

НАУЧНО-ТЕХНИЧЕСКИЙ
ЖУРНАЛ

Основан в 1968 г.

Свидетельство
о государственной
регистрации
КВ № 9148 от 10.09.2004

ISSN 0555-2656

Журнал включен
в список
специальных изданий
ВАК Украины
по техническим наукам

БИОНИКА ИНТЕЛЛЕКТА

ИНФОРМАЦИЯ, ЯЗЫК, ИНТЕЛЛЕКТ

№ 1 (64)

2006

Гл. редактор М. Ф. Бондаренко, д-р техн. наук, профессор,
заслуженный деятель науки и техники Украины,
академик Академии высшей школы

Редакционная коллегия: Ю. П. Шабанов-Кушнарченко (отв. редактор),
д-р техн. наук, профессор
Г. Г. Четвериков (зам. отв. редактора),
д-р техн. наук, профессор
Р. П. Базилевич, д-р физ.-мат. наук, профессