $S(T_1, T_2) = \frac{(l_1 + l_2 + l_3)}{(w_0 + w_1) - (l_1 + l_2 + l_3)}$
III	$a_1 \leq a_2 \leq b_2 \leq b_1$ , $d_1 \leq e_1 \leq d_2 \leq e_2$	$l_1 = (b_1 - a_2)h_1$ ; $l_2 = (b_2 - b_1)h_2$ ; $h_1 = \frac{b_1 - a_2}{(b_1 - e_1) + (b_2 - e_2)}$ ; $h_2 = \frac{b_2 - b_1}{(b_2 - e_2) - (b_1 - e_1)}$ ; $S(T_1, T_2) = \frac{(l_1 + l_2)}{(w_0 + w_1) - (l_1 + l_2)}$
IV	$a_1 \leq a_2 \leq b_1 \leq b_2$ , $d_1 \leq d_2 \leq e_2 \leq e_1$	$h = \frac{b_2 - b_1}{(b_2 - e_2) - (b_1 - e_1)}$ ; $l = (b_2 - b_1)h$ ; $S(T_1, T_2) = \frac{2w_1 - l}{2w_0 + l}$
V	$a_1 \leq a_2 \leq b_1 \leq b_2$ , $d_1 \leq d_2 \leq e_1 \leq e_2$	$S(T_1, T_2) = \frac{e_1 - d_2 + b_1 - a_2}{e_2 - d_1 + b_2 - a_1}$
VI	$a_1 \leq a_2 \leq b_1 \leq b_2$ , $d_1 \leq e_1 \leq d_2 \leq e_2$	$h = \frac{b_1 - a_2}{b_1 - e_1 + b_2 - e_2}$ ; $S(T_1, T_2) = h / \left( \frac{w_0 + w_1}{b_1 - a_2} - h \right)$
VII	$a_2 \leq a_1 \leq b_1 \leq b_2$ , $d_1 \leq d_2 \leq e_2 \leq e_1$	$h_1 = \frac{a_2 - a_1}{(b_1 - e_1) - (b_2 - e_2)}$ ; $h_2 = \frac{b_2 - b_1}{(b_2 - e_2) - (b_1 - e_1)}$ ; $l_1 = (a_2 - a_1)h_1$ ; $l_2 = (b_1 - b_2)h_2$ ; $S(T_1, T_2) = \frac{w_1 + (l_1 + l_2)}{w_0 - (l_1 + l_2)}$
VIII	$(b_1 \leq a_2)$	$S(T_1, T_2) = 0$

Окончательно, получим

$$S(T_1, T_2) = \frac{|T_1 \cap T_2|}{|T_1| + |T_2| - |T_1 \cap T_2|} = \frac{\frac{1}{2}(l_1 + l_2)}{\frac{1}{2}(w_0 + w_1) - \frac{1}{2}(l_1 + l_2)} = \frac{(l_1 + l_2)}{(w_0 + w_1) - (l_1 + l_2)}.$$

В рамках исследования рассмотрены все возможные случаи взаимного расположения нечетких чисел и для каждого случая построена функция подобия (табл. 1).

Полученные результаты в дальнейшем использовались для исследования алгоритма нечеткой классификации.

Заметим, что если для каждой пары объектов  $x_i$  и  $x_j$  из множества  $X = \{x_1, \dots, x_N\}$ , заданных нечеткими числами  $T_i$  и  $T_j$  соответственно, определить значение функции подобия  $S(T_i, T_j)$ , то можно построить матрицу  $S = \{S_{ij}\}_{N \times N}$ , где  $S_{ij} = S(T_i, T_j) \in [0, 1]$ , которая задает нечеткое бинарное отношение — симметричное и рефлексивное исходя из свойств функции подобия, а следовательно,  $S$  — отношение сходства. Тогда, применяя операцию  $(\max - \min)$ -транзитивного замыкания, можно перейти от  $S$  к транзитивному отношению  $\hat{S}$ , которое является рефлексивным, симметричным и транзитивным, т. е. нечеткой эквивалентностью, или подобием. Однако, помимо  $(\max - \min)$ -транзитивного замыкания можно рассматривать замыкания, порожденные другими типами композиции.

### Транзитивность нечетких отношений

Рассмотрим более подробно свойство транзитивности для нечетких отношений. Пусть  $R_1$  и  $R_2$  — нечеткие бинарные отношения, заданные на множестве  $X$ , т. е.  $R_1, R_2 \in F(X \times X)$ ;  $\mu_{R_1}(x, y)$ ,  $\mu_{R_2}(x, y)$  — их функции принадлежности.

Говорят, что нечеткое отношение  $R_2$  включает в себя нечеткое отношение  $R_1$ , если для нечетких множеств  $R_1$  и  $R_2$  выполнено  $R_1 \subseteq R_2$ . Это означает, что имеет место неравенство  $\mu_{R_1}(x, y) \leq \mu_{R_2}(x, y)$  для любых  $x, y \in X$ .

$(\max - t)$ -композицией нечетких отношений  $R_1$  и  $R_2$  называется нечеткое отношение  $R_1 \overset{t}{\circ} R_2$  с функцией принадлежности

$$\mu_{R_1 \overset{t}{\circ} R_2}(x, y) = \max_{z \in X} t(\mu_{R_1}(x, z), \mu_{R_2}(z, y)),$$

где  $t: [0, 1]^2 \rightarrow [0, 1]$  — треугольная норма, которая, по сути, является операцией типа умножения и определяет пересечение нечетких множеств.

Простейшим представителем треугольных норм является  $\min(x, y)$ -операция, которая впервые использовалась для определения пересечения нечет-

ких множеств. Заметим, что определяемое таким образом пересечение не удовлетворяет закону коммутативности, поэтому алгебра  $\langle F(U), \neg, \cap = \min, \cup = \max \rangle$  уже не является булевой. Однако, как показывают исследования, операция  $\min(x, y)$  в качестве пересечения имеет определенные преимущества по сравнению с другими  $t$ -нормами. Вместе с тем, для нее характерны и недостатки: трудная интерпретируемость результатов; жесткость, которая проявляется в том, что некоторое изменение входных данных (иногда существенное) может не оказать влияния на результат. Для повышения чувствительности модели к исходным данным желательно использовать другие  $t$ -нормы.

В рамках проведенного исследования рассматривались  $t$ -нормы, представленные в табл. 2.

В работах [2, 4] показано, что справедливы следующие включения:

$$а) R_{\gamma}^2 \subseteq R_{\text{отн}}^2 \subseteq R_{\text{алг}}^2 \subseteq R_{\text{лог}}^2;$$

$$б) R_{\text{др}}^2 \subseteq R_{\text{огр}}^2 \subseteq R_{\text{алг}}^2 \subseteq R_{\text{лог}}^2,$$

где степень отношения определяется с помощью соответствующей композиции. Важно, что полученные результаты вычислительного эксперимента полностью зависят от этих соотношений.

Таблица 2

Название нормы	Функция принадлежности
Логическая	$\mu_{t_{\text{лог}}(R_1, R_2)}(x, y) = \min\{\mu_{R_1}(x, y), \mu_{R_2}(x, y)\}$
Алгебраическая	$\mu_{t_{\text{алг}}(R_1, R_2)}(x, y) = \mu_{R_1}(x, y)\mu_{R_2}(x, y)$
$\lambda$ -параметрическая, $\lambda \in (0, 1)$	$\mu_{t_{\lambda}(R_1, R_2)}(x, y) = \max\{0, (1 - \lambda)(\mu_{R_1}(x, y) + \mu_{R_2}(x, y) - 1) + \lambda xy\}$
Ограниченная	$\mu_{t_{\text{огр}}(R_1, R_2)}(x, y) = \max\{\mu_{R_1}(x, y) + \mu_{R_2}(x, y) - 1, 0\}$
$\gamma$ -параметрическая, $\gamma > 0$	$\mu_{t_{\gamma}(R_1, R_2)}(x, y) = \frac{\mu_{R_1}(x, y)\mu_{R_2}(x, y)}{\gamma + (1 - \gamma)(\mu_{R_1}(x, y) + \mu_{R_2}(x, y) - \mu_{R_1}(x, y)\mu_{R_2}(x, y))}$
Относительная	$\mu_{t_{\text{норм}}(R_1, R_2)}(x, y) = \frac{\mu_{R_1}(x, y)\mu_{R_2}(x, y)}{\mu_{R_1}(x, y) + \mu_{R_2}(x, y) - \mu_{R_1}(x, y)\mu_{R_2}(x, y)}$
Драстическая	$\mu_{t_{\text{др}}(R_1, R_2)}(x, y) = \begin{cases} \mu_{R_1}(x, y), & \text{если } \mu_{R_2}(x, y) = 1, \\ \mu_{R_2}(x, y), & \text{если } \mu_{R_1}(x, y) = 1, \\ 0, & \text{иначе} \end{cases}$



Особый интерес представляют параметрические  $t$ -нормы. Это объясняется тем, что подбирая параметр, можно в большой степени обеспечить адекватность модели и настроиться на конкретную информационную среду.  $\lambda$ -параметрическая норма впервые была введена в работе [4]. При исследовании моделей нечетких систем эта норма проявила лучшие аппроксимационные свойства.

Нечеткое бинарное отношение  $R$  называется  $(\max - t)$ -транзитивным, если  $R \circ^t R \subseteq R$ , т. е.

$$\mu_R(x, y) \geq \max_z t(\mu_R(x, z), \mu_R(z, y)).$$

Если отношение не является транзитивным, то с помощью транзитивного замыкания

$$\hat{R} = R \cup R_2 \cup R_3 \cup \dots$$

его можно обеспечить таким свойством.

Заметим, что при транзитивном замыкании сохраняются лишь некоторые свойства отношений и среди них — рефлексивность и симметричность [2]. Тогда если найти транзитивное замыкание рефлексивного и симметричного отношения, то получим минимальное отношение, включающее исходное, которое помимо упомянутых свойств будет обладать еще и заданным типом транзитивности.

В приложениях чаще всего используется  $(\max - \min)$ -транзитивность. Если отношение рефлексивно, симметрично и обладает данным типом транзитивности, то оно называется *нечетким отношением эквивалентности*, или *отношением подобия*. Любой  $\alpha$ -уровень ( $\alpha$ -срез) нечеткого отношения подобия является обычным отношением эквивалентности и, следовательно, определяет разбиение множества объектов на непересекающиеся классы эквивалентности. Из вложенности  $\alpha$ -уровней нечетких отношений следует и вложенность разбиений множества  $X$ , соответствующих различным  $\alpha$ -уровням, причем с уменьшением  $\alpha$  происходит укрупнение классов эквивалентности.

Из включений а) и б) следует, что если нечеткое отношение  $R$  обладает свойством  $(\max - \min)$ -транзитивности, то оно является транзитивным во всех других смыслах. Обратное неверно.

Для нечетких отношений помимо  $(\max - t)$ -транзитивности можно рассматривать и другие типы транзитивности. На наш взгляд, особый интерес представляет  $\alpha$ -транзитивность, определяемая следующим образом: отношение  $R$  называется транзитивным, если для любых  $x, y, z$  из того, что  $\mu_R(x, z) \geq \alpha$  и  $\mu_R(z, y) \geq \alpha$  следует, что  $\mu_R(x, y) \geq \alpha$ .

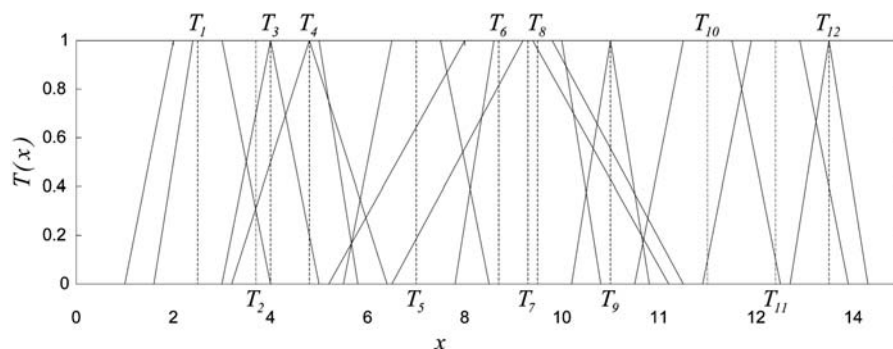


Рис. 3. Множество нечетких чисел

Справедливо следующее **утверждение**: если нечеткое отношение  $R$  обладает свойством  $(\max - \min)$ -транзитивности, то оно является также и  $\alpha$ -транзитивным для всех  $\alpha \in (0, 1]$ .

Данное свойство обуславливает тот факт, что граф, соответствующий обычному отношению  $R_\alpha$  уровня  $\alpha$  с характеристической функцией

$$h_{R_\alpha}(x, y) = \begin{cases} 1, & \mu_R(x, y) \geq \alpha; \\ 0, & \text{иначе,} \end{cases}$$

распадается на сильносвязные компоненты. Каждая сильная компонента соответствует классу эквивалентности. Именно нечеткое отношение — рефлексивное, симметричное и  $(\max - \min)$ -транзитивное имеет смысл называть нечетким отношением эквивалентности. Другие типы транзитивности порождают нечеткое отношение, которое разложимо на пересекающиеся максимальные подотношения подобия. В случае же  $(\max - \min)$ -транзитивности эти подотношения являются классами подобия, которым при фиксированном значении уровня  $\alpha \in (0, 1]$  соответствуют классы эквивалентности (или сильные компоненты в графе).

Для других типов  $(\max - t)$ -транзитивности связь с  $\alpha$ -транзитивностью не имеет места.

Согласно теореме декомпозиции [1], любое нечеткое отношение  $R$  можно представить через систему обычных отношений уровня  $\alpha$  по формуле

$$\mu_R(x, y) = \max_\alpha \{\alpha h_{R_\alpha}(x, y)\},$$

где  $h_{R_\alpha}(x, y)$  — характеристическая функция обычного отношения  $R_\alpha$  уровня  $\alpha$ . Декомпозиционные деревья, построенные на основе различных типов композиции и отражающие нечеткую классификацию объектов, оценки которых заданы нечеткими числами (рис. 3), приведены на рис. 4. Сравнение результатов нечеткой классификации — существенная проблема. Для ее решения нами использовались декомпозиционные деревья.

## Система показателей для сравнения результатов нечеткой классификации

Декомпозиционное дерево  $T$  представляет собой иерархическую структуру, каждый ярус которой образован вершинами, соответствующими классам эквивалентности объектов заданного множества, а переход от  $k$ -го яруса к  $(k + 1)$ -му соответствует переходу от одного разбиения к другому. Ярусы будем считать одинаковыми, если они соответствуют одному и тому же разбиению. Число ярусов определяется числом значений уровня  $\alpha_0 < \alpha_1 < \dots < \alpha_n$ , рассматриваемых в данной задаче, при этом ярус уровня  $\alpha_0$  содержит единственную вершину, которая соответствует исходному множеству объектов, объединенных в один класс. Эта вершина называется *корнем декомпозиционного дерева*.

Имеет место следующее **утверждение**: в декомпозиционном дереве, порожденном (max—min)-композицией, число классов эквивалентности является неубывающей функцией от значения уровня.

Значение каждого уровня  $\alpha$  формирует свое представление о схожести объектов заданного множества, при этом по мере возрастания  $\alpha$  это представление все более уточняется (рис. 4).

В рамках алгоритма нечеткой классификации величина  $(1 - \alpha)$  играет роль порога классифика-

ции, так что если транзитивное расстояние между объектами не превышает этой величины, то объекты оказываются в одном классе. В связи с этим параметр  $\alpha$  можно интерпретировать как *индекс чувствительности* разбиения. Таким образом, разбиение  $(k + 1)$ -го яруса индуцировано "уточненным отношением эквивалентности" по отношению к  $k$ -му.

Поскольку декомпозиционное дерево отражает процесс нечеткой классификации, когда при переходе от яруса к ярусу формируются разбиения, содержащие все большее число классов, то будем считать, что декомпозиционное дерево — это дерево, растущее от корня вверх. Поэтому связь между вершинами соседних ярусов — ориентированная, обозначается дугой и означает, что класс, соответствующий начальной вершине этой дуги, включает все вершины класса, соответствующего конечной вершине дуги. Заметим, что при необходимости декомпозиционное дерево будем рассматривать как неорграф, игнорируя ориентацию дуг. Это означает, что ориентированный граф заменяется его основанием (соответствующим неориентированным графом).

Пусть  $x$  — некоторая вершина декомпозиционного дерева,  $d^+(x)$  — полустепень захода,  $d^-(x)$  — полустепень исхода. Заметим, что для каждой вершины декомпозиционного дерева, за исключением корневой  $r$ , выполняется условие  $d^-(x) = 1$ . Вершина соответствует сформированному классу разбиения при данном значении порога классификации (или индекса чувствительности). Для корневой вершины  $d^+(r) = 0$ , т. е. она содержит единственную выходящую дугу и инициализирует процесс классификации. Вершины  $x, y$ , которых  $d^+(x) = 1$  и  $d^-(y) = 0$ , — это вершины, содержащие только входящие дуги и формирующие верхний ярус декомпозиционного дерева, на котором представлен результат разбиения на классы в случае, если порог классификации принимает минимальное значение.

Верхний ярус декомпозиционного дерева содержит максимальное число классов эквивалентности. Класс, содержащий единственный объект из  $X$ , будем называть *тривиальным классом*. Если разбиение содержит только тривиальные классы, то будем называть его *тривиальным разбиением*. Заметим, что тривиальные классы могут появиться на любом ярусе, за исключением корневого.

**Утверждение**: если для нечеткого отношения подобия  $R \subseteq \text{Fuzzy}(X^2)$  существует обычное отношение эквивалентности  $\tilde{R} \subseteq X^2$ , такое что  $\tilde{R} \subseteq R$ , то в результате нечеткой классификации невозможно получить тривиальное разбиение  $R$ , иначе тривиальное разбиение существует.

В дальнейшем будем рассматривать деревья, формирующие тривиальное разбиение заданного множества объектов.

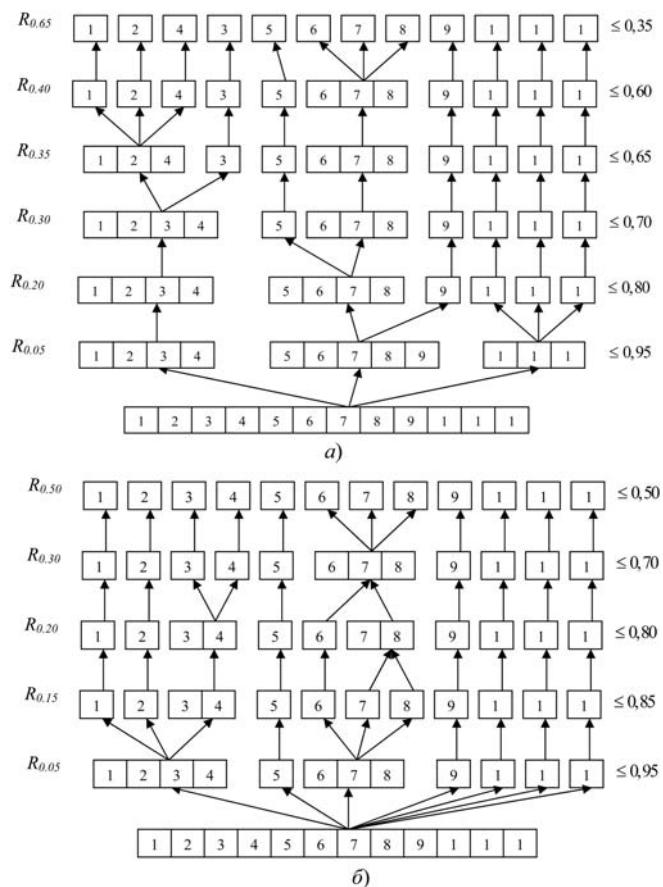


Рис. 4. Декомпозиционные деревья, полученные на основе: а — (max — min)-композиции; б — (max —  $\cdot$ )-композиции

В декомпозиционном дереве можно выделить *растущее дерево* (РД), которое является его подграфом и характеризуется следующими свойствами:

а) существует единственная вершина  $r$ , у которой  $d^+(r) \leq 1$  и  $d^-(r) > 1$ , которую будем называть *корнем растущего дерева*. Данная вершина инициализирует активный рост декомпозиционного дерева, и именно с нее "начинается" классификация;

б) в растущем дереве выделим вершины, называемые его *конечными вершинами*, такие что каждая вершина соответствует впервые появившемуся тривиальному классу, содержащему объект, который не участвует в дальнейшей классификации. Соответствующий уровень тривиального класса будем называть *весом объекта*. Очевидно, что чем больше вес, тем большее значение индекса чувствительности порождает тривиальный класс с этим объектом, т. е. распознает этот объект.

Заметим, что корень растущего дерева может совпадать с корнем декомпозиционного дерева. Появление конечных вершин растущего дерева свидетельствует о завершении процесса классификации, при которой каждая вершина образует класс. Заметим, что конечные вершины растущего дерева могут появиться на различных ярусах.

*Верхней вершиной растущего дерева*  $s$  назовем вершину, которая соответствует минимальному значению уровня, формирующего тривиальное разбиение. Соответствующие значения уровней будем называть *уровнем корня*  $\alpha_r$  *растущего дерева* и *уровнем вершины*  $\alpha_s$ . Промежуток  $Z_{act} = [\alpha_r, \alpha_s]$  можно рассматривать как *зону роста* декомпозиционного дерева, которая содержит последовательные значения уровней, таких что при переходе от уровня к уровню формируются ярусы с новыми классами объектов. Величину  $h = \alpha_s - \alpha_r$  будем называть *высотой растущего дерева*.

Растущее дерево назовем *правильным*, если его основание (т. е. соответствующий неориентированный граф) не содержит циклов. Декомпозиционное дерево, которое содержит правильное растущее дерево, будем также называть *правильным декомпозиционным деревом*.

Появление циклов свидетельствует о том, что на некотором ярусе появляются вершины  $x$  с  $d^+(x) > 1$ , т. е. происходит объединение появившихся ранее классов. Соответствующие им значения уровней будем называть *критическими*. Минимальное  $\underline{\alpha}$  и максимальное  $\bar{\alpha}$  значения уровней, которым соответствуют ярусы, содержащие вершины с  $d^+(x) > 1$ , назовем соответственно *нижней* и *верхней* границами *зоны неустойчивой классификации*  $Z = [\underline{\alpha}, \bar{\alpha}]$ . Наличие зоны неустойчивой классификации свидетельствует о необходимости дополнительного исследования для тех индексов чувствительности, которые ее образуют, и разработки рекомендаций по выбору параметров классификации. В рамках проведенного исследования оказалось, что зона неустойчивой классификации появляется не всегда.

В правильных растущих деревьях появление тривиального класса свидетельствует о том, что существует такое значение уровня (индекса чувствительности), начиная с которого данный объект не участвует в дальнейшей классификации. В неправильных растущих деревьях может случиться так, что объект попал в тривиальный класс, но на следующих ярусах этот класс объединился с другим, а затем вновь рассматриваемый объект образовал тривиальный класс. Таким образом, каждому объекту  $x$  можно поставить в соответствие, по крайней мере, промежуток  $[\underline{v}(x), \bar{v}(x)]$ , так что  $\underline{v}(x)$  — минимальное значение, а  $\bar{v}(x)$  — максимальное значение уровней, при которых данный объект образовал тривиальный класс, при этом  $\underline{v}(x)$  будем называть *весом объекта*, поскольку данная величина определяет минимальное значение индекса чувствительности, позволяющее максимально отличить объект от остальных.

Заметим, что введенные характеристики декомпозиционного дерева позволяют однозначно его описывать и строить процедуры сравнения результатов нечеткой классификации.

### Результаты вычислительного эксперимента

Для исследования результатов нечеткой классификации на основе функции подобия была написана компьютерная программа, которая позволяет строить декомпозиционное дерево на основе разложения рефлексивного, симметричного и  $(\max - t)$ -транзитивного отношения  $S$  на обычные уровневые отношения. Проведенный вычислительный эксперимент позволяет сделать следующие выводы.

1. Правильное декомпозиционное дерево может быть получено только с использованием  $t_{\log}$ -нормы (рис. 4, а), в остальных случаях дерево является неправильным. Данный факт подтверждает теоретический вывод о том, что только  $(\max - \min)$ -композиция позволяет от отношения  $S$  перейти к нечеткому отношению эквивалентности  $\hat{S}$ , во всех остальных случаях формируется отношение толерантности, которое допускает разложение лишь на максимальные подотношения подобия [1]. В правильном декомпозиционном дереве каждому ярусу соответствует свое отношение эквивалентности и свое разбиение заданного множества объектов. С увеличением  $\alpha$  при переходе от одного яруса к другому происходит как бы уточнение предыдущего отношения эквивалентности.

Отношение толерантности характеризуется тем, что нарушается последовательность включений уровневых отношений, что приводит к появлению зон неустойчивой классификации, которые и обуславливают "неправильность" декомпозиционного дерева. В неправильных декомпозиционных деревьях объекты, распределенные по классам, среди

которых возможны и тривиальные, могут объединяться вновь в разных комбинациях.

2. Пусть  $\hat{S}_{t_{type}}$  — отношение подобия/толерантности, порожденное с помощью  $t$ -нормы данного типа  $type \in \{\text{лог}, \text{алг}, \text{огр}, \text{др}\}$ . Если для каждого отношения  $\hat{S}_{t_{type}}$  существуют минимальные значения уровня  $\alpha_{type}$ , порождающие разбиение объектов заданного множества  $X$  на данное число классов  $NumKl$ , то эти значения упорядочены следующим образом:

$$\alpha_{др} \leq \alpha_{огр} \leq \alpha_{алг} \leq \alpha_{лог},$$

причем в этом неравенстве возможно отсутствие уровней для тех отношений, для которых не существует заданного числа классов  $NumKl$ . Если величину  $\alpha$  интерпретировать как степень чувствительности (чем больше значение, тем в большей степени учитываются свойства объектов), то полученный результат означает, что для разбиения множества  $X$  на заданное число классов  $NumKl$  именно  $t_{лог}$ -норма (а следовательно, (max – min)-композиция) предполагает самое большое значение индекса чувствительности. Вместе с тем, можно рассуждать в терминах порога классификации  $1 - \alpha_{type} = \varepsilon_{type}$ . Для порога классификации имеем неравенство

$$\varepsilon_{др} \geq \varepsilon_{огр} \geq \varepsilon_{алг} \geq \varepsilon_{лог},$$

которое означает, что минимальное значение порога для разбиения на  $NumKl$  классов обеспечивает  $t_{лог}$ -норма.

Аналогичное исследование было проведено для  $\gamma$ -параметрической треугольной нормы. Оно показало, что с возрастанием  $\gamma$  индекс чувствительности уменьшается для заданного числа классов.

3. Пусть  $NumKl_{type}(\alpha)$  — число классов, содержащихся на ярусе, который соответствует уровню  $\alpha$ , при использовании  $t$ -нормы данного типа  $type \in \{\text{лог}, \text{алг}, \text{огр}, \text{др}\}$ . Зафиксируем некоторое значение  $\alpha^* \in (0, 0,5]$ , тогда

$$\begin{aligned} NumKl_{огр}(\alpha^*) &\geq NumKl_{алг}(\alpha^*) \geq \\ &\geq NumKl_{отн}(\alpha^*) \geq NumKl_{лог}(\alpha^*). \end{aligned}$$

При  $\alpha \in (0,5, 1]$  все типы  $t$ -нормы дают одно и то же максимальное число классов. Данный вывод позволяет при заданном значении индекса чувствительности  $\alpha^*$  подобрать приемлемое число классов или варьировать это число за счет выбора  $t$ -нормы.

4. Установлено, что часто используемые в приложениях расстояния Хемминга и Евклида позволяют осуществлять разбиение на классы при значениях  $\alpha \geq 0,5$  для всех типов  $t$ -норм. Это означает, что для растущих декомпозиционных деревьев в этом случае для уровня корня имеет место  $\alpha_r \geq 0,5$ . При использовании функции подобия декомпозиционное дерево растет практически от нулевых значений уровня, при этом для уровня вершины

справедливо  $\alpha_s = 0,5$  для всех типов  $t$ -норм. Использование функции подобия позволяет уже при малых значениях индекса чувствительности осуществить классификацию объектов. Это означает, что функция подобия позволяет улучшить "распознающие" свойства (max – min)-композиции. Это отражается в том, что зона роста декомпозиционного дерева во всех примерах значительно больше для функции подобия, чем в декомпозиционных деревьях, полученных с использованием функций расстояния. В случае  $\gamma$ -параметрической  $t$ -нормы высота растущего дерева для всех значений параметров одинаковая, причем  $\alpha_r = 0,5$  и  $\alpha_s = 0,5$ .

5. Пусть  $\gamma_1$  — некоторое значение параметра  $\gamma$ -параметрической  $t$ -нормы и  $R_\alpha(\gamma_1)$  — разбиение, соответствующее уровню  $\alpha$ , тогда если такое же разбиение существует для значения параметра  $\gamma_2 > \gamma_1$  и соответствующее ему значение уровня есть  $\alpha'$ , тогда  $\alpha' \leq \alpha$ . Это означает, что с возрастанием параметра  $\gamma$  для "распознавания" такого же разбиения будет достаточно меньшее значение индекса чувствительности.

6. Вес  $w(x)$  объекта  $x$  — это минимальное значение уровня  $\alpha$ , при котором объект  $x$  образует тривиальный класс. По сути, вес — это требование к чувствительности, которая необходима для "распознавания" индивидуальности объекта и помещения его в отдельный класс. Тогда с возрастанием параметра  $\gamma$   $\gamma$ -параметрической нормы вес фиксированного объекта не возрастает. Это означает, что при увеличении параметра  $\gamma$  возрастают шансы получения тривиального класса, образованного данным объектом. И, таким образом, параметр  $\gamma$  можно интерпретировать как "силу удара", который влечет за собой разбиение множества объектов на отдельные классы. С ростом  $\gamma$  уже для малых значений уровней возможно формирование тривиальных классов, причем число классов на нижних ярусах декомпозиционного дерева очень резко возрастает.

Заметим, что оценка качества разбиения на классы предполагает чисто формальный подход и является для исследователя только вспомогательным средством. Основная роль принадлежит держателю анализу результатов классификации. Выбор лучшего варианта разбиения облегчается в значительной мере серьезной подготовительной работой, в частности, выбором признаков, характеризующих классифицируемые объекты.

#### Список литературы

1. Кофман А. Введение в теорию нечетких множеств. М.: Радио и связь, 1982. 432 с.
2. Каплиева Н. А., Леденева Т. М. Исследование различных типов транзитивности в приложении к нечеткой классификации // Вестник Воронежского государственного университета. Сер. Физика. Математика. 2006. № 2. С. 206—216.
3. Пегат А. Нечеткое моделирование и управление. М.: БИНОМ, Лаборатория знаний. 2009. 798 с.
4. Леденева Т. М. Моделирование процесса агрегирования целей в целенаправленных системах. Воронеж: Изд-во ВГТУ, 1999. 155 с.

**А. В. Савченко**, канд. техн. наук, ст. преподаватель,  
Национальный исследовательский университет,  
Высшая школа экономики, Нижний Новгород,  
e-mail: avsavchenko@hse.ru

## Трехпороговая система для распознавания изображений на основе метода направленного перебора

*Ставится и решается задача автоматического распознавания изображений из большой базы данных на основе метода направленного перебора альтернатив и принципа минимума информационного рассогласования Кульбака—Лейблера. Исследуется возможность повышения точности распознавания за счет применения трех решающих правил: для ускорения вычислительной процедуры распознавания на основе метода направленного перебора, отбраковки сомнительных с точки зрения надежности решений и выбора решения, наиболее отличающегося от остальных. Представлены программа и результаты экспериментального исследования в задаче распознавания людей по фотографиям. Показано, что предложенный подход характеризуется повышенной степенью точности и вычислительной эффективности.*

**Ключевые слова:** автоматическое распознавание изображений, метод направленного перебора, принцип минимума информационного рассогласования

### Введение

Метод направленного перебора [1] альтернатив (МНП) является, как известно, эффективным средством решения задач широкого круга в области автоматического распознавания изображений (АРИ). Этот метод, основанный на теоретико-информационном подходе и отталкивающийся от метрических свойств решающей статистики минимального информационного рассогласования (МИР) [2, 3], позволяет учесть характерную особенность задачи АРИ — число эталонов в реальных базах данных составляет сотни и тысячи единиц. Однако ситуация резко усложняется на практике при неизбежном стремлении исследователя по возможности гарантировать достаточную степень надежности каждого принятого им решения [4]. Подобная проблема особенно актуальна в задачах АРИ с неизвестными заранее условиями (варьирующееся расстояние до объекта, ракурс, освещение, характер искажений и др.). К сожалению, существующая теория не дает на такие вопросы строгих ответов. Поэтому чаще исследователи полагаются почти исключительно на свой опыт и интуицию и вводят фиксированное пороговое значение для отказа в выдаче решения при превышении минимальным рассогласованием

порога. К сожалению, указанное решение не является оптимальным в теоретико-вероятностном смысле [5], поэтому в настоящей работе реализуется идея обнаружения сомнительных с точки зрения надежности решений по очевидному признаку: одновременной близости входного изображения по отношению к более чем одному эталону из базы данных. В рамках предлагаемой трехпороговой системы эта идея сочетается с МНП в целях повышения скорости вычислений без потери качества распознавания образов. Полученные результаты и сделанные по ним выводы рассчитаны на широкий круг специалистов в области распознавания образов.

### Задача автоматического распознавания изображений

Пусть задано множество из  $R > 1$  полутоновых изображений

$$X_r = \|x_{uv}^r\|, \quad u = \overline{1, U}, \quad v = \overline{1, V}.$$

Здесь  $r$  — номер эталона в базе изображений;  $U$  — высота изображения;  $V$  — его ширина;  $x_{uv}^r \in \{0, 1, \dots, x_{\max}\}$  — интенсивность точки изображения с координатами  $(u, v)$ ;  $x_{\max}$  — максимальное значение интенсивности.

Предполагается, что каждому эталону  $X_r$  присвоен некий номер  $c(r) \in \{1, \dots, C\}$  класса, где  $C$  — число различных классов. При этом каждый класс характеризуется тем, что принадлежащие ему объекты обладают некоторой общностью или сходством в характеристиках. То общее, что объединяет объекты в класс, и называют *образом* [6]. Задача состоит в том, чтобы отнести вновь поступающее (на вход) изображение  $X = \|x_{uv}\|$  к одному из  $C$  таких классов. Это типичный пример задачи распознавания образов с обучением [7].

Процедуры построения решающих правил для поставленной задачи в общем случае делятся на детерминированные и статистические. В настоящее время наиболее часто используется первый, детерминистский, подход. В его рамках в пространстве объектов определяется некое расстояние (мера близости) между любыми парами объектов. Зачастую для АРИ [8] применяется критерий, основанный на  $l_1$ -метрике:

$$\rho_1(X, X_r) = \frac{1}{UV} \sum_{u=1}^U \sum_{v=1}^V |x_{uv} - x_{uv}^r| \rightarrow \min. \quad (1)$$

К сожалению, подобный подход не всегда позволяет получить удовлетворительные результаты. Это обстоятельство связано, во-первых, с известной [8, 9] вариативностью зрительных образов, а, во-вторых, с наличием во входном изображении  $X$  помех, таких как неопределенная заранее интенсивность источников освещения или просто случайное искажение некоторых точек изображения.

Указанные проблемы при детерминистском подходе обычно решаются путем добавления новых изображений в базу данных, что приводит, в свою очередь, к резкому увеличению ее объема. Во всех перечисленных случаях на помощь приходит второй, статистический, подход [3, 9]. Например, в работе [1] предложено использование рассогласования Кульбака—Лейблера для сравнения гистограмм цветности изображения [10]. Поставим каждому эталону  $X_r$  в соответствие гистограмму распределения цвета

$$H_r = [h_1^r, h_2^r, \dots, h_{x_{\max}}^r];$$

$$h_x^r = \frac{1}{UV} \sum_{u=1}^U \sum_{v=1}^V \delta(x_{uv}^r - x), \quad (2)$$

где  $\delta(x)$  — дискретная дельта-функция. Такая же процедура построения гистограммы  $H$  применяется и для входного изображения  $X$ . Гистограмма  $H_r$  определяет собой распределение одномерной дискретной случайной величины — цвета пикселя. Задача АРИ сводится в таком случае к проверке  $R$  гипотез  $W_r$  о распределении  $H_r$  сигнала изображения на входе  $H$ . А ее решение может быть получено с использованием оптимального в байесовском смысле критерия, основанного на принципе МИР:

$$\rho_{KL}(X, X_r) = \sum_{x=1}^{x_{\max}} h_x \ln \frac{h_x}{h_x^r} \rightarrow \min. \quad (3)$$

Таким образом, процедура АРИ реализуется по схеме многоканальной обработки, в которой число каналов  $R$  определяется числом изображений-эталонов.

### Алгоритм с обнаружением ошибок

Как говорилось во введении, ситуация резко усложняется при неизбежном стремлении проконтролировать и по возможности гарантировать достаточную степень надежности каждого принятого решения (1), (3). И здесь выясняется, что критерий МИР содержит важную информацию именно в отношении надежности принимаемых решений [4].

Прежде всего, отметим метрические свойства решающей статистики МИР [3]  $\rho_{KL}(X/X_r) \geq 0$  с равенством ее нулю лишь в идеальном случае совпадения входного и эталонного сигналов. Поэтому для отбраковки сомнительных с точки зрения надежности решений может применяться традиционный для детерминированного подхода критерий вида

$$\min_{r \in \{1, \dots, R\}} \rho_{KL}(X/X_r) > \rho_1 = \text{const}. \quad (4)$$

Здесь  $\rho_1$  — порог для допустимой величины информационного рассогласования (ВИР) на множестве объектов одного класса за счет известной их

вариативности. Значение такого порога нетрудно установить, зафиксировав ошибку  $\alpha$  первого рода

$$\alpha = P\{\rho_{KL}(X/X_r) > \rho_1 | W_r\}. \quad (5)$$

Здесь под  $P\{\rho_{KL}(X/X_r) > \rho_1 | W_r\}$  понимается условная вероятность того, что рассогласование  $\rho_{KL}(X/X_r)$  превышает порог  $\rho_1$  при справедливости гипотезы  $W_r$  — входное изображение  $X$  принадлежит классу, определяемому эталоном  $X_r$ . Строгий теоретико-вероятностный анализ выражения (5) в общем случае наталкивается на непреодолимые трудности вычислительного характера. Поэтому в практических приложениях для определения порога  $\rho_1$  по имеющимся эталонам из базы данных используется уравнение

$$\alpha = \frac{1}{R} \sum_{r=1}^R H(\min_{c(v)=c(r), v \neq r} \rho_{KL}(X_r/X_v) - \rho_1). \quad (6)$$

Здесь  $H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$  — функция Хэвисайда.

### Досрочный останов перебора

Во многих наиболее важных с точки зрения практики приложениях распознавания изображений число имеющихся в базе данных эталонов весьма велико:  $R \gg 1$ . Известный способ снижения вычислительных затрат АРИ связан с использованием отмеченных выше метрических свойств статистики МИР [4]. Оптимизационная процедура вида (3) преобразуется к упрощенному (в его практической реализации) виду [1]

$$W_v: \rho_{KL}(X/X_v) < \rho_0 = \text{const}. \quad (7)$$

Здесь  $\rho_0$  — порог ВИР на множестве объектов разных классов. Значение такого порога можно установить, зафиксировав ошибку  $\beta$  второго рода

$$\beta = P\{\rho_{KL}(X/X_v) < \rho_0 | W_v\}. \quad (8)$$

Здесь  $P\{\rho_{KL}(X/X_v) < \rho_0 | \bar{W}_v\}$  — условная вероятность того, что рассогласование  $\rho_{KL}(X/X_v)$  меньше порога  $\rho_0$  при справедливости гипотезы  $\bar{W}_v$ , обратной к  $W_v$  — входное изображение  $X$  не принадлежит классу, определяемому эталоном  $X$ . На практике имеем следующее уравнение для определения порога  $\rho_0$ :

$$\beta = \frac{1}{R} \sum_{r=1}^R H(\rho_0 - \min_{c(v) \neq c(r)} \rho_{KL}(X_r/X_v)). \quad (9)$$

По своей сути выражение (7) с использованием порога ВИР, определенного из уравнения (9), представляет собой условие "останова" при переборе альтернатив в рамках проверочной процедуры по критерию МИР (3). Естественным развитием этой идеи служит описанный далее метод направленного перебора альтернатив [1], в котором метрические свойства статистики МР (3) используются в полной степени.

## Метод направленного перебора

Следуя определению (3) составим  $(R \times R)$ -матрицу  $\Pi = \|\rho_{ij}\|$  значений  $\rho_{ij} = \rho_{KL}(X_i/X_j)$ . После этого по произвольной первой контрольной выборке  $X_1, \dots, X_N$  из заданного множества альтернатив  $\{X_r\}$  при условии  $N \ll R$  получим ранжированный по критерию МИР (5) ряд данных  $\{X_{i_j}\}$  и, наконец, найдем из него первый локальный оптимум  $X_{i_N}$ . На этом завершается первый этап вычислений. На втором этапе для выделенного изображения-эталона  $X_{i_N}$  по матрице  $\Pi$  найдем множество из  $M < R$  изображений  $X^{(M)} = \{X_{i_{N+1}}, \dots, X_{i_{N+M}}\}$ , находящихся от изображения  $X_{i_N}$  на "расстоянии" (3), не превышающем  $\rho_{KL}(X/X_{i_N})$ :

$$(\forall X_i \notin X^{(M)})(\forall X_j \in X^{(M)})\Delta\rho(X_i) \geq \Delta\rho(X_j). \quad (10)$$

Здесь  $\Delta\rho(X_i) = |\rho_{KL}(X_i/X_{i_N}) - \rho_{KL}(X/X_{i_N})|$ . Добавим к этому множеству еще один,  $(M + 1)$ -й элемент  $X_{i_{N+M+1}}$  из числа не попавших в состав контрольной выборки по результатам предыдущего этапа вычислений. Этим мы вносим в поисковую процедуру определенный элемент случайности как способ достижения глобального оптимума за конечное число этапов вычислений. В результате получаем вторую контрольную выборку изображений-эталонов  $\{X_{i_N}, \dots, X_{i_{N+M+1}}\}$  для анализа. Далее все вычисления первого этапа циклически повторяются до тех пор, пока на некотором  $K$ -м этапе для элемента  $X^*$  не будет выполнено условие останова (7).

## Оптимальный алгоритм

Традиционно [8, 9, 11] на практике в качестве критерия АРИ применяется выражение вида (4). Однако с точки зрения задачи статистической классификации [3, 5] замена решающего правила (3) на (4) приводит к потере свойства оптимальности. Действительно, условие (4) эквивалентно сравнению максимальной апостериорной вероятности принадлежности объекта к эталонному классу с некоторым наперед заданным порогом. Очевидно, что это решение не является оптимальным, так как оно не эквивалентно критерию, основанному на отношении правдоподобия [7]. Для формулировки оптимального критерия, основанного на сравнении рассогласования с порогом, введем следующее обозначение для наиболее близкого (в смысле информационного рассогласования Кульбака—Лейблера) к входному изображению  $X$  эталона:

$$r^* = \arg \min_{r \in \{1, \dots, R\}} \rho_{KL}(X/X_r). \quad (11)$$

Тогда оптимальное в смысле леммы Неймана—Пирсона решение задачи АРИ даст следующее выражение [3]:

$$\left( \min_{r \in \{1, \dots, R\}, c(r) \neq c(r^*)} \rho_{KL}(X/X_r) - \rho_{KL}(X/X_{r^*}) \right) > \rho_2 \geq 0. \quad (12)$$

Значение порога  $\rho_2$  также несложно подобрать опытным путем. Выражение (12) имеет простую физическую интерпретацию — решение АРИ принимается в пользу класса  $c(r^*)$ , если рассогласование до эталона из этого класса меньше рассогласований до эталонов из остальных классов на величину, большую наперед заданного значения  $\rho_2$ .

Заметим, что использование критерия (12) не отменяет выражений (4) и (7). Действительно, каждое из пороговых значений служит для достижения своей цели. Порог  $\rho_0$  используется для сокращения перебора,  $\rho_1$  — для отказа ото всех альтернатив,  $\rho_2$  — для принятия оптимального в смысле Неймана—Пирсона решения задачи статистической классификации.

На рис. 1 приведена общая схема предлагаемой трехпороговой системы распознавания образов, являющейся расширением разработанного нами ранее устройства для распознавания изображений [12].

Изображение для распознавания  $X$  поступает на вход блока измерения сходства, который вычисляет информационное рассогласование Кульбака—Лейблера (3) между  $X$  и эталоном  $X_r$ , выбранным блоком выбора эталонов. Вычисленное расстояние  $\rho_{KL}(X/X_r)$  подается на вход блока принятия решения об останове, а также сохраняется в блоке оперативной памяти с тем, чтобы выбрать два наилучших из числа проверенных эталонов (с наименьшими расстояниями до входного изображения  $X$ ). Блок при-

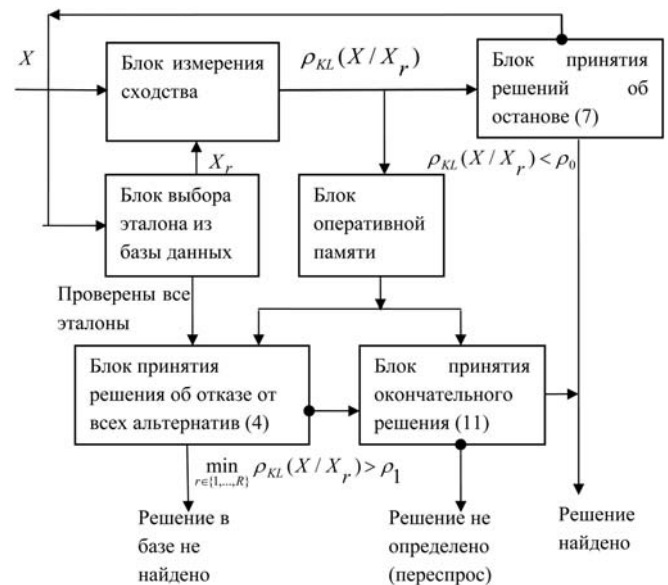


Рис. 1. Схема трехпорогового устройства для распознавания изображений

нятия решения об останове сравнивает расстояние  $\rho_{KL}(X/X_r)$  с порогом  $\rho_0$ . Если  $\rho_{KL}(X/X_r) < \rho_0$ , то выдается решение  $X_r$ . В противном случае с инвертированного выхода на управляющий вход блока выбора эталонов поступает команда на выбор следующего эталона из базы данных. Если для всех эталонов расстояние до входного изображения превышает порог  $\rho_0$ , то в блоке принятия решений об отказе от всех альтернатив проверяется условие (4). Если минимальное рассогласование больше порога  $\rho_1$ , делается вывод об отсутствии входного изображения среди эталонов. В противном случае в блоке принятия решений по правилу (12) вычисляется разность между двумя наименьшими рассогласованиями, сохраненными в оперативной памяти. Если эта разность превышает порог  $\rho_2$ , принимается решение АРИ в пользу класса  $r^*$  (11), иначе делается вывод о невозможности надежного различения, поэтому требуется выполнить переспрос — повторить процедуру АРИ для нового входного изображения.

### Результаты экспериментальных исследований

Рассмотрим применение предложенного подхода (4)—(12) для задачи АРИ к реальным базам данных фотографий людей AT&T (ORL)<sup>1</sup>, Yale<sup>2</sup> и Essex<sup>3</sup>. Во всех экспериментах в качестве предварительной обработки изображений для выделения лиц использовалась библиотека OpenCV<sup>4</sup>. После выделения лиц все изображения нормировались по яркости [9] для ослабления проблемы вариативности освещения входных объектов. Распознавание проводилось средствами системы распознавания лиц [13].

В ходе эксперимента проведен сравнительный анализ предложенной трехпороговой схемы АРИ (рис. 1) с традиционными критериями, основанными на минимизации расстояний, такими как (1), (3) (в них альтернатива об отказе распознавания отсутствует), и с общепринятыми устройствами АРИ с одним порогом на минимальное рассогласование (4). Порог  $\rho_2$  определялся так, чтобы вероятность ошибки (сумма вероятностей ошибок первого и второго рода при альтернативе — отказе дать надежное решение) была минимальна.

Вначале процедура АРИ проводилась для большой базы данных Essex. Объем множества эталонов после кластеризации [14] составил  $R = 900$ , для тестирования использовались другие 1200 фотографий. Вероятность точного распознавания для критерия (3) составила 99,25 %, для критерия (4) — 99,5 %. Для предложенного устройства (рис. 1) точность классификации также составила 99,5 %. Одна-

<sup>1</sup> AT&T Face database, [www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html](http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html)

<sup>2</sup> Yale Face database, <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

<sup>3</sup> Essex Face database, <http://cswww.essex.ac.uk/mv/allfaces/index.html>

<sup>4</sup> OpenCV, <http://opencv.willowgarage.com/wiki/>

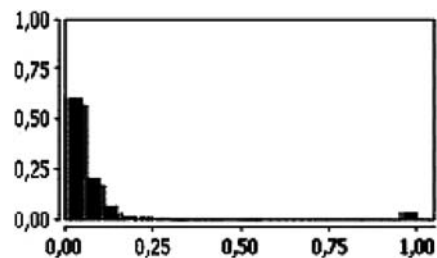


Рис. 2. Гистограмма числа вычислений рассогласований относительно объема базы  $R$

ко, когда для распознавания были поданы фотографии из другой базы данных Yale, применение условия (4) позволило в 91 % случаев сделать вывод об отсутствии изображений в базе; в остальных случаях предложенное устройство (рис. 1) предложило задействовать переспрос. А использование организации вычислений по МНП за счет дополнительного порога  $\rho_0$  на досрочный останов перебора (7)—(8) позволило в среднем сократить объем вычислений в ходе АРИ на 89 %. Гистограмма числа вычислений ВИР согласно МНП показана на рис. 2. Здесь с вероятностью 65 % число вычислений рассогласований (1) по МНП не превысило 5 % от  $R$ . Без применения условия останова (7) такого ускорения алгоритмов АРИ без упрощения модели изображения добиться невозможно.

Во втором эксперименте исследовался предлагаемый подход для базы данных Yale. В качестве множества эталонов использовались 15 изображений (по одной фотографии каждого человека), а тестирование качества распознавания проводилось на остальных 181 фотографии тех же людей. Тем самым достигались наиболее жесткие условия для последующего распознавания. Для критерия (3) точность правильной классификации составила 86,7 %, а для сравнения с одним порогом — 88,5 %. При применении предложенного подхода (4)—(12) точность повысилась до 90,1 %. Для этой базы использованные МНП позволило в среднем сократить объем вычислений в ходе АРИ на 40 %.

В последнем эксперименте использовалась база данных AT&T. В качестве множества эталонов использовались  $R = 110$  изображений  $C = 40$  различных людей, а тестирование качества распознавания проводилось на остальных 291 фотографии тех же людей. Для критерия (3) точность правильной классификации составила 90,75 %. Для традиционной однопороговой схемы аналогичный показатель повысился до 91,25 %. При применении описанного в данной работе подхода (4)—(12) точность стала еще выше — 92,5 %. А среднее число вычислений по МНП (7)—(10) составило здесь 35,5 % от числа эталонов  $R$ .

Все результаты для трех баз данных сведены в табл. 1.

В заключение предыдущие эксперименты по сравнению предлагаемой трехпороговой схемы АРИ



Таблица 1

## Оценки вероятности ошибки (рассогласование Кульбака)

Критерий АРИ	Essex	Yale	AT&T
Решающая статистика (3)	0,75 %	13,3 %	9,25 %
Сравнение с порогом (4)	0,5 %	11,5 %	8,75 %
Трехпороговая схема (4)—(12)	0,5 %	9,9 %	7,5 %

Таблица 2

## Оценки вероятности ошибки (рассогласование Евклида)

Критерий АРИ	Essex	Yale	AT&T
Решающая статистика (1)	2,55 %	15,75 %	32,25 %
Сравнение с порогом (4)	2,3 %	15,15 %	24,5 %
Трехпороговая схема (4)—(12)	1,7 %	7,3 %	17,25 %

с распространенными аналогами были повторены для более традиционного подхода к АРИ, основанного на непосредственном сопоставлении матриц интенсивности (1). Результаты приведены в табл. 2. Здесь качество распознавания заметно ниже сравнения гистограмм интенсивности (2), (3) (за исключением базы данных Yale), но в этих условиях выигрыш в точности предлагаемого подхода в сравнении с традиционными методами еще более заметен.

По результатам проведенного экспериментального исследования можно сделать следующие выводы. Во-первых, точность классификации для традиционных критериев (1), (3), в которых отсутствует альтернатива об отказе распознавания, ниже, чем для критериев вида (4). Во-вторых, не менее традиционный критерий сравнения расстояния с порогом (4) не обеспечивает максимальной точности. И, в-третьих, предложенный подход (4)—(12) показал наиболее высокие результаты в смысле точности и объема вычислений при распознавании.

### Заключение

Вопрос о повышении вычислительной эффективности алгоритмов распознавания изображений при наличии базы данных эталонов объемом в тысячи единиц в настоящее время вызывает повышенный интерес. Общепринято [8, 11, 15] введение правил вида (7) для досрочного останова перебора. Однако исследователи зачастую забывают, что переход от решающих правил вида (3) к сравнению рассогласования с порогом (4), (7) может привести к потере свойства статистической оптимальности (в смысле леммы Неймана—Пирсона).

В настоящей работе показано, что использование основанного на лемме Неймана—Пирсона критерия (12) позволяет существенно повысить надежность классификации. При этом подчеркивается, что решающее правило (12) не заменяет сопоставление рассогласований с порогом (4), (7). Наоборот, в рамках трехпорогового устройства (см. рис. 1) каждое из указанных правил используется для достижения своей цели. Так, показано, что условие (7)

является основным для ускорения вычислительной процедуры классификации — метода направленного перебора [1, 14]; условие (4) служит для отбраковки объектов, не принадлежащих ни к одному классу [4]; а условие (12) позволяет выбрать решения (3), которые наиболее различаются (в смысле рассогласования с входным объектом) с другими эталонами из базы. Проведенные эксперименты наглядно показывают, что этот подход может давать хорошие результаты на основе не только информационного рассогласования Кульбака—Лейблера (3), но и традиционной, "евклидовой" метрики  $l_1$  (1).

Таким образом, благодаря проведенному исследованию получены оценки преимуществ разработанного подхода (3)—(12) к повышению скорости и надежности АРИ. Показано, что разработанный подход способен в значительной степени ослабить, а в ряде случаев и преодолеть (без потерь в точности) проблему недостаточной вычислительной эффективности существующих алгоритмов типа "ближайших соседей" [15] для широкого круга задач распознавания образов.

### Список литературы

1. Савченко А. В. Метод направленного перебора альтернатив в задаче распознавания полутоновых изображений // Автометрия. 2009. № 45. С. 90—98.
2. Савченко В. В., Пономарев Д. А. Оптимизация фонетической базы данных по группе дикторов на основе критерия МИР // Информационные технологии. 2009. № 12. С. 7—12.
3. Кульбак С. Теория информации и статистика. М.: Наука, 1967. 408 с.
4. Савченко В. В., Савченко А. В. Принцип минимального информационного рассогласования в задаче распознавания дискретных объектов // Известия вузов России. Радиоэлектроника. 2005. Вып. 3. С. 10—18.
5. Боровков А. А. Математическая статистика. Дополнительные главы. М.: Наука, 1984. 615 с.
6. Цыпкин Я. З. Адаптация и обучение в автоматических системах. М.: Наука, 1968. 400 с.
7. Орлов А. И. О развитии математических методов теории классификации // Заводская лаборатория. 2009. Т. 75. № 7. С. 51—63.
8. Горелик А. Л., Гуревич И. Б., Скрипкин В. А. Современное состояние проблемы распознавания: некоторые аспекты. М.: Радио и связь, 1985. 160 с.
9. Koutroumbas K., Theodoridis S. Pattern Recognition, Boston: Academic Press, 2008. 840 p.
10. Wong K. M., Cheung C. H., Liu T. S. Dominant Color Image Retrieval Using Merged Histogram // Proc. IEEE ISCAS. 2003. Vol. 2. P. 908—911.
11. Пат. 6882741 US, 6G 06 K 9/00. Facial image recognition apparatus / Dobashi, Hirasawa, Okazaki, etc, патентообладатель Kabushiki Kaisha Toshiba. — 09/814012; заявл. 22.03.2001; опубл. 19.04.2005. Бюл. № 30.
12. Пат. 88171 РФ, 7G 06 K 9/00. Устройство для распознавания изображений / Савченко А. В.; заявитель и патентообладатель Савченко А. В. — 2009127049/22; заявл. 14.07.2009; опубл. 27.10.2009. Бюл. № 30.
13. Савченко А. В. Автоматизированная система распознавания людей по фотографиям лиц — Программа для ЭВМ / Свид-во о гос. регистрации № 2009616508 по заявке 2009615314 от 28.09.2009.
14. Савченко А. В. Распознавание изображений методом направленного перебора с применением редукции множества альтернатив // Системы управления и информационные технологии. 2009. № 3(37). С. 40—47.
15. Толчеев В. О. Анализ точностных характеристик модифицированного метода ближайшего соседа // Информационные технологии. 2006. № 4. С. 52—58.

**В. Н. Якимов**, д-р техн. наук, проф.,  
**И. С. Мошков**, аспирант,  
 Самарский государственный  
 технический университет,  
 e-mail: moshkoff@list.ru

## Методика анализа сложных составных терминов в тексте на естественном языке

*Анализируются особенности текстов на естественном языке, описывающих таксономическую структуру. Основной упор сделан на классификацию элементов, из которых состоят термины в тексте. В основу методики анализа терминов заложены критерии принадлежности, по которым можно классифицировать тот или иной элемент сложного составного термина.*

**Ключевые слова:** естественный язык, анализ текста, таксономическая структура

### Введение

В соответствии с увеличением потока информации усложняются задачи автоматизации обработки данных, поступающих из различных текстовых документов. Поэтому актуальность разработки новых и совершенствования известных инструментов для извлечения информации из текста постоянно растет. Одним из способов применения данных инструментов является оценка знаний, содержащихся в тексте [1, 2], которая заключается в сравнении структуры знаний некоторого субъекта с эталоном и может использоваться как средство автоматической обработки результатов открытого тестирования [1]. Однако некоторые особенности текста на естественном языке (неполнота, избыточность, противоречивость) создают трудности в процессе создания инструмента для полноценного анализа текста [4].

Таким образом, возникает потребность в разработке формальных способов анализа текста, которые бы, с одной стороны, позволили проводить автоматический анализ текста, необходимого для оценки знаний, а с другой стороны, позволили упростить анализ за счет введения допустимых ограничений, сохраняющих необходимый уровень качества анализа. Одним из таких ограничений является использование в качестве анализируемого материала текста, описывающего таксономическую структуру. Это обусловлено тем, что практически в любой области науки и техники с точки зрения обеспечения системности требуется проводить структурирование и классификацию имеющихся знаний [1].

Вместе с тем, существующие исследования показывают, что есть взаимосвязь между умением строить правильную классификацию понятий определенной предметной области и умением аргументировано принимать адекватные решения в данной предметной области [1, 4].

### Анализ структурных особенностей текста, описывающего таксономию

Для того чтобы сформулировать требования к формальному аппарату анализа, поделим высказывание на естественном языке (ЕЯ), описывающее таксономию, на отдельные части и определим функции, которые они выполняют в тексте, а также возможные способы их нахождения. Ниже будем использовать высказывание  $\varphi = \langle Obj, L, K, T \rangle$ , где *Obj* — множество простых, сложных или составных терминов; *L* — связи между терминами; *K* — критерии деления терминов; *T* — метаязыковые конструкции, описывающие качественные особенности таксономии. В данном случае под "простым термином" понимается отдельное слово или словосочетание, обозначающее специальное понятие, характерное для определенной предметной области. "Сложный термин" представляет собой термин, построенный на основе более простого термина. Под "составным термином" понимается термин, состоящий из объединения двух или более терминов.

Для определенности в качестве примера будем использовать следующее высказывание: "По химической классификации нефть делится на три основные группы: парафиновые нефти, нафтеновые нефти, ароматические нефти".

В высказываниях на ЕЯ, описывающих таксономию, можно выделить четыре функциональных элемента:

1. *Obj* — описание элементов классификаций (сложных составных терминов (ССТ)). В предложении ССТ, как правило, являются подлежащими и дополнениями, описывающими основную сущность термина, и согласованными с ними дополнениями и определениями, которые задают уточнения месторасположения элемента среди других элементов в таксономии. Таким образом, при условии нахождения главного слова ССТ уже на этапе синтаксического анализа можно определить его границы и структуру. ССТ являются основными элементами таксономии, причем существует некоторая "признаковая" часть, которая позволяет отделить значение одного ССТ от другого. Под значением термина понимается семантическое определение объекта предметной области. Нахождение ССТ в тексте — сравнительно несложная задача. Однако в русском языке, как правило, употребляются осколочные выражения — когда одна или несколько частей ССТ не описаны явно, а лишь подразумева-

ются. Поэтому поиск ССТ в высказывании усложняется, и возникает дополнительная задача восстановления в высказывании полного описания ССТ. Для подтверждения правильности найденной части термина на этапе синтаксического анализа используется формальное описание предметной области экспертом (эталонная таксономия). В используемом примере это слова и словосочетания, называющие термины, относящиеся к классу нефтяных веществ: "нефть", "парафиновые нефти" и т. п.

2. *L* — описание связей между ССТ. В предложении *L* обычно выражаются глагольной группой, а также системой падежей и знаков препинания. Так как в таксономии существуют два вида связей: родитель — дочерний элемент и дочерний элемент — родитель, то для распознавания связей можно использовать лексические шаблоны. При построении субъективной таксономии (таксономии, извлекаемой из анализируемого высказывания) данный элемент носит функциональный характер, связывая извлеченные термины. Существует также тип связей между ССТ, которые создают более сложный термин, выполняя операции сочетания двух ССТ. Причем один из терминов становится частью характеристики другого термина, соответственно, эти термины могут не иметь видовой или композиционной связи. В используемом примере описанием связи является глагольная группа "делится на ...".

3. *K* — критерии деления ССТ. В предложении эти критерии обычно выражаются дополнениями. Критерии деления объясняют принцип деления и сопоставления различных элементов таксономии. В используемом примере критерием является словосочетание "по химической классификации".

4. *T* — описание каких-либо структурных особенностей таксономии. В используемом примере описанием структурных особенностей является словосочетание "три основные группы", что означает намерение описания в субъективной структуре трех групп.

Термины, встречающиеся в высказывании, описывающем таксономию, могут находиться как в отношении вид — подвид, так и быть независимыми в данном отношении. На рис. 1 приведена структурная схема фразы используемого примера.

Для большинства ССТ, встречающихся в таксономических текстах, характерны три составные части [4]. Поэтому зададим структуру термина *obj* как вектор  $obj = \langle o, P, obj' \rangle$ , где *obj* — корневой элемент; *P* — множество признаков корневого элемента; *obj'* — внутренний термин, зависимый от корневого элемента. Для наглядности введем при-

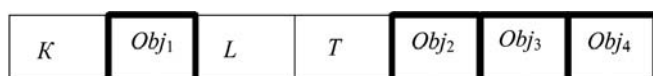


Рис. 1. Структурная схема фразы

мер: "Повреждения рельсов делятся на изгибы, повреждения в шейке, изломы по всему сечению и дефекты подошвы. Изломы бывают поперечными с видимыми пороками и без видимых пороков". Выделим три основные части ССТ.

1. Корневой элемент *o* (ядро ССТ), который на семантическом уровне является классом терминов в эталонной таксономии, куда входит множество зависимых элементов. Элементы данного множества разделяют, используя в их описании различного рода признаки. На синтаксическом уровне это слово, которому подчиняется остальная часть описания термина. Это также означает, что остальная часть грамматически согласована с корневым элементом.

В используемом примере можно выделить два класса терминов:

- "повреждения", "изгибы", "изломы" относятся к одному классу понятий, объединяемых словом "повреждения";
- "рельс", "подошва", "шейка" относятся к классу понятий, объединяемых словом "рельс".

2. Признаковая часть *P*, которая на семантическом уровне является суммой всех признаков, являющихся одним из способов определения занимаемого места среди множества элементов некоторого класса термина. На синтаксическом уровне, как правило, они являются определениями (прилагательными, причастными оборотами, согласованными второстепенными предложениями). Кроме того, в признаковую часть могут входить ССТ, связанные с ядром предложно-падежной конструкцией. В используемом примере признаком является слово "поперечные", относящееся к корневому элементу "излом".

3. Субъект *obj'*, который на семантическом уровне является значением, описываемым фразой, подчиненным ядру. С одной стороны, он является частью родительского термина, а с другой стороны, — самостоятельным значимым термином. Имеет такую же структуру, как и весь ССТ, причем корневой элемент субъекта синтаксически согласован с корневым элементом данного термина. При этом каждый внутренний термин может относиться к различным классам предметной области.

Схематично структура ССТ показана на рис. 2.

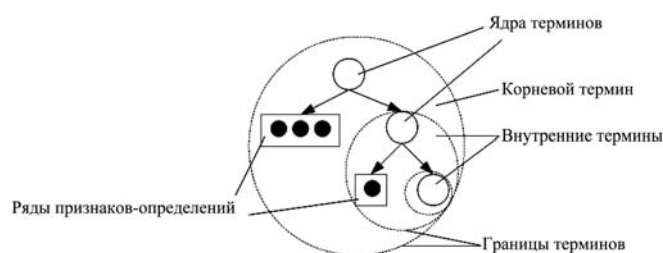


Рис. 2. Пример структуры сложного составного термина

## Построение формального аппарата описания ССТ

Определим входное высказывание  $\varphi$  как упорядоченное множество слов  $\varphi = (sw_1, \dots, sw_n)$ , где  $n$  — число слов в высказывании. У каждого слова есть собственные морфологические характеристики и синтаксическая роль относительно других слов в высказывании. Также каждое слово имеет собственное значение, которое может выражаться в тексте одним словом. Кроме этого, слово может быть частью словосочетания в высказывании, имеющем другое значение и не зависящее от данного слова.

На начальном этапе анализа необходимо определять для слов фразы соответствующие им части речи. Для этого зададим множество основных частей речи  $\eta$ , которое необходимо при анализе границ терминов и их связей  $\eta = \{\eta_{\text{сущ}}, \eta_{\text{прил}}, \eta_{\text{глагол}}, \eta_{\text{пред}}\}$  и введем функцию  $F_\eta(sw)$ , которая возвращает элемент множества  $\eta$  для слова  $sw$ .

Зададим способы определения морфологических характеристик слов, описывающих ССТ. Существует два основных способа морфологического анализа: на основе словаря и на основе морфемного анализа [5]. Для достижения поставленных целей был использован подход на основе создания таблицы всех словоформ, так как он проще в реализации, а предметная область описывается конечным набором слов. Таким образом, словарь  $D$  для слов, описывающих ССТ, представляется как множество словообразующих парадигм слова  $D = \{dw_1, \dots, dw_n\}$ , где  $dw$  — словообразующая парадигма, определяемая как  $dw = \{de_1, \dots, de_m\}$ , а  $de_i^{dw_i}$  — один из видов представления слова в зависимости от рода, числа и падежа. Введем множества, определяющие набор морфологических характеристик для существительных и прилагательных: множество падежей  $\sigma = \{\sigma_0, \dots, \sigma_6\}$ , множество рода  $\tau = \{\tau_0, \tau_1, \tau_2\}$  и множество числа  $\mu = \{\mu_0, \mu_1\}$ .

Введем функции  $F_\sigma:(de) \rightarrow \sigma$ ,  $F_\tau:(de) \rightarrow \tau$ ,  $F_\mu:(de) \rightarrow \mu$ , позволяющие сопоставлять словоформ

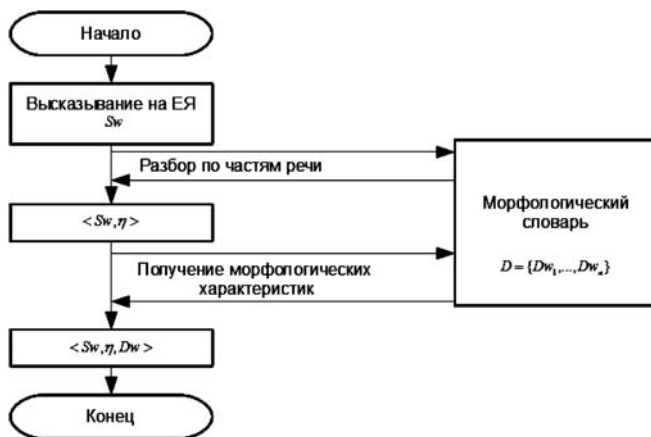


Рис. 3. Схема получения информации о слове в процессе морфологического анализа

му  $de$  слова  $sw_i$  с ее соответствующими характеристиками  $\sigma$ ,  $\tau$ ,  $\mu$  путем перебора множества словоформ. Кроме этого, функция  $F_d$  ставит в соответствие слову  $sw$  в высказывании одну из словоформ морфологического словаря  $de$ . Таким образом, для каждого слова  $sw$ , описывающего термин, можно получить необходимые морфологические характеристики. Схема получения информации о слове в процессе морфологического анализа приведена на рис. 3.

Для того чтобы получить представление о структуре текста и входящих в него терминов, необходимо оперировать с синтаксическими характеристиками. Причем существует взаимосвязь между синтаксической ролью в предложении и местоположением в структуре ССТ. Поэтому введем предикат  $F_{lsync}$ , который определяет лингвистическую согласованность двух слов:

$$F_{lsync}(sw_i, sw_j) \rightarrow \{0, 1\}. \quad (1)$$

Для слов, описывающих ССТ, это означает, что

$$\begin{aligned} S_{w_i} = \sigma_{w_j} \cap \tau_{w_i} = \tau_{w_j} \cap \mu_{w_i} = \mu_{w_j} \rightarrow \\ \rightarrow F_{lsync}(sw_i, sw_j) = 1. \end{aligned}$$

На основе предиката, описанного в работе [1], можно задать предикат определения синтаксического подчинения, который позволит преобразовать упорядоченное множество слов в таксономическую структуру:

$$F_{sl}(sw_i, sw_j) \rightarrow \{0, 1\}.$$

Специальные предикаты позволяют сделать предположения о семантической роли слова, опираясь на синтаксическую информацию. Однако особенности русского языка требуют нескольких критериев определения семантической роли, в том числе на основе заданных (эталонных) значений слова и словосочетания. Для критериев при необходимости можно определять степень значимости и порог реагирования. Введем множество критериев принадлежности  $Kr$ , элементами которого являются предикаты, определяющие принадлежность слова к определенной семантической роли:

$$Kr = \{kr_o^{syn}, kr_p^{syn}, kr_{sub}^{syn}, kr_o^{sem}, kr_p^{sem}, kr_{sub}^{sem}\},$$

где  $kr_o^{syn}$  — синтаксический (полученный на основе синтаксической информации) критерий ядра термина;  $kr_p^{syn}$  — синтаксический критерий признака;  $kr_{sub}^{syn}$  — синтаксический критерий субъекта;  $kr_o^{sem}$  — семантический (полученный на основе значения слова в эталоне) критерий ядра термина;  $kr_p^{sem}$  — семантический критерий признака;  $kr_{sub}^{sem}$  — семантический критерий субъекта.

В общем случае ядро является существительным и не имеет синтаксических зависимостей от других элементов термина. Внутри фразы оно не имеет зависимостей от подлежащего и дополнения. Следовательно, можно обобщить критерий  $kr_o^{syn}$  для слова  $sw_k \in \varphi$ :

$$kr_o^{syn} = 1 \leftrightarrow (F_\sigma(F_d(sw_k)) = \sigma_0) \cup (F_\eta(F_d(sw_k)) = \eta_{сущ}) \cup \left( \bigcup_{i=1}^n (F_{sl}(sw_k, sw_i)) \cup F_\sigma(sw_i) = \sigma_0 \neq 0 \right).$$

Признаки не имеют зависимых слов, поэтому являются терминальными элементами, и критерий  $kr_p^{syn}$  для слова  $sw_k \in \varphi$  задается как

$$kr_p^{syn} = 1 \leftrightarrow (F_\eta(F_d(sw_i)) = \eta_{прил}) \cup \left( \bigcup_{j=1}^n F_{sl}(sw_i, sw_j) = 0 \right).$$

Элемент термина — субъект  $s$  в общем случае является дополнением в косвенном падеже, основным признаком этого элемента является отсутствие подчиненного слова. Поэтому критерий  $kr_{sub}^{syn}$  для слова  $sw_k \in \varphi$  задается как

$$kr_{sub}^{syn} = 1 \leftrightarrow \left( \bigcup_{j=1}^n (F_\sigma(F_d(Sw_k)) \neq \sigma_0) \cup (F_\eta(F_d(Sw_k)) = \eta_{сущ}) \cup F_{sl}(Sw_i, Sw_j) = 0 \right).$$

Дополнение, которое зависит от ядра, и вместе с тем имеет другое зависимое дополнение, образует новый термин  $obj'$  и становится его ядром. При этом как ядро  $o$ , так и простейший элемент  $s$  могут иметь неограниченное множество признаков  $P$ .

Полученные синтаксические критерии являются общими, их можно делить на составные высказывания и вводить систему их значимости. Таким образом, уже на этапе синтаксического анализа можно найти во фразе  $\varphi$  слова, относящиеся к множеству терминов  $obj$  и задать их структуру. На рис. 4 по-



Рис. 4. Схема поиска ССТ в предложении

казана общая схема построения структуры сложного термина.

По смыслу, который несет набор слов, выделяют следующие уровни: слова, словосочетания, предложения, абзаца и текста в целом. Поэтому эталонная система значений должна быть многоуровневой.

Зададим систему значений на уровне слова и построим на этой системе значений множество необходимых для анализа уровней. Так как каждое слово  $sw$  является текстовым выражением определенного значения, то зададим систему, хранящую значения  $sem$  вводимого текста. Для сопоставления множества значений и множества их текстовых выражений введем функцию  $F_{sem}: Sw \rightarrow Sem$  получения значения текстового представления. То есть, если полностью задана система значений, то должно выполняться условие:

$$\forall sw \exists sem, F_{sem}(sw) = sem.$$

При этом данная функция возвращает одно наиболее вероятное значение. Реализация данной функции возможна, так как для составных частей терминов не так ярко выражена проблема омонимии. Причем множество  $Sem$  можно описывать сложной системой значений, которая используется при оценке качества описанной таксономии, так как необходимо учитывать семантические связи между словами.

Для того чтобы оперировать с различными ССТ и его частями, объединим множество значений эталона в необходимую структуру. Так как структура эталонных знаний базируется на структуре субъективных знаний, изложенных в тексте, то обобщим рекурсивную структуру ССТ:

$$obj = \langle P_{obj}, o_{obj}, obj' \rangle.$$

Если термин  $obj$  имеет внутренний термин  $obj'$  со схожей структурой с родительским термином, то имеет собственное ядро  $o_{obj}$ , однако в косвенном падеже, так как оно подчинено родительскому ядру  $o_{obj}$ . Внутренний термин также может иметь свой внутренний термин  $obj''$ , если же его нет, то имеем ядро  $s$ , для которого нет подчиненных слов. Таким образом, получается система следующего вида:

$$obj' = \begin{cases} \langle P_{obj'}, o_{obj'}, obj'' \rangle, & \text{если } obj'' \neq \emptyset \\ \langle P_{obj'}, s_{obj'} \rangle, & \text{если } obj'' = \emptyset, s_{obj'} \neq \emptyset \\ \emptyset, & \text{если } s_{obj'} = \emptyset. \end{cases}$$

Исходя из структуры термина, зададим структуру хранения терминов в эталонной базе знаний. База знаний должна содержать термины, которые образуют таксономическую структуру. Каждый ССТ делится на элементы, являющиеся значениями, для которых задаются возможные текстовые выражения. Подобное деление позволяет задавать отдельное семантическое значение не только для слова, но и для словосочетания (рис. 5). Это позво-

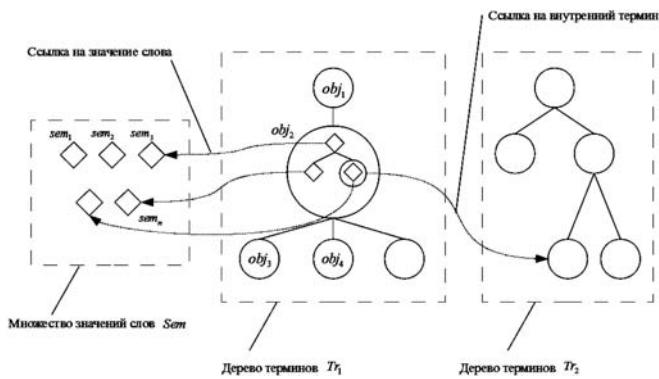


Рис. 5. Схематическое представление хранения значений в эталонной базе знаний

ляет адекватно реагировать на различные именованя одного и того же ССТ.

Введем понятие класса терминов  $\Omega$ , в который входят все термины с одинаковым ядром:

$$\Omega = \{obj_0, \dots, obj_i, \dots, obj_n | o_{obj_i} = o_{obj_j}, i, j = 0 \dots n\}.$$

Так как все термины класса имеют одинаковое ядро, то найденное во фразе ядро будет ассоциироваться с данным классом понятий. Следовательно, если ожидается соответствие между субъективными и эталонными знаниями, то, в первую очередь, в связи с ядром во фразе будут ожидаться элементы ядра в эталонной базе для данного класса.

Выделим семантические критерии, которые позволяют определить местоположения термина во фразе, а также семантическую роль слова. Термин должен присутствовать в эталонной таксономии как класс понятий  $\Omega$ , т. е. является ядром одной из семантик, причем конкретное семантическое значение определяется зависимыми элементами. Таким образом, семантический критерий для термина формулируется как

$$kr_o^{sem}(sem) = 1 \leftrightarrow sem \exists \Omega = \{obj | o^{obj} = sem\}.$$

Если термин  $obj$  содержит в качестве субъекта внутренний термин  $obj'$ , то в эталонной базе знаний должны присутствовать описания обоих терминов, причем в описание общего термина  $obj$  включена ссылка на описание внутреннего термина  $obj'$  как его субъекта  $s^{obj}$ . При этом оба этих термина могут быть как из независимых таксономий, так и из одной таксономии. Таким образом, семантический критерий для субъекта формулируется как

$$kr_{sub}^{sem}(sem) = 1 \leftrightarrow (sem \exists \Omega = \{obj | o^{obj} = sem\}) \cup (sem \exists \Omega^{sem^{obj}} = \{obj | s^{obj} = sem\}).$$

Для подтверждения того, что значение  $sem$  слова  $sw$  является признаком  $P_{obj}$  некоторого термина

$obj$ , нужно найти в эталонной базе знаний множество терминов  $Obj$ , к которым он принадлежит. Среди этого множества терминов предполагается такой, что его появление не нарушает последовательности описания таксономии:

$$kr_p^{sem}(sem) = 1 \leftrightarrow sem \exists Obj = \{obj | P^{obj} = sem\}.$$

Таким образом, введено множество критериев  $Kr$ , позволяющих определить семантическую роль слова, входящего в описание ССТ. Применяя критерии на этапах анализа текста, можно выделить из текста находящиеся в нем термины.

## Заключение

Проанализирована структура высказываний, которые встречаются при описании таксономии в тексте на естественном языке. На основе анализа типичных высказываний для текста данного типа построен формальный аппарат, учитывающий основные особенности способов описания термина в высказываниях. Введены синтаксические и семантические критерии, позволяющие проводить автоматически разбор высказывания и принимать решения о семантической роли того или иного слова высказывания. Разработанный формальный аппарат анализа текстов на естественном языке, описывающих таксономическую структуру, можно рассматривать как методику организации, хранения и управления знаниями, полученными из текста.

Практическая реализация разработанной методики нашла отражение в реализации экспертной системы для тестирования профессиональных знаний студентов старших курсов Лечебного факультета медицинского вуза. Применение системы подтвердило целесообразность преобразования линейного текста в таксономическую структуру на основе предложенной методики для построения структуры субъективных знаний, а также использования полученной структуры в процессе подготовки и оценки знаний специалиста.

## Список литературы

1. Гаврилова Т. А. Базы знаний интеллектуальных систем. СПб.: Питер, 2000. 384 с.
2. Якимов В. Н., Мошков И. С. Определение объектов и их характеристик в процессе обработки текстовой информации // Ресурс- и энергосберегающие технологии и оборудование, экологически безопасные технологии: Материалы Девятой междунар. науч.-техн. конф., Минск, 24–26 ноября 2010 г. Ч. 2. Минск: Белорусский гос. техн. ун-т, 2010. С. 334–337.
3. Знаков В. В. Понимание в познании и общении. Самара: СамГПУ, 2000. 188 с.
4. Солсо Р. Л. Когнитивная психология. М.: Тривола, 1996. 600 с.
5. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы. М.: Академия, 2006. 303 с.

УДК 004.8

**Ю. Г. Бояринов**, канд. техн. наук, доц.,  
**В. В. Борисов**, д-р техн. наук, проф.,  
Филиал ГОУВПО "МЭИ(ТУ)" в г. Смоленске,  
e-mail: BYG@yandex.ru

## Анализ систем и процессов на основе нечетких полумарковских моделей

*Предложена нечеткая полумарковская модель, в которой для задания нечетких вероятностей состояний системы и времен пребывания системы в соответствующих состояниях используются нечеткие отображения на основе нечетких продукционных или реляционных моделей. Разработан метод анализа систем и процессов на основе нечетких полумарковских моделей, позволяющий решать нечеткие оптимизационные задачи. Метод позволяет снизить трудоемкость решения оптимизационных задач за счет компактификации представления и оперирования в нечетком признаковом пространстве и использования аппроксимационных свойств нечетких моделей в разработанной постановке. Представлен пример использования метода для определения оптимального значения нечеткого параметра периодичности проведения проверок системы.*

**Ключевые слова:** нечеткая полумарковская модель, нечеткое отображение, нечеткая оптимизационная задача

### Введение

Одним из направлений анализа систем и процессов в различных состояниях, позволяющим учесть факторы неопределенности анализируемых переменных и случайности событий, является использование полумарковских моделей. Однако вероятностный подход, традиционно используемый для учета стохастической неопределенности, тем не менее, в полумарковских моделях сложных систем не всегда применим вследствие недостатка статистики и экспертного характера большей части информации, а также эвристического, разнокачественного и нечеткого описания параметров, состояний и воздействий на системы и процессы.

Перспективным направлением решения указанных проблем при анализе систем и процессов является использование нечетких полумарковских моделей [1–4].

Вместе с ограничениями существующих нечетких полумарковских моделей предлагаемые методы построения нечетких полумарковских моделей не учитывают наличия избыточности ресурса как необходимого условия функционирования системы, а также отсутствуют постановки и методы решения оптимизационных задач с использованием нечетких полумарковских моделей.

В первом разделе статьи представлена информация для описания полумарковских процессов, во втором — рассмотрена разработанная нечеткая полумарковская модель, в третьем — предложен метод анализа систем и процессов на основе разработанной модели и приведен пример определения оптимального значения нечеткого параметра периодичности проведения проверок сложной системы.

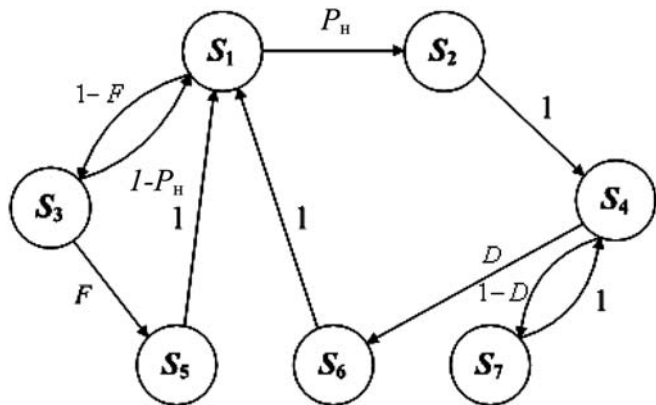
### 1. Использование полумарковских моделей для анализа систем и процессов

Системы, обладающие некоторым уровнем избыточности, с одной стороны, позволяют противостоять отказам, а с другой стороны, могут накапливать соответствующее этому уровню число отказов элементов (неисправностей), не приводящее к отказу системы в целом. Для восстановления этой избыточности в системе проводится "обслуживание", заключающееся в контроле состояния системы и в проведении восстановительных мероприятий.

Основными характеристиками контроля системы являются его *периодичность* и *достоверность контроля*. Взаимное влияние этих характеристик на эффективность функционирования системы обуславливает существование оптимальной периодичности проведения проверок состояния системы. Так, слишком частые проверки приводят к "недоиспользованию" системы, в то же время редкие проверки обуславливают накопление неисправностей элементов системы, что приводит к "недоиспользованию" системы из-за простоев, связанных с ее восстановлением.

Описание процесса функционирования системы представляется полумарковской моделью, структура которой показана на рисунке.

Здесь  $S_1$  — состояние системы при отсутствии неисправностей (желаемое состояние);  $S_2$  — состояние системы при наличии неисправностей элементов, не приводящих к отказу системы;  $S_3$  — состояние проверки исправной системы;  $S_4$  — состояние проверки неисправной системы;  $S_5$  — состояние повторной проверки после ложной реги-



Структура полумарковской модели процесса функционирования системы с учетом достоверности контроля

страции неисправности;  $S_6$  — состояние восстановления исправности системы;  $S_7$  — состояние функционирования системы с необнаруженной неисправностью;  $P_H$  — вероятность возникновения неисправности в системе в диапазоне  $[0, T_{\Pi}]$ ,  $P_H = F_H(T_{\Pi})$ ,  $F_H(T_{\Pi}) = \text{Вер}(t_H < T_{\Pi})$ ,  $F_H$  — функция распределения случайной величины  $t_H$  момента возникновения неисправности;  $T_{\Pi}$  — период проверки состояния системы;  $D$  — вероятность обнаружения неисправности в системе;  $F$  — вероятность ложной регистрации неисправности в системе.

С учетом того, что  $P_H$  является также и вероятностью перехода из состояния  $S_1$  в состояние  $S_2$  за один шаг для вложенной марковской цепи, для нее должно выполняться условие отсутствия последствия как основное условие марковости переходов из состояния в состояние.

Полумарковский процесс задается с помощью матрицы  $\mathbf{F}(t)$  условных функций распределения продолжительности пребывания в состояниях, матрицы  $\mathbf{W}$  переходных вероятностей вложенной марковской цепи и начального состояния процесса, из которого он стартует:

$$\mathbf{F}(t) = \begin{pmatrix} 0 & F_{12}(t) & F_{13}(t) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & F_{24}(t) & 0 & 0 & 0 \\ F_{31}(t) & 0 & 0 & 0 & F_{35}(t) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & F_{46}(t) & F_{47}(t) \\ F_{51}(t) & 0 & 0 & 0 & 0 & 0 & 0 \\ F_{61}(t) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & F_{74}(t) & 0 & 0 & 0 \end{pmatrix};$$

$$\mathbf{W} = \begin{pmatrix} 0 & F_H(T_{\Pi}) & 1 - F_H & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 - F & 0 & 0 & 0 & F & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & D & 1 - D \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Компоненты матрицы  $\mathbf{F}(t)$  запишутся в следующем виде:

$$F_{12}(t) = \begin{cases} 0, & t \leq 0; \\ \frac{F_H(t)}{F_H(T_{\Pi})}, & 0 < t < T_{\Pi}; \\ 1, & t \geq T_{\Pi}; \end{cases} \quad F_{13}(t) = \begin{cases} 0, & t < T_K; \\ 1, & t \geq T_K; \end{cases}$$

$$F_{24}(t) = \begin{cases} 0, & t \leq 0; \\ \frac{F_H(T_{\Pi}) - F_H(T_{\Pi} - t)}{F_H(T_{\Pi})}, & 0 < t < T_{\Pi}; \\ 1, & t \geq T_{\Pi}; \end{cases}$$

$$F_{13}(t) = \begin{cases} 0, & t < T_{\Pi}; \\ 1, & t \geq T_{\Pi}; \end{cases}$$

$$F_{31}(t) = F_{35}(t) = F_{46}(t) = F_{47}(t) = \begin{cases} 0, & t < T_K; \\ 1, & t \geq T_K; \end{cases}$$

$$F_{51}(t) = \begin{cases} 0, & t < T_{\text{ПК}}; \\ 1, & t \geq T_{\text{ПК}}; \end{cases} \quad F_{61}(t) = \begin{cases} 0, & t < T_B; \\ 1, & t \geq T_B; \end{cases}$$

$$F_{74}(t) = \begin{cases} 0, & t < T_{\Pi}; \\ 1, & t \geq T_{\Pi}. \end{cases}$$

Здесь  $T_K$  — математическое ожидание продолжительности контроля;  $T_{\text{ПК}}$  — математическое ожидание продолжительности повторного контроля в случае ложной регистрации неисправности;  $T_B$  — математическое ожидание продолжительности восстановления системы.

Безусловные функции распределения  $F_i(t)$  определяются в соответствии с выражением

$$F_i(t) = \sum_{j=1}^7 F_{ij}(t)w_{ij},$$

где  $w_{ij}$  — вероятность перехода вложенной марковской цепи за один шаг.

В рассматриваемом случае получим

$$F_1(t) = \begin{cases} 0, & t \leq 0; \\ F_H(t), & 0 < t < T_{\Pi}; \\ 1, & t \geq T_{\Pi}; \end{cases} \quad (1)$$

$$F_2(t) = F_{24}(t); \quad F_3(t) = F_4(t) = F_{31}(t); \quad F_5(t) = F_{51}(t); \\ F_6(t) = F_{61}(t); \quad F_7(t) = F_{74}(t).$$

Средние времена пребывания системы в соответствующих состояниях определяются по формуле

$$m_i = \int_0^{\infty} [1 - F_i(t)] dt.$$



Тогда

$$m_1 = \int_0^{T_{\Pi}} [1 - F_H(t)] dt = T_{\Pi} - \int_0^{T_{\Pi}} F_H(t) dt;$$

$$m_2 = \frac{1}{F_H(T_{\Pi})} \int_0^{T_{\Pi}} F_H(T_{\Pi} - t) dt, \quad (2)$$

$$m_3 = m_4 = T_K; m_4 = T_K; m_5 = T_{\text{ПК}}; m_6 = T_B; m_7 = T_{\Pi}.$$

Финальное распределение вероятностей состояний вложенной марковской цепи определяется в результате решения системы уравнений в матричном виде:

$$\mathbf{P} = \mathbf{P}\mathbf{W},$$

где  $\mathbf{P}$  — вектор-строка ( $P_1, \dots, P_7$ ).

Эта система уравнений является линейно зависимой, так как для нее выполняется условие нормировки. Для рассматриваемого случая получим:

$$\begin{cases} P_2 = F_H(T_{\Pi})P_1; \\ P_3 = [1 - F_H(T_{\Pi})]P_1; \\ P_4 = \frac{1 - F_H(T_{\Pi})}{D} P_1; \\ P_5 = [1 - F_H(T_{\Pi})]FP_1; \\ P_6 = [1 - F_H(T_{\Pi})]P_1; \\ P_7 = \frac{1-D}{D} [1 - F_H(T_{\Pi})]P_1; \\ 1 = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7. \end{cases}$$

Решения этой системы уравнений имеют вид

$$P_1 = \frac{1}{2\left[1 + \frac{F_H(T_{\Pi})}{D}\right] + F[1 - F_H(T_{\Pi})]} = \frac{1}{A};$$

$$P_2 = \frac{F_H(T_{\Pi})}{A}; P_3 = \frac{1 - F_H(T_{\Pi})}{A};$$

$$P_4 = \frac{1 - F_H(T_{\Pi})}{AD}; P_5 = \frac{F[1 - F_H(T_{\Pi})]}{A};$$

$$P_6 = \frac{1 - F_H(T_{\Pi})}{A}; P_7 = \frac{F_H(T_{\Pi})(1 - D)}{AD}.$$

В стационарном режиме распределение вероятностей состояний для полумарковского процесса в целом определяется из выражения

$$\pi_i = \frac{P_i m_i}{\sum_{j=1}^7 P_j m_j}.$$

В соответствии с этим выражением вероятность пребывания системы в состоянии  $S_1$  рассчитывается следующим образом:

$$\pi_1 = \frac{T_{\Pi} - \int_0^{T_{\Pi}} F_H(t) dt}{(T_{\Pi} + T_K)\left[1 - F_H(T_{\Pi}) + \frac{F_H(T_{\Pi})}{D}\right] + T_{\text{ПК}}[1 - F_H(T_{\Pi})]F + T_B F_H(T_{\Pi})}. \quad (3)$$

## 2. Нечеткая полумарковская модель

В предлагаемой нечеткой полумарковской модели функционирования системы реализован предложенный в работе [5] способ введения нечеткости, заключающийся в следующем:

- вероятности состояний и времен пребывания системы в соответствующих состояниях представляются нечеткими числами (нечеткими множествами);
- для задания нечетких переменных, характеризующих вероятности состояний и времен пребывания системы в соответствующих состояниях, используются нечеткие отображения в виде нечетких продукционных или реляционных моделей;
- операции суммирования, произведения и деления над вероятностями состояний и времен пребывания системы в соответствующих состояниях заменяются на расширенные нечеткие операции суммирования, произведения и деления нечетких чисел соответственно.

Рассмотрим основные этапы построения предлагаемой нечеткой полумарковской модели.

**Этап 1.** Задание входных нечетких параметров системы.

Для задания терм-множеств нечетких параметров и нечетких переменных и построения их логико-лингвистических шкал целесообразно использовать типовые  $L$ – $R$ -функции, например, колоколообразного типа. Так, например, для нечеткого параметра  $\tilde{T}_{\Pi}$  значение  $C_{\tilde{T}_{\Pi}}$  задается следующим образом:

$$\mu_{C_{\tilde{T}_{\Pi}}}(\tilde{T}_{\Pi}) = \exp\left[-\frac{1}{2}\left(\frac{\tilde{T}_{\Pi} - a}{b}\right)^2\right],$$

где  $a, b$  — параметры функции принадлежности колоколообразного типа.

**Этап 2.** Задание нечетких переменных  $\tilde{P}_i$  на базовых множествах значений вероятностей  $P_i$  состояний системы.

При этом значения вероятностей  $P_i$  состояний системы являются базовыми множествами, на которых определены функции принадлежности нечетких множеств, число которых для каждого состояния системы может быть различным. Для наглядности будем использовать по три нечетких множества

{Низкий, Средний, Высокий} для задания каждой нечеткой переменной  $\tilde{P}_i$  соответственно: для  $\tilde{P}_1$  —  $H_{P1}, C_{P1}, B_{P1}$ ; для  $\tilde{P}_2$  —  $H_{P2}, C_{P2}, B_{P2}$ ; для  $\tilde{P}_3$  —  $H_{P3}, C_{P3}, B_{P3}$ .

**Этап 3.** Задание нечетких переменных  $\tilde{m}_i$  на базовых множествах значений вероятностей  $m_i$  времен пребывания системы в соответствующих состояниях.

Значения вероятностей времен  $m_i$  пребывания системы в соответствующих состояниях являются базовыми множествами, на которых определены функции принадлежности нечетких множеств нечетких переменных  $\tilde{m}_i$ .

Так же как и для нечетких переменных  $\tilde{P}_i$  задаются нечеткие множества для переменных  $\tilde{m}_i$  соответственно: для  $\tilde{m}_1$  —  $H_{m1}, C_{m1}, B_{m1}$ ; для  $\tilde{m}_2$  —  $H_{m2}, C_{m2}, B_{m2}$ ; для  $\tilde{m}_3$  —  $H_{m3}, C_{m3}, B_{m3}$ .

**Этап 4.** Задание отображений для нечетких переменных  $\tilde{m}_i$ , характеризующих нечеткие вероятности времен пребывания системы в соответствующих состояниях.

В общем случае нечеткие отображения для нечетких переменных  $\tilde{m}_i$  задаются следующим образом:

$$\tilde{m}_i = \tilde{\varphi}_i(\tilde{T}_K, \tilde{T}_\Pi, \tilde{T}_B, \tilde{T}_{PK}, \tilde{F}_H(\tilde{T}_\Pi)), i = 1, \dots, N. (4)$$

Для рассматриваемого на рисунке примера в соответствии с выражениями (1) и (2) зададим следующие нечеткие отображения для нечетких переменных  $\tilde{m}_i$ :

$$\begin{aligned} \tilde{m}_1 &= \tilde{\varphi}_1(\tilde{T}_\Pi), \tilde{m}_2 = \tilde{\varphi}_2(F_H(\tilde{T}_\Pi)), \tilde{m}_3 = \tilde{\varphi}_3(\tilde{T}_K), \\ \tilde{m}_4 &= \tilde{\varphi}_4(\tilde{T}_K), \tilde{m}_5 = \tilde{\varphi}_5(\tilde{T}_{PK}), \tilde{m}_6 = \tilde{\varphi}_1(\tilde{T}_B), \\ \tilde{m}_7 &= \tilde{\varphi}_7(\tilde{T}_\Pi), \end{aligned}$$

где  $\tilde{\varphi}_i$  — нечеткое отображение, заданное одним из известных способов: нечеткими продукционными правилами, нечеткими отношениями.

Рассмотрим пример реализации отображения  $\tilde{\varphi}_3$  для переменной  $\tilde{m}_3$  на основе нечеткой продукционной модели SISO-типа (SingleInputs, SingleOutputs), в предположении, что входной параметр  $T_K$  также является нечетким и задается соответствующими нечеткими множествами  $\tilde{T}_K$  —  $H_K, C_K, B_K$ ; для  $\tilde{T}_B$  —  $H_B, C_B, B_B$ :

П<sub>1</sub>: ЕСЛИ  $\tilde{T}_K$  есть  $H_K$ , ТО  $\tilde{m}_3$  есть  $H_{m3}$ ;

П<sub>2</sub>: ЕСЛИ  $\tilde{T}_K$  есть  $C_K$ , ТО  $\tilde{m}_3$  есть  $C_{m3}$ ;

П<sub>3</sub>: ЕСЛИ  $\tilde{T}_K$  есть  $B_K$ , ТО  $\tilde{m}_3$  есть  $B_{m3}$ .

Далее для переменной  $\tilde{m}_3$  в рамках сформированной нечеткой продукционной модели реализуется алгоритм нечеткого вывода Мамдани.

**Этап 5.** Задание отображений для нечетких переменных  $\tilde{P}_i$ , характеризующих нечеткие вероятности состояний системы.

В общем случае нечеткие отображения для нечетких переменных  $\tilde{P}_i$  определяются следующим образом:

$$\tilde{P}_i = \tilde{\phi}_i(\tilde{T}_\Pi, \tilde{F}, \tilde{F}(T_\Pi), \tilde{D}), i = 1, \dots, N, (5)$$

где  $\tilde{\phi}_i$  — нечеткое отображение, реализуемое, например, на основе нечетких продукционных моделей MISO-типа (MultipleInputs, SingleOutputs).

Далее для переменных  $\tilde{P}_i$  в рамках сформированных нечетких продукционных моделей реализуется алгоритм нечеткого вывода Мамдани.

Если переменная  $T_\Pi$  является четкой, то для реализации нечеткого отображения  $\tilde{\phi}_i$  целесообразно использовать нечеткую функцию от этой четкой переменной  $T_\Pi$ , формирующей образ ее четкой области определения в соответствующем нечетком множестве  $\tilde{\phi}_i: T_\Pi \rightarrow \tilde{P}_i$ .

Следует отметить, что так как понятия нечеткой функции четкой переменной и нечеткого отношения соответствуют друг другу в математическом смысле, то нечеткая функция  $\tilde{\phi}_i$  может быть интерпретирована как нечеткое отношение  $\tilde{R}$  и определена следующим образом:

$$\begin{aligned} \text{для } \forall (t_\Pi, p_i) \in T_\Pi \times P_i; \\ \mu_{\tilde{\phi}}(P_i) = \mu_{\tilde{R}}(t_\Pi, p_i). \end{aligned}$$

**Этап 6.** Определение значений нечетких переменных, характеризующих распределение вероятностей состояний для нечеткого полумарковского процесса.

В основе предлагаемой нечеткой полумарковской модели лежит следующее выражение, вычисляемое либо на основе нечеткого интервального метода, либо с использованием принципа расширения Л. Заде:

$$\tilde{\pi}_i = (\tilde{P}_i \tilde{\otimes} \tilde{m}_i) \tilde{\int} \left( \bigotimes_{j=1}^3 (\tilde{P}_j \tilde{\otimes} \tilde{m}_j) \right), (6)$$

где  $\tilde{\pi}_i$  — нечеткая переменная, характеризующая распределение нечетких вероятностей состояний для нечеткого полумарковского процесса ( $i = 1, \dots, N$ );  $\tilde{m}_i$  и  $\tilde{P}_i$  — нечеткие переменные, задаваемые отображениями (4) и (5);  $\tilde{\otimes}$ ,  $\tilde{\int}$  — операции нечеткого сложения, умножения и деления соответственно [5].

### 3. Метод анализа систем и процессов на основе нечетких полумарковских моделей

Помимо основных задач анализа систем и процессов с использованием нечетких полумарковских моделей, заключающихся, прежде всего, в нахождении распределения нечетких вероятностей состояний для нечеткого полумарковского процесса, разработанная модель позволяет решать нечеткие оптимизационные задачи.

Постановка и решение нечеткой оптимизационной задачи в общем случае сводится к определению входных нечетких параметров системы, обеспечивающих формирование требуемого (экстремального) значения выходного нечеткого параметра.

Следует отметить, что решение нечеткой оптимизационной задачи в указанной постановке тесно связано с задачей нечеткого обратного вывода.

Существуют различные подходы к выполнению задачи нечеткого обратного вывода, наиболее известным из которых является метод, предложенный Е. Санчесом в работе [6], в основе которого лежит решение нечетких уравнений. Однако данному методу присущи ограничения на параметры нечеткого отображения.

Постановка нечеткой оптимизационной задачи с использованием нечеткой полумарковской модели формулируется следующим образом.

Пусть заданы нечеткие отображения для нечетких переменных  $\tilde{m}_i$ , характеризующих нечеткие вероятности времен пребывания системы в соответствующих состояниях, на основе выражения (4).

Пусть заданы нечеткие отображения для нечетких переменных  $\tilde{P}_i$ , характеризующих нечеткие вероятности состояний системы, в соответствии с выражением (5).

Пусть заданы нечеткие переменные  $\tilde{\pi}_i$ , характеризующие распределение нечетких вероятностей состояний для нечеткого полумарковского процесса, в соответствии с выражением (6).

Таким образом, необходимо для нечетких выходных переменных  $\tilde{\pi}_i$  ( $i = 1, \dots, N$ ) найти такие значения нечетких входных параметров системы  $\{\tilde{T}_k, \tilde{T}_\pi, \tilde{T}_b, \tilde{T}_{pk}, \tilde{D}, \tilde{F}, \tilde{F}_n(\tilde{T}_\pi)\}$  как по отдельности, так и в их сочетании, которые обеспечивали бы формирование требуемых (экстремальных) значений этих нечетких выходных переменных.

Рассмотрим далее в качестве примера постановку и решение оптимизационной задачи определения оптимального значения  $T_\pi^*$  нечеткого параметра

$\tilde{T}_\pi$  системы при отсутствии неисправностей, т. е. находящейся в состоянии  $S_1$ :

$$\left\{ \begin{array}{l} \tilde{m}_i = \tilde{\phi}_i(\tilde{T}_k, \tilde{T}_\pi, \tilde{T}_b, \tilde{T}_{pk}, \tilde{F}(\tilde{T}_\pi)), i = 1, \dots, N; \\ \tilde{P}_i = \tilde{\phi}_i(\tilde{T}_\pi, \tilde{F}, \tilde{F}(\tilde{T}_\pi), \tilde{D}), i = 1, \dots, N; \\ \tilde{\pi}_1 = (\tilde{P}_1 \otimes \tilde{m}_1) \tilde{\gamma} \left( \bigoplus_{j=1}^N (\tilde{P}_j \otimes \tilde{m}_j) \right); \\ \tilde{\pi}_1 \xrightarrow{\tilde{T}_\pi = T_\pi^*} \max; \\ T_\pi^* = \arg \max_{\tilde{T}_\pi} (\tilde{\pi}_1). \end{array} \right.$$

Предлагаемый метод состоит из двух основных этапов.

На первом этапе задаются отображения (4) и (5) нечетких переменных  $\tilde{m}_i$  и  $\tilde{P}_i$  ( $i = 1, \dots, N$ ) на основе нечетких продукционных или реляционных моделей. После этого для рассматриваемой задачи определения оптимальной периодичности проведения проверок системы данный этап завершается построением модели оценки нечеткой вероятности  $\tilde{\pi}_1$  нахождения системы в состоянии  $S_1$  в соответствии с выражением (6).

На втором этапе, собственно, реализуется решение нечеткой оптимизационной задачи с использованием построенной модели.

Для этого сначала интервал базового множества, на котором задан нечеткий параметр  $T_\pi$ , разбивается на  $N$  отсчетов с шагом дискретизации

$$\gamma = \frac{T_{\pi(\max)} - T_{\pi(\min)}}{N},$$

где  $T_{\pi(\min)}$ ,  $T_{\pi(\max)}$  — нижняя и верхняя границы базового множества нечеткого параметра  $T_\pi$ .

Затем для каждого из отсчетов  $T_{\pi j}$  ( $j = 1, \dots, N$ ) при фиксированных значениях остальных входных нечетких параметров системы  $\{\tilde{T}_k, \tilde{T}_b, \tilde{T}_{pk}, \tilde{D}, \tilde{F}\}$  определяются значения нечетких переменных  $\tilde{m}_i, T_{\pi j}$  и  $\tilde{P}_i, T_{\pi j}$  ( $i = 1, \dots, N$ ) на основе нечетких отображений (4) и (5) соответственно.

Далее на основе полученных значений  $\tilde{m}_i, T_{\pi j}$  и  $\tilde{P}_i, T_{\pi j}$  с использованием модели (6) находятся значения выходной нечеткой переменной  $\tilde{\pi}_1, T_{\pi j}$  ( $j = 1, \dots, N$ ) нечеткой полумарковской модели.

На заключительном шаге этапа на основе полученных значений выходной нечеткой переменной

$\tilde{\pi}_{i, T_{pj}}$  ( $j = 1, \dots, N$ ) определяется значение  $T_{pj}^*$ , удовлетворяющее следующему условию:

$$T_{pj}^* = \arg \max_{T_{pj}} (\tilde{\pi}_{i, T_{pj}}).$$

Предлагаемый метод позволяет также найти сочетания значений нечетких параметров системы, обеспечивающих формирование требуемых значений нечетких выходных переменных.

В дополнение к вышесказанному следует отметить, что созданный метод позволяет существенно снизить трудоемкость решения оптимизационных задач за счет, во-первых, компактификации представления и оперирования в нечетком признаковом пространстве, во-вторых, использования аппрок-

симационных свойств нечетких моделей в разработанной постановке.

#### Список литературы

1. **Bhattacharyya M.** Fuzzy Markovian decision process // Fuzzy Sets and Systems. 1998. Vol. 99. P. 273–282.
2. **Praba B., Sujatha R., Srikrishna S.** Fuzzy reliability measures of fuzzy probabilistic semi-Markov model // Int. Journal of Recent Trend in Engineering. 2009. Vol. 2, N 2. P. 25–29.
3. **Praba B., Sujatha R., Srikrishna S.** A study on homogeneous fuzzy semi-Markov model // Applied Mathematical Sciences. 2009. Vol. 3, N 50. P. 2453–2467.
4. **Бояринов Ю. Г., Борисов В. И., Мищенко В. И., Дли М. И.** Метод построения нечеткой полумарковской модели функционирования сложной системы // Программные продукты и системы. 2010. № 3(91). С. 26–31.
5. **Moore R. E.** Interval analysis. New Jersey: Prentice Hall, 1966.
6. **Sanchez E.** Resolution of composite fuzzy relation equations // Inform. Contr. 1976. Vol. 30. P. 38–48.

УДК 004.021

**И. А. Кузьмина**, аспирант,  
МГТУ им. Н. Э. Баумана,  
e-mail: kuzminainna@yandex.ru

## Генетические методы синтеза сетей энергоснабжения

*Предлагаются алгоритмы синтеза структуры и распределения трафика в сетях энергоснабжения. Для решения используется сочетание эвристических и генетических методов.*

**Ключевые слова:** сети энергоснабжения, оптимизация проектирования, перспективное развитие, генетические методы, метод комбинирования эвристик

### Введение

В настоящее время невозможно представить себе жизнь современного города без надежного функционирования сети энергоснабжения. Сетью энергоснабжения является предназначенная для передачи и распределения электрической энергии совокупность подстанций, распределительных узлов и соединяющих их линий электропередач.

При проектировании сетей энергоснабжения в пределах большого города важно обеспечить экономное использование дорогостоящих узлов сети (подстанций, кабельных линий и т. д.), обеспечить надежное снабжение электрической энергией всех потребителей, а также снизить затраты на обслуживание самой сети.

Решение данной задачи подразумевает рассмотрение следующих вопросов:

1. Выбор топологии сети — системы линий связи между узлами и местоположение новых подстанций сети. При этом следует минимизировать финансовые затраты на построение и обслуживание сети.

2. Распределение трафика между кабельными линиями. От интенсивности потока мощности зависят требования к пропускной способности линий и их сечению, а следовательно, и к стоимости их прокладки.

Задачи определения топологии сетей различного рода и распределения в них трафика относятся к сложным задачам структурного синтеза. Ввиду больших вычислительных затрат, необходимых для решения задач практической размерности, применение к ним точных методов дискретной оптимизации не представляется возможным. В связи с чем рациональным является применение генетических методов, так как данный подход может обеспечить получение решений, близких к оптимальным, при приемлемых затратах машинного времени.

Задача проектирования отказоустойчивых сетей синхронной иерархии, заключающаяся в синтезе топологии сети передачи данных и распределении трафика в ней, рассматривалась в работе [1]. Задачи формулировались как задачи оптимизации, для их решения использовались элементы эвристических и генетических методов.

Однако в предложенном в работе [1] варианте решения задачи используются хромосомы излишней длины. Длину хромосомы можно уменьшить за счет применения иного варианта переноса трафика из хорд в ребра каркаса. Кроме того, имеются еще три особенности, отсутствующие в сетях син-

хронной иерархии, рассмотрение которых необходимо при решении задачи синтеза энергосети [4], а именно:

- необходимость наличия не менее двух независимых маршрутов питания каждой подключаемой подстанции от двух распределительных подстанций (РП);
- каждая кабельная линия в случае рассмотрения сети энергоснабжения может передавать поток только в одном направлении;
- в ходе решения возможно возникновение необходимости возведения новых узлов сети — РП.

В данной статье предлагается подход к решению задач синтеза структуры и распределения трафика сетей энергоснабжения. Задачи формулируются как задачи оптимизации, для их решения используются элементы эвристических и генетических методов.

В первой части статьи дана постановка задачи синтеза. Далее описываются алгоритмы отдельных этапов решения задачи, предлагаемые эвристические критерии выбора варианта переноса маршрута и генетический метод поиска оптимального варианта.

### Постановка задачи

Рассматривается следующая постановка задачи. Заданы:

- $\mathbf{R}$  — множество узлов (вершин) сети типа РП;  $\mathbf{T}$  — множество узлов (вершин) сети типа трансформаторная подстанция (ТП);  $\mathbf{Y}$  — множество узлов типа ТП, неподключенных к существующей сети,  $\mathbf{Y} \subset \mathbf{R}$ ;  $N_R, N_T, N_Y$  — мощности множеств  $\mathbf{R}, \mathbf{T}$  и  $\mathbf{Y}$  соответственно. В общем случае линия состоит из двух параллельно проложенных кабельных линий; каждая линия связи, соединяющая узлы  $i$  и  $j$ , характеризуется пропускной способностью  $S_{\max ij}$ ;
- трафик  $\mathbf{S}$  передачи электроэнергии от узла типа РП к узлу типа ТП, трафик между узлами может проходить через несколько транзитных узлов; в рассматриваемой задаче трафик  $\mathbf{S}$  представлен в виде матрицы, ее элемент  $S_{ij}$  есть трафик между узлами  $i$  и  $j$ , равный сумме всех потоков, проходящих через  $(i, j)$  кабельную линию соответственно;
- матрица  $\mathbf{C}$  стоимостей передачи электроэнергии,  $C_{ij}$  — стоимость передачи единицы электроэнергии между узлами  $i$  и  $j$ , пропорциональная расстоянию между узлами  $i$  и  $j$ .

Нужно минимизировать общие затраты на эксплуатацию и построение сети

$$Z = \sum_{\substack{i=N_R+N_T+N_Y+N_J \\ j=N_T+N_Y \\ i=0, j=0}} C_{ij} S_{ij} + N_R C_R \rightarrow \min, \quad (1)$$

где  $N_R$  — число РП, "возводимых" в ходе решения задачи для обеспечения надежного функциониро-

вания сети энергоснабжения;  $C_R$  — стоимость строительства одной РП.

Варьируемыми параметрами в данной задаче являются трафик  $S_{ij}$  между узлами  $i$  и  $j$  соответственно и число возводимых РП  $N_R$ .

Описанная задача включает в себя ряд подзадач.

Во-первых, это задачи синтеза варианта сети, т. е. варианта подключения новых узлов к сети энергоснабжения с оценкой качества синтезированного варианта по критерию минимизации затрат на строительство новых узлов сети.

Во-вторых, в общую задачу входит задача выбора оптимального варианта распределения трафика и выбор основного маршрута питания для каждой подключенной трансформаторной подстанции из нескольких возможных вариантов. При этом в качестве целевой функции используется критерий равномерности загрузки кабельных линий и отсутствия перегрузок.

Рассмотрим алгоритм последовательного решения указанных задач более подробно.

### Классификация ребер графа

Алгоритм синтеза сети начинается с процедуры классификации (ранжирования) ребер графа.

Следует отметить, что построение сети энергоснабжения происходит не "с нуля", а на основе дополнения существующей сети энергоснабжения новыми элементами сети.

Учитывая данную особенность, примем, что ребрами первого ранга являются ребра уже существующей сети энергоснабжения. Ребро второго ранга — это хорда  $(i, j)$  такая, что имеются вершина  $k$  и ребра  $(j, k)$  и  $(k, j)$ , являющиеся ребрами первого ранга. Хорда  $p$ -го ранга — это хорда  $(i, j)$  такая, что имеются вершина  $k$  и ребра  $(i, k)$  и  $(k, j)$ , являющиеся ребрами ранга меньше  $p$ , причем одно из них имеет ранг  $p - 1$ .

Для учета направленности кабельных линий предполагаем для ребер, существование которых невозможно в связи с существующим встречным потоком,  $p = \infty$ .

Ребра, соединяющие подключаемые ТП с существующей сетью энергоснабжения, имеют ранг  $p = 0$ .

Результаты классификации хорд составляют матрицу  $\mathbf{Q}$ , ее элемент  $q_{ij} = p_{ij}$ , где  $p_{ij}$  — ранг связи  $(i, j)$ .

Матрица  $\mathbf{Q}$  классификации ребер используется далее в основной части алгоритма синтеза сети, в которой осуществляется перенос трафика из хорд в ребра сети (существующие кабельные линии). Перенос выполняется последовательно для всех хорд, соединяющих неподключенные подстанции сети с распределительными подстанциями. Вариант переноса трафика осуществляется по эвристическим правилам, описанным ниже.

Поток из хорды  $(i, j)$  переносится в хорды или ребра  $(i, k)$   $(l, j)$  и хорду  $(k, l)$ , если:

$$1. q_{ik} = 1, q_{lj} \leq 1, q_{kl} < p_{ij}, C_{lk} < C_{ij}.$$

2.  $q_{ik} = 1, q_{lj} \leq 1, C_{ik} + C_{lj} + C_{ij}/2 \leq C_{ij}$ .
3.  $q_{ik} \leq 1, q_{lj} \leq 1, k = l$ ; если указанный вариант не существует, то среди всех возможных вариантов выбирается хорда минимальной длины ( $C_{kl} = \min$ ) и строится новое ребро ( $k, l$ ).
4.  $q_{ik} = 1, q_{lj} \leq 1, q_{ik} = \min_{m=0..M}(q_{im})$ , где  $M$  — число ребер, исходящих из  $i$ -го узла.
5.  $q_{ik} = 1, q_{lj} \leq 1, C_{lj} = \min_{m=0..T}(C_{Tj})$ , где  $T$  — число ребер, входящих в  $j$ -й узел.

Очевидно, что здесь приведены лишь несколько возможных вариантов эвристик.

### Учет пропускной способности линий при выборе маршрута распределения трафика

Как было сказано ранее, обеспечение надежного энергоснабжения при условии неисправности любой из линий связи, требует питания каждой трансформаторной подстанции по двум независимым маршрутам от двух распределительных подстанций.

Существующий в нормальном режиме трафик требуется обеспечить при условии резервного питания одной из подстанций.

При этом затраты должны быть минимизированы.

Требуется не только удовлетворение экономического критерия минимизации затрат, но и выполнения условия отсутствия перегрузки кабельных линий.

На предыдущем шаге алгоритма для каждой подключаемой подстанции сети было несколько вариантов питания. На данном этапе требуется определить, питание от какой из распределительных подстанций является основным, а от какой — резервным.

Для принятия решения можно использовать эвристические критерии, например, для каждой подстанции выбрать маршрут минимальной длины или включающий в себя наименьшее число подстанций (для минимизации потерь электроэнергии при ее транспортировке).

Однако при таком подходе возможно пересечение маршрутов и увеличение потока через кабельную линию выше допустимого значения.

Тогда для всех указанных случаев следует выбрать в качестве основных иные маршруты передачи данных (для всех подстанций, маршрут питания которых проходит через перегруженные участки), либо перераспределить поток с перегруженных линий на другие линии сети посредством переноса маршрутов питания ряда подстанций на другие линии связи.

Пример фрагмента сети энергоснабжения приведен на рис. 1 (см. третью сторону обложки).

### Генетический алгоритм поиска решения

Для решения задачи синтеза структуры сети энергоснабжения методами генетического поиска необходимо определить, какие параметры сети будут являться генами хромосомы.

Положим, что хромосома будет состоять из  $P = N_1 + N_2$  генов, где  $N_1$  — число хорд (несуществующих линий связи);  $N_2$  — число существующих линий связи сети. Каждый ген соответствует трафику между парой узлов  $i$  и  $j$ , а значением гена является номер линии, в которую перенесен данный трафик. Таким образом, пространство поиска характеризуется числом возможных случайно генерируемых вариантов, равным  $P^P$ , причем подавляющее большинство вариантов распределения трафика окажется некорректным вследствие нарушения ограничений связности маршрутов передачи данных. Очевидно, что решение оказывается практически неосуществимым.

### Метод комбинирования эвристик

Эффективность метода комбинирования эвристик для решения сложных задач структурного синтеза, в том числе подобных рассматриваемой задаче синтеза топологии и распределения трафика в сетях, доказана в ряде работ [2, 3].

Применение данного метода существенно повышает эффективность поиска по сравнению с использованием традиционной для генетических методов постановки задачи. В этом случае мощность множества возможных вариантов существенно меньше и не превышает  $h^P$ , где  $P = N_3 + 1$ ;  $h$  — число эвристик;  $N_3$  — число вариантов присоединения трансформаторных подстанций к распределительным узлам сети. При этом  $N_3$  существенно меньше  $N_1$ , так как число возможных вариантов подключения ограничивается максимальной удаленностью подстанций друг от друга.

Это связано с тем, что при данном подходе хромосома состоит из  $P = N_3 + 1$  генов. Такое уменьшение размера хромосомы в сравнении с классическим вариантом связано с тем, что каждый ген в данном случае представляет собой способ выполнения очередного шага переноса трафика из хорд, соединяющих подключаемые подстанции с распределительными подстанциями сети, в существующие линии сети и другие хорды. Значениями генов (аллелями) являются номера эвристик, выбираемых для выполнения очередного шага переноса трафика.

После того как определено множество маршрутов питания сети, необходимо для каждой подключаемой подстанции выбрать из них основной и резервные маршруты. Выбор указанных маршрутов следует проводить по эвристическим критериям.

Предложенный алгоритм решения задачи схематично изображен на рис. 2.

После определения согласно выбранным эвристикам структуры сети проводится проверка способности линий электропередач обеспечить передачу возросших потоков мощности и рассчитывается целевая функция (в случае удовлетворения значений потоков имеющимся ограничениям). Тем самым

получен и оценен очередной экземпляр в популяции хромосом в рамках генетического алгоритма.

В случае превышения значений передаваемых потоков над значениями пропускной способности некоторых линий электропередачи, следует применить к данному потоку процедуру локального поиска посредством операций кроссовера или мутации, тем самым перенаправив поток из перегруженного участка по другим линиям сети, т. е. применить так называемый гибридный алгоритм, включающий как процедуры генетического, так и локального поиска. Недостатком указанного метода является то, что при каждом выполнении операций кроссовера и мутации требуется проверка ограничений пропускных способностей линий генерируемых потоков (вариантов сети), следовательно, общее число обращений к процедуре будет весьма значительным. Одной из проблем генетического поиска является разработка алгоритмов, эффективных как по количеству проводимых вычислений, так и по степени приближения к оптимальному решению задачи. Отсутствие необходимости расчетов значения целевой функции "неработоспособных" вариантов сети существенно повышает скорость работы алгоритма за счет уменьшения количества проводимых вычислений.

### Заключение

В работе предложены алгоритмы синтеза структуры сети энергоснабжения и распределения трафика передачи потоков мощности при подключении новых узлов к существующей энергосети. Приведены алгоритмы синтеза и распределения потока, основанные на классификации ребер сети с последующим переносом потоков мощности. Основой

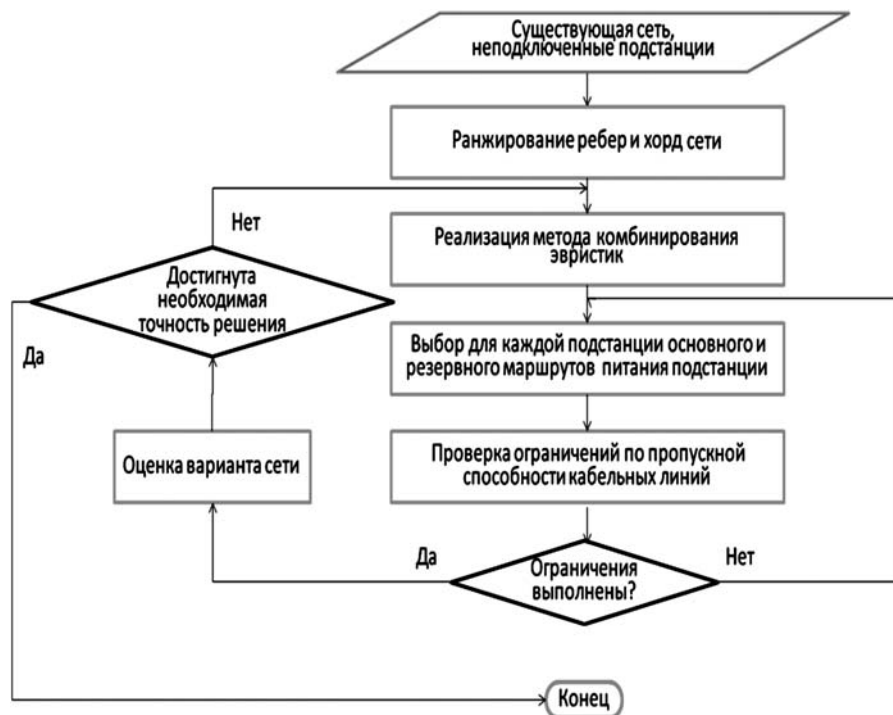


Рис. 2. Блок-схема алгоритма решения задачи синтеза сети и распределения потока

рассматриваемого в статье подхода является метод комбинирования эвристик. Предложенные в статье подходы могут быть также использованы для решения задач синтеза иных сетей, характеризующихся ограничениями пропускных способностей линий и направленностью передачи потоков.

### Список литературы

1. Норенков И. П., Трудоношин В. А., Федорук В. Г. Задачи синтеза сетей синхронной иерархии // Информационные технологии, 2003. № 10. С. 17–21.
2. Норенков И. П. Эвристики и их комбинации в генетических методах дискретной оптимизации // Информационные технологии, 1999. № 1. С. 2–7.
3. Филиппова А. С. Моделирование эволюционных алгоритмов решения задач прямоугольной упаковки на базе технологии блочных структур // Приложение к журналу "Информационные технологии", 2006. № 6. 32 с.
4. Булатов Б. Г. САПР и модели оптимального развития энергосистем. Учебное пособие по курсовому проектированию. — Челябинск: Изд. ЮУрГУ, 2004. 47 с.

**Д. В. Ефанов,**

канд. техн. наук, доц., вед. науч. сотр.,

Федеральное государственное  
унитарное предприятие

"Центральный научно-исследовательский институт  
экономики, информатики и систем управления",

**А. М. Ибрагимов,** аспирант НИЯУ "МИФИ",

e-mail: ibragimov\_arsen@mail.ru

## Динамические условные знаки и способы их реализации в современных геоинформационных системах

*Предложен обобщенный способ формализации широкого спектра динамических условных знаков (по сути, дается формальное определение класса таких знаков), а также логическая архитектура их представления в современных географических информационных системах.*

**Ключевые слова:** динамические условные знаки, географическая информационная система, ГИС, представление информации на картах, метеорологическая обстановка, графическое представление географических данных, нанесение динамических условных знаков, представление динамических условных знаков в СУБД, архитектура динамических условных знаков, формализация динамических условных знаков

### Введение

В последнее время значительно возросли объемы разнородной информации, которую необходимо обрабатывать в оперативном режиме. Ужесточились также и требования к скорости ее анализа и скорости реакции на события, связанные с подобной информацией. При этом обычно источники, предоставляющие необходимые данные, не сгруппированы в одном месте, а распределены по огромной территории (иногда десятки тысяч километров). Все это требует от различных государственных служб (Министерство Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий; Федеральная служба по гидрометеорологии и мониторингу окружающей среды и др.) создания распределенных глобальных автоматизированных систем управления (РГ АСУ).

Чтобы обеспечить указанным выше службам возможность оперативного наблюдения и координации сил при ликвидации чрезвычайных ситуаций, составлении карт метеорологической обстановки определенных районов и т. д., РГ АСУ подобного рода должны иметь в своем составе географическую информационную систему (ГИС).

Проблема заключается в том, что все современные ГИС строились в расчете на использование либо в задачах географии, геодезии, геологии, топографии и т. д., либо в задачах, связанных с сетями и планированием маршрутов, а поэтому приспособлены для работы только со статическими, почти не меняющимися во времени данными, такими как рельеф местности, архитектурные сооружения, дорожные сети, и т. д. [1]. Действительно, скорость модификации таких объектов крайне мала и измеряется годами, а порой — десятилетиями. В противоположность этому, для оперативных служб необходимо решать задачи в режиме реального времени.

### Постановка задачи

Под картографическим условным знаком понимаются графические символы, с помощью которых на карте обозначают вид объектов, их местоположение, форму, размеры, качественные и количественные характеристики [2].

Условные знаки подразделяются на три основные группы [2]:

- точечные знаки представляют собой просто картинку (векторные или растровые), которые обозначают, что в данном месте карты находятся некоторые объекты (нефтяные вышки, месторождения полезных ископаемых, АЭС и т. д.). Иногда размер картинки может отражать некоторые параметры, например мощность АЭС, всегда немасштабные, т. е. их размер никак не изменяется при изменении масштаба карты;
- линейные (одномерные) знаки — это по существу статические векторные линии с повторяющимися вдоль них узорами различного вида (пунктир, штрих-пунктир и т. д.). Подобные знаки используются, например, для изображения транспортных систем. Они масштабны по длине, но немасштабны по ширине;
- площадные (двумерные) знаки служат для представления статических сплошных объектов, таких как страны, провинции, здания и т. д. Они могут быть заполнены различными видами заливок и штриховок, всегда масштабны и позволяют точно определить площадь объектов.



## Формализация динамических знаков

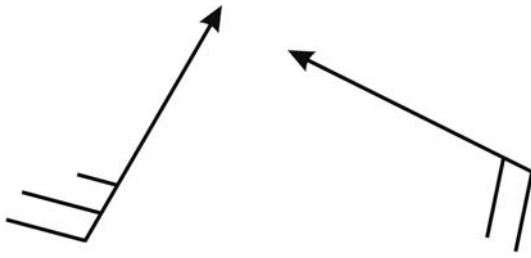


Рис. 1. Стрелки, указывающие направление и силу ветра (справа 10 м/с, слева — 12 м/с)

Для координации усилий на местности на карту наносится специальная символика, при работе с которой необходима поддержка динамических условных знаков. Поясним суть указанной проблемы. Все современные ГИС реализуют на базовом уровне поддержку только трех указанных выше статических типов знаков [1]. Их вполне достаточно для полного решения задач географии, топографии и т. д. [2, 3]. Однако в других прикладных научных областях зачастую необходимо использовать знаки с более широкими возможностями, например, в метеорологии — линии холодных и теплых фронтов; стрелки, показывающие направления распространения лесных пожаров или направление и силу ветра и т. д.

В связи с этим рассмотрим пример одного из простейших метеорологических знаков, который невозможно эффективно реализовать с использованием указанных выше трех стандартных типов (рис. 1). Этот знак представляет собой совокупность стрелки, показывающей направление ветра и одного или нескольких перьев разной длины, задающих его силу (каждое длинное перо добавляет 5 м/с, короткое — 2 м/с) [4].

Очевидно, что даже при наличии возможности вращения знака на карте (и тем самым возможности указания направления ветра), полная его реализация стандартными средствами современных ГИС невозможна, так как потребовала бы предварительного создания и использования в качестве шаблона всех возможных типов хвостов (всех возможных значений силы ветра). Очевидно, что такой подход является недопустимым. Кроме того, более сложные типы знаков (см. ниже), используемых на отечественных картах, невозможно реализовать стандартными средствами даже с использованием подобных ухищрений.

Сформулируем следующие задачи, связанные с реализацией динамических знаков:

- разработать обобщенный способ формализации широкого спектра динамических знаков (формально определить динамический условный знак);
- разработать логическую архитектуру представления динамических условных знаков в современных ГИС.

В целях разъяснения предложенного способа формализации воспользуемся условным знаком более общего вида (рис. 2), чем был рассмотрен ранее.

Данный знак обычно используется на исторических картах военных действий для обозначения направлений удара войск [5]. Его основой являются несколько главных линий (на рисунке  $A_1B_1A_2$  и  $A_1B_2A_3$ ), которые соединяют начальную ( $A_1$ ) и конечные ( $A_2$  и  $A_3$ ) точки удара. Вспомогательные точки  $B_1$  и  $B_2$  задают соответственно кривизну линий  $A_1B_1A_2$  и  $A_1B_2A_3$ , а меняя положение вспомогательной точки  $C$ , можно указать толщину всей стрелки и прямую ( $A_1C$ ), на которой лежат ее исходные точки (обозначена на рисунке пунктирной линией). Сбоку от стрелки размещается дополнительная текстовая информация, например, название наступающей армии.

Условный знак на рис. 2 содержит в себе все типы элементов, необходимые для построения динамического знака произвольного вида:

- опорные точки ( $A_1$ ,  $A_2$  и  $A_3$ ), указывающие направление удара;
- вспомогательные точки ( $B_1$ ,  $B_2$  и  $C$ ), задающие некоторую дополнительную информацию (в данном случае кривизну линий  $A_1B_1A_2$  и  $A_1B_2A_3$ , толщину всей стрелки и расположение ее исходных точек);
- набор свойств, связанных со знаком (в данном случае текст надписи сбоку от стрелки);
- функция отображения  $f_R$ , позволяющая по вышеприведенным данным построить изображение динамического знака.

На представленной основе можно построить динамический знак любой сложности. Действительно, каждый такой знак должен содержать некоторое множество точек, которые характеризуют его основ-

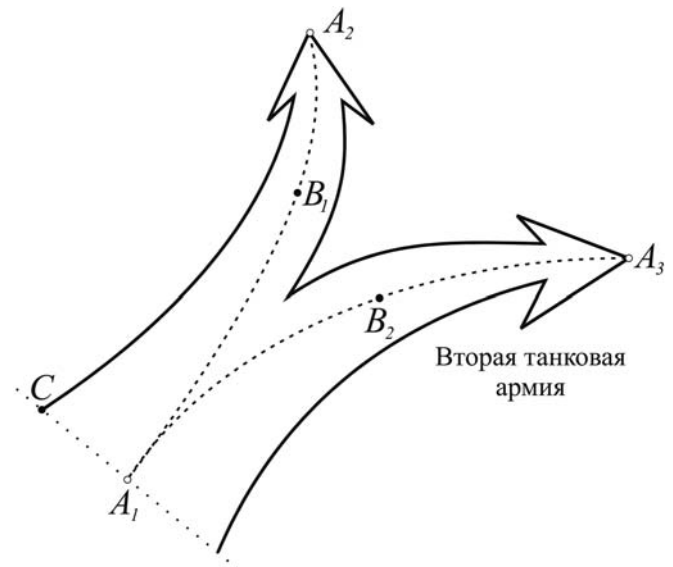


Рис. 2. Динамический условный знак, обозначающий направления удара

ные свойства (опорные точки), дополнительную форму или способ привязки к карте (вспомогательные точки). Кроме того, со знаком может быть связана дополнительная информация различного рода (набор свойств), такая как цвет, текст надписей и т. д. Для рисования знака на основе всех перечисленных выше данных используется функция  $f_R$ .

Таким образом, приходим к следующему формальному определению динамического знака произвольного вида.

*Динамическим условным знаком назовем следующую совокупность:*

$$S = \langle O, D, C, f_R(O, D, C) \rangle,$$

где  $O = \{o_i\}_{i=1}^n$  — конечное упорядоченное множество опорных точек знака;  $D = \{d_i\}_{i=1}^m$  — конечное упорядоченное множество вспомогательных точек знака;  $C = \{c_i\}_{i=1}^k$  — конечное упорядоченное множество свойств, связанных со знаком;  $f_R(O, D, C)$  — некоторая функция отображения, позволяющая по заданным множествам  $O$ ,  $D$  и  $C$  получить визуальное представление знака.

### Представление динамических знаков в базе географических данных

База географических данных (БГД) содержит в себе все данные, необходимые ГИС для работы с конкретной предметной областью, поэтому совершенно естественно будет использовать ее и для хранения полной информации о динамических условных знаках.

Некоторая совокупность таких знаков, нанесенных на карту, представляет собой законченное описание ситуации из конкретной предметной области (например, полную метеорологическую обстановку конкретного региона или план тушения лесных пожаров в данном округе). Каждую такую совокупность знаков естественно хранить на отдельном слое карты. Кроме того, слои со знаками не должны закрываться географической информацией, а значит, их всегда следует располагать выше всех остальных слоев карты.

Удобно для каждого слоя с динамическими условными знаками завести в БГД отдельную таблицу стандартной структуры, содержащую всю информацию о размещенных на нем знаках. Данный подход позволяет минимизировать накладные расходы, а также дает возможность более прозрачно организовать структуру хранения данных. Каждый конкретный знак при этом будет задаваться отдельной строкой таблицы.

Условимся хранить географическую точку как совокупность двух пространственных координат  $x$  и  $y$  в некоторой географической проекции.

Для представления слоя с динамическими условными знаками в БГД предлагается использовать

следующую структуру минимально необходимых полей таблицы.

1. *Идентификатор условного знака SID (signidentifier)* позволяет выделить из таблицы строку, содержащую информацию о конкретном условном знаке. Значения поля должны быть уникальными как минимум в пределах одной таблицы, но лучше, если они будут уникальными среди всех динамических условных знаков данной БГД. Из вышесказанного следует, что поле SID можно использовать в качестве первичного ключа таблицы. Кроме того, это поле позволяет связать с каждым динамическим знаком некоторые дополнительные данные, хранящиеся в других таблицах. Наиболее подходящим типом данных для рассматриваемого поля является целое.

2. *Тип условного знака S\_TYPE (signtype)* представляет собой тип шаблона динамического условного знака, на основе которого строится его конкретный вид (экземпляр), записанный в строке таблицы. По сути, тип шаблона определяется конкретным видом функции отображения  $f_R$ , которая используется для его рисования. Это следует из того, что именно последняя определяет способ рисования знака, размерности множеств опорных  $O$  и вспомогательных  $D$  точек, а также размеры и типы элементов множества свойств  $C$ . Для представления в таблице рассматриваемого поля удобно использовать целый тип данных, однако для этого необходимо предварительно инъективно отобразить каждый возможный тип динамического знака  $f_R$  на некоторое целое число.

3. *Множество опорных точек S\_BASE\_PT (signbasepoints)* естественно было бы представить массивом, содержащим последовательно значения координат  $x_1, y_1, \dots, x_n, y_n$ . Однако такой тип данных поддерживается не всеми современными СУБД, поэтому самым удобным в данном случае оказывается бинарное представление вида  $\tilde{x}_1 \oplus \tilde{y}_1 \oplus \dots \oplus \tilde{x}_n \oplus \tilde{y}_n$ , где  $\tilde{x}_i$  и  $\tilde{y}_i$  — соответственно бинарное представление координат  $i$ -й точки, а  $\oplus$  — оператор сцепления бинарных представлений, подобный оператору конкатенации строк. Для быстрой упаковки и распаковки данных, представленных таким способом, необходимо реализовать соответственно встраиваемые функции PT\_ARRAY\_TO\_BIT() и PT\_ARRAY\_FROM\_BIT().

Другим вариантом представления множества опорных точек в БГД является их представление в виде бинарного массива, формат которого идентичен формату массива точек, используемого исходной ГИС. Рассмотрим преимущества данного способа. Обычно каждый слой данных в ГИС хранится с использованием своей собственной проекции (однако проекции разных слоев могут не совпадать). При использовании вместе нескольких слоев для их отображения выбирается единый тип про-

екции, а это значит, что слои, использующие тип проекции, отличный от выбранного, должны быть конвертированы. Так как функции конвертации между различными проекциями в ГИС уже реализованы, то использование второго способа представления массива опорных точек позволяет не выполнять ту же самую работу заново. Кроме того, к этим массивам применимы и различные инструменты географической обработки, позволяющие, например, измерять расстояния между точками и т. д.

4. *Множество вспомогательных точек  $S\_SUB\_PT$  (signsubsidiarypoints)* можно хранить в БГД, используя те же способы, которые были описаны выше для основных точек.

5. *Множество свойств  $S\_PROP$  (signproperties)* позволяет хранить различные дополнительные данные, связанные с динамическим знаком. Их эффективное представление возможно только в бинарном виде и разрабатывается в зависимости от типов данных, которые необходимо использовать. В целом же подход напоминает предлагавшийся выше первый способ представления множества опорных точек.

Следует отметить, что нет необходимости заводить отдельные поля для хранения размерностей множеств основных и опорных точек, а также множества свойств, так как все они определяются типом динамического условного знака  $f_R$ .

## Заключение

В статье дается подробное описание обобщенного способа формализации динамических знаков, а также логической архитектуры их представления в современных географических информационных системах.

По мнению авторов, приведенная формализация может с успехом использоваться для реализации широкого спектра визуальных редакторов условных знаков, позволяющих в режиме реального времени и с минимальными затратами усилий наносить оперативную обстановку для координации различного рода действий (планы тушения лесных пожаров, ликвидации чрезвычайных ситуаций, взаимодействия войск при наступлении и т. д.).

## Список литературы

1. ArcGIS Web-based Help. URL: <http://resources.arcgis.com/content/web-based-help>
2. Берлянт А. М. Картография. М.: Аспект Пресс, 2002. 336 с.
3. Условные знаки, образцы шрифтов и сокращения для топографических карт масштабов 1:25 000, 1:50 000, 1:100 000 / Под ред. В. А. Бабичева и др. М.: Военно-топографическое управление генерального штаба, 1983.
4. Васильев К. П. Что должен знать судоводитель о картах погоды и состояния моря. Л.: Гидрометеоздат, 1980.
5. Помбрик И. Д., Шевченко Н. А. Карта офицера. М.: Воениздат, 1985. 176 с.

**ФОРУМ  
БЕЗОПАСНОСТИ И СВЯЗИ**

**Казань**

**1-3  
февраля  
2012**

**17-я специализированная выставка  
БЕЗОПАСНОСТЬ**

**6-я специализированная выставка  
ОХРАНА И БЕЗОПАСНОСТЬ ТРУДА**

**15-я специализированная выставка  
СВЯЗЬ**

Выставочный центр "Казанская ярмарка"  
Россия, 420059, г. Казань, Оренбургский тракт, 8  
Тел./факс: (843) 570-51-16, 570-51-11-круглосуточно  
E-mail: [pdv@expokazan.ru](mailto:pdv@expokazan.ru)  
[http:// www.exposvyaz.ru](http://www.exposvyaz.ru), [www.exposecurity.ru](http://www.exposecurity.ru), [www.expokazan.ru](http://www.expokazan.ru)

ВЫСТАВОЧНЫЙ ЦЕНТР  
ISO - 9001  
КАЗАНСКАЯ  
ЯРМАРКА

**О. Е. Архипова**<sup>1</sup>, канд. техн. наук, ст. науч. сотр.,  
**В. В. Бойко**<sup>3</sup>, канд. техн. наук, глав. специалист,  
**Г. В. Ковалева**<sup>1</sup>, канд. биол. наук, ст. науч. сотр.,  
**В. А. Москаленко**<sup>4</sup>,

канд. техн. наук, глав. специалист,

**Т. Т. Тарасова**<sup>2</sup>, канд. соц. наук, ст. науч. сотр.,  
 e-mail: arkhipova@ssc-gas.ru

<sup>1</sup> Институт Аридных зон,

Южный научный центр РАН,

<sup>2</sup> Институт социально-экономических  
и гуманитарных исследований,

Южный научный центр РАН,

<sup>3</sup> Комитет охраны окружающей среды  
Администрации Ростовской области,

<sup>4</sup> ФГУ Земельная кадастровая палата  
по Ростовской области

## Методические подходы к оценке природных и антропогенных изменений Азово-Черноморского побережья

*Цель исследования — определение рекреационной емкости, выявление факторов социально-экологической чувствительности и оценка устойчивости развития прибрежных экосистем южных морей. Задача решается на основе использования геоинформационного подхода к созданию систем мониторинга социально-экологического состояния рекреационной зоны и оценки влияния увеличения рекреационной емкости курортов на выявленные факторы. Представленная статья описывает методологический подход определения уровня допустимых рекреационных нагрузок на побережье Азовского и Черного морей, опирающийся на использование методов системного анализа, математического моделирования, ГИС-технологий.*

**Ключевые слова:** системный подход, устойчивое развитие, прибрежная зона, информационные технологии, математическое моделирование, индикаторы

### Введение

В современных исследованиях российских и зарубежных ученых представлены различные методы решения проблемы оценки допустимых нагрузок на прибрежные экосистемы, на природные ресурсы, как основной потребляемый ресурс туристической отрасли [12, 13, 15, 19, 20]. Однако проблема выработки научных подходов, выбора системы индикаторов и эффективных технологий не решена полностью ни в одной, даже высокоразвитой стране мира. Для решения задач управления прибрежной зоной во многих странах Европы и в России стали применять так называемый индикаторный подход,

который заключается в комплексном анализе параметров, характеризующих различные сферы человеческой и природной активности.

На сегодняшний момент разработано и активно используется большое число индикаторов устойчивого развития прибрежных зон и внутренних территорий. Необходимо отметить, что проблемы прибрежных зон Российской Федерации достаточно близки к проблематике развитых прибрежных зарубежных стран. Поэтому ориентация на мировой опыт интегрированного управления прибрежными экосистемами является вполне оправданной. Проводимое исследование опирается на концепты и положения теории территориальных рекреационных систем [12, 15].

### 1. Постановка задачи

В мировой практике широко и достаточно давно используется понятие "прибрежная зона". Предложено множество разных определений. Но при всем их многообразии их объединяет одна общая идея: *прибрежная зона это пространство, где с особой интенсивностью осуществляется взаимодействие человека с окружающей средой* [22].

Прибрежная зона сама по себе является ценным природным объектом. Многообразие местообитаний на суше и в прибрежной акватории моря обеспечивает высокое разнообразие животного и растительного мира, включая редкие виды, занесенные в Красную Книгу.

Необходимость оптимизации развития современной рекреационной отрасли и малоизученность ее социальной составляющей обусловили выбор предметной области исследования. Основным потенциалом рекреационного комплекса является высокое качество окружающей среды. Тем не менее, расширение курорта и увеличение его пропускной способности ведет к увеличению интенсивности рекреационного давления, а также к росту сельского хозяйства и промышленного производства на территории курортной зоны. Негативное воздействие нагрузки на экосистемы проявляется только в том случае, если ее уровень значительно больше, чем возможности среды справиться с последствиями, не выходя за рамки приемлемых изменений [20]. Бесконтрольный процесс развития рекреационных зон создает потенциальную угрозу природным территориям, оказывает огромное давление на прибрежные экосистемы. Результатом этого является нарушение природных экосистем, загрязнение окружающей среды, что в конечном итоге приводит к снижению стоимости курортных и рекреационных систем.

Предлагаемый в статье подход к оценке природных и антропогенных изменений природного комплекса позволит определить закономерности их формирования и развития, взаимосвязи между ними, выявить антропогенные изменения не только в отдельных компонентах, но и экосистеме в целом.

## 2. Решение проблемы

Рекреационная сфера авторами рассматривается как система, состоящая из следующих взаимосвязанных подсистем: природно-ресурсный потенциал курорта; рекреационная инфраструктура; социальная среда (рис. 1). Рекреационная нагрузка оказывает влияние на все элементы системы. Привлекательный природный потенциал вызывает рост рекреационной инфраструктуры и увеличивает приток рекреантов. Это, в свою очередь, приводит к увеличению нагрузки на природные ресурсы. Определенная оптимальной емкости позволяет выявить сбалансированное развитие всего комплекса без оказания негативного воздействия на среду.

Решение этой задачи включает разработку методологической основы оценки рекреационной емкости и допустимой нагрузки. Созданная на ее основе интегрированная система оценки устойчивого развития прибрежных экосистем южных морей будет включать в себя систему оценки текущего состояния прибрежных экосистем, комплекс математических моделей для оценки устойчивости развития рекреационной прибрежной зоны, определения оптимальной рекреационной емкости, не приводящей к ухудшению состояния прибрежных экосистем.

Предлагаемая методология информационного обеспечения комплексного мониторинга и прогноза устойчивого развития курорта на основе комплекса математических моделей позволит оценить влияние на экосистему принятых решений по развитию курорта.

**Предметная область исследования** — приморские рекреационные районы Краснодарского и Ростовского макрорайонов, т. е. районы, относящиеся к прибрежным экосистемам Черного и Азовского морей [7].

Применение предлагаемого подхода к задаче оценки рекреационной емкости курортов Азовского и Черноморского побережий позволит определить предельно допустимые нагрузки на прибрежные экосистемы Азовского и Черного морей.



Рис. 1. Рекреационная нагрузка и рекреационная емкость природно-рекреационного комплекса

Курортно-рекреационная зона Черноморского и Азовского побережий состоит из трех взаимосвязанных систем: природные ландшафты, речные экосистемы и морские экосистемы. Образующиеся отходы сельскохозяйственного и промышленного производства рекреационной инфраструктуры загрязняют все компоненты окружающей природной среды, участвуют в природных циклах миграции и могут быть перенесены в моря и реки. В связи с этим необходимо объединить три основные области — **исследование состояния окружающей среды на суше, изучение рек и прибрежной части морей**, состояние которых зависит от потока материала с прилегающих территорий. Основные методы экологического исследования следующие: методы оценки качества воздушной среды; геохимические методы оценки качества среды; метод интегральной оценки состояния водной экосистемы.

С точки зрения социально-экономических процессов прибрежная зона — естественный ресурс для развития рекреационно-туристической индустрии. Развитие курорта способствует увеличению числа рабочих мест за счет развития инфраструктуры отдыха и развлечений, способствует увеличению доходов местного населения, предоставляющего различные услуги отдыхающим. В связи с этим важным элементом системы оценки рекреационной емкости прибрежной территории также являются социологические методы.

**Методы оценки воздушной среды.** В качестве "управляемых" факторов риска здоровью человека определен круг факторов, характеризующих состояние окружающей среды курорта: качество атмосферного воздуха; микроклиматический комфорт; акустический комфорт.

Для оценки влияния качества воздушной среды построена система критериев. При этом использовались два вида критериев:

- векторный критерий, компонентами которого являются различные характеристики окружающей среды; каждая компонента вектора меняется при изменении антропогенной нагрузки на курорт, либо при изменении варианта его развития;
- скалярный критерий, являющийся интегральной обобщенной характеристикой качества окружающей среды на единичном интервале  $[0, 1]$ .

В качестве интегрального критерия построен индекс, который называется экологическим индексом качества окружающей среды или экоиндексом ( $EI$ ). Составляющими экоиндекса или частными критериями являются индекс комфортности микроклимата ( $IK$ ), индекс чистоты атмосферного воздуха ( $IA$ ), индекс акустического комфорта ( $IAK$ ):

$$EI = \{IK, IA, IAK\}. \quad (1)$$

Одной из проблем интегральных оценок является необходимость учитывать неодинаковую важность частных параметров. Как правило, эту проблему решают введением коэффициентов весомости па-

раметров. Каждому параметру методом экспертных оценок присваивается весовой коэффициент, отражающий его важность в конкретной ситуации.

Рассмотрим механизм построения частных критериев, составляющих экоиндекс для рекреационной зоны, на примере индекса комфортности микроклимата. Среди рекреационных ресурсов, в первую очередь, рассматривается климат, поскольку основные рекреационные центры являются климатическими курортами. Рекреационная оценка климата состоит в изучении зависимости состояния организма человека от воздействия климатопогодных факторов. Для определения степени воздействия этих факторов строится индекс комфортности, отражающий отклик организма человека на изменение метеорологических факторов, составляющими компонентами которого являются индекс комфортности температуры ( $IKT$ ), индекс комфортности влажности ( $IKW$ ), индекс комфортности скорости ветра ( $IKV$ ):

$$IK = \{IKT, IKW, IKV\}. \quad (2)$$

Выявление самого факта влияния состояния окружающей среды и особенно степени влияния на показатели состояния здоровья возможно лишь тогда, когда известны уровни этих показателей в условиях отсутствия фактора риска для здоровья. На практике это осуществляется путем изучения аналогичных показателей состояния здоровья у населения контрольных районов, имеющих такое качество среды, при котором показатели состояния здоровья населения можно принять за относительную "норму". При этом микроклиматические факторы окружающей среды находятся в "зоне комфорта", санитарно-гигиенические факторы не превышают предельно-допустимых концентраций и уровней. Индекс качества окружающей среды по отдельному фактору для контрольного района принимается за единицу. В общем случае индекс — это отношение показателя в расчетном районе к его значению в контрольном районе.

Описанная система критериев дает возможность получить совокупность оценок, характеризующих состояние окружающей среды рекреационной зоны, оценить различные варианты планов развития курорта и выбрать наилучший вариант из числа допустимых [14].

**Методы оценки геохимического загрязнения.** В основе геохимических исследований рекреационной зоны лежит оценка распределения тяжелых металлов и других загрязняющих веществ, концентрирующих загрязнение: почвенный, снеговой и растительный покров, подземные и поверхностные воды. Происходящие в почве процессы определяют химический состав поверхностных и грунтовых вод и в значительной мере газовый состав в приземной атмосфере. Примеси загрязняющих веществ осаждаются на растительный покров и с осадками попадают в почву, поэтому почвы считаются индикатором оценки качества рекреационной зоны [3].



Рис. 2. Схема миграционной модели

В основе миграционной геохимической модели лежит схема миграции загрязняющих веществ (рис. 2).

Динамика загрязняющих веществ в атмосфере описывается следующей моделью:

$$\begin{cases} \frac{dM_a}{dt} = V_p - V_o - V_g - V_r - V_b; \\ \frac{dM_p}{dt} = V_o - V_f - V_u, \end{cases} \quad (3)$$

где  $M_a$  — запас загрязняющих веществ в атмосфере;  $M_p$  — запас загрязняющих веществ в почве;  $V_p$  — масса поступивших в атмосферу загрязняющих веществ;  $V_o$  — масса загрязняющих веществ, осажденных на поверхность земли;  $V_g$  — масса загрязняющих веществ, участвующая в глобальном рассеивании загрязняющих веществ;  $V_r$  — масса загрязняющих веществ, связанных растительностью;  $V_b$  — масса загрязняющих веществ, поглощенных человеком и биотой;  $V_f$  — масса загрязняющих веществ, поступающих с поверхности земли в грунтовые воды за счет процесса фильтрации;  $V_u$  — масса загрязняющих веществ, вынесенных с поверхности земли поверхностным стоком (ливневые и талые воды).

Поступление загрязняющих веществ в атмосферу определяется из выражения

$$V_p = V_t + V_d + V_m, \quad (4)$$

где  $V_t$  — масса загрязняющих веществ, поступающая с техногенными выбросами в атмосферу;  $V_d$  — масса загрязняющих веществ, поступающая в атмосферу в результате процесса дефляции;  $V_m$  — масса загрязняющих веществ, поступающая в атмосферу в результате сжигания растительной биомассы и мусора.

Частным случаем миграционной модели загрязнения урбанизированной территории для оценки динамики распространения загрязняющих веществ является балансовая геохимическая модель загрязнения курортной территории, использующая имею-

щиеся данные о многолетних наблюдениях загрязнения почвы, поверхностных вод, атмосферных выпадений на территории курорта. Опираясь на полученные в результате расчета данные, можно проводить анализ, синтез эколого-геохимической информации, прогнозировать ситуацию, осуществлять выбор стратегий управления [8, 9].

**Методы оценки качества водной среды.** Водная экосистема — природный объект, являющийся неделимым и взаимосвязанным целым среды и биоты. Для того чтобы охарактеризовать состояние водной экосистемы, необходимо знать показатели воды, как среды обитания и показатели биотической (организменной) части экосистемы. Конкретному состоянию водной экосистемы соответствует определенный уровень способности к самоочищению. Он может быть охарактеризован сочетанием показателей среды и биоты, а их значения, в свою очередь, выражаются в определенных интервалах значений [5, 6]. На способность к самоочищению и, следовательно, на состояние водной экосистемы, влияет множество факторов (как внешнего воздействия, так и внутренних закономерностей саморазвития его водной экосистемы). Интенсивность их воздействия отражается не только на обилии развивающихся в водоеме организмов, но и на их видовом составе. Именно эти характеристики используют в биоиндикационных методах [6, 18].

Биотическая часть экосистемы организована в виде трофической пирамиды. В водной экосистеме — это водоросли, которые создают органическое вещество для формирования второго трофического уровня — беспозвоночных животных, которые, в свою очередь, являются объектом питания для рыб и морских млекопитающих. Соответственно, зная состояние первого трофического уровня, можно предполагать состояние всей трофической пирамиды, что подтверждается многолетними исследованиями [5].

Среди водорослей известны индикаторы различных параметров среды (рН среды, солености, органического загрязнения и пр.) [10]. Индикаторными являются сравнительно небольшое число видов, но они самые распространенные и массовые. Оценка состояния водных экосистем, основанная на биоиндикации с помощью водорослей, позволяет сравнивать водные объекты разного типа, расположенные в различных регионах [6]. Интегральным показателем воздействия суммы загрязнений является органическое загрязнение. Для оценки степени органического загрязнения водоемов и водотоков (около 3900 индикаторных таксонов) в России и странах ближнего зарубежья наиболее широко применяется метод Пантле—Бука в модификации Сладечека [16]. Индекс органического загрязнения по сообществу водорослей рассчитывают с использованием формулы

$$S = \frac{\sum sh}{\sum h}, \quad (5)$$

где  $S$  — степень сапробности сообщества водорослей;  $s$  — сапробное значение организма-сапробионта;  $h$  — частота встречаемости сапробионта в пробе.

Еще один метод оценки органического загрязнения, основанный на экологической характеристике сообществ диатомовых водорослей, был разработан японскими специалистами [23—25]. Предложенный ими индекс носит название "индекс диатомового комплекса органического загрязнения воды" (*DAIpo* — *diatom assemblage index to organic water pollution*). По системе Т. Ватанабе [23—25] для каждого вида диатомовых водорослей определен индекс толерантности  $D$  (*tolerance index for organic water pollution of diatom taxon*). Порядок значения величины  $D$  является показателем степени толерантности видов диатомовых водорослей к органическому загрязнению (чем меньше  $D$ , тем более устойчив организм к загрязнению). Чем меньше *DAIpo*, тем больше органическому загрязнению подвергался водоем.

Используя метод Т. Ватанабе (около 1000 индикаторных таксонов) индекс органического загрязнения можно рассчитать по формуле

$$DAIpo = 100 - \sum_{i=1}^m S_i - 0,5 \sum_{j=1}^n E_j, \quad (6)$$

где  $\sum_{i=1}^m S_i$  — сумма относительных встречаемостей сапрофилов на станции, %;  $m$  — число сапрофилов на станции;  $\sum_{j=1}^n E_j$  — сумма относительных встречаемостей эврисапробов на станции, %;  $n$  — число эврисапробов.

Метод Пантле—Бука дает большое количество информации по оценке органического загрязнения водоемов и водотоков, причем в многолетнем аспекте [5]. Сопоставление индексов, рассчитанных обоими методами, значительно расширяет возможности для интерпретации данных, получаемых по разным системам оценки качества вод.

**Методы оценки состояния социальной среды рекреационных зон.** Рекреационная сфера является одним из важнейших факторов устойчивого развития большинства стран и регионов мира. Вместе с тем, подавляющее большинство исследователей рекреации и туризма отмечают, что Россия все еще остается в стороне от мировых тенденций развития рекреационной сферы [4, 21].

Кластерный анализ специализированных и обобщающих работ показал, что социальные аспекты рекреационной сферы прибрежных зон южных морей России так и не стали предметом самостоятельного исследования. Рекреационная система — сложная социально управляемая система, в центре которой стоит человек, а точнее — различные группы отдыхающих и население рекреационных зон, взаимосвя-

занных главной целевой функцией — наиболее полное удовлетворение рекреационных потребностей.

Функционирование и развитие всех составных частей рекреационной системы осуществляется на основе различных видов социального взаимодействия. Поэтому в исследовании применяются положения классической и современных социологических теорий социального действия, а также методологический и методический арсенал, накопленный в рамках социологии досуга и свободного времени, социологии социальных изменений, социологии рекреации. Кроме того, анализ демографических, этносоциальных, миграционных процессов рекреационных территорий осуществляется с применением методов демографии, этносоциологии и миграциологии.

Под социальными аспектами рекреационной сферы понимается социальная подсистема рекреационной территориальной системы. Авторами разработана методика социологического мониторинга для оценки состояния социальной среды рекреационных зон южных морей.

Как показали проведенные в 2008—2009 гг. социологические исследования на территории г. Сочи, резкий наплыв мигрантов (особенно иностранных) в курортный город, проблемы, связанные с началом строительства олимпийских объектов, грузино-осетинский конфликт, сохранение нестабильной обстановки в соседней Абхазии повысили уровень социальной и межэтнической напряженности, что вызвало, в свою очередь, сокращение в 2008 г., по оценкам сочинских экспертов, объема потока рекреантов на 30 %. Поэтому устойчивое развитие рекреационной сферы невозможно без учета и мониторинга ее социальных аспектов [17].

### Заключение

На основе анализа мирового опыта интегрированного управления прибрежными системами предлагается комплексный подход к оценке природных и антропогенных изменений природного комплекса, который позволит определить закономерности их формирования, развития отдельных подсистем, взаимосвязи между ними, выявление антропогенных изменений не только в отдельных компонентах, но и в экосистеме в целом.

В ходе исследования определены основные методы оценки качества природной среды, состояния прибрежных морских экосистем.

*Исследование проводится при частичной финансовой поддержке Программы фундаментальных исследований Президиума РАН и проекта РФФИ № 09-05-00194.*

### Список литературы

1. **Абакумов В. А.** Научные основы контроля качества вод по гидробиологическим показателям // Научные основы контроля качества вод по гидробиологическим показателям: Тр. Всесоюз. конференции. Л.: Наука, 1981. С. 1—5.
2. **Алимов А. Ф.** Введение в продукционную гидробиологию. Л.: Гидрометеиздат, 1989. 152 с.
3. **Архипова О. Е., Бойко В. В., Сладкова Ю. М.** Геоинформационное моделирование экологического состояния Ростована-Дону // Материалы XXXV школы-семинара "Математическое моделирование в проблемах рационального природопользования". Ростов н/Д.: Изд-во СКНЦ ВШ, 2007. С. 49—52.
4. **Барзыкин Ю. А.** Туризм в системе социально-экономических отношений: основы, функции, пути развития. Автореф. на соискание уч. степени канд. эк. наук. М., 2007. С. 2.
5. **Барнинова С. С., Медведева Л. А., Анисимова О. В.** Водоросли-индикаторы в оценке качества вод окружающей среды. М.: ВНИИприроды, 2000. 150 с.
6. **Барнинова С. С., Медведева Л. А., Анисимова О. В.** Биоразнообразие водорослей-индикаторов окружающей среды. Тель-Авив: PiliesStudio, 2006. 498 с.
7. **Битюков Н. И., Анисимов В. И., Сычев С. Л.** Рекреационные ресурсы Северного Кавказа: учеб. пособие. Сочи: СГУТиКД, 2009. 330 с.
8. **Бойко В. В.** Модель загрязнения урбанизированной территории // Изв. вузов. Сев.-Кав. регион. 2003. № 6. С. 50—52.
9. **Бойко В. В., Приваленко В. В., Шустова В. Л.** Эколого-геохимический мониторинг урбанизированной территории. Модельный подход (на примере г. Ростована-Дону) // Изучение зообентоса шельфа. Информационное обеспечение экосистемных исследований. Апатиты: Изд. КНЦ РАН, 2004. С. 151—163.
10. **Брагинский Л. П.** Оценка качества вод природных водоемов по токсикологическим показателям // Научные основы контроля качества вод по гидробиологическим показателям: Тр. Всерос. конф. СПб.: Наука, 1998. С. 201—206.
11. **Зайканов В. Г., Минакова Т. В.** Методологические основы комплексной геоэкологической оценки районов. М.: Институт геоэкологии РАН, 2008. С. 62—67.
12. **Ивонин И. М.** Реэкология: учеб. пособие. Ростов-на-Д.: Изд-во ЮФУ, 2008. 240 с.
13. **Лукашина Н., Трунов А.** Основы экологии и природных ресурсов рекреации. Сочи: СНИЦ РАН. 1999. С. 115—144.
14. **Москаленко В. А.** Система моделей для исследования качества воздушной среды рекреационной зоны. Автореф. на соискание уч. степени канд. эк. наук. Ростов-на-Дону, 1989.
15. **Рященко С. В.** Рекреационная география России. Иркутск: Изд-во Иркутского гос. ун-та, 2005. 236 с.
16. **Сладечек В.** Общая биологическая схема качества воды // Санитарная и техническая гидробиология. М.: Наука, 1967. С. 26—31.
17. **Тарасова Т. Т.** Социальные аспекты развития рекреационной сферы Черноморского побережья (на примере сочинской рекреационной зоны) / Труды XXXVII конференции "Математическое моделирование в проблемах рационального природопользования", 7—12 сентября 2009 г. // Экология. Экономика. Информатика. Ростов-на-Дону: Изд-во СКНЦ ВШ, 2009. С. 236—241.
18. **Унифицированные** методы исследования качества вод. Ч. 3. Методы биологического анализа вод. М.: СЭВ, 1977. 91 с.
19. **Чепижко А. В.** Проблемы развития зон рекреации Черноморского побережья Украины. URL: <http://www.ecologylife.ru/tourism-kurort/1633.html>
20. **Экологические** последствия туризма. URL: <http://www.gdrc.org/uem/ecotour/envi/one.html>
21. **Юрик Р. А.** Анализ современного состояния Российского рынка туристских услуг // Маркетинг в России и за рубежом. 2005. № 2.
22. **The tourism industry** // A Report for the world summit on sustainable development Johannesburg. September 2002. New York: United Nations, 2002.
23. **Watanabe T., Asai K., Houki A.** Numerical estimation to organic pollution of flowing water by using the epilithic diatom assemblage — Diatom Assemblage Index (DAI<sub>po</sub>) // The Science of the Total Environment. 1986. V. 55. P. 209—218.
24. **Watanabe T., Asai K., Houki A.** Numerical index of water quality using diatom assemblages // In M. Yasuno and A. Witton (eds.). Biological monitoring of environmental pollution. Tokyo: Univ. Press, 1988. P. 179—192.
25. **Watanabe T., Asai K., Houki A.** Numerical water quality monitoring of organic pollution using diatom assemblages // Proc. of the 9th International Diatom Symposium at 1986. Koeltz Scientific Books, Koenigstein, West Germany, 1988. P. 123—141.



УДК 621.391; 519.21; 519.245

**Г. Ф. Малыхина**, д-р техн.наук, проф., зав. каф.,  
Санкт-Петербургский государственный  
политехнический университет,  
e-mail: v\_m\_malykhin@mail.ru

## Статистическая модель для группы параметров, контролируемых в интеллектуальной измерительной системе

*Предложена вероятностная модель распределения группы коррелированных показателей для совокупных измерений при многопараметрическом контроле. Модель представляет собой семейство многомерных обобщенных экспоненциальных распределений (МОЭР). Предложено обоснование, аналитическая форма и метод идентификации МОЭР по экспериментальным данным.*

**Ключевые слова:** измерительная процедура, идентификация объекта, модель, интеллектуальность, принятие решений, метрологические характеристики, отношение правдоподобия, двухфазный поток, установка "Фактор"

Современное развитие измерительной техники направлено на повышение уровня интеллектуальности микропроцессорных приборов и информационно-измерительных систем (ИИС) на основе автоматизации измерительной процедуры и обработки результатов измерения. Это достигается использованием все более сложных алгоритмов реализации многофункциональных процедур измерения, адекватного выбора алгоритма на основе идентификации объекта измерения, настройки параметров выбранного алгоритма, использования итеративных методов измерения и автоматизированного развития алгоритма измерения на основе самообучения.

Последовательность уровней интеллектуальности средств измерения [1] основана на ряде признаков, отражающих степень использования одного или нескольких названных принципов. Такая классификация, организованная по принципу возрастания уровня интеллектуальности, показывает, что ИИС выше второго уровня включает процедуру принятия решений на основе получаемой информации об объекте измерения.

Метрологические характеристики ИИС существенно зависят от вида формализованного правила принятия решений, поэтому представляются важными задача адекватного отражения процедур принятия решений в математической модели измерения, анализ метрологических характеристик процедур принятия решений и синтез алгоритмов измерения, обеспечивающих требуемые метрологические характеристики.

### Метрологические характеристики, связанные с процедурами принятия решений

Рассмотрим статистический подход к анализу процесса реализации принятия решений о принадлежности результатов измерений к одному из классов. Необходимость использования статистического метода обусловлена наличием помех, влияющих на результаты измерения параметров объекта, стремлением расширить диапазон измерения и уменьшить влияние мешающих факторов. Для этого в алгоритм обработки включается осреднение измеряемой величины в классах, сформированных на основе процедур принятия решений.

Определение влияния характеристик процедуры классификации на средние значения результатов измерения параметров в каждой классификационной группе можно выполнить на основе использования байесовского классификатора с решающей функцией

$$d_i(\mathbf{x}) = P(\mathbf{x}|\omega_i)P(\omega_i),$$

где  $P(\mathbf{x}|\omega_i)$  — плотность распределения вектора  $\mathbf{x}$  измеряемых величин в классе  $\omega_i$ ,  $P(\omega_i)$  — априорная вероятность принадлежности  $\mathbf{x}$  к классу  $i$ .

В случае нормального распределения измеряемых величин плотности вероятностей в классах  $\omega_i$  и  $\omega_j$  имеют вид

$$P(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i)\right];$$

$$P(\mathbf{x}|\omega_j) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_j|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m}_j)^T \mathbf{C}_j^{-1} (\mathbf{x} - \mathbf{m}_j)\right],$$

где  $\mathbf{m}_i$ ,  $\mathbf{m}_j$  — векторы математических ожиданий (средние значения) вектора  $\mathbf{x}$  измеряемых параметров

ров в классах  $\omega_i$  и  $\omega_j$ ;  $\mathbf{C}_i$  и  $\mathbf{C}_j$  — ковариационные матрицы;  $|\mathbf{C}_i|$  и  $|\mathbf{C}_j|$  — определители этих матриц.

Для правила принятия решения о разделении классов  $\omega_i$  и  $\omega_j$  используется скалярный критерий — логарифмическое отношение правдоподобия

$$u_{ij}(\mathbf{x}) = \ln\left(\frac{P(\mathbf{x}|\omega_i)}{P(\mathbf{x}|\omega_j)}\right).$$

При равенстве ковариационных матриц  $\mathbf{C}_i = \mathbf{C}_j = \mathbf{C}$  этот критерий принимает вид

$$u_{ij}(\mathbf{x}) = \mathbf{x}^T \mathbf{C}^{-1}(\mathbf{m}_i - \mathbf{m}_j) - 1/2 \cdot (\mathbf{m}_i + \mathbf{m}_j)^T \mathbf{C}^{-1}(\mathbf{m}_i - \mathbf{m}_j). \quad (1)$$

Принадлежность вектора измеряемой величины  $\mathbf{x}$  к классу  $\omega_i$  определяется согласно правилу  $u_{ij}(\mathbf{x}) > \alpha$ ,

к классу  $\omega_j$  — по правилу  $u_{ij}(\mathbf{x}) < \alpha$ , где  $\alpha = \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right)$ .

Из соотношения (1) следует, что логарифм отношения правдоподобия является линейной комбинацией компонент вектора  $\mathbf{x}$  и, следовательно, он тоже подчиняется нормальному распределению. Математическое ожидание  $m\{u_{ij}\}$  логарифма отношения правдоподобия имеет вид

$$m\{u_{ij}\} = 1/2 \cdot (\mathbf{m}_i - \mathbf{m}_j)^T \mathbf{C}^{-1}(\mathbf{m}_i - \mathbf{m}_j) = 1/2 \cdot r_{ij},$$

где  $r_{ij} = (\mathbf{m}_i - \mathbf{m}_j)^T \mathbf{C}^{-1}(\mathbf{m}_i - \mathbf{m}_j)$  — расстояние Махаланобиса между плотностями распределения  $P(\mathbf{x}|\omega_i)$  и  $P(\mathbf{x}|\omega_j)$ . Дисперсия логарифма отношения правдоподобия

$$D\{u_{ij}\} = m_i\{(u_{ij} - m\{u_{ij}\})^2\} = r_{ij},$$

где  $m_i$  — символ процедуры усреднения в классе  $\omega_i$ .

Таким образом, для  $\mathbf{x} \in \omega_i$  логарифм отношения правдоподобия подчиняется нормальному распределению  $N(r_{ij}/2; r_{ij})$ . Подобным образом для  $\mathbf{x} \in \omega_j$  соответствующий логарифм отношения правдоподобия подчиняется нормальному распределению  $N(-r_{ij}/2; r_{ij})$ . Следовательно, среднее значение  $\bar{u}$  логарифма отношения правдоподобия в области ошибочных решений  $u_{ij} > \alpha$  — при условии, что  $\mathbf{x} \in \omega_j$  определяется выражением

$$\begin{aligned} \bar{u}(u_{ij} > \alpha|\omega_j) &= \\ &= \int_{\alpha}^{\infty} \frac{u_{ij}}{(2\pi r_{ij})^{1/2}} \exp\left[-\frac{(u_{ij} + (1/2)r_{ij})^2}{2r_{ij}}\right] du_{ij} = \\ &= (r_{ij}/2) \left[ 1 - \Phi\left(\frac{\alpha + (r_{ij}/2)}{\sqrt{r_{ij}}}\right) \right] + r_{ij} F\left(\frac{\alpha + (r_{ij}/2)}{\sqrt{r_{ij}}}\right), \end{aligned}$$

где  $1 - \Phi\left(\frac{\alpha + (r_{ij}/2)}{\sqrt{r_{ij}}}\right)$  — вероятность ошибочного

решения  $P(u_{ij} > \alpha|\omega_j)$ ;  $F(y) = \frac{\exp\left(\frac{-y^2}{2}\right)}{\sqrt{2\pi}}$ .

Среднее значение  $\bar{u}$  логарифма отношения правдоподобия в области ошибочных решений  $u_{ij} < \alpha$  при условии, что  $\mathbf{x} \in \omega_i$ , составляет

$$\begin{aligned} \bar{u}(u_{ij} < \alpha|\omega_i) &= \int_{-\infty}^{\alpha} \frac{u_{ij}}{(2\pi r_{ij})^{1/2}} \exp\left[-\frac{(u_{ij} - (1/2)r_{ij})^2}{2r_{ij}}\right] du_{ij} = \\ &= (r_{ij}/2) \left[ 1 - \Phi\left(\frac{\alpha - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) \right] + r_{ij} F\left(\frac{\alpha - (r_{ij}/2)}{\sqrt{r_{ij}}}\right), \end{aligned}$$

где  $\Phi\left(\frac{\alpha - (r_{ij}/2)}{\sqrt{r_{ij}}}\right)$  — вероятность ошибочного ре-

шения  $P(u_{ij} < \alpha|\omega_i)$ , а  $F(y) = \frac{\exp\left(\frac{-y^2}{2}\right)}{\sqrt{2\pi}}$ .

Погрешность определения среднего значения вектора измеряемых параметров обусловлена ошибочными решениями при классификации результатов измерения и, таким образом, зависит от решающего правила. Значение погрешности можно определить, сделав оценку для смещения среднего значения логарифма отношения правдоподобия.

Математическое ожидание логарифма отношения правдоподобия, апостериорно вычисленное при  $u_{ij} > \alpha$ , определяется по формуле

$$\begin{aligned} m\{u_{ij}|\Omega_i\} &= \frac{P(\omega_i)}{P(\Omega_i)} m\{u_{ij}\} + \frac{P(\omega_j)}{P(\Omega_i)} \bar{u}\{u_{ij} > \alpha|\omega_j\} + \\ &+ \frac{P(\omega_i)}{P(\Omega_i)} \bar{u}\{u_{ij} < \alpha|\omega_i\}, \end{aligned}$$

где  $P(\Omega_i) = P(u_{ij} > \alpha)$  — вероятность принятия решения о принадлежности  $\mathbf{x}$  к классу  $\omega_i$ . Следовательно, погрешность имеет систематический характер и включает мультипликативную и аддитивную составляющие.

Мультипликативная составляющая определяется выражением

$$\frac{P(\omega_i)}{P(\Omega_i)} = \frac{P(\omega_i)}{\left[ P(\omega_i) \Phi\left(\frac{\alpha - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) + P(\omega_j) \Phi\left(\frac{\alpha - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) \right]}.$$

Аддитивная составляющая определяется соотношением

$$\Delta u_{ij} = -\frac{\sqrt{r_{ij}}}{P(\Omega_i)} \left\{ P(\omega_j) F\left(\frac{\alpha + (r_{ij}/2)}{\sqrt{r_{ij}}}\right) + P(\omega_i) F\left(\frac{\alpha - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) \right\} + \frac{\sqrt{r_{ij}}}{P(\Omega_i)} \left\{ (P(\omega_j) - P(\omega_i)) \Phi\left(\frac{\alpha + (r_{ij}/2)}{\sqrt{r_{ij}}}\right) + P(\omega_i) F\left(\frac{\alpha - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) \right\}.$$

В случае, когда априорные вероятности соответствующих классов равны, т. е.  $P(\omega_i) = P(\omega_j)$ , пороговая величина  $\alpha = 0$ , мультипликативная погрешность равна нулю, а аддитивная определяется по формуле

$$\Delta u_{ij} = r_{ij} \Phi\left(-\frac{\sqrt{r_{ij}}}{2}\right) - \sqrt{r_{ij}} 2F\left(-\frac{\sqrt{r_{ij}}}{2}\right).$$

Направление смещения среднего значения вектора измеряемых параметров задается соотношением

$$\mathbf{V} = \frac{\mathbf{m}_j - \mathbf{m}_i}{|\mathbf{m}_j - \mathbf{m}_i|}.$$

Используя уравнение вектора в пространстве измеряемых параметров

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{V} \cdot t,$$

где  $\mathbf{x}_0$  — вектор начальных значений;  $\mathbf{V}$  — единичный вектор, указывающий направление;  $t$  — длина вектора, можно получить зависимость между смещением центра решающей функции и смещением компонент вектора измеряемых параметров в виде

$$\Delta u_{ij} = \Delta t \mathbf{V}^T \cdot \mathbf{C} \cdot (\mathbf{m}_j - \mathbf{m}_i).$$

Отсюда получается также оценка в виде

$$\Delta t = \frac{\Delta u_{ij}}{\mathbf{V}^T \cdot \mathbf{C} \cdot (\mathbf{m}_j - \mathbf{m}_i)} = \frac{\Delta u_{ij}}{r_{ij}} \cdot |\mathbf{m}_j - \mathbf{m}_i|; \Delta \mathbf{x} = \mathbf{V} \cdot \Delta t.$$

При этом систематическая погрешность оценки центра распределения измеряемых параметров в классе  $\omega_i$  при условии  $P(\omega_i) = P(\omega_j)$  и  $\alpha = 0$  определяется по формуле

$$\Delta \mathbf{x} = (\mathbf{m}_j - \mathbf{m}_i) [2\sqrt{r_{ij}}] F\left(-\frac{\sqrt{r_{ij}}}{2}\right) - \Phi\left(-\frac{\sqrt{r_{ij}}}{2}\right).$$

Таким образом, погрешность среднего значения измеряемых параметров в классах, сформирован-

ных на основе процедур принятия решений, имеет систематический характер. При нормальном распределении измеряемых параметров погрешность зависит от расстояния Махаланобиса и априорных вероятностей классов. Погрешность может быть вычислена с использованием решающих функций по приведенным формулам.

### Улучшение метрологических характеристик

Для уменьшения погрешности среднего значения измеряемых параметров в сформированных классах предлагается использовать интерквантильные оценки. С этой целью для процедуры принятия решения по значению логарифма отношения правдоподобия введены два порога  $\alpha_1$  и  $\alpha_2$ , причем возможно введение порогов на априорное и на апостериорное распределения логарифма отношения правдоподобия.

Если значения  $\alpha_1$  и  $\alpha_2$  соответствуют  $q$  %-квантили априорного распределения  $u_{ij}$ , то величина порогов  $\alpha_1 = \alpha_2 = \alpha_0$  определяется из условия

$$\Phi\left(\frac{\alpha_0 + (r_{ij}/2)}{\sqrt{r_{ij}}}\right) = \frac{q}{2}.$$

При этом погрешность смещения среднего для класса  $\omega_i$  отношения правдоподобия определяется по формуле

$$\Delta u = \frac{1}{P(\Omega_i)} \left\{ (r_{ij}/2) \Phi\left(\frac{\alpha_0 - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) - (r_{ij}/2) F\left(\frac{\alpha_0 - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) \right\},$$

где  $P(\Omega_i) = (1 - q) + \Phi\left(\frac{\alpha_0 - (r_{ij}/2)}{\sqrt{r_{ij}}}\right)$ .

Если значения  $\alpha_1$  и  $\alpha_2$  соответствуют  $q$  %-квантили апостериорного распределения измеряемых параметров, отнесенных к классу  $\omega_i$ , то величина порогов определяется из условия

$$\frac{P(\omega_i)}{P(\Omega_i)} \Phi\left(\frac{\alpha_1 - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) = \frac{P(\omega_i)}{P(\Omega_i)} \left\{ 1 - \Phi\left(\frac{\alpha_2 - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) \right\} + \frac{P(\omega_i)}{P(\Omega_i)} \Phi\left(\frac{\alpha_2 - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) = \frac{q}{2},$$

где  $P(\Omega_i) = P(\omega_i) \Phi\left(\frac{\alpha_1 - (r_{ij}/2)}{\sqrt{r_{ij}}}\right) + P(\omega_j) \Phi\left(\frac{\alpha_1 - (r_{ij}/2)}{\sqrt{r_{ij}}}\right)$ .

Погрешность за счет смещения среднего значения для логарифма отношения правдоподобия в случае  $P(\omega_i) = P(\omega_j)$  определяется по формуле

$$\begin{aligned} \Delta u_{ij} = & (r_{ij}/2) \left\{ \Phi \left( \frac{\alpha_2 - (r_{ij}/2)}{\sqrt{r_{ij}}} \right) + \Phi \left( \frac{\alpha_1 - (r_{ij}/2)}{\sqrt{r_{ij}}} \right) \right\} \times \\ & \times \Phi \left( \frac{\alpha_2 - (r_{ij}/2)}{\sqrt{r_{ij}}} \right) \Phi \left( \frac{\alpha_1 - (r_{ij}/2)}{\sqrt{r_{ij}}} \right) \left\} \frac{1}{P(\Omega'_i)} + \\ & + \sqrt{r_{ij}} \left\{ F \left( \frac{\alpha_2 - (r_{ij}/2)}{\sqrt{r_{ij}}} \right) + F \left( \frac{\alpha_1 - (r_{ij}/2)}{\sqrt{r_{ij}}} \right) \right\} \times \\ & \times F \left( \frac{\alpha_2 - (r_{ij}/2)}{\sqrt{r_{ij}}} \right) + F \left( \frac{\alpha_1 + (r_{ij}/2)}{\sqrt{r_{ij}}} \right) \left\} \frac{1}{P(\Omega'_i)}, \end{aligned}$$

где  $P(\Omega'_i) = (1 - q) + \Phi \left( \frac{\alpha_2 - (r_{ij}/2)}{\sqrt{r_{ij}}} \right)$ .

Таким образом, использование интерквантильных оценок позволяет уменьшить погрешность, связанную с принятием ошибочных решений.

### Область применения

Приведенные теоретические положения использованы при разработке микропроцессорной установки "Фактор", предназначенной для измерения относительной доли свободного газа в двухфазном потоке [2].

Измерение относительной доли свободного газа ( $\phi$ ) осуществляется косвенным методом путем измерения плотности жидкости ( $\rho_{ж}$ ) и плотности смеси ( $\rho_{см}$ ), включающей жидкость и свободный газ согласно соотношению

$$\phi = 1 - \frac{\rho_{см}}{\rho_{ж}}.$$

Измерение средней плотности газожидкостной смеси и плотности жидкого компонента осуществляется с помощью радиоизотопного первичного измерительного преобразователя (РПП). При этом используется свойство двухфазного газожидкостного потока с малой долей свободного газа, заключающееся в том, что в потоке существуют перемежающиеся интервалы присутствия и отсутствия свободного газа. На интервалах отсутствия свобод-

ного газа РПП измеряет плотность жидкостного компонента потока ( $\rho_{ж}$ ). На всех других интервалах наблюдения — плотность смеси двух компонентов ( $\rho_{см}$ ).

Для обнаружения интервалов отсутствия свободного газа разработан алгоритм, который заключается в измерении текущей плотности потока дискретно во времени с постоянным интервалом, равным 4 мс, и формировании выборки заданного объема. Решающая функция строится на основе трех параметров, полученных на выборке измерений плотности, к которым относятся среднее значение, дисперсия, автоковариация. Области принятия решения формируются на основе обучающей статистики, полученной в результате стендовых испытаний установки "Фактор".

В процессе измерения анализируемая выборка "скользит" по статистике, смещаясь на один отсчет. Решение об отсутствии свободного газа принимается при условии, что такое решение принято для  $P\%$  выборок, в которые входило данное измерение. На основе принятых решений формируется класс отсчетов плотности, соответствующих отсутствию свободного газа в потоке. Оценкой средней плотности в одном цикле измерения служит медиана. Окончательная оценка средней плотности в классе определяется как среднее значение медиан отдельных циклов.

Установка "Фактор" разработана в ЦНИИРТК и предназначена для измерения относительной доли свободного газа в товарной нефти. Приведенная погрешность измерения составляет 4 % — в диапазоне измерения от 0,002 до 0,05. Имеется модификация алгоритма измерения относительной доли свободного газа в двухфазном потоке, которая использует рассмотренный метод и базируется на совместном применении вейвлет-преобразования и нейронной сети для обработки измерительной информации [3].

### Список литературы

1. Цветков Э. И. Процессорные измерительные средства. Л.: Энергоатомиздат. 1989. 224 с.
2. Кратиров Д. В., Казаков А. Н., Малыхина Г. Ф. и др. Способ измерения параметров газожидкостного потока. Патент на изобретение № 94007263/25(007607), (Решение от 15.05.1996).
3. Кратиров Д. В., Меркушева А. В. Алгоритм, основанный на вейвлет-преобразовании и нейронной сети, для бесконтактного измерения параметров газожидкостного потока. Датчики и системы // Сб. докл. междунар. конф. Т. 3. СПб.: Изд. СПбГТУ, 2002. С. 51—55.

# ЖУРНАЛ В ЖУРНАЛЕ

## НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ

№ 11

НОЯБРЬ

2011

### Главный редактор:

ГАЛУШКИН А. И.

### Редакционная коллегия:

АВЕДЬЯН Э. Д.  
БАЗИЯН 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.

### Четырбоцкий А. Н.

Параметрическая идентификация радиальных базисных функций нейронных сетей методами глобальной оптимизации. . . . . 54

### Потапов В. И.

Модели и задачи оценки надежности нейронной системы при обучении и переобучении нейронной сети после отказов . . . . . 59

### Красовская Л. В.

Имитация обработки действий на нечетких семантических сетях в алгоритмах самообучения интеллектуальных систем . . . . . 64

### Куравский Л. С., Мармалюк П. А., Панфилова А. С., Ушаков Д. В.

Исследование факторных влияний на развитие психологических характеристик с применением нового подхода к оценке адекватности моделей наблюдениям . . . . . 67

**А. Н. Четырбоцкий,**

д-р физ.-мат. наук, вед. науч. сотр.,  
Дальневосточный геологический институт  
ДВО РАН,  
e-mail: Chetyrbotsky@yandex.ru

## Параметрическая идентификация радиальных базисных функций нейронных сетей методами глобальной оптимизации

*Рассматривается задача аппроксимации функции, которая на допустимом множестве аргументов задается дискретным набором значений. Для ее решения используется аппарат радиальных базисных функций нейронных сетей (RBFNN). Алгоритм процедуры обучения сети здесь реализован на базе методов глобальной оптимизации. В зависимости от режима его работы совокупность искомым неизвестных может содержать параметры радиальных базисных функций, процедуры обучения, координаты центров и, в ряде случаев, координаты точек обучения нейронной сети. Обсуждаются вопросы построения начального приближения и числа скрытых нейронов. Особенности разработанных подходов и методик их численной реализации демонстрируются экспериментальными исследованиями.*

**Ключевые слова:** радиальные базисные функции, задача поиска минимума, методы глобальной оптимизации

### Введение

При решении разнообразных прикладных задач часто встречается ситуация, когда вместо дискретного набора значений некоторого объекта требуется адекватное ему функциональное представление. В терминах методов идентификации системы следует выполнить построение такой цифровой модели, в которой установлены закономерности между входными и выходными переменными рассматриваемого объекта [5]. Решение подобных проблем обычно выполнялось методами полиномиальных и дробно-рациональных приближений. Однако встречаются значительные трудности их использования для аппроксимации функций с конечной, притом небольшой, гладкостью [3]. Учет отмеченного и отрицательный опыт работы различных модификаций тригонометрических и полиномиальных рядов при построении кривых и поверхностей по разбросанным данным привел Р. Харди к необходимости создания новых методов решения таких задач [13]. В предложенной им концепции значение функции в произвольной точке  $n$ -мерного пространства  $X = \{X \in D \subset R^n\}$  задается ее линейным разложе-

нием по базисным функциям, аргументом каждой из них является мера близости между этой точкой и одним из элементов дискретного набора  $C \equiv \{C_j \in D \subset R^n, j = 1 \div J\}$  (далее скрытых нейронов RBFNN). Базисом разложения выступают так называемые радиальные базисные функции (RBF), значимо отличные от нуля только в некоторой окрестности соответствующего им элемента этого набора.

Аппарат этих методов успешно применяется для решения различных прикладных задач [14, 12, 3, 4, 9], что обусловлено универсальностью используемых механизмов аппроксимации, способностью обучения и адаптации к изменяющимся условиям внешней среды [8]. Отмечается также независимость работы этих методов от масштабов расчетных сеток (эти методы являются бессеточными) и бесконечная дифференцируемость результатов аппроксимации, поэтому возможной становится "подгонка" значений функций и частных производных [2]. Последнее обстоятельство выгодно отличает эти методы от других методов аппроксимации функций.

Характерной особенностью аппарата RBFNN является требование оценки числа скрытых нейронов и зависимость его эффективности от большого числа свободных параметров, значения которых доставляют минимум некоторому функционалу. Для расчетов обычно используют весьма "чувствительные" к начальным распределениям различные модификации градиентных методов. В этой ситуации существуют проблемы задания начального приближения и того факта, что для вычисления градиента функционала требуются значительные вычислительные ресурсы. За рамками исследований также остаются вопросы статистической оценки параметров.

Здесь рассматривается постановка такой задачи, обсуждается выбор процедуры оценки параметров RBFNN, выполнено экспериментальное исследование разработанных алгоритмов.

### 1. Постановка задачи

Радиальные базисные функции были разработаны Р. Харди [13] для решения задач аппроксимации функции, которая в области определения своих аргументов задается дискретным набором ее значений. Формальная постановка задачи записывается следующим образом: по заданному дискретному набору значений некоторой функции  $\{f_i \equiv f(x_i), x_i \in D \subset R^n, i = 1 \dots I\}$  следует так подобрать зависящее от набора параметров  $\varepsilon$  соотношение

$$F(X, \varepsilon) \equiv \sum_{j=1}^J a_j \phi(\|X - C_j\|, \varepsilon) + p(X), \quad (1)$$

чтобы оно при  $X = \{x_i \in D \subset R^n, i = 1 \dots I\}$  в наибольшей степени, в смысле определенного критерия, отвечало бы указанному набору. Другими словами, требуется так выбрать радиальные базисные функции  $\phi(\|X - C_j\|, \varepsilon)$  и их параметры, а также подобрать веса  $\{a_j, j = 1 \dots J\}$  и коэффициенты некоторого полинома  $p(X)$ , чтобы соотношение (1) в узлах  $\{x_i \in D \subset R^n, i = 1 \dots I\}$  обучения RBF как можно в большей степени соответствовало бы  $\{f_i \equiv f(x_i), x_i \in D \subset R^n, i = 1 \dots I\}$ . Здесь введены следующие обозначения:  $D \subset R^n$  —  $n$ -мерная область определения функции;  $\{a_j, j = 1 \dots J\}$  — набор искоемых весов;  $J$  — число центров для RBF или число скрытых нейронов (определяется при вычислительных экспериментах для каждого конкретного случая);  $\phi(\|X - C_j\|, \varepsilon)$  — радиальная базисная функция с центром в точке  $C_j$ ;  $C \equiv \{C_j \in D \subset R^n, j = 1 \dots J\}$  — центры радиальных функций;  $\varepsilon \equiv \{\varepsilon_g, g = 1 \dots G\}$  — параметры RBF и настраиваемые при обучении параметры соответствующих вычислительных алгоритмов,  $G$  — их общее число;  $\|\circ\|$  — введенная на  $R^n$  некоторая метрика (в большинстве случаев — евклидова метрика);  $p(X)$  — многочлен определенной степени. В дальнейшем везде рассматривается ситуация  $p(X) \equiv \text{const}$ , и тогда полагается  $\text{const} \equiv a_{J+1}$ . Из чего следует увеличение на 1 размерности набора  $a$ :  $a \equiv \{a_j, j = 1 \dots J + 1\}$ , где в зависимости от ситуации на его элементы могут налагаться некоторые ограничения (в частности, сумма элементов полагается равной 0 или 1).

Представители семейства  $\phi(\|X - C_j\|, \varepsilon)$  радиально изменяются вокруг соответствующих им центров и значимо отличаются от нуля лишь в некоторой их окрестности. Обычно они представлены соотношениями (при записи функций используется общепринятая терминология и обозначения) [3]: 1) мультиквадрики (MQ)  $\phi(r, \varepsilon) \equiv \sqrt{1 + (\varepsilon r)^2}$ ; 2) обратные мультиквадрики (IMQ)  $\phi(r, \varepsilon) \equiv 1/\sqrt{1 + (\varepsilon r)^2}$ ; 3) обратные квадрики (IQ)  $\phi(r, \varepsilon) \equiv 1/[1 + (\varepsilon r)^2]$ ; 4) обобщенные мультиквадрики (GMQ)  $\phi(r, \varepsilon, \beta) \equiv [1 + (\varepsilon r)^2]^\beta$ ; 5) гауссианы (GA)  $\phi(r, \varepsilon) \equiv \exp[-(\varepsilon r)^2]$ . Здесь  $r = \{r_{i,j} \equiv \|x_i - C_j\|, i = 1 \dots I, j = 1 \dots J\}$  — мера близости между элементами области  $D \subset R^n$  и центрами RBF;  $\varepsilon$  — коэффициент, определяющий гладкость RBF. Необходимо заметить, что обобщенной формой записи первых четырех типов RBF является выражение

$$\phi(x_i, C_j, \varepsilon, \gamma) \equiv \left[ \sum_{l=1}^n \varepsilon_l (x_{i,l} - C_{j,l})^2 + \varepsilon_{n+1} \right]^\gamma, \quad (2a)$$

и для GA выражение

$$\phi(x_i, C_j, \varepsilon) \equiv \exp \left[ - \sum_{l=1}^n \varepsilon_l (x_{i,l} - C_{j,l})^2 \right]. \quad (2б)$$

Параметры этих соотношений  $\{C\}, \{\varepsilon = \{\varepsilon_l > 0, l = 1 \dots n\}, \gamma\}$  подлежат определению на основании заданного набора экспериментальных данных.

Мерой соответствия между  $\{f_i \equiv f(x_i), x_i \in D \subset R^n, i = 1 \dots I\}$  и  $\{F(x_i), x_i \in D \subset R^n, i = 1 \dots I\}$  выступает функционал

$$\Phi(a, P) = \sum_{i=1}^I \left[ f_i - \sum_{j=1}^J a_j \phi(\|x_i - C_j\|, \varepsilon) - a_{J+1} \right]^2, \quad (3)$$

где  $P \equiv \{C, \{\varepsilon\}, \{x\}\}$  — совокупность определяющих его параметров. Здесь рассматривается случай, когда его аргументами, в том числе, могут являться координаты точек  $\{x_i \in D \subset R^n, i = 1 \dots I\}$  обучения нейронной сети. Такая ситуация типична тогда, когда требуется заменить высокочемкую в вычислительном плане сложную аналитическую конструкцию/функцию более простой. В указанной обстановке допустимая потеря в точности сопутствует, как правило, существенному снижению объема вычислений.

Задача поиска минимума (3) принимает вид

$$\min_P \min_a \Phi(a, P). \quad (4)$$

Ее решение следует итерационной схеме, каждая итерация которой насчитывает два этапа: на первом этапе при фиксированных центрах  $C$  и  $\varepsilon$  оценке подлежат линейные веса выходного слоя, а на втором этапе они сами становятся объектами вычислений. В рамках первого этапа решение (4) отвечает системе линейных относительно набора параметров  $a$  уравнений:

$$Aa = d, \quad (5)$$

где элементы матрицы  $A = \{A_{ij}\}_{i=1 \dots I, j=1 \dots J+1}$  и вектора  $d$  в рамках принятых здесь положений и обозначений представлены выражениями

$$A_{ij} \equiv \begin{cases} \phi(\|x_i - C_j\|, \varepsilon) & \text{при } i \leq I, j \leq J, \\ 1 & \text{при } i \leq I, j = J + 1, \end{cases} \quad (6)$$

$$d \equiv \{f_i \equiv f(x_i), x_i \in D \subset R^n, i = 1 \dots I\}.$$

Согласно (5) и (6), возбуждения  $x_i$  генерируют в скрытом слое нейронной сети сигналы  $\phi_i \equiv \{\phi(\|x_i - C_1\|, \varepsilon), \phi(\|x_i - C_2\|, \varepsilon), \dots, \phi(\|x_i - C_J\|, \varepsilon), 1\}$ , где 1 есть единичный сигнал и  $i = 1 \dots I$ . Сигналу  $\phi_i$  отвечает выходной сигнал  $F_i \equiv \phi_i a^T$ .

После нахождения  $P$  на предыдущем этапе и подстановки его значения в выражение (6) рассчитывают веса RBFNN. Тогда решение (5) принимает вид

$$a = (A'A)^{-1}A'd, \quad (7)$$

где  $A'$  — транспонированная матрица  $A$ . Вырожденность  $(A'A)^{-1}$  указывает на отсутствие единственности решения (5), что является недопустимым. Такая ситуация происходит, если:  $J > I$  (число столбцов  $A$  превышает число ее строк [10]), число скрытых слоев RBFNN больше числа пар обучения; существует линейная зависимость между строками/столбцами этой матрицы.

Поиск решения (4) после подстановки  $a$  составляет второй этап итерационной схемы обучения RBF. В задачах такого класса допускается наличие у функционала (3) большого числа локальных минимумов. Поэтому результат их поиска и объем соответствующих вычислений непосредственно определяется характером используемой процедуры и выбором начального приближения. При использовании градиентных методов происходит сходимость поиска к расположенному в окрестности начального приближения локальному минимуму. Тогда на каждой итерации для нахождения шага градиентной процедуры требуется расчет элементов соответствующего градиента функционала, что связано с большими объемами вычислений.

Для выполнения второго этапа естественным видится применение современных способов решения подобных задач, а именно методов глобальной оптимизации. В этом случае поиск минимума на допустимом множестве искомым переменных сводится к определенному перебору локальных решений. Среди них выделяются методы, основанные на поведенческой и эволюционной стратегии коллективного поведения самоорганизующихся живых и неживых систем [14, 6]. В алгоритмах данного класса методов для улучшения глобального поиска минимума применяется локальная оптимизация. Следует отметить, что такие алгоритмы обуславливают сходимость генерируемой ими последовательности точек к глобальному оптимальному решению. Механизм их работы состоит в предварительной генерации совокупности частиц (строк-стрингов или хромосом [7]). Далее формируется итерационный процесс, на каждой стадии которого определенным образом пересчитываются их координаты. Для  $m$ -й итерации простейший алгоритм подобной процедуры (так называемого алгоритма PSO "Particle Swarm Optimization" [15]) записывается в виде

$$\begin{aligned} V_{k,j}^{(m+1)} &= \beta_1 P_{k,j}^{(m)} + \beta_2 \text{rand}(\tilde{P}_{k,j} - P_{k,j}^{(m)}) + \\ &+ \beta_3 \text{rand}(P_j^* - P_{k,j}^{(m)}); \quad (8) \\ P_{k,j}^{(m+1)} &= P_{k,j}^{(m)} + V_{k,j}^{(m+1)}, \end{aligned}$$

где  $\{V_{k,j}^{(m)}, k = 1 \dots K, j = 1 \dots J\}$  — сдвиг значения  $j$ -го признака частицы (согласно общепринятой терминологии, элементы искомого множества именуются частицами [17]) на  $m$ -й итерации;  $K$  — общее число частиц;  $\beta_1, \beta_2, \beta_3$  — эмпирические параметры алгоритма;  $\text{rand}$  — равномерно распределенное на отрезке  $[0, 1]$  случайное число;  $\tilde{P}_{k,j}$  — лучшая, согласно (4), позиция  $k$ -й частицы за текущие  $m$  итераций;  $P_k^{(m)} \equiv \{P_{k,j}^{(m)}, j = 1 \dots J\}$  — координаты частиц на  $m$ -й итерации;  $P^* \equiv \{P_j^*, j = 1 \dots J\}$  — координата частицы с наилучшим, в смысле (4), за текущие  $m$  итераций поиском значением целевой функции.

## 2. Построение начального приближения и интерпретация оценок

Для решения задач поиска экстремума следует указать границы допустимой многомерной области, внутри которой должно находиться это решение. Сходимость результатов поиска к корректному минимуму достигается обычно назначением довольно произвольных границ, вне которых алгоритмы "вязнут в болоте" не относящихся к делу областей параметрического пространства [1]. При проведении вычислительных экспериментов границы изменения параметров функций (2а) и параметры алгоритма (8) задавались случайным образом. Поскольку координаты центров принадлежат области определения объекта исследования, то для задания их начального приближения достаточно в этом качестве указать некоторые узлы регулярной сетки или сгенерировать их случайным образом.

Результаты решения (4) позволяют получить не сами значения искомым параметров, а только их оценки. Действительно, вместо точных значений отдельных характеристик исследуемого процесса из экспериментов обычно известны только их измеряемые с определенной погрешностью приближенные значения [11]. Такая ситуация обусловлена рядом причин: выборочный характер исследований; высокая погрешность измерительных приборов; многофакторность самого процесса; невозможность измерения части переменных объекта исследований непосредственно прямым способом. Поэтому для корректного построения RBFNN следует не только получить оценки параметров, но также и исследовать их надежность. В такой ситуации мерой надежности обычно выступают рассчитанные по исходным данным средние квадратичные отклонения (СКО) параметров.

Если решение задачи (4) уже найдено, то расчет СКО весов RBFNN и остальных параметров предлагается выполнять согласно этапам ее обучения. Поскольку веса  $a$  являются коэффициентами линейной регрессионной модели, то в рамках поло-



жений этой модели матрица их ковариации определяется выражением [10]

$$S^{(a)} = (A'A)^{-1} \Phi(a^*, P^*) / (I - J - 1), \quad (9)$$

где  $a^*$  — решение (7) при  $P^*$ . Квадратные корни ее диагональных элементов есть СКО весов RBFNN. Элементы  $S^{(a)}$  могут оказаться полезными для оценки числа скрытых нейронов RBFNN. Действительно, равенство нулю некоторого элемента  $a$  указывает на сокращение числа скрытых нейронов. Проверка гипотезы  $H_0: a_j = 0$  выполняется с помощью  $t$ -статистики:

$$t = a_j / \sqrt{S_{j,j}^{(a)}}. \quad (10)$$

Аналогичным образом проверяются и другие подобные гипотезы, в частности, равенство всех элементов  $a$  одному и тому же значению.

При расчете матрицы ковариации остальных параметров RBFNN учитывается сходимость последовательности точек  $\{P_k^{(m)}, k = 1 \dots K, m = 1 \dots M\}$  многомерного пространства к решению (4). Здесь  $M$  — минимум между заданным максимумом общего числа итераций и номером итерации, где происходит стабилизация решения (4). В окрестности  $P^*$  совокупность  $\{P_k^{(M)}\}_{k=1 \dots K}$  представляет статистическую выборку решений, которую можно использовать для расчетов различных статистических характеристик.

### 3. Численные эксперименты

Для оценки работоспособности рассмотренного алгоритма использовались двумерные варианты функций из работы [16] (табл. 1).

Точки обучения RBFNN располагались в узлах регулярной сетки области определения функции (строка  $I$ , табл. 2). Во всех случаях используется

представление RBF (2a). Начальное распределение центров RBF для функции де Джонга задавалось случайным образом, для остальных функций — в узлах регулярных сеток (строка  $J$ ). Число частиц в алгоритме глобальной оптимизации задавалось равным 40, а максимальное число итераций — 10. Оценка погрешности аппроксимации на регулярной сетке  $40 \times 40$  здесь оценивается мерой их среднего квадратичного относительного несоответствия

$$E = \sqrt{\frac{\sum_{i=1}^{40} \sum_{j=1}^{40} [F(x_i, y_j) - f(x_i, y_j)]^2}{\sum_{i=1}^{40} \sum_{j=1}^{40} f^2(x_i, y_j)}}, \quad (11)$$

Таблица 1

Двумерные варианты функций

Название функции	Функция	Область определения
Функция де Джонга	$f(x, y) = x^2 + y^2$	$x \in [-1, 1], y \in [-1, 1]$
Функция Франке	$f(x, y) = 0,75[\exp(-\varphi/4) + \exp(-\psi)] + 0,5\exp(-\eta) + 0,2\exp(-\omega)$ $\varphi = (9x - 2)^2 + (9y - 2)^2$ $\psi = (9x + 1)^2/49 - (9y + 1)/10$ $\eta = (9x - 7)^2 + (9y - 3)^2$ $\omega = (9x - 4)^2 - (9y - 7)^2$	$x \in [0, 1], y \in [0, 1]$
Функция Экли	$f(x, y) = 20 + \exp(1) - 20\exp(\varphi) - \exp(\psi)$ $\varphi = -0,2\sqrt{(x^2 + y^2)}/2$ $\psi = [\cos(2\pi x) + \cos(2\pi y)]/2$	$x \in [-20, 20], y \in [-20, 20]$
Функция Шекеля	$f(x, y) = \sum_{l=1}^3 [b_l^{(1)} + (x - b_l^{(2)})^2 + (y - b_l^{(3)})^2]^{-1}$ $b^{(1)} = (1, 2, 2), b^{(2)} = (2, 10, 18), b^{(3)} = (10, 15, 4)$	$x \in [0, 20], y \in [0, 20]$

Таблица 2

Значения оценок параметров

Оценки	Функция де Джонга	Функция Франке	Функция Экли	Функция Шекеля
$\varepsilon_1$	97,266 ± 15,697	87,102 ± 13,438	87,521 ± 15,811	75,202 ± 14,233
$\varepsilon_2$	66,708 ± 17,038	87,258 ± 15,351	61,352 ± 15,009	60,177 ± 16,375
$\varepsilon_3$	2,427 ± 0,425	2,227 ± 0,228	2,643 ± 0,393	2,645 ± 0,346
$\gamma$	0,560 ± 0,039	0,551 ± 0,038	0,562 ± 0,031	0,579 ± 0,032
$\beta_1$	0,893 ± 0,117	0,881 ± 0,132	0,968 ± 0,135	0,939 ± 0,126
$\beta_2$	2,287 ± 0,551	2,430 ± 0,541	2,407 ± 0,550	1,985 ± 0,488
$\beta_3$	1,907 ± 0,386	2,765 ± 0,431	2,378 ± 0,450	2,062 ± 0,524
$I$	3 × 3	8 × 8	10 × 10	12 × 12
$J$	3	3 × 3	6 × 6	6 × 6
$E$	0,073/0,994	0,158/0,985	0,064/0,984	0,258/0,944

Примечание: в первых шести строках указаны доверительные интервалы оценок; в строке  $I$  указан способ начального размещения центров RBF (в первом случае центры размещались случайным образом, в остальных случаях — на регулярной сетке); в строке  $E$  — первое число есть значение (10), а второе — коэффициент корреляции между  $\{f(x_i, y_j), i = 1 \dots 40, j = 1 \dots 40\}$  и  $\{F(x_i, y_j), i = 1 \dots 40, j = 1 \dots 40\}$ , рассматриваемые как векторы.

где  $f(x_i, y_j)$  — значение тестовой функции в узлах регулярной сетки  $(x_i, y_j)$  ее области определения;  $F(x_i, y_j)$  — соответствующее значение RBFNN аппроксимации. Мерой подобия форм поверхностей здесь выступает коэффициент корреляции между  $\{f(x_i, y_j), i = 1..40, j = 1..40\}$  и  $\{F(x_i, y_j), i = 1..40, j = 1..40\}$ , которые в такой ситуации рассматриваются как векторы.

Результаты экспериментов представлены на рисунке (см. четвертую сторону обложки).

Анализ элементов табл. 2 показывает узость границ доверительных интервалов, что соответствует высокому уровню надежности рассчитанных параметров. Имеет место также низкая погрешность аппроксимации и высокая степень подобия форм сопоставляемых поверхностей. Отмеченные факты и рассмотрение случаев на рисунке отражает хорошую эффективность рассмотренного алгоритма обучения RBFNN и высокую степень соответствия между функциями и отвечающих им образами, которые строятся на основании RBF-аппроксимации.

### Заключение

Полученные здесь результаты показывают высокую эффективность использования методов глобальной оптимизации для режима обучения RBFNN. В этой ситуации эмпирические параметры алгоритма поиска также подлежат оцениванию. Предлагаемая методика может быть использована для построения различных эмпирических моделей по соответствующим экспериментальным данным.

### Список литературы

1. Бард Й. Нелинейное оценивание параметров. М.: Статистика, 1979. 349 с.
2. Василенко В. А. Сплайн-функции: теория, алгоритмы, программы. Новосибирск: Наука, 1983. 214 с.
3. Васильев А. Н., Тархов Д. А. Нейросетевое моделирование. Принципы. Алгоритмы. Приложения. СПб.: Изд-во Политех. ун-та, 2009. 527 с.
4. Горбаченко В. И. Нейрокомпьютеры в решении краевых задач теории поля. М.: ИПРЖ, 2000. 336 с.
5. Гроп Д. Методы идентификации систем. М.: Мир, 1979. 302 с.
6. Карпенко А. П., Селиверстов Е. Ю. Глобальная оптимизация методом роя частиц. Обзор // Информационные технологии. 2010. № 2. С. 25—34.
7. Орлянский И. В. Современные подходы к построению методов глобальной оптимизации // Исследовано в России. 2002. URL: <http://zhurnal.apc.relarn.ru/articles/2002/189.pdf>.
8. Осовский С. Нейронные сети для обработки информации. М.: Финансы и статистика, 2002. 344 с.
9. Питенко А. А. Нейросетевой анализ в геоинформационных системах. Красноярск: Изд-во ИВМ СО РАН, 2000. 97 с.
10. Пао С. Р. Линейные статистические методы и их применение. М.: Наука, 1968. 548 с.
11. Четырбоцкий А. Н. Параметрическая идентификация крупномасштабной модели эволюции морского ледяного покрова (на примере ледяного покрова Японского моря) // Информационные технологии. 2006. № 11. С. 12—20.
12. Fasshauer G. E. Solving differential equations with radial basis functions: multilevel methods and smoothing // Advances in Computational Mathematics, 1999. N 11(2—3). P. 139—159.
13. Hardy R. L. Multiquadric equations of topography and other irregular surfaces // J. Geophys. Res. 1971. Vol. 76(8). P. 1905—1915.
14. Kansa E. J. Multiquadrics — a scattered data approximation scheme with applications of computational fluid dynamics — II: Solutions to parabolic hyperbolic and elliptic differential equations // Comput. Math. Appl. 1990. Vol. 19. P. 147—161.
15. Kennedy J. Particle swarm optimization // In Proceeding of IFFF International conference on Neural Networks. 1995. P. 1942—1948.
16. Tang R., Yao X., Suganthan P. N., MacNish C., Chen J. P., Chen C. M., Yang Z. Benchmark Functions for the CEC'2008 Special Session and Competition on Large Scale Global Optimization // Nature Inspired Computation and Applications Laboratory. USTC, China, 2007.
17. Weise T. Global Optimization Algorithms: Theory and Application. URL: <http://www.it-weise.de/projects/book.pdf>. 2009. 820 p.

## ИНФОРМАЦИЯ

16—19 апреля 2012 г. в Москве в ИПУ им. В. А. Трапезникова РАН состоится  
3-я Российская конференция с международным участием

## "Технические и программные средства управления, контроля и измерения"

### Тематические направления:

- Теория, методы исследования и проектирования, опыт применения технических средств управления, контроля и измерения (УКИ).
- Теория, алгоритмы и программное обеспечение систем УКИ.
- Анализ состояния, тенденций и перспектив развития систем УКИ.
- Создание средств систем УКИ на основе новых современных технологий.
- Компьютерное и физическое моделирование систем УКИ.
- Теоретические и прикладные аспекты оценивания и повышения качества систем УКИ, метрологическое обеспечение, испытание и сертификация.
- Вопросы обучения в области средств систем УКИ.

Подробную информацию о конференции см. на сайте: <http://cmm.ipu.ru>

**В. И. Потапов**, д-р техн. наук, проф., зав. каф.,  
Омский государственный  
технический университет,  
e-mail: ivt@omgtu.ru

## Модели и задачи оценки надежности нейронной системы при обучении и переобучении нейронной сети после отказов\*

*Рассматриваются вероятностные модели для решения задач надежности нейронной системы при обучении и переобучении искусственной нейронной сети после отказов с участием человека-оператора. Приводятся формулы для вычисления функциональной готовности нейронной системы.*

**Ключевые слова:** математическая модель, нейронная сеть, нейронная система, надежность, обучение нейронной сети

### Введение

В данной работе сделана попытка рассмотреть и обосновать некоторые математические модели поведения нейрокомпьютерной системы (НКС) в процессе обучения (настройки) и переобучения (перенастройки) нейронной сети после отказов.

В связи с тем, что в процессе обучения и переобучения нейронной сети (НС) нейрокомпьютерной системы может принимать участие человек-оператор, который также оказывает влияние на характеристики надежности нейронной системы, в некоторых моделях будет учитываться конечная надежность оператора.

Практически во всех известных работах, посвященных исследованию надежности искусственных нейронных сетей и НКС, влияние человеческого фактора на надежность таких систем либо не рассматривалось, либо молчаливо полагалось, что человек, в той или иной мере участвующий в процессе настройки (обучения) и восстановлении нейронной сети после отказов (переобучении), является абсолютно надежным [1, 3].

На самом деле любая нейронная система с программно-аппаратной системой управления процессом обучения и переобучения после отказов НС представляет собой человекомашинный комплекс, в котором человек может выполнять различные функции: оператора (настройщика), ремонтника, оператора и ремонтника, эргатического резерва (резервного компонента в системе) и другие, в за-

висимости от поставленных целей и специфики решаемых нейронной системой задач. Поэтому логично рассматривать две модели надежности любой нейронной системы: одну — когда человек-оператор абсолютно надежный, а вторую — когда человек-оператор обладает конечной надежностью.

### Модели для вычисления функциональной готовности нейронной системы при обучении и переобучении НС после отказов при абсолютной надежности системы обучения

При подготовке нейронов, в частности нейрокомпьютерной, системы к функционированию проводят обучение ее НС на реализацию заданных функций, определяемых типом решаемой задачи, т. е. выполняют так называемую "настройку" нейронной системы. В дальнейшем, если нет особой необходимости, оба термина "нейронная система" и "нейрокомпьютерная система" будем считать эквивалентными. Так же эквивалентными будем считать термины "обучение" — "подготовка" и "переобучение" — "восстановление".

Аналогичные действия (дообучение, переобучение) имеют место в отказоустойчивых нейрокомпьютерных системах с аппаратной и логической избыточностью при логической перестройке и восстановлении функциональных возможностей НС после отказов нейронов в процессе выполнения задания.

При начальной настройке (обучении) НС и ее логической перестройке после отказов (переобучении) выполняется ряд операций, связанных с изменением весов входов и порогов срабатывания нейронов, в соответствии с выбранным алгоритмом, ориентированным, как правило, на минимизацию времени настройки и восстановления функциональных свойств НС нейрокомпьютерной системы. Так как время выполнения операций настройки и перестройки логики НС после отказов в общем случае является случайной величиной, то и общее время подготовки НКС к функционированию, т. е. к выполнению задания, будет также случайной величиной.

В соответствии со сказанным под функциональной готовностью НКС  $P_{\text{фг}}(t)$  будем понимать вероятность того, что система окажется в работоспособном состоянии в произвольно выбранный момент времени после начала настройки или логической перестройки после отказа НС нейрокомпьютера. Вопросам расчета функциональной готовности НКС при обучении и переобучении НС после отказов в литературе уделено недостаточно внимания. Поэтому в данной работе делается попытка в какой-то мере восполнить указанный пробел.

Для простоты будет полагать, что процесс настройки или логической перестройки НС после от-

\*Данная работа выполнена при финансовой поддержке гранта РФФИ, проект № 11-08-00271.

каза в соответствии с выбранным алгоритмом, т. е. процесс подготовки к функционированию нейронной системы, представляет собой линейную последовательность выполнения одной операции за другой. Очевидно, что на практике возможны и другие, отличные от линейных, алгоритмы выполнения операций обучения и переобучения НС нейрокompьютера, которые в данной работе не рассматриваются. Будем также полагать, что рассматриваемый процесс подготовки НС нейронной системы к функционированию может быть аппроксимирован марковским случайным процессом с конечным числом состояний. Граф процесса подготовки к функционированию НКС для случая линейной последовательности операций обучения представлен на рис. 1.

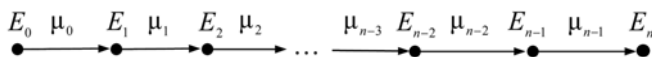


Рис. 1. Граф процесса подготовки НКС к функционированию

В приведенных на графе обозначениях  $E_i$  ( $i = 0, 1, \dots, n$ ) — состояние нейронной системы на  $i$ -м шаге процесса подготовки к функционированию;  $\mu_i$  ( $i = 0, 1, \dots, n - 1$ ) — интенсивность выполнения  $i$ -й операции подготовки.

Обозначив через  $p_i(t)$  вероятность нахождения НКС в состоянии  $E_i$  ( $i = 0, 1, \dots, n$ ), нетрудно составить систему дифференциальных уравнений Колмогорова, описывающих поведение рассматриваемой системы:

$$\begin{aligned} p'_0(t) &= -\mu_0 p_0(t); \\ p'_k(t) &= \mu_{k-1} p_{k-1}(t) - \mu_k p_k(t), \\ k &= 1, 2, \dots, (n - 1); \\ p'_n(t) &= \mu_{n-1} p_{n-1}(t). \end{aligned} \quad (1)$$

Начальные условия при решении системы уравнений (1) имеют вид  $p_0(0) = 1, p_1(0) = p_2(0) = \dots = p_n(0) = 0$ . Очевидно, что  $\sum_{i=0}^n p_i(t) = 1$ . Отсюда следует, что функциональная готовность НКС определяется выражением

$$P_{\text{фг}}(t) = p_n(t) = 1 - \sum_{i=0}^{n-1} p_i(t).$$

Выполняя преобразование Лапласа и полагая, что все  $\mu_i$  ( $0 \leq i \leq n - 1$ ) различны, получим из системы уравнений (1) выражение для функциональной готовности НКС в следующем виде:

$$P_{\text{фг}} = 1 - \prod_{l=0}^{n-1} \mu_l \sum_{\substack{i=0 \\ j \neq i}}^{n-1} \frac{e^{-\mu_i t}}{\mu_i \prod_{j \neq i} (\mu_j - \mu_i)}, \mu_i \neq 0.$$

Среднее время подготовки системы к функционированию в рассматриваемом случае определяется выражением

$$T_{\text{фг}} = \sum_{i=0}^{n-1} \frac{1}{\mu_i}, \mu_i \neq 0.$$

Решение системы уравнений (1) при произвольных  $\mu_i$  (в том числе равных между собой  $\mu_i = \text{const}$  ( $0 \leq i \leq n - 1$ )) с помощью ПЭВМ не представляет сложности. Однако в реальных условиях интенсивности выполнения операций подготовки НКС к работе могут быть не постоянными величинами, а изменяющимися во времени, т. е. на дугах графа, изображенного на рис. 1, вместо  $\mu_i = \text{const}$  следует указывать реальные интенсивности переходов  $\mu_i(t)$ . Тогда система дифференциальных уравнений, соответствующая реальному процессу настройки (перестройки после отказов) нейронной системы, принимает следующий вид:

$$\begin{aligned} p'_0(t) &= -\mu_0(t) p_0(t), \\ p'_k(t) &= \mu_{k-1}(t) p_{k-1}(t) - \mu_k(t) p_k(t), \\ k &= 1, 2, \dots, (n - 1), \\ p'_n(t) &= \mu_{n-1}(t) p_{n-1}(t). \end{aligned} \quad (2)$$

В работе [1] показано, что приближенное решение с заданной точностью подобных систем дифференциальных уравнений с переменными коэффициентами возможно методом дискретизации и целочисленного программирования, и даны примеры решения таких систем уравнений, описывающих поведение НКС. В данной работе для решения рассматриваемой задачи воспользуемся изложенным в [2] методом определения закона распределения времени перехода системы из нулевого состояния в поглощающее путем замены реальной динамической системы статистически эквивалентной.

Известно [2], что модель реальной динамической системы с переменными во времени интенсивностями переходов можно приближенно представить моделью, статистически эквивалентной на фиксированном интервале времени ( $\tau, t$ ), если каждый локальный переход в ней из одного состояния в другое заменить эквивалентной группой переходов, имеющих постоянные во времени интенсивности, таким образом, чтобы результирующая условная вероятность перехода в этой группе была достаточно близка к условной вероятности перехода в реальной системе. Таким образом, задача сводится к синтезу такой эквивалентной системы с фиксированным конечным числом состояний, которая описывается дифференциальными уравнениями с постоянными интенсивностями переходов. При этом в качестве критерия такой эквивалентности можно, например, взять ошибку в оценке времени "жизни" системы или другой критерий.

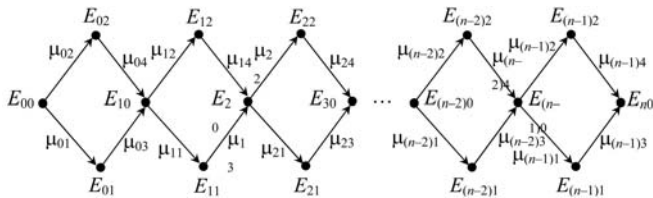


Рис. 2. Граф переходов статистически эквивалентной системы

Для получения системы, статистически эквивалентной рассматриваемой (рис. 1), в исходной системе заменим каждый переход из состояния  $E_i$  в состояние  $E_{i+1}$ , имеющий интенсивность  $\mu_i(t)$ , четырьмя переходами с постоянными интенсивностями  $\mu_{i1}, \mu_{i2}, \mu_{i3}, \mu_{i4}$  ( $0 \leq i \leq n-1$ ). Полученный в результате такой замены граф переходов эквивалентной системы изображен на рис. 2.

Обозначим  $p_{i0}(t)$  — вероятность нахождения системы в состоянии  $E_{i0}$  ( $0 \leq i \leq n$ );  $p_{i1}(t), p_{i2}(t)$  — вероятности нахождения системы соответственно в состояниях  $E_{j1}$  и  $E_{j2}$  ( $0 \leq j \leq n-1$ ). Тогда система дифференциальных уравнений, описывающих поведение эквивалентной системы, соответствующая графу на рис. 2, примет вид

$$\begin{aligned}
 p'_{00}(t) &= -(\mu_{01} + \mu_{02})p_{00}(t); \\
 p'_{k0}(t) &= \mu_{(k-1)3}p_{(k-1)1}(t) + \\
 &+ \mu_{(k-1)4}p_{(k-1)2}(t) - (\mu_{k1} + \mu_{k2})p_{k0}(t); \\
 k &= 1, 2, \dots, (n-1); \\
 p'_{l1}(t) &= \mu_{l1}p_{l0}(t) - \mu_{l3}p_{l1}(t); \\
 p'_{l2}(t) &= \mu_{l2}p_{l0}(t) - \mu_{l4}p_{l2}(t); \\
 l &= 0, 1, 2, \dots, (n-1); \\
 p'_{n0}(t) &= \mu_{(n-1)3}p_{(n-1)1}(t) + \mu_{(n-1)4}p_{(n-1)2}(t).
 \end{aligned} \tag{3}$$

В качестве примера положим, что

$$\mu_{l1} = \mu_{l3} = \mu_{l4} = \tilde{\mu}_l; \mu_{l2} = \alpha_l \tilde{\mu}_l; \alpha_l > 0; 0 \leq l \leq (n-1).$$

Тогда систему уравнений (3) можно переписать в следующем виде:

$$\begin{aligned}
 p'_{00}(t) &= -\tilde{\mu}_0(1 + \alpha_0)p_{00}(t); \\
 p'_{k0}(t) &= \tilde{\mu}_{k-1}p_{(k-1)1}(t) + \\
 &+ \tilde{\mu}_{k-1}p_{(k-1)2}(t) - \tilde{\mu}_k(1 + \alpha_k)p_{k0}(t); \\
 k &= 1, 2, \dots, (n-1); \\
 p'_{l1}(t) &= \tilde{\mu}_l p_{l0}(t) - \tilde{\mu}_l p_{l1}(t); \\
 p'_{l2}(t) &= \alpha_l \tilde{\mu}_l p_{l0}(t) - \tilde{\mu}_l p_{l2}(t); \\
 l &= 0, 1, \dots, (n-1); \\
 p'_{n0}(t) &= \tilde{\mu}_{n-1}p_{(n-1)1}(t) + \tilde{\mu}_{n-1}p_{(n-1)2}(t)
 \end{aligned} \tag{4}$$

с начальными условиями

$$\begin{aligned}
 p_{00}(\tau) &= 1; p_{k0}(\tau) = p_{l0}(\tau) = p_{n0}(\tau) = 0 \\
 (1 \leq k \leq n-1, 0 \leq l \leq n-1).
 \end{aligned}$$

Решение системы уравнений (4) не представляет трудностей, если известны интенсивности переходов  $\tilde{\mu}_l$  ( $0 \leq l \leq n-1$ ) в эквивалентной системе. В работе [2] показано, что зная математическое ожидание  $T_l$  и второй начальный момент  $\alpha_{2l}$  для локального перехода в реальной системе, можно определить соответствующие интенсивности переходов в эквивалентной системе по следующим формулам:

$$\tilde{\mu}_l = \frac{1}{T_l} \left( \frac{2 + \alpha_l}{1 + \alpha_l} \right), \alpha_l = \frac{-(3 - 2\omega_l) \pm \sqrt{2\omega_l - 3}}{2 - \omega_l}, \omega_l = \frac{\alpha_{2l}}{T_l^2}.$$

Параметры  $T_l$  и  $\alpha_{2l}$  могут быть определены по формулам

$$T_l = \int_0^{\infty} \int_{\tau}^{\tau+\tilde{\tau}} e^{-\tau} \mu_1(t) dt d\tilde{\tau}, \alpha_{2l} = \int_0^{\infty} \int_{\tau}^{\tau+\tilde{\tau}} e^{-\tau} \mu_1(t) dt d\tilde{\tau}.$$

Решение рассматриваемой задачи имеет смысл при выполнении условий:  $\alpha_l > 0$  и  $\omega_l \geq 3/2$ .

В общем случае систему дифференциальных уравнений (4) целесообразно интегрировать численными методами на ПЭВМ. При этом функциональная готовность  $P_{\text{фг}}(t)$  нейрокомпьютерной системы после начала ее настройки или перестройки (восстановления) после отказа определяется выражением

$$\begin{aligned}
 P_{\text{фг}}(t) &= p_{n0}(t) = \\
 &= 1 - \left[ p_{00}(t) + \sum_{l=0}^{n-1} [p_{l1}(t) + p_{l2}(t)] + \sum_{k=1}^{n-1} p_{k0}(t) \right].
 \end{aligned}$$

В частном случае (когда известно, что все  $\tilde{\mu}_l$  ( $0 \leq l \leq n-1$ ) различны), воспользовавшись преобразованием Лапласа, для вычисления  $P_{\text{фг}}(t)$  НКС можно получить следующую формулу для вычисления функциональной готовности НКС:

$$P_{\text{фг}}(t) = 1 - \prod_{i=0}^{2n-1} M_i \sum_{j=0}^{2n-1} \frac{e^{-M_j t}}{M_j \prod_{\substack{k=0 \\ k \neq j}}^{2n-1} (M_k - M_j)},$$

где  $M_l = \tilde{\mu}_l(1 + \alpha_l)$  ( $0 \leq l \leq n-1$ ) для ( $0 \leq l \leq n-1$ );

$$M_l = \tilde{\mu}_l \quad (0 \leq l \leq n-1) \text{ для } (n \leq l \leq 2n-1);$$

$$M_l \neq 0 \text{ для } (0 \leq l \leq 2n-1).$$

### Математические модели для вычисления функциональной готовности нейронной системы при обучении и переобучении НС после отказов при конечной надежности системы обучения

Развивая изложенные выше идеи, будем считать, что в процессе обучения и переобучения после отказов НС нейрокомпьютерной системы, осуществляемого системой обучения, как правило, в авто-

матризованном режиме принимает участие человек-оператор, который может ошибаться, делать сбой и даже отказывать, а система обучения и переобучения НС имеет конечную надежность. В этом случае, как было сказано выше, нейронную систему следует рассматривать как человекомашинный комплекс и при составлении ее математической модели учитывать человеческий фактор, т. е. возможные ошибки и сбои оператора (интенсивность которых на 2–3 порядка выше интенсивности сбоев технических средств [4, 5]), которые могут быть в процессе настройки НС устранены.

**Первая математическая модель.** Будем рассматривать процесс подготовки НКС к работе как марковский процесс в предположении, что аппаратно-программная часть системы подготовки имеет конечную надежность, а оператор абсолютно надежный.

Обозначим:  $E_i$  ( $i = 0, 1, \dots, n$ ) — состояние НКС на  $i$ -м шаге процесса подготовки к функционированию при обучении или переобучении после отказов НС системы;  $\mu_i = \text{const}$  ( $i = 0, 1, \dots, n - 1$ ) — интенсивность выполнения  $i$ -й операции подготовки;  $E_{ij}$  ( $i = 0, 1, \dots, n - 1; j = 1, 2, \dots, n$ ) — состояние НКС на  $i$ -м шаге подготовки при наличии отказа в аппаратно-программной части системы подготовки;  $\lambda$  — интенсивность отказов аппаратно-программной части системы подготовки НС, а  $\mu$  — интенсивность ее восстановления после отказа.

Тогда поведение рассматриваемой нейронной системы может быть описано (при условии марковского процесса) графом переходов, представленным на рис. 3.

Пусть  $p_i(t)$  — вероятность нахождения системы в состоянии  $E_i$  ( $i = 0, 1, \dots, n$ );  $p_{ij}(t)$  — вероятность на-

хождения системы в состоянии  $E_{ij}$  ( $i = 0, 1, \dots, n - 1; j = 1, 2, \dots, n$ ). Тогда по известным правилам [6] нетрудно составить систему дифференциальных уравнений, описывающих поведение рассматриваемой нейронной системы, при сделанных допущениях:

$$\begin{aligned} p'_0(t) &= \mu p_{01}(t) - (\mu_0 + \lambda)p_0(t); \\ p'_k(t) &= \mu_{k-1}p_{k-1}(t) + \mu p_{k,k+1}(t) - \\ &- [\mu_k p_k(t) + \lambda p_{k,k+1}(t)]; \\ k &= 1, 2, \dots, n - 1 \\ p'_n(t) &= \mu_{n-1}p_{n-1}(t). \end{aligned} \quad (5)$$

Решая систему уравнений (5) аналитически [7] либо численными методами на компьютере при начальных условиях  $p_0(0) = 1; p_i(0) = 0$  ( $1 \leq i \leq n$ );  $p_{ij}(0) = 0$  ( $i = 0, 1, \dots, n - 1; j = 1, 2, \dots, n$ ), легко вычислить функциональную готовность рассматриваемой нейронной системы по формуле

$$P_{\text{фр}}(t) = 1 - \left[ \sum_{i=0}^{n-1} p_i(t) + \sum_{i=0,1,\dots,n-1} \sum_{j=1,2,\dots,n} p_{ij}(t) \right].$$

**Вторая математическая модель.** Используя предположения и обозначения, введенные при рассмотрении первой модели, будем полагать, что оператор, участвующий в обучении и переобучении НС после отказов, имеет конечную надежность. Это означает, что в процессе обучения и переобучения после отказов НС нейрокомпьютерной системы он может допускать ошибки (сбои — временные отказы) с интенсивностью  $\zeta$ , до устранения которых — восстановления работоспособности оператора с интенсивностью  $\nu$ , рассматриваемая нейронная система будет находиться в одном из "отказовых" состояний  $E_{ij}^*$  ( $i = 0, 1, \dots, n - 1; j = 1, 2, \dots, n$ ). Вероятность нахождения системы в состоянии  $E_{ij}^*$  обозначим  $p_{ij}^*(t)$ .

Поведение такой системы описывается графом, представленным на рис. 4.

Очевидно, что в рамках сделанных предположений система дифференциальных уравнений, описывающих поведение такой НКС, будет иметь следующий вид:

$$\begin{aligned} p'_0(t) &= \mu p_{01}(t) + \zeta p_{01}^*(t) - (\mu_0 + \lambda + \nu)p_0(t); \\ p'_k(t) &= \mu_{k-1}p_{k-1}(t) + \mu p_{k,k+1}(t) + \\ &+ \zeta p_{k,k+1}^*(t) - [\mu_k p_k(t) + \\ &+ \lambda p_{k,k+1}(t) + \nu p_{k,k+1}^*(t)]; \\ k &= 1, 2, \dots, n - 1, \\ p'_n(t) &= \mu_{n-1}p_{n-1}(t). \end{aligned} \quad (6)$$

Решение системы уравнений (6) с начальными условиями  $p_0(0) = 1; p_i(0) = 0$  ( $1 \leq i \leq n$ );  $p_{ij}(0) = 0, p_{ij}^*(0) = 0$  ( $i = 0, 1, \dots, n - 1;$

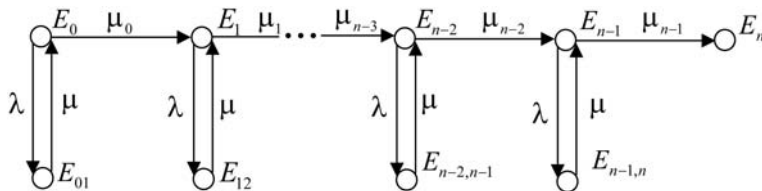


Рис. 3. Граф процесса подготовки к функционированию НКС при абсолютно надежном операторе и конечной надежности аппаратно-программной части системы подготовки к работе

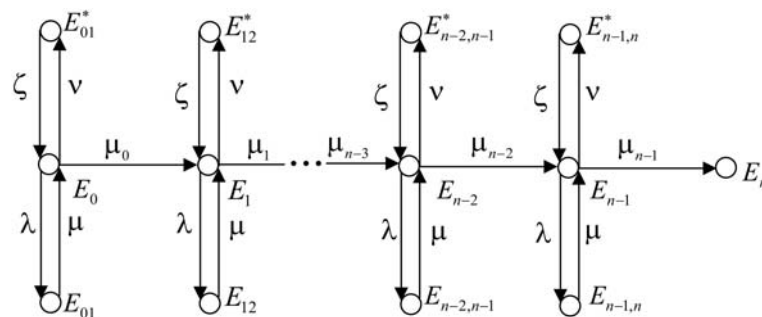


Рис. 4. Граф процесса подготовки к функционированию НКС при конечной надежности оператора и аппаратно-программной части системы подготовки к работе

$j = 1, 2, \dots, n$ ), так же как и системы уравнений (5), не представляет трудности.

Для рассматриваемой модели нейрокомпьютерной системы функциональная готовность вычисляется по формуле

$$P_{\text{ФГ}}(t) = 1 - \left[ \sum_{i=0}^{n-1} p_i(t) + \sum_{j=1, 2, \dots, n} p_{ij}(t) + \sum_{i=0, 1, \dots, n-1} \sum_{j=1, 2, \dots, n} p_{ij}^*(t) \right].$$

**Третья математическая модель.** При рассмотрении первых двух моделей полагали, что интенсивности выполнения каждой  $i$ -й операции подготовки к работе или восстановлению работоспособности НКС после отказов  $\mu_i$  ( $i = 0, 1, \dots, n-1$ ) являются постоянными величинами. В работе [8] показано, что математическая модель для вычисления функциональной готовности НКС более адекватна реальным условиям, если считать, что интенсивности выполнения операций подготовки и восстановления работоспособности НКС после отказов нейронной сети являются функциями времени, т. е.  $\mu_i(t) \neq \text{const}$  ( $0 \leq i \leq n-1$ ). В этом случае задача может быть сведена, как указывалось выше, к построению статистически эквивалентной системы [2], где каждый локальный переход из одного состояния в другое так заменен эквивалентной группой переходов, имеющих постоянные во времени интенсивности, что результирующая условная вероятность перехода в этой группе достаточно близка к условной вероятности перехода в реальной системе.

С учетом сказанного при условии конечной надежности оператора и аппаратной части системы подготовки НКС к работе, используя ранее введенные обозначения и обозначения, сделанные при рассмотрении первой и второй модели, граф состояний, описывающий поведение рассматриваемой модели нейронной системы, можно представить в виде рис. 5.

Система дифференциальных уравнений, описывающих поведение рассматриваемой системы, с уче-

том сделанных предположений, соответствующая графу на рис. 5, может быть записана в следующем виде:

$$\begin{aligned} p'_{00}(t) &= \mu p_0(t) + \zeta p_0^*(t) - \lambda p_0(t) - \\ &\quad - \nu p_0^*(t) - \mu_{01} p_{01}(t) - \mu_{02} p_{02}(t); \\ p'_{10}(t) &= \mu p_1(t) + \zeta p_1^*(t) + \mu_{03} p_{01}(t) + \mu_{04} p_{02}(t) - \\ &\quad - [\mu_{11} p_{11}(t) + \mu_{12} p_{12}(t) + \lambda p_1(t) + \nu p_1^*(t)]; \\ p'_{20}(t) &= \mu p_2(t) + \zeta p_2^*(t) + \mu_{13} p_{11}(t) + \mu_{14} p_{12}(t) - \\ &\quad - [\mu_{21} p_{21}(t) + \mu_{22} p_{22}(t) + \lambda p_2(t) + \nu p_2^*(t)]; \\ &\dots\dots\dots \\ p'_{(n-2)0}(t) &= \mu p_{n-2}(t) + \zeta p_{n-2}^*(t) + \\ &\quad + \mu_{(n-1)1} p_{(n-2)1}(t) + \mu_{(n-1)2} p_{(n-2)2}(t) - \\ &\quad - [\mu_{(n-1)1} p_{(n-1)1}(t) + \mu_{(n-1)2} p_{(n-1)2}(t) + \\ &\quad + \lambda p_{n-2}(t) + \nu p_{n-2}^*(t)]; \\ p'_{(n-1)0}(t) &= \mu p_{n-1}(t) + \zeta p_{n-1}^*(t) + \\ &\quad + \mu_{(n-1)3} p_{(n-1)1}(t) + \mu_{(n-1)4} p_{(n-1)2}(t) - \\ &\quad - [\mu_{n1} p_{n1}(t) + \mu_{n2} p_{n2}(t) + \lambda p_{n-1}(t) + \nu p_{n-1}^*(t)]; \\ p'_{n0}(t) &= \mu_{n3} p_{n1}(t) + \mu_{n4} p_{n2}(t). \end{aligned} \quad (7)$$

Решение системы уравнений (7) в аналитическом виде ввиду громоздкости не представляется возможным, поэтому для ее решения следует использовать численные методы, легко реализуемые на современных профессиональных ПЭВМ.

Для вычисления функциональной готовности рассматриваемой нейронной системы используется следующее выражение:

$$P_{\text{ФГ}}(t) = 1 - \left[ \sum_{i=0, 1, \dots, n} (p_{i0}(t) + p_{i1}(t) + p_{i2}(t)) + \sum_{i=0, 1, \dots, n-1} (p_i(t) + p_i^*(t)) \right],$$

где вероятности  $p_{ij}(t)$ ,  $p_i(t)$ ,  $p_i^*(t)$  соответствуют состояниям НКС, обозначенным одноименными вершинами графа на рис. 5.

Следует отметить, что рассмотренные простейшие модели строились в предположении, что интенсивности отказов и восстановления работоспособности человека-оператора являются постоянными величинами. Однако психическое состояние и внешние факторы, оказывающие влияние на поведение человека, выполняющего интеллектуальную сложную работу, связанную с обучением и переобучением искусственной нейронной системы, влияют на интенсивность его сбоев и отказов и, как следствие, на интенсивность восстановления работоспособности. Поэтому

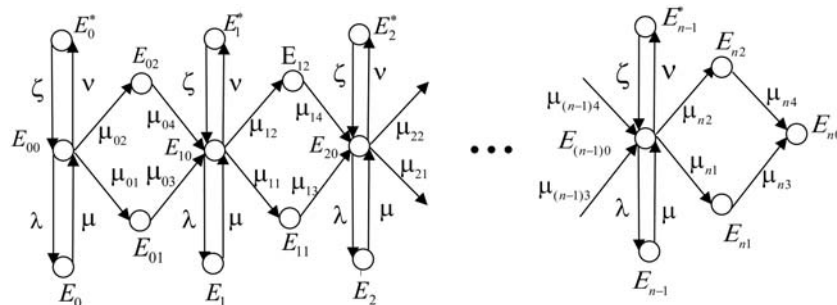


Рис. 5. Граф процесса подготовки к функционированию НКС при конечной надежности оператора и аппаратно-программной части системы подготовки к работе при зависимой от времени интенсивности выполнения операций подготовки

данный фактор следует учитывать при построении более сложных моделей, адекватных конкретным условиям работы человека-оператора в составе автоматизированного комплекса контроля и управления процессом обучения и переобучения нейронной сети НКС после отказов.

#### Список литературы

1. **Потапов И. В.** Надежность нейронных систем. Модели и задачи. Омск: Изд-во ОмГТУ, 2007. 240 с.
2. **Васильев Б. В., Козлов Б. А., Ткаченко Л. Г.** Надежность и эффективность радиоэлектронных устройств. М.: Сов. радио, 1964. 368 с.

3. **Потапов В. И., Потапов И. В.** Математические модели, методы и алгоритмы оптимизации надежности и технической диагностики искусственных нейронных сетей. Омск: Изд-во ОГУП Омская областная типография, 2004. 220 с.
4. **Галушкин А. И.** Теория нейронных сетей. Нейрокомпьютеры и их применение. М.: ИПРЖ "Радиотехника", 2000. 416 с.
5. **Литвинов В. А., Крамаренко В. В.** Контроль достоверности и восстановление информации в человеко-машинных системах. Киев: Техніка, 1986. 200 с.
6. **Вентцель Е. С.** Исследование операций. М.: Сов. радио, 1972. 550 с.
7. **Козлов Б. А., Ушаков И. А.** Справочник по расчету надежности аппаратуры радиоэлектроники и автоматики. М.: Сов. радио, 1975. 472 с.
8. **Потапов И. В.** Модели для расчета надежности нейрокомпьютерной системы при настройке и восстановлении после отказов // Омский научный вестник, 2006. № 3(36). С. 129–132.

УДК 681.51:519

**Л. В. Красовская,**  
канд. техн. наук, ст. преподаватель,  
Дагестанский государственный  
технический университет

## Имитация отработки действий на нечетких семантических сетях в алгоритмах самообучения интеллектуальных систем

*Разработана методика сравнения нечетко представленных отношений в модели проблемной среды (ПС), отличающаяся от известных тем, что она позволяет выполнять несмещенную оценку равенства нечетко выраженных показателей и тем самым повысить достоверность сравнения. Описаны алгоритмы самообучения, позволяющие формировать программы целесообразного поведения в различных проблемных средах, отличающиеся от известных, имитацией отработки пробующих действий на нечетких семантических сетях, что дает возможность исключить влияние интеллектуальных систем на ПС в процессе изучения закономерностей среды.*

**Ключевые слова:** интеллектуальные системы, проблемная среда, нечеткая семантическая сеть, множество вершин и ребер, характеристик, терм, алгоритмы самообучения

Характерной особенностью интеллектуальных систем (ИС), способных функционировать в условиях неопределенности, является то, что знания таких систем должны быть не только структурированы, но и представлены безотносительно к конкретным условиям функционирования. Для описания ситуаций проблемной среды (ПС) безотносительно к конкретной области можно использовать

нечеткие семантические сети (НСС) активного и пассивного типа [1].

Формально нечеткая семантическая сеть является ориентированным нечетким мультиграфом  $G_1 = (V_1, E_1)$ , где  $V_1 = v_i, i = 1, n_{-1}$  и  $E_1 = e_i, i = 1, n$  — соответственно множество вершин и ребер. Вершины  $v_i \in V_1$  биективно соответствуют объектам ПС, ребра — отношениям, складывающимся в среде между этими объектами. Вершины  $v_i \in V_1$  могут быть двух типов: свободные  $v_i^*$  и занятые  $v_i$ . Каждая свободная (активная) вершина  $v_i^* \in V_1$  определяется множеством характеристик  $X_i$ , которыми должны обладать конкретные объекты  $o_{i1} \in O$ , чтобы была разрешена пометка этой вершины их именами в конкретной (текущей) ситуации ПС. После выполнения такой пометки активная вершина  $v_i^*$  становится пассивной  $v_i$  и определяется множеством характеристик  $X_{i1}$  конкретного объекта, которым она помечена. Иными словами, активная вершина  $v_i^* \in V_1$  помечается объектом  $o_i(X_i) \in O$ , если выполняется условие  $X_i \subset X_{i1}$ , где запись  $o_{i1}(X_{i1})$  означает, что объект  $o_{i1}$  описывается множеством характеристик  $X_{i1}$ .

Ребра  $e_i \in E_1$  или отношения между объектами ПС задаются парами  $\langle \mu(x_i), T_j \rangle$ , где  $T_j$  — нечеткое значение (терм) лингвистической переменной  $T_j \in T^*$ ;  $\mu(x_i) \in [0, 1]$  — степень принадлежности количественного значения лингвистической переменной  $T_j \in T^*$  к интервалу численных значений термина  $T_j$ ;  $T^* \subset \{T_j\}$  — множество лингвистических переменных, биективно соответствующих семантическому определению различных отношений.

В рассмотренном случае при описании НСС ограничения, определяемые элементами терм-множества лингвистической переменной, и накладываемые на базовые переменные  $x_i \in U_i$ , задаются четко и вычисляются исходя из функционального назначения и возможностей ИС. Иначе говоря, множество  $U$  разбивается на  $i = \overline{1, k}$  непересекающихся



открытых справа интервалов согласно заданному на его элементах отношению эквивалентности "находиться внутри интервала  $J$ ".

Для перехода от количественных значений отношений между объектами ПС, измеренных с помощью информационно-измерительной ИС и определяемых базовыми значениями  $x_i$  лингвистических переменных, к качественным их значениям, т. е. к одному из термов  $T_j$ , можно использовать преобразования следующего вида:

$$F: i \rightarrow \begin{cases} T_j, & \text{если } 0 \leq x_i < x_i^*; \\ T_j', & \text{если } x_i^* < x_i^{**} \leq x_{i+1}^*; \\ T_j'', & \text{если } x_{ik-1}^* < x_{i+1}^* \leq x_{ik}^*, \end{cases} \quad (1)$$

где  $x_i^*$  и  $x_{i+1}^*$ ,  $i = \overline{1, k}$  — соответственно нижняя и верхняя граница числовых значений  $x_i^{**}$  терма  $T_j'$ ;  $[x_i^{**}]$  — середина интервала числовых значений этого терма;  $x_{ik-1}^*$ ,  $x_{ik}^*$  — соответственно нижняя и верхняя граница числовых значений терма  $T_j''$ ;  $x_{i+1}^*$  — середина интервала числовых значений терма  $T_j''$ .

Степень принадлежности  $\mu(x_i)$  значений базовой переменной  $x_i$  к множеству числовых значений терма  $T_j$  может вычисляться согласно следующему характеристическому уравнению:

$$\mu(x_i) = \begin{cases} \frac{x_i - x_i^*}{x_i^{**} - x_i^*}, & \text{если } x_i \in [x_i^*, x_i^{**}]; \\ \frac{x_i - x_i^{**}}{x_{i+1}^* - x_i^{**}}, & \text{если } x_i \in [x_i^{**}, x_{i+1}^*]. \end{cases} \quad (2)$$

Для сравнения двух значений отношений между собой, заданных тройками  $\langle x_i, \mu(x_i), T_j \rangle$  и  $\langle x_i', \mu(x_i'), T_j' \rangle$  введем характеристику степени их равенства (близости)  $\rho(x_i, x_i')$ , которую можно вычислять следующим образом:

$$\rho(x_i, x_i') = \begin{cases} \text{а) } 1, & \text{если } (|x_i - x_i'| < \varepsilon_0) \& (T_j = T_j'); \\ \text{б) } \mu(x_i) \leftrightarrow \mu(x_i'), & \text{если } (|x_i - x_i'| > \varepsilon_0) \& \\ & \& (T_j = T_j') \& ((x_i, x_i') \in [x_i^*, x_i^{**}] \vee \\ & \vee (x_i, x_i') \in [x_i^{**}, x_{i+1}^*]); \\ \text{в) } (\mu(x_i) \leftrightarrow (1 - \mu(x_i'))), & \text{если } |x_i - x_i'| > \varepsilon_0 \& \\ & \& (T_j = T_j') \& ((x_i, x_i') \notin [x_i^*, x_i^{**}] \vee \\ & \vee (x_i, x_i') \notin [x_i^{**}, x_{i+1}^*]); \\ \text{г) } 0, & \text{если } T_j \neq T_j', \end{cases} \quad (3)$$

где  $\varepsilon_0$  — параметр, задающий приведенное значение точности сравнения величины отношения;  $x_i'$  — базовая переменная к множеству числовых значений терма  $T_j'$ ;  $\leftrightarrow$  — операция расплывчатой эквивалентности, определяемая по формуле  $\min(\max(\mu(x_i), (1 - \mu(x_i'))), \max(\mu(x_i'), (1 - \mu(x_i'))))$  [2];  $\&$  — конъюнкция, обозначающая одновременность выполняющихся условий.

Предложенное выражение для определения степени близости позволяет избежать искажения получаемого результата, когда степени принадлежности сравниваемых значений отношений равны, но находятся по разные стороны от центра тяжести функции принадлежности.

Выражение (3) можно обосновать следующим образом. Два количественных значения отношения равны между собой, если они попадают в интервал числовых значений одного и того же терма  $T_j$  в окрестность одной и той же точки, определяемую значением параметра  $\varepsilon_0$  (случай 3а), два количественных значения  $x_i$  и  $x_i'$  нечетко равны между собой, если они принадлежат интервалу числовых значений одного и того же терма  $T_j$ . Причем, если оба значения степени принадлежности сравниваемых значений попадают в одну и ту же половину интервала числовых значений терма, то степень сравнения вычисляется по (3б), в противном случае, она определяется по (3в). Значения  $x_i$  и  $x_i'$  не равны между собой, если они попадают в интервалы числовых значений различных термов лингвистической переменной  $T_j$  (случай 3г). Следовательно, два значения одного и того же отношения равны при  $\rho(x_i, x_i') = 1$ . Эти значения отношения являются нечетко равными, если  $\rho(x_i, x_i') > 0$  и они не равны в случае, когда  $\rho(x_i, x_i') = 0$ .

Рассмотрим ПС как множество взаимосвязанных между собой объектов и независимых от ИС событий  $Q = \{q_{i_1}\}$ ,  $i_1 = \overline{1, n_1}$ . В каждый дискретный момент времени  $t$  среду можно охарактеризовать текущей ситуацией  $s_i^t \in S$ , определяемой текущими состояниями находящихся в ней объектов и характером отношений между этими объектами. Часть ситуаций  $S$  — множество  $C_2 = \{c_{i_2}\}$ ,  $i_2 = \overline{1, n_2}$  будем называть стандартными (безусловными). Эти ситуации определяют различные цели и подцели условного функционирования ИС и вызывают у нее при восприятии соответствующие стандартные реакции (СР), связанные с достижением заданной цели  $S_{\text{цел}}$ . Для имитации отработки действий

на НСС каждое из них определяется с помощью следующего формата описания: <<имя действия> <НСС, определяющая допустимые условия обработки действия>> <НСС, описывающая результат обработки>>, которое будем называть фреймом действия (ФД).

Первая часть — <имя действия> является идентификатором действия. Вторая часть — <условия, выполнение которых в ПС требуется для успешной обработки действия> представляет собой активную НСС, формальное описание которой является мультиграфом  $G_1 = (V_1, E_1)$ , где  $V_1$  — множество свободных вершин, каждая из которых помечается списком характеристик  $X_j$ , которыми должны обладать объекты, чтобы было допустимым выполнение над ними действия ФД.

Третья часть ФД — <результат обработки действия> — представляет собой НСС, получаемую из сети  $G_1$  после обработки действия этого фрейма.

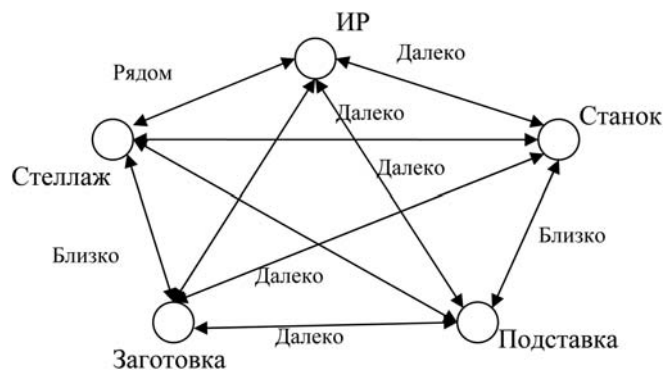
Имитация обработки действий проводится следующим образом. На первом этапе определяются все действия, которые можно непосредственно выполнить в ПС согласно содержанию второй части соответствующих действий ФД. Затем выбирается конкретное действие для обработки, и осуществляется имитация его обработки на НСС, определяющей текущие условия функционирования. В НСС, определяющей текущую ситуацию ПС по содержанию третьей части ФД, вносятся соответствующие изменения значений отношений между объектами среды, которые получаются в результате непосредственной обработки действия в ПС. Если в результате имитации обработки действия получается ситуация, которая приближается к целевой ситуации по своему содержанию, то формируется звено в цепи поведения в форме имплицативного решающего правила  $S_{\text{тек}} \& b_j^1 \rightarrow S'$ , где приведенная запись означает, что при восприятии текущей ситуации ПС  $S_{\text{тек}}$  обработка действия  $b_j^1$  приводит к ее преобразованию в результирующую ситуацию  $S'_{\text{тек}}$ . Причем степень близости  $\rho(S_{i+1\text{тек}}, S_{\text{цел}}) > \rho(S_{i\text{тек}}, S_{\text{цел}})$ , т. е. действие  $b_j^1$  преобразует ситуацию  $S_{i\text{тек}}$  в ситуацию  $S_{i+1\text{тек}}$ ; между вновь полученной и целевой ситуациями наблюдается меньшее число различий, чем между целевой и исходной ситуациями. В результате формируется модель целесообразного поведения следующего вида:

$$L(x) = S_{T_{i+1}} \& b_j \rightarrow S_{T_{i+1}}^2 \& b_{j+1} \rightarrow \dots \rightarrow S_{T_{i+n}} \& b_{j+n} \rightarrow S_{\text{цел}}$$

Полученная в процессе самообучения модель поведения закрепляется окончательным образом

достижением цели после ее непосредственной реализации в ПС.

С учетом выражений (1) и (2) мультиграф, биективно соответствующий описанной ситуации ПС, представлен на рисунке.



Аналитически этот граф можно описать следующим образом:

$$\{ИР (\{ \{ \{ \{ \langle \langle 0,2/рядом \rangle \langle заготовка \rangle \rangle \& \langle \langle 0,2/рядом \rangle \langle стеллаж \rangle \rangle \& \langle \langle 1/далеко \rangle \langle станок \rangle \rangle \& \langle \langle 1/далеко \rangle \langle подставка \rangle \rangle \} \} \& \{ станок (\{ \{ \langle \langle 0,2/рядом \rangle \langle подставка \rangle \rangle \& \langle \langle 1/далеко \rangle \langle стеллаж \rangle \rangle \} \} \& \{ подставка (\{ \{ \langle \langle 0,2/рядом \rangle \langle станок \rangle \rangle \& \langle \langle 1/далеко \rangle \langle заготовка \rangle \rangle \} \} \} \}$$

где & — конъюнкция, обозначающая одновременность происходящего события; рядом, близко, далеко — термы лингвистической переменной "расстояние". При аналитическом описании НСС выражение, заключенное в фигурные скобки, называется отдельным фрагментом сети, а имена объектов, с которых начинаются фрагменты, определяются как ключевые понятия фрагментов. Фрагменты НСС, необходимые для описания ПС, достаточного для принятия решений, можно определить эвристическим путем. Например, первый базовый фрагмент строится относительно ключевой вершины сети, помеченной понятием ИР. Следующие необходимые фрагменты формируются относительно объектов, входящих в структуру заданного на текущий момент времени целевого условия, но над которыми ИС непосредственных действий не выполняет. Третий вид фрагментов строится в процессе функционирования относительно объектов, используемых ИС в качестве вспомогательных инструментов.

Таким образом, процесс самообучения ИС можно рассматривать как автоматическую генерацию графа при условии, что ей априорно неизвестны результаты, к которым могут привести обрабатываемые действия.

#### Список литературы

1. Берштейн Л. С., Мелехин В. Б. Планирование поведения интеллектуального робота. М.: Энергоатомиздат, 1994. 238 с.
2. Мелехов А. Н., Берштейн Л. С., Коровин С. П. Экспертные соответствующие системы с нечеткой логикой. М.: Наука, 1991. 270 с.

**Л. С. Куравский**, д-р техн. наук, проф., декан,  
**П. А. Мармалюк**, науч. сотр.,  
**А. С. Панфилова**, программист,  
 Московский городской  
 психолого-педагогический университет,  
 e-mail: l.s.kuravsky@gmail.com  
**Д. В. Ушаков**, д-р псих. наук, проф., зав. лаб.,  
 Институт психологии РАН,  
 dv.usakov@gmail.com.

## Введение

Измеряемые психологами параметры, как правило, не представляют исследуемые характеристики в форме, удобной для непосредственной интерпретации и психологической диагностики. Как правило, в случае многомерных измерений исследователи стараются выявить несколько латентных факторов, отвечающих за изменчивость наблюдаемых переменных, а затем определить их природу, чтобы использовать далее полученную информацию для анализа собранных данных. При этом для обеспечения минимальных потерь полезной информации, параметры, которые легко измерять, могут заменяться на факторы, которые легко интерпретировать. При последующем анализе определяются функциональные зависимости наблюдаемых переменных от выявляемых факторов. В итоге выявляется вся структура причинных связей между факторами и наблюдаемыми переменными, а также, если потребуется, непосредственные значения факторов, необходимые для идентификации индивидуальных различий испытуемых.

Для решения таких задач разработаны эмпирические математические модели и соответствующие методы многомерного статистического анализа [10–12, 17]. Наиболее приемлемыми в рассматриваемой ситуации являются исследовательские (эксплораторные) и проверочные (конфирматорные) факторные модели и методы их исследования. Оба подхода работают с выборочными матрицами ковариаций или корреляций наблюдаемых переменных. Исследовательский анализ предполагает наличие некоторого неизвестного заранее числа некоррелированных факторов с неопределенной интерпретацией, тогда как при конфирматорном факторном анализе факторы, их интерпретация и причинные связи с наблюдаемыми переменными, а также корреляционные связи между латентными факторами определяются постановкой задачи. Конфирматорные модели позволяют с помощью несложной процедуры оценивать статистическую значимость каждого своего компонента.

Факторный анализ, как общенаучный метод, эффективно применяется для решения задач из самых разных прикладных областей. Однако традиционные формы факторного анализа имеют ряд существенных недостатков:

- для определения свободных параметров факторных моделей в общем случае необходимо численно решать трудоемкую задачу многомерной оптимизации, в результате решения которой находится один из локальных минимумов критерия, зависящий от начального приближения (глобальный минимум, как правило, определить нельзя);

## Исследование факторных влияний на развитие психологических характеристик с применением нового подхода к оценке адекватности моделей наблюдения\*

*Представлен новый метод выявления и исследования факторов, определяющих развитие психологических характеристик. Предложенный подход опирается на возможности вейвлет-преобразований и факторного анализа. В соответствии с ним, значения коэффициентов, полученные в результате дискретного вейвлет-преобразования временного ряда наблюдаемого процесса и соответствующие различным периодам наблюдений, рассматриваются как значения наблюдаемых переменных в последующем конфирматорном факторном анализе, который, в свою очередь, используется для выявления динамики факторных влияний и оценки показателей взаимодействий между факторами. Идентификация свободных параметров факторной модели (как правило, факторных дисперсий и ковариаций) выполняется с помощью новой прямой (неитерационной) процедуры, опирающейся на метод максимального правдоподобия, что является альтернативой традиционному поиску локального решения задачи многомерной численной оптимизации. Рассматривается метод оценки статистической значимости компонентов факторной модели, а также новый способ оценки степени адекватности произвольных факторных моделей, который опирается на метод Монте-Карло и возможности самоорганизующихся карт Кохонена. Этот способ позволяет избежать жестких ограничений, налагаемых на вероятностные распределения результатов наблюдений, присущие традиционной процедуре идентификации свободных параметров модели.*

**Ключевые слова:** факторный анализ, путевая диаграмма, критерий адекватности модели

\* Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 10-06-00423).

- эти методы обычно требуют определенных вероятностных распределений результатов наблюдений, степень согласованности с которыми сложно проверить на практике ввиду ограниченности выборки исследуемых данных.

Следует отметить, что симплекс-метод конфирматорного факторного анализа [13], который применяется для анализа лонгитюдных данных, допускает сравнительно небольшой запас свободных параметров и позволяет исследовать взаимодействие факторов между смежными контрольными точками. Он применим только в случае ковариационных и корреляционных матриц наблюдаемых переменных, имеющих симплексную структуру.

Чтобы преодолеть указанные проблемы, был разработан представленный далее новый подход [3–5, 7, 15, 16], опирающийся на возможности вейвлет-преобразований, факторного анализа и нейронных сетей. К его особенностям и преимуществам относятся:

- возможность нахождения оценок свободных параметров модели прямыми методами, обеспечивающими однозначное оптимальное решение (в случае модели дисперсионных составляющих);
- гибкие средства для исследования взаимодействия факторов;
- возможность анализа взаимного влияния факторов с помощью методов структурного моделирования;
- применимость для анализа ковариационных и корреляционных матриц произвольной структуры;
- большой запас свободных параметров факторных моделей;
- критерий адекватности моделей, не налагающий ограничения на распределения наблюдаемых характеристик.

### 1. Основные этапы анализа

Основные этапы предлагаемого подхода представлены на рис. 1. Выборки коэффициентов, полученных в результате дискретных вейвлет-преоб-

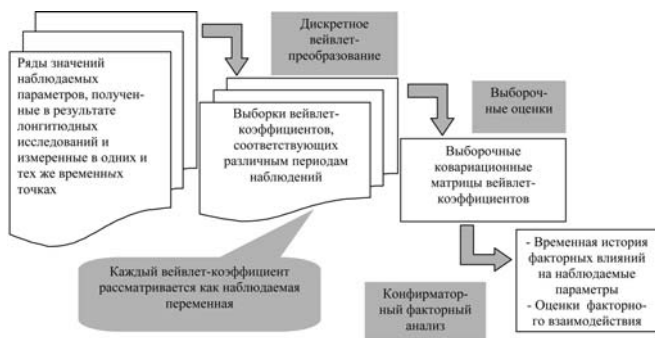


Рис. 1. Основные этапы анализа

разований временных рядов значений исследуемого параметра и соответствующих различным периодам наблюдений, в последующем конфирматорном факторном анализе рассматриваются как значения наблюдаемых переменных, которые позволяют выявить временную историю факторных влияний и оценки факторного взаимодействия.

Представление данных, полученное с помощью вейвлет-преобразований, дает возможность выявить различия в характеристиках процесса в различных шкалах, что особенно важно при большом числе контрольных моментов времени, связанных с проведением измерений. Идентификация свободных параметров факторной модели (обычно корреляций или ковариаций) выполняется с помощью новой прямой (неитерационной) процедуры, которая опирается на метод максимального правдоподобия и является альтернативой традиционному итерационному способу решения задач локальной оптимизации.

### 2. Альтернативные факторные модели

Для создания факторных моделей применяют модели путевых коэффициентов и дисперсионных составляющих [18], включая их модификации. Примеры типовых вариантов этих моделей показаны на рис. 2 и 3.

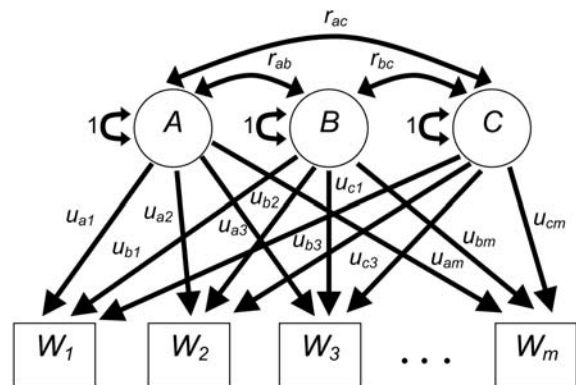


Рис. 2. Модель путевых коэффициентов, представленная путевой диаграммой

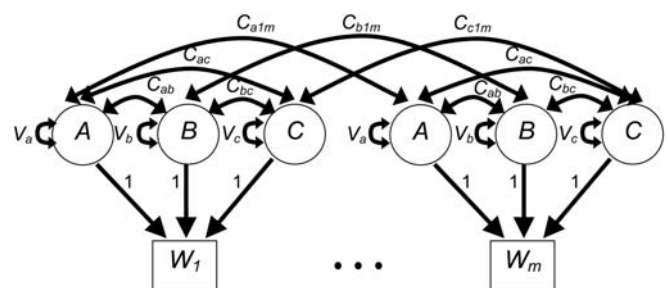


Рис. 3. Модель дисперсионных составляющих, представленная путевой диаграммой

В случае модели путевых коэффициентов аналитические выражения для ковариаций и дисперсий вейвлет-коэффициентов  $W_i$  нелинейны:

$$\text{Cov}(W_i, W_j) = \sum_k \sum_l r_{kl} u_{ki} u_{lj};$$

$$\text{Var}(W_i) = \sum_k \sum_l r_{kl} u_{ki} u_{li};$$

где  $k$  и  $l$  — индексы факторов;  $u_{**}$  — факторные нагрузки;  $r_{**}$  — корреляции между факторами, что не позволяет получать простые и однозначные оценки свободных параметров модели. Подобные выражения в случае модели дисперсионных составляющих являются линейными:

$$\text{Cov}(W_i, W_j) = \sum_k C_{kij},$$

$$\text{Var}(W_i) = \sum_k V_k + \sum_k \sum_l C_{kl},$$

где  $k$  и  $l$  — индексы факторов;  $V_*$  — дисперсии;  $C_{**}$  и  $C_{***}$  — ковариации между факторами.

Отражая важные для исследования особенности прикладных задач, модель дисперсионных составляющих принимает различные частные формы. Например, при изучении факторных влияний в различных условиях может быть полезен одновременный анализ различных групп моделей (рис. 4).

### 3. Идентификация свободных параметров модели: проблемы

При идентификации свободных параметров факторных моделей методом максимального правдоподобия в качестве минимизируемого критерия используется функция

$$F = [\ln|\Sigma| - \ln|S| + \text{tr}(S\Sigma^{-1}) - q](N - 1),$$

где  $S$  — выборочная ковариационная матрица наблюдаемых переменных;  $\Sigma$  — прогнозируемая ковариационная матрица наблюдаемых переменных;  $|\Sigma|$  и  $|S|$  — определители матриц  $\Sigma$  и  $S$ ;  $\text{tr}(S\Sigma^{-1})$  — след матрицы  $(S\Sigma^{-1})$ ;  $N$  — объем выборки, использованной для вычисления матрицы  $S$ ;  $q$  — число наблюдаемых переменных [6]. Элементы прогнозируемой ковариационной матрицы представляют собой аналитические выражения относительно свободных параметров модели. В случае многомерного нормального распределения наблюдаемых переменных значения критерия  $F$  описываются распределением  $\chi^2$ .

Для определения свободных параметров модели путевых коэффициентов необходимо численно решить итерационными методами достаточно трудоемкую задачу локальной многомерной оптимизации. Из этого в общем случае вытекает невозможность определения глобального минимума, так как в результате решения находится один из локальных

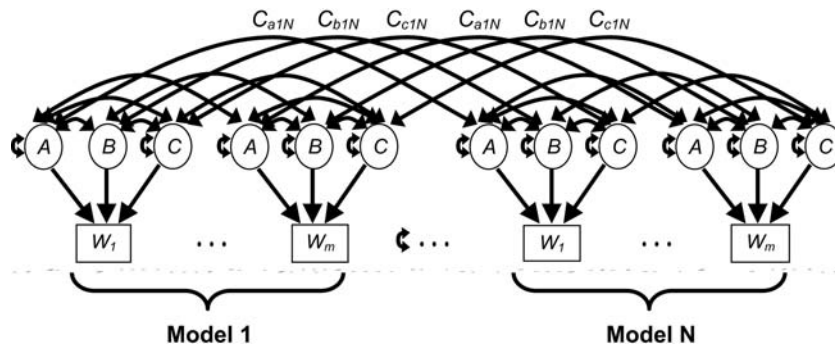


Рис. 4. Изучение факторных влияний в различных условиях: одновременный анализ различных групп моделей ( $C_{***}$  — ковариации между факторами)

минимумов, выбор которого зависит от начального приближения. Таким образом, получаемое решение неоднозначно.

Альтернативная процедура решения [3] включает в себя следующие этапы:

- составление переопределенной системы линейных алгебраических уравнений, выражая аналитически выборочные дисперсии и ковариации через аналогичные свободные параметры факторной модели;
- решение полученной системы прямым (неитерационным) методом, используя метод максимального правдоподобия<sup>1</sup>;
- проверка адекватности полученных моделей наблюдениям, опираясь на статистические критерии согласия.

Чтобы избежать составления сложных для решения нелинейных систем уравнений относительно коэффициентов корреляции и факторных нагрузок, используется путевая модель дисперсионных составляющих, в которой факторные нагрузки равны единице. Каждой наблюдаемой дисперсии и ковариации ставится в соответствие алгебраическое уравнение, которое связывает ее выборочную оценку с соответствующей прогнозируемой величиной, выраженной аналитически через неопределенные дисперсии и ковариации латентных переменных. В результате получается линейная система, число уравнений которой равно числу наблюдаемых дисперсий и ковариаций.

Представим полученную переопределенную систему  $n$  уравнений в матричной форме:

$$Ax = b,$$

где  $A$  — матрица системы, коэффициенты которой определяются факторной моделью;  $b$  — вектор-столбец  $n$  выборочных дисперсий и ковариаций, определяемых результатами наблюдений;  $x$  — вектор-столбец  $m$  искоемых дисперсий и ковариаций латентных переменных.

Рассмотрим вектор  $\varepsilon = Ax_* - b$ , представляющий полученную методом наименьших квадратов

<sup>1</sup> В другой форме, чем при традиционном подтверждающем факторном анализе.

невязку псевдорешения  $\mathbf{x}_*$  переопределенной системы. Полагая в общем случае, что компоненты вектора невязок коррелированы, выразим их невырожденную ковариационную матрицу как  $\sigma^2 \mathbf{V}$ .

Если рассматриваемая система невырождена ( $\text{rank } \mathbf{A} = m$ , где  $m$  — число свободных параметров модели), вектор невязки имеет многомерное нормальное распределение, а  $\mathbf{x}_* = (\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{V}^{-1} \mathbf{b}$  — псевдорешение, полученное методом наименьших квадратов, то это псевдорешение является *оценкой максимального правдоподобия*, а статистика  $X^2 = (\mathbf{b} - \mathbf{A}\mathbf{x}_*)^T \mathbf{V}^{-1} (\mathbf{b} - \mathbf{A}\mathbf{x}_*) / \sigma^2$  имеет распределение  $\chi^2$  с  $n - m$  степенями свободы [1].

Указанная статистика  $X^2$  позволяет, при заданных выше предположениях, проверить *гипотезу о представимости* выборочных дисперсий и ковариаций, составляющих вектор  $\mathbf{b}$ , дисперсиями и ковариациями латентных переменных, содержащихся в исследуемой модели. Область принятия гипотезы есть  $X^2 \leq \chi_{n-m; \alpha}^2$ , где  $\alpha$  — уровень значимости критерия.

При реализации данного подхода удобно сделать следующие упрощающие предположения, обусловленные особенностями искомого решения:

- компоненты вектора невязок  $\boldsymbol{\varepsilon}$  являются некоррелированными;
- значения средних квадратических отклонений различных компонентов вектора невязок  $\boldsymbol{\varepsilon}$  составляют одну и ту же фиксированную долю (процент)<sup>2</sup> от соответствующих компонентов вектора  $\mathbf{b}$ .

Чтобы обеспечить сопоставимость оценок, указанная доля (процент) подбирается так, чтобы равенство  $X^2 = \chi_{n-m; \alpha}^2$  выполнялось при  $\alpha = 0,05$ .

Для оценки степени допустимости вычисленной характеристики для нее удобно установить разумное критическое значение, например 0,1. Таким образом, вместо уровня значимости появляется новый показатель — *критический процент*.

Преимуществом предложенного подхода является то, что он не сводит решение к трудоемкой процедуре многомерной локальной оптимизации, которая не гарантирует нахождение глобальных минимумов, обеспечивая при этом однозначность результата. Рассмотренный метод решения позволяет выявлять характер взаимных связей между свободными параметрами факторной модели путем прямого расчета по матричным формулам, численно вычисляя оценки одних характеристик для заданного множества сочетаний значений других.

В ряде моделей для обеспечения корректного результата на оцениваемые дисперсии и ковариации следует налагать дополнительные условия в форме неравенств, выражающие положительность значений дисперсий и корректность соотношений

между связанными дисперсиями и ковариациями. В этом случае в качестве решения принимается вектор  $\mathbf{x}_c$ , удовлетворяющий указанным условиям и ближайший к найденному псевдорешению  $\mathbf{x}_*$  в евклидовой метрике. Для поиска вектора  $\mathbf{x}_c$  решается несложная задача линейной оптимизации. Гипотеза о представимости выборочных дисперсий и ковариаций, составляющих вектор  $\mathbf{b}$ , дисперсиями и ковариациями латентных переменных, составляющими вектор  $\mathbf{x}_c$ , проверяется, как и ранее, с помощью статистики

$$X^2 = (\mathbf{b} - \mathbf{A}\mathbf{x}_c)^T \mathbf{V}^{-1} (\mathbf{b} - \mathbf{A}\mathbf{x}_c) / \sigma^2,$$

имеющей распределение  $\chi^2$  с  $n - m$  степенями свободы.

Как и в традиционном конфирматорном факторном анализе, рассматриваемый метод дает возможность строить заключения о статистической значимости различных компонентов модели, используя статистические критерии согласия.

Для этого следует сравнить статистики  $X^2$  для двух моделей: полной модели, содержащей исследуемый компонент, и упрощенной модели, в которой этот компонент отсутствует. Гипотезу о том, что полная модель согласуется с результатами наблюдений, будем обозначать как  $H_f$ . Выявление степени значимости исследуемого компонента осуществляется, если отвергать гипотезу  $H_f$  нет оснований. Сначала следует оценить свободные параметры упрощенной модели. Полученное значение статистики  $X^2$  для упрощенной модели сравнивается с аналогичной характеристикой для полной модели.

Поскольку разность указанных статистик асимптотически распределена как  $\chi^2$  с числом степеней свободы, равным разности в числах степеней свободы полной и упрощенной моделей, эта разность используется для проверки нулевой гипотезы  $H_r$  о том, что упрощенная модель согласуется с результатами наблюдений, против альтернативной гипотезы  $H_f$ .

Если гипотеза  $H_r$  не отвергается при заданном уровне значимости, то исследуемый компонент признается статистически незначимым, и делается вывод о том, что имеющиеся данные не свидетельствуют о его влиянии на данную характеристику. Если гипотеза  $H_r$  отвергается (а гипотеза  $H_f$  принимается), то можно говорить о влиянии исследуемого компонента на эту характеристику.

Однако корректное использование метода максимального правдоподобия для идентификации значений свободных параметров и оценки степени адекватности модели, описанное выше для традиционного и альтернативного вариантов конфирматорного факторного анализа, предполагает проверку многомерной нормальности наблюдаемых переменных или компонентов вектора невязки. Эта процедура является трудоемкой и зачастую невозможной ввиду малой выборки исследуемых данных.

<sup>2</sup> Гипотеза о пропорциональности.

Кроме того, при традиционном конфирматорном факторном анализе критерий максимального правдоподобия является очень чувствительным к объему выборки: сравнительно небольшие отклонения от прогнозируемых характеристик приводят к существенному ухудшению согласования модели с наблюдениями.

Для преодоления указанных проблем предлагается новый критерий адекватности факторных моделей [7, 15, 16], который использует метод Монте-Карло и возможности самоорганизующихся карт признаков (сетей Кохонена) [2, 14].

#### 4. Оценка степени адекватности моделей наблюдениям

Оценка степени адекватности моделей наблюдениям основана на сравнении вектора невязки псевдорешения и случайной выборки векторов невязок  $\epsilon_r$  для его окрестности, в которых определенный процент случайных компонентов выходит за границы доверительных интервалов. Векторы невязок  $\epsilon_r$  вычисляются с использованием рассматриваемой факторной модели. Предлагаются две разновидности нового критерия адекватности: для оценки адекватности линейных моделей дисперсионных составляющих и нелинейных моделей путей коэффициентов.

##### 4.1. Оценка степени адекватности наблюдениям в случае моделей дисперсионных составляющих

Выборка случайных векторов невязок  $\epsilon_r$  используется для обучения самоорганизующихся карт признаков подходящей размерности в целях получения выборок евклидовых расстояний между векторами невязок, используемыми в качестве входных данных, и центрами нейронов-победителей обучен-

ной сети. Учитывая достаточно высокую размерность векторов невязок, характерную для практических задач, можно говорить о том, что распределение полученных евклидовых расстояний близко к нормальному. Выборочные оценки средних и дисперсий этих расстояний идентифицируют указанное распределение и позволяют оценивать вероятности превышения расстояния между полученным вектором невязки псевдорешения и соответствующим центром нейрона-победителя, что дает возможность судить о степени адекватности модели наблюдениям. Схема вычисления предлагаемых оценок представлена на рис. 5.

Самоорганизующиеся карты признаков Кохонена, для обучения которых генерируются последовательности векторов с заданным средним процентным соотношением<sup>3</sup> случайных компонентов, выходящих за границы доверительных интервалов, позволяют выявить характерную структуру отклонений дисперсий и ковариаций наблюдаемых переменных от полученных оценок. Сопоставление оценок адекватности модели при различных процентных соотношениях позволяет выявить наиболее правдоподобную покомпонентную структуру статически значимых отклонений от вектора невязки псевдорешения.

Сгенерированные дисперсии и ковариации формируют вектора невязок  $\epsilon_r = Ax_* - Ax_r$ , необходимые для обучения карт Кохонена. Процедура их вычисления показана на рис. 6. Генерация обучающих выборок позволяет рассматривать представленный подход как одну из форм метода Монте-Карло.

Выборочные оценки дисперсий вычисляются с помощью следующей формулы, полученной из выражения для распределения выборочной дисперсии нормально распределенной случайной величины<sup>4</sup>:

$$V_s = \frac{\chi_{N-1}^2}{N-1} V,$$

где  $V_s$  — это оценка выборочной дисперсии;  $N$  — размер генерируемой выборки векторов невязки;  $V$  — дисперсия, входящая в состав найденного оптимального псевдорешения;  $\chi_{N-1}^2$  — генерируемый случайный элемент, распределенный как  $\chi^2$  с  $N-1$  степенями свободы.

Формируемые оценки ковариаций подсчитываются через соответствующие значения выборочных корреляций ис-

<sup>3</sup> От 0 до 100 % через определенные интервалы.

<sup>4</sup> Таким образом, формируемые выборки значений компонентов псевдорешения считаются нормально распределенными. Это предположение не накладывает каких-либо ограничений на наблюдаемые данные.

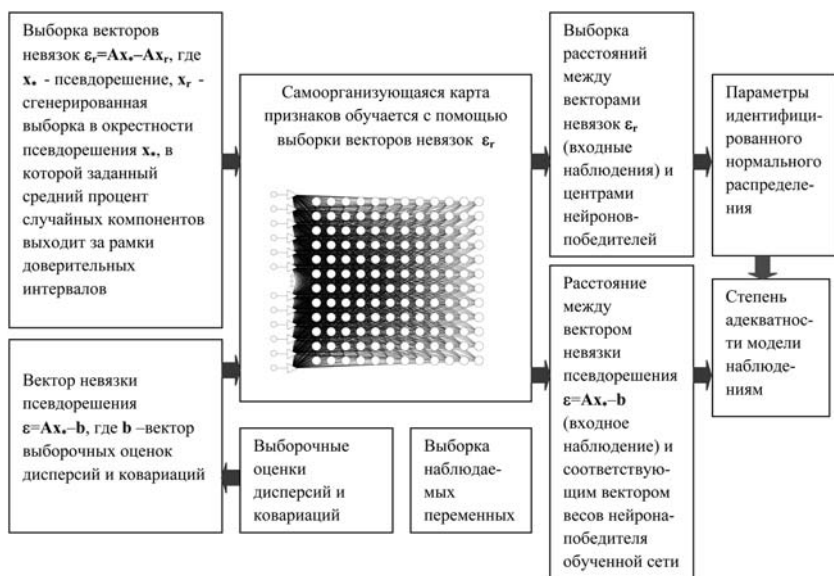


Рис. 5. Процедура оценки степени адекватности линейной факторной модели наблюдениям с помощью самоорганизующихся карт признаков и метода Монте-Карло

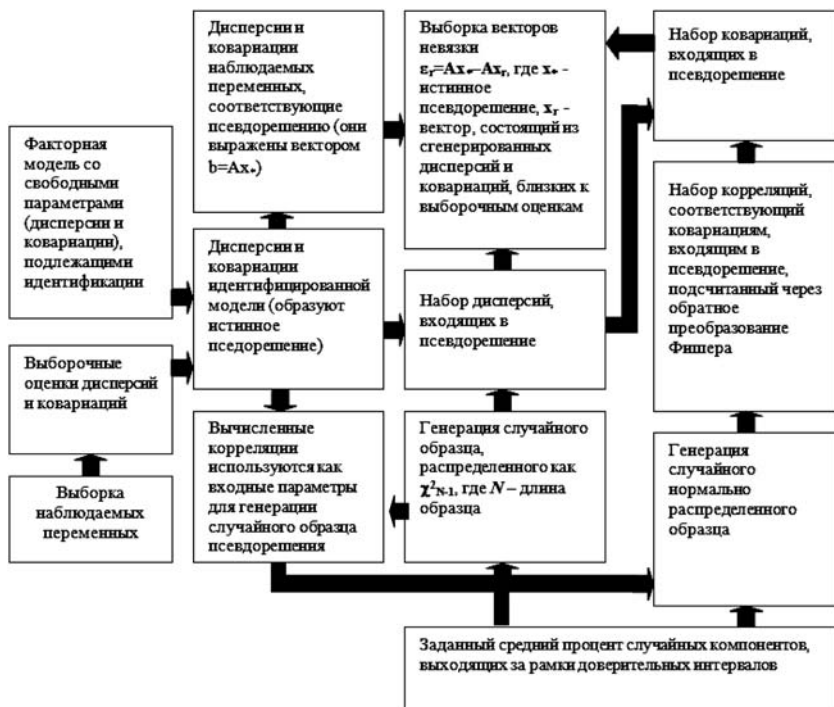


Рис. 6. Генерация выборки векторов невязок  $\varepsilon_r = Ax_* - Ax_r$ , где  $x_*$  — псевдорешение;  $x_r$  — вектор, состоящий из генерированных дисперсий и ковариаций, близких к выборочным оценкам

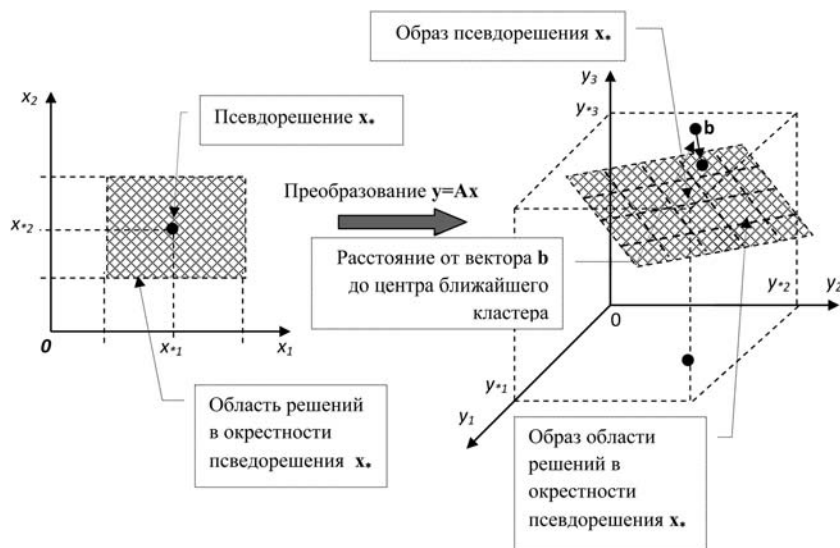


Рис. 7. Геометрическая интерпретация линейного преобразования области допустимых решений

ходя из предположения, что их преобразование Фишера аппроксимируется нормальным распределением, а именно, распределение статистики  $z = \frac{1}{2} \ln \frac{1+r}{1-r}$ , где  $r$  — выборочная корреляция, может быть аппроксимирована нормальным распределением со средним значением  $z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$ , где  $\rho$  — значение корреляции, и дисперсией  $\frac{1}{N-3}$ .

Значения преобразования Фишера программно вычисляются для каждой ковариации, входящей в псевдорешение, а корреляции  $\rho$  заменяются соответствующими результирующими корреляциями. Необходимые корреляции восстанавливаются с помощью обратного преобразования Фишера вышеупомянутых сгенерированных значений. После восстановления они конвертируются в ковариации, входящие в состав псевдорешения.

В приведенной выше процедуре самоорганизующиеся карты Кохонена обеспечивают кластеризацию векторов невязок, которые связывают точку, заданную радиус-вектором  $Ax_*$ , и точки, заданные радиус-векторами  $Ax_r$ , с центрами радиальных базисных элементов, представляющих вычисленные центры кластеров. Размеры соответствующей топологической карты расширяются до тех пор, пока для генерируемой выборки векторов  $x_r$  или не перестанут уменьшаться наибольшие частоты выигрышей на элементах, или не перестанет возрастать число нейронов-победителей. В этом случае, с точки зрения карты Кохонена, расстояния между векторами невязок  $\varepsilon_r$  и ближайшими к ним центрами кластеров уже не представляют геометрическое расположение точек, заданных радиус-векторами  $Ax_r$ . Поэтому значения этих расстояний полагаются находящимися ниже "порога чувствительности" карт Кохонена и, следовательно, выражающими допустимые незначимые вариации ("шум") расстояний от невязок до ближайших центров кластеров, которые не обусловлены геометрическим расположением.

Геометрическая иллюстрация, поясняющая суть рассмотренной процедуры, приведена на рис. 7.

#### 4.2. Оценка степени адекватности в случае произвольных факторных моделей

Многие практические задачи не могут быть решены в рамках моделей дисперсионных составляющих. Далее рассматривается предназначенная для их решения модификация представленного подхода, позволяющая работать с произвольными факторными моделями без взаимных влияний между переменными, образующих контуры обратной связи<sup>5</sup>. Достигнутое расширение возможностей, как пра-

<sup>5</sup> Включая модели путевых коэффициентов.



вило, приводит к нелинейности уравнений решаемой системы и, следовательно, к потере возможности идентификации свободных параметров посредством простой неитерационной процедуры.

В соответствии с предложенным подходом следует:

- выразить аналитически наблюдаемые дисперсии и ковариации через свободные параметры используемой факторной модели, составив переопределенную систему алгебраических уравнений, которая в общем случае является нелинейной;
- вычислить псевдорешение полученной системы уравнений с помощью подходящего численного метода;
- оценить степень адекватности полученного псевдорешения результатам наблюдений с помощью релевантного статистического критерия согласия.

Указанная переопределенная система уравнений может быть представлена в следующем виде:

$$\mathbf{F}(\mathbf{x}) = \mathbf{b},$$

где  $\mathbf{F}(\mathbf{x})$  —  $n$ -мерный нелинейный оператор, действующий на  $m$ -мерный вектор  $\mathbf{x}$  свободных параметров модели, компоненты которого являются аналитическими выражениями прогнозируемых дисперсий и ковариаций наблюдаемых переменных, через  $m$  свободных параметров рассматриваемой факторной модели;  $\mathbf{b}$  — вектор-столбец, составленный из  $n$  выборочных оценок дисперсий и ковариаций наблюдаемых переменных.

Вектор  $\boldsymbol{\varepsilon} = \mathbf{F}(\mathbf{x}_*) - \mathbf{b}$  представляет невязку псевдорешения  $\mathbf{x}_*$ , где

$$\|\mathbf{F}(\mathbf{x}_*) - \mathbf{b}\| = \min_{\mathbf{x} \in \mathbf{X}} \|\mathbf{F}(\mathbf{x}) - \mathbf{b}\|,$$

$\|\cdot\|$  — евклидова норма;  $\mathbf{X}$  — область допустимых значений для  $\mathbf{x}$ .

Для вычисления данного псевдорешения может применяться любой подходящий численный метод нелинейной многомерной локальной оптимизации, в котором критерий минимизации представлен евклидовой нормой невязки. В частности, для этого приемлемы градиентные методы.

На рис. 8 представлена схема рассмотренного выше метода оценки степени адекватности вычисленного псевдорешения наблюдениям, модифицированная для случая произвольных моделей. Как и ранее, расчет основан на сравнении вектора невязки псевдорешения  $\boldsymbol{\varepsilon} = \mathbf{F}(\mathbf{x}_*) - \mathbf{b}$  и случайной выборки векторов невязок  $\boldsymbol{\varepsilon}_r = \mathbf{F}(\mathbf{x}_r) - \mathbf{F}(\mathbf{x}_*)$ , где  $\mathbf{x}_r$  — сгенерированный случайный вектор, находящийся в заданной окрестности псевдорешения  $\mathbf{x}_*$ .

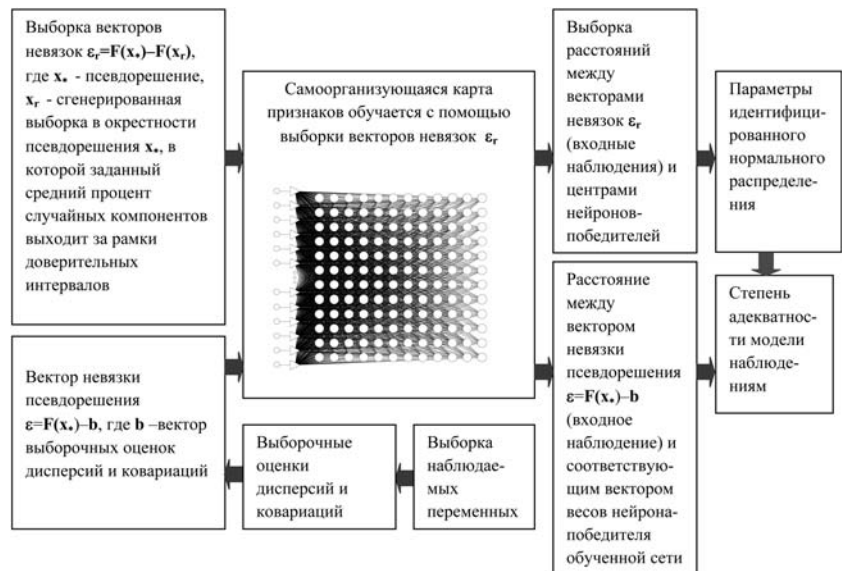


Рис. 8. Оценка степени адекватности нелинейной факторной модели с помощью самоорганизующихся карт признаков и метода Монте-Карло

Векторы  $\mathbf{x}_r$  могут быть распределены произвольным образом, однако на практике удобно генерировать их нормально распределенными, варьируя соответствующее стандартное отклонение и число компонентов, выходящих за границы доверительных интервалов. Случайные выборки векторов невязок  $\boldsymbol{\varepsilon}_r$  применяют для обучения самоорганизующихся карт признаков подходящей размерности, формируя выборки евклидовых расстояний между векторами невязок  $\boldsymbol{\varepsilon}_r$ , использованных в качестве входных данных сети, и центрами нейронов-победителей. Как указано в разд. 4.1, распределения этих расстояний близки к нормальным. Оценки их средних и дисперсий идентифицируют распределения, позволяя вычислять вероятности превышения расстояния между полученным вектором невязки псевдорешения и соответствующим центром нейрона-победителя и на этой основе судить о степени адекватности модели наблюдениям.

Для получения информации о наиболее правдоподобных отклонениях идентифицируемых параметров от их оценок, полученных с помощью заданной факторной модели, для обучения самоорганизующихся карт признаков генерируются серии выборок с различными стандартными отклонениями и средними процентами случайных компонентов, выходящих за границы доверительных интервалов.

Для удобства дальнейшего анализа стандартные отклонения генерируемых выборок поддерживают для всех параметров равными заданному постоянному проценту от соответствующих средних значений. Сопоставление рассмотренных выше вероятностных мер адекватности моделей для различных стандартных отклонений и средних процентов компонентов, выходящих за границы доверительных интервалов, позволяет выявлять наиболее правдо-

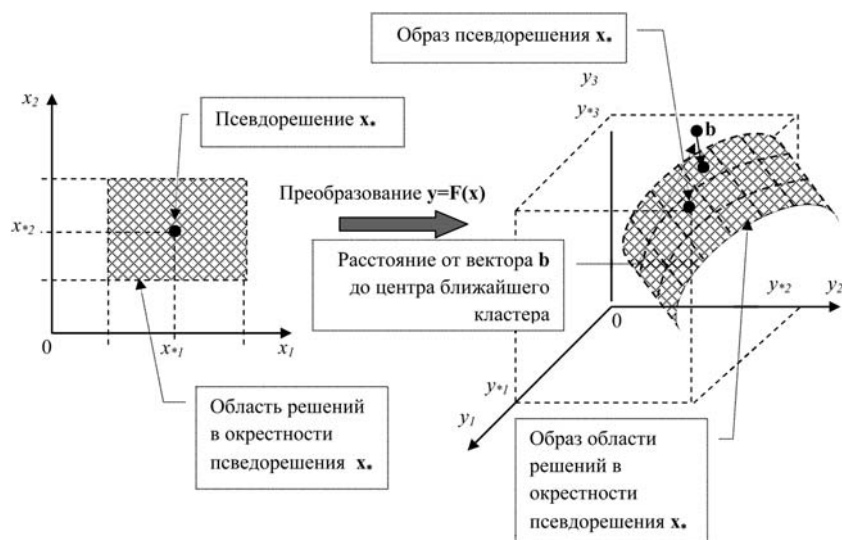


Рис. 9. Геометрическая интерпретация нелинейного преобразования области допустимых решений

подобные сочетания достигнутой точности идентификации, представляемой оценкой стандартного отклонения, и структуры значимых погрешностей в определении компонентов полученного псевдорешения. Геометрическая иллюстрация, поясняющая суть процедуры в случае нелинейного оператора  $F(x)$ , приведена на рис. 9.

Предлагаемый подход позволяет исследовать значимость различий между наиболее правдоподобными вариантами факторных моделей, используя технику проверки статистических гипотез. Свободные параметры этих моделей предварительно идентифицируются с помощью рассмотренного выше метода. Сравнение опирается на сопоставление отношений  $r = \sigma/m$ , где  $\sigma$  — наиболее правдоподобное стандартное отклонение сгенерированных нормально распределенных значений свободного параметра модели;  $m$  — соответствующее среднее значение распределения. Отношение  $r$  при генерации выборок поддерживается равным для всех параметров одной модели, но может различаться для разных моделей, которые, в общем случае, могут иметь и разные средние процентные соотношения компонентов, лежащих вне доверительных интервалов.

Пусть указанные отношения для сравниваемых моделей равны, соответственно,  $r_1 = \sigma_1/m_1$  и  $r_2 = \sigma_2/m_2$ , причем  $r_1 \leq r_2$ . Сравнение моделей выполняется при одном и том же относительном стандартном отклонении  $\sigma_* = r_1 = r_2 m_2$ , когда среднее значение  $m_1$  выбирается в качестве единицы измерения (приравнивается к 1), и опирается на оценку вероятности получения приведенного среднего  $m_2 = r_1/r_2$ , а именно: для случайной величины  $X$  вычисляется вероятность  $P(m_2 \leq X \leq 1) = \Phi(1) - \Phi(m_2)$  пребывания в интервале  $[m_2; m_1 = 1]$ , где  $\Phi$  — функция нормального распределения со средним

значением 1 и стандартным отклонением  $\sigma_*$ . Если эта вероятность превышает заданный критический уровень, который обычно принимается равным 0,05, различие между моделями рассматривается как статистически значимое, в противном случае — как незначимое.

Рассматриваемый подход позволяет определять объемы выборок, необходимых для проверки гипотез о том, что расстояние между векторами невязки псевдорешения  $\epsilon$  и соответствующими центрами нейрон-победителей обученной сети равно определенному значению для заданных уровней значимости и мощности выборки. Расчетная формула выводится из сравнения соответствующих областей принятия гипотезы [9]:

$$N = \left( \frac{z_{1-\alpha/2} + z_{1-\beta}}{d_{\text{norm}}} \right)^2,$$

где  $z_{1-\alpha/2}$  и  $z_{1-\beta}$  — квантили стандартного нормального распределения порядков  $1 - \alpha/2$  и  $1 - \beta$ , соответственно;  $\alpha$  — уровень значимости;  $\beta$  — вероятность ошибки второго рода;  $d_{\text{norm}}$  — отношение отклонения истинного математического ожидания исследуемого расстояния от значения, постулируемого гипотезой, к стандартному отклонению распределения данного расстояния.

Преимущества представленного метода:

- применимость для произвольных факторных моделей без взаимных влияний между переменными, образующих контуры обратной связи;
- отсутствие ограничений, налагаемых на распределения наблюдаемых переменных и компонентов вектора невязки (для которых ранее требовалось проверять соответствие многомерному нормальному распределению);
- наличие простой процедуры оценки вероятностей появления статистических ошибок 2-го рода;
- возможность выявления наиболее правдоподобных сочетаний достигнутой точности идентификации и структуры значимых погрешностей в определении компонентов полученного псевдорешения;
- возможность получения оценок адекватности моделей с заданной степенью точности путем соответствующего увеличения объема генерируемых случайных выборок компонентов псевдорешения.

Представленные методы реализованы в среде графического программирования *LabVIEW*. Работа самоорганизующихся карт признаков моделировалась с помощью пакета *STATISTICA Neural Networks*.

## 5. Пример практического применения

Особенности рассмотренного подхода далее иллюстрируются результатами исследования гнездовой и иерархической факторных моделей интеллекта [8], представленных на рис. 10 и 11.

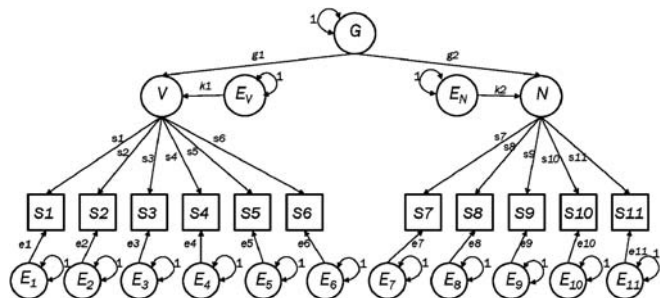


Рис. 10. Иерархическая факторная модель интеллекта ( $G, V, N, E_V, E_N, E_1, E_2, \dots, E_{11}$  — латентные факторы;  $S_1, S_2, \dots, S_{11}$  — наблюдаемые переменные)

Прогнозируемые ковариационные матрицы для наблюдаемых переменных этих моделей приведены в табл. 1 и 2. Каждый представленный аналитическим выражением элемент матрицы после приравнивания его выборочной оценке ковариации или дисперсии соответствующих наблюдаемых пере-

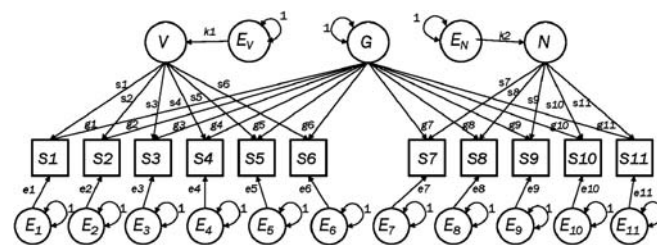


Рис. 11. Гнездовая факторная модель интеллекта ( $G, V, N, E_V, E_N, E_1, E_2, \dots, E_{11}$  — латентные факторы;  $S_1, S_2, \dots, S_{11}$  — наблюдаемые переменные)

Таблица 1

Прогнозируемая ковариационная матрица для наблюдаемых переменных иерархической факторной модели

	$S1$	$S2$	$S3$	$S4$	$S5$	$S6$	$S7$	$S8$	$S9$	$S10$	$S11$
$S1$	$s_1^2 * (g_1^2 + k_1^2) + e_1^2$										
$S2$	$s_1 * s_2 * (g_1^2 + k_1^2)$	$s_2^2 * (g_1^2 + k_1^2) + e_2^2$									
$S3$	$s_1 * s_3 * (g_1^2 + k_1^2)$	$s_2 * s_3 * (g_1^2 + k_1^2)$	$s_3^2 * (g_1^2 + k_1^2) + e_3^2$								
$S4$	$s_1 * s_4 * (g_1^2 + k_1^2)$	$s_2 * s_4 * (g_1^2 + k_1^2)$	$s_3 * s_4 * (g_1^2 + k_1^2)$	$s_4^2 * (g_1^2 + k_1^2) + e_4^2$							
$S5$	$s_1 * s_5 * (g_1^2 + k_1^2)$	$s_2 * s_5 * (g_1^2 + k_1^2)$	$s_3 * s_5 * (g_1^2 + k_1^2)$	$s_4 * s_5 * (g_1^2 + k_1^2)$	$s_5^2 * (g_1^2 + k_1^2) + e_5^2$						
$S6$	$s_1 * s_6 * (g_1^2 + k_1^2)$	$s_2 * s_6 * (g_1^2 + k_1^2)$	$s_3 * s_6 * (g_1^2 + k_1^2)$	$s_4 * s_6 * (g_1^2 + k_1^2)$	$s_5 * s_6 * (g_1^2 + k_1^2)$	$s_6^2 * (g_1^2 + k_1^2) + e_6^2$					
$S7$	$s_1 * s_7 * g_1 * g_2$	$s_2 * s_7 * g_1 * g_2$	$s_3 * s_7 * g_1 * g_2$	$s_4 * s_7 * g_1 * g_2$	$s_5 * s_7 * g_1 * g_2$	$s_6 * s_7 * g_1 * g_2$	$s_7^2 * (g_2^2 + k_2^2) + e_7^2$				
$S8$	$s_1 * s_8 * g_1 * g_2$	$s_2 * s_8 * g_1 * g_2$	$s_3 * s_8 * g_1 * g_2$	$s_4 * s_8 * g_1 * g_2$	$s_5 * s_8 * g_1 * g_2$	$s_6 * s_8 * g_1 * g_2$	$s_7 * s_8 * (g_2^2 + k_2^2)$	$s_8^2 * (g_2^2 + k_2^2) + e_8^2$			
$S9$	$s_1 * s_9 * g_1 * g_2$	$s_2 * s_9 * g_1 * g_2$	$s_3 * s_9 * g_1 * g_2$	$s_4 * s_9 * g_1 * g_2$	$s_5 * s_9 * g_1 * g_2$	$s_6 * s_9 * g_1 * g_2$	$s_7 * s_9 * (g_2^2 + k_2^2)$	$s_8 * s_9 * (g_2^2 + k_2^2)$	$s_9^2 * (g_2^2 + k_2^2) + e_9^2$		
$S10$	$s_1 * s_{10} * g_1 * g_2$	$s_2 * s_{10} * g_1 * g_2$	$s_3 * s_{10} * g_1 * g_2$	$s_4 * s_{10} * g_1 * g_2$	$s_5 * s_{10} * g_1 * g_2$	$s_6 * s_{10} * g_1 * g_2$	$s_7 * s_{10} * (g_2^2 + k_2^2)$	$s_8 * s_{10} * (g_2^2 + k_2^2)$	$s_9 * s_{10} * (g_2^2 + k_2^2)$	$s_{10}^2 * (g_2^2 + k_2^2) + e_{10}^2$	
$S11$	$s_1 * s_{11} * g_1 * g_2$	$s_2 * s_{11} * g_1 * g_2$	$s_3 * s_{11} * g_1 * g_2$	$s_4 * s_{11} * g_1 * g_2$	$s_5 * s_{11} * g_1 * g_2$	$s_6 * s_{11} * g_1 * g_2$	$s_7 * s_{11} * (g_2^2 + k_2^2)$	$s_8 * s_{11} * (g_2^2 + k_2^2)$	$s_9 * s_{11} * (g_2^2 + k_2^2)$	$s_{10} * s_{11} * (g_2^2 + k_2^2)$	$s_{11}^2 * (g_2^2 + k_2^2) + e_{11}^2$

Прогнозируемая ковариационная матрица для наблюдаемых переменных гнездовой факторной модели

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
S1	$s_1^2 * k_1^2 + g_1^2 + e_1^2$										
S2	$s_1 * s_2 * k_1^2 + g_1 * g_2$	$s_2^2 * k_1^2 + g_2^2 + e_2^2$									
S3	$s_1 * s_3 * k_1^2 + g_1 * g_3$	$s_3 * s_2 * k_1^2 + g_3 * g_2$	$s_3^2 * k_1^2 + g_3^2 + e_3^2$								
S4	$s_1 * s_4 * k_1^2 + g_1 * g_4$	$s_4 * s_2 * k_1^2 + g_4 * g_2$	$s_4 * s_3 * k_1^2 + g_4 * g_3$	$s_4^2 * k_1^2 + g_4^2 + e_4^2$							
S5	$s_1 * s_5 * k_1^2 + g_1 * g_5$	$s_5 * s_2 * k_1^2 + g_5 * g_2$	$s_5 * s_3 * k_1^2 + g_5 * g_3$	$s_5 * s_4 * k_1^2 + g_5 * g_4$	$s_5^2 * k_1^2 + g_5^2 + e_5^2$						
S6	$s_1 * s_6 * k_1^2 + g_1 * g_6$	$s_6 * s_2 * k_1^2 + g_6 * g_2$	$s_6 * s_3 * k_1^2 + g_6 * g_3$	$s_6 * s_4 * k_1^2 + g_6 * g_4$	$s_6 * s_5 * k_1^2 + g_6 * g_5$	$s_6^2 * k_1^2 + g_6^2 + e_6^2$					
S7	$g_1 * g_7$	$g_2 * g_7$	$g_3 * g_7$	$g_4 * g_7$	$g_5 * g_7$	$g_6 * g_7$	$s_7^2 * k_2^2 + g_7^2 + e_7^2$				
S8	$g_1 * g_8$	$g_2 * g_8$	$g_3 * g_8$	$g_4 * g_8$	$g_5 * g_8$	$g_6 * g_8$	$s_7 * s_8 * k_2^2 + g_7 * g_8$	$s_8^2 * k_2^2 + g_8^2 + e_8^2$			
S9	$g_1 * g_9$	$g_2 * g_9$	$g_3 * g_9$	$g_4 * g_9$	$g_5 * g_9$	$g_6 * g_9$	$s_7 * s_9 * k_2^2 + g_7 * g_9$	$s_8 * s_9 * k_2^2 + g_8 * g_9$	$s_9^2 * k_2^2 + g_9^2 + e_9^2$		
S10	$g_1 * g_{10}$	$g_2 * g_{10}$	$g_3 * g_{10}$	$g_4 * g_{10}$	$g_5 * g_{10}$	$g_6 * g_{10}$	$s_7 * s_{10} * k_2^2 + g_7 * g_{10}$	$s_8 * s_{10} * k_2^2 + g_8 * g_{10}$	$s_9 * s_{10} * k_2^2 + g_9 * g_{10}$	$s_{10}^2 * k_2^2 + g_{10}^2 + e_{10}^2$	
S11	$g_1 * g_{11}$	$g_2 * g_{11}$	$g_3 * g_{11}$	$g_4 * g_{11}$	$g_5 * g_{11}$	$g_6 * g_{11}$	$s_7 * s_{11} * k_2^2 + g_7 * g_{11}$	$s_8 * s_{11} * k_2^2 + g_8 * g_{11}$	$s_9 * s_{11} * k_2^2 + g_9 * g_{11}$	$s_{10} * s_{11} * k_2^2 + g_{10} * g_{11}$	$s_{11}^2 * k_2^2 + g_{11}^2 + e_{11}^2$

менных, принадлежащих множеству  $\{S_1, S_2, \dots, S_{11}\}$ , дает нелинейное уравнение для  $n$ -мерного нелинейного оператора  $F(x)$ , примененного к  $m$ -мерному вектору идентифицируемых свободных параметров  $g_1, g_2, \dots, g_{11}, k_1, k_2, s_1, s_2, \dots, s_{11}, e_1, e_2, \dots, e_{11}$ , где  $n = 66, m = 26$  для первой модели и  $m = 35$  — для второй (см. разд. 4.2).

На рис. 12 и 13 (см. третью сторону обложки) показаны вычисленные с помощью представленного выше метода распределения вероятностей превышения расстояния между вектором невязки псевдорешения и соответствующим центром нейрона-победителя самоорганизующейся карты признаков для различных сочетаний стандартных отклонений от компонентов вектора псевдорешения и средних процентов компонентов, выходящих за рамки доверительных интервалов.

Показатели  $r_1$  и  $r_2$  равны 0,05 и 0,1 соответственно. Поскольку  $m_2 = 0,5$  и  $\sigma_* = 0,05$ , вероятность

$P(m_2 \leq X \leq 1) = \Phi(1) - \Phi(0,5)$  достигнутого отклонения среднего значения равна 0,5. Следовательно, можно сделать вывод, что иерархическая модель согласуется с наблюдениями значимо лучше, чем гнездовая модель при уровне значимости 0,05.

### Заключение

1. Разработан метод исследования динамики психологических характеристик, использующий возможности вейвлет-преобразований и обучаемых факторных структур. Его особенностью является составление переопределенной системы линейных или нелинейных алгебраических уравнений относительно свободных параметров факторной модели с последующим поиском псевдорешения.

2. Для оценки степени адекватности факторных моделей наблюдениям предложен новый подход, опирающийся на метод Монте-Карло и возможно-

сти самоорганизующихся карт Кохонена. Он позволяет избежать жестких ограничений, налагаемых на вероятностные распределения результатов наблюдений, которые присущи традиционной процедуре идентификации свободных параметров факторных моделей.

3. Преимуществами этого подхода являются:

- применимость для произвольных факторных моделей без взаимных влияний между переменными, образующих контуры обратной связи;
- отсутствие ограничений, налагаемых на распределения наблюдаемых переменных и компонентов вектора невязки (для которых ранее требовалось проверять соответствие многомерному нормальному распределению);
- наличие простой процедуры оценки вероятностей появления статистических ошибок 2-го рода;
- возможность выявления наиболее правдоподобных сочетаний достигнутой точности идентификации и структуры значимых погрешностей в определении компонентов полученного псевдорешения;
- возможность получения оценок адекватности моделей с заданной степенью точности путем соответствующего увеличения объема генерируемых случайных выборок компонентов псевдорешения.

#### Список литературы

1. **Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф.** Справочник по теории вероятностей и математической статистике. М.: Наука, 1985.
2. **Куравский Л. С., Баранов С. Н., Малых С. Б.** Нейронные сети в задачах прогнозирования, диагностики и анализа данных. М.: РУСАВИА, 2003.
3. **Куравский Л. С., Корниенко П. А.** Применение нейронных сетей для идентификации локусов количественных признаков в психогенетике // *Нейрокомпьютеры: разработка и применение*. 2007. № 4. С. 15—31.
4. **Куравский Л. С., Мармалюк П. А., Абрамочкина В. И., Петрова Е. А.** Применение факторного анализа результатов

вейвлет-преобразований для исследования динамики психологических характеристик // *Экспериментальная психология*. 2009. № 1. С. 97—111.

5. **Куравский Л. С., Юревич А. В., Мармалюк П. А., Иванова Е. Г.** Факторный анализ показателей нравственного состояния общества в европейских странах // *Психологическая наука и образование*. 2010. № 1. С. 94—110.

6. **Лоули Д., Максвелл А.** Факторный анализ как статистический метод. М.: Мир. 1967.

7. **Мармалюк П. А.** Оценка степени адекватности факторных моделей с помощью самоорганизующихся карт признаков Кохонена // *Нейрокомпьютеры: разработка и применение*. 2010. № 10. С. 53—62.

8. **Ушаков Д. В.** Интеллект: структурно-динамическая теория. М.: Изд. ИП РАН, 2003.

9. **Bendat J. S., Piersol A. G.** Random data. Analysis and measurement procedures. New York: John Wiley & Sons, 1986.

10. **Bishop Y. M. M., Fienberg S. E., Holland P. W.** Discrete multivariate analysis: Theory and practice. Cambridge: MA, MIT Press, 1975.

11. **Bollen K. A.** Structural equations with latent variables. New York: John Wiley, 1989.

12. **Goldstein H.** Multilevel statistical models. 3rd ed. — London: Arnold, 2003.

13. **Jöreskog K. G.** Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*. 1970. Vol. 23. P. 121—145.

14. **Kohonen T.** Self-organized formation of topologically correct feature maps. *Biological Cybernetics*. 1982. Vol. 43. P. 59—69.

15. **Kuravsky L. S., Baranov S. N., Baranov N. I.** Wavelet-based confirmatory factor analysis for monitoring of system factors: estimating goodness-of-fit measures with the aid of self-organizing feature maps // In: Proc. 6th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies. — Dublin, Ireland. June 2009. P. 224—245.

16. **Kuravsky L. S., Baranov S. N., Marmalyuk P. A.** Estimation of goodness-of-fit measures accompanying the identification of factor models // In: Proc. 7th International Conference on Condition Monitoring & Machinery Failure Prevention Technologies. — Stratford-upon-Avon, England, June 2010.

17. **Loehlin J. C.** Latent variable models: An introduction to factor, path, and structural analysis. Hillsdale, NJ: Erlbaum, 1987.

18. **Neale M. C., Cardon L. R.** Methodology for genetic studies of twins and families. Dordrecht (Netherlands): Kluwer Academic Publishers, 1992.

# CONTENTS

**Zhuravleva L. M.** *Prognoses from Moore's and Development of Information-Processing and Information-Transmitting Devices* . . . . . 2

This paper deals with generations of electronic engineering and optoelectronics hardware components. Prognoses from Moore's law are matched with the results achieved in speed increase of functional operations of microprocessors and optical communication and the reasons for its disarrangement are analysed. The author draws conclusions about updating prospects of information-processing and information-transmitting devices.

**Keywords:** prognoses, speed, microprocessors, optical communication

**Saak A. E.** *Dispatching Algorithms in Grid Systems Based on Array of Demands Quadratic Typification* . . . . . 9

Linear and triode polyhedron of coordinate resource rectangles i. e. user's demands are suggested as the model of multiple service for Grid technology tasks. It is introduced the notions of fast and slow alteration monotonicity of polyhedron model parameters. It is defined quadratic typification of polyhedrons of planar elements with the help of this difference and other array features. It is introduced the objective measures and multipurpose criterion of ordering heuristics of mentioned arrays as the dispatching technology of Grid technology computing facilities. It is suggested the algorithms of frame distribution of arrays based on ring, angular and inverse localization depending on quadratic typification.

**Keywords:** Grid system, multiprocessor computer system, dispatching, quadratic type of user's demands array, criteria heuristics of symmetrization of resource shell and resource measure and fullness of shell by user's demands

**Ledeneva T. M., Nguen Ngok Chui.** *The Influence of Likelihood Function to Results of Fuzzy Classification.* . . . . 14

The results of the research concerning fuzzy classification algorithm, in which likelihood function has been used in order to form similarity relations, are considered in this paper.

**Keywords:** fuzzy numbers, likelihood function, fuzzy equivalence relation, transitivity, decomposition tree, fuzzy classification

**Savchenko A. V.** *Image Recognition Three-Threshold System Based on Directed Enumeration Method* . . . . . 21

The problem of image recognition from large database is put and solved using directed alternatives' enumeration method and minimum discrimination informative criterion. The possibility to increase recognition accuracy at the expense of three decision rules application (to choice decision most different from others based on directed enumeration method, reject unreliable decision and prevent exhaustive search) is investigated. The experimental program and results in a problem of human face recognition have been produced. It is shown that the proposed approach is characterized by increased accuracy and computing efficiency.

**Keywords:** automatic image recognition, directed enumeration method, minimum discrimination information principle

**Yakimov V. N., Moshkov I. S.** *Methods of Compound Terms Analysis in Natural Language Text Processing.* . . . . 26

Article under the title "Methods of compound terms analysis in natural language text processing" examines features of natural language texts, describing taxonomic structure. The emphasis is made on classification of elements that are parts of compound terms found in the text. Methods are based on criteria that help to classify a particular element of compound term.

**Keywords:** natural language, text processing, taxonomic structure

**Boyarinov Yu. G., Borisov V. V.** *System and Process Analysis on the Base of Fuzzy Semi-Markov Model.* . . . . 31

The paper deals with fuzzy semi-Markov model in which the fuzzy display on the base of product or relational models are used for obtaining a fuzzy state system probabilities and being in appropriate state time. The method of system analysis and process has been developed on the base of fuzzy semi-Markov models which allows to solve fuzzy optimization tasks. This method can decrease laboriousness solving optimization tasks on account of approximation properties of fuzzy models. The example of using the method for determination of optimal fuzzy parameter value frequency when testing the system is given.

**Keywords:** fuzzy semi-Markov model, fuzzy display, fuzzy optimization tasks

**Kuzmina I. A.** *Synthesis Electronetwork Task Solution* . . . . . 36

Algorithms for the electronetworks structure synthesis and traffic distribution are given. The combination of heuristic and genetic methods are using.

**Keywords:** electronetworks, design optimization, perspective development, genetic methods, combination of heuristic method

**Efanov D. V., Ibragimov A. M.** *Dynamic Conventional Signs and Their Realization in Modern Geographical Information Systems* . . . . . 40

This article reports formal description mode of wide specter of dynamic conventional signs (It provides formal definition of these signs), and their logical architecture in modern geographic information systems.

**Keywords:** dynamic conventional signs, geographic information systems, GIS, meteorological situation, architecture of dynamic conventional signs, formalization of dynamic conventional signs

**Arhipova O. E., Boyko V. V., Kovaleva G. V., Moskalenko V. A., Tarasova T. T.** *Methodological Approaches to the Assessment of Natural and Man-Made Changes the Coast of Azov and Black Seas* . . . . . 44

This work describes the methodological approach to determine the level of permissible recreational loads on the coastal ecosystems of South Seas. Select subject area studies due to the necessity of optimization of modern recreational sector and poorly studied its social component. Methodological framework of the project is a multidisciplinary integrated approach, which accumulates the most heuristic and tested theoretical models explaining the current trends of development of processes recreation. Recreational sphere is considered as a system composed of interconnected subsystems. The approach relies on the use of systems analysis, mathematical modeling, GIS-technologies. The purpose of work is to develop recommendations for the determination of recreational load, its impact on coastal ecosystems.

**Keywords:** system approach, sustainable development, coastal zone, information technology, mathematical modeling, indicator

**Malychina G. F.** *Statistical Model for Parameter Collection Being Controlled Using Intellectual Measuring System* . . . . . 49

Probability distribution model for the group of correlated parameters for totality measurements in the case of many-parametric control is investigated. The model base is multidimensional generalized exponential distribution. The basis, analytical form and method of identification on the basis of experimental data are investigated.

**Keywords:** measurement procedure, object identification, model, intellectuality, decision-making, metrological characteristics, likelihood ratio, bi-phase flow device "Factor"

**Chetyrbotsky A. N.** *Parametrical Identification of Radial Basic Functions Neural Networks Methods of Global Optimization* . . . . . 54

It is considered to a problem of approximation of function which on admissible set of arguments is set by a discrete set of values. For its decision the device of radial basic functions of neural networks (RBFNN) is used. The algorithm of procedure of training of a network here is realized on the basis of methods of global optimization. Depending on a mode of its work set of required unknown persons can contain parameters of radial basic functions, procedures of training, coordinates of the centers and, in some cases, coordinates of points of training of a neural network. Questions of construction of initial approximation and number hidden neurons are discussed. Features of the developed approaches and techniques of their numerical realization are shown by experimental researches.

**Keywords:** radial basic functions, a problem of search of a minimum, methods of global optimization

**Potapov V. I.** *Modes and Problems of Reliable Estimation of Neuron System by Teaching and Reteaching of Neuron Net after Reject* . . . . . 59

Variable models for solving problems of reliable estimation of neuron system by teaching and reteaching of neuron net after rejection with man-operator are considered in this article. Some formulas for calculating functional prepared neuron system are given.

**Keywords:** mathematical model, neuron net, neuron system, reliability, teaching of neuron net

**Krasovskaya L. V.** *Imitation of Working off of Actions on Indistinct Semantic Networks in Algorithms of Self-Training of Intellectual Systems* . . . . . 64

It is designed strategy of comparison ill-defined presented relations in models PA, differing from the known that that it allows to execute unremovable estimation of equality of ill-defined denominated factors and hereunder raise validity of comparison. Described algorithms of self-training, allowing form program of expedient behaviour in different problem-solving ambiances, differing from the known, imitation an work trying actions on ISN that enables to exclude the influence SS on PA in process of study of regularities of ambience.

**Keywords:** intellectual systems, problem-solving ambience, ill-defined semantic network, ensemble of tops and ribs, features, therms, algorithms of self-training

**Kravsky L. S., Marmalyuk P. A., Panfilova A. S., Ushakov D. V.** *Studying Factor Influences on Psychological Characteristics Development with the Aid of a New Approach to Estimation of Goodness-of-Fit Measure* . . . . . 67

A new technology for revealing and studying factors responsible for evolution of psychological characteristics is presented. It combines capabilities of wavelet transforms and factor analysis. According to the proposed approach, the samples of coefficients resulted from discrete wavelet transform of initial parameter time series under study and responsible for different observation periods are considered as values of observed variables in the subsequent confirmatory factor analysis to reveal time history of factor influences and estimates of factor interaction. Identification of free factor model parameters (usually factor variances and covariances) is carried out by a new direct (noniterative) procedure based on the maximum likelihood method, which is an alternative to traditional local iterative solution of optimization problems. A statistical method to check significance of factor model components and a new technique of goodness-of-fit measure estimation in case of unrestricted factor models, which is based on the Monte-Carlo approach and capabilities of self-organizing feature maps (Kohonen networks) and makes it possible to avoid tight restrictions on observed data inherent in the traditional factor model identification procedure, is discussed.

**Keywords:** factor analysis, path diagram, model goodness-of-fit criterion

---



---

**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 *Н. В. Комиссарова*.

Сдано в набор 08.09.2011. Подписано в печать 17.10.2011. Формат 60×88 1/8. Бумага офсетная. Печать офсетная.  
Усл. печ. л. 9,8. Уч.-изд. л. 11,06. Заказ 774. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати,  
телерадиовещания и средств массовых коммуникаций.  
Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Отпечатано в ООО "Подольская Периодика"  
142110, Московская обл., г. Подольск, ул. Кирова, 15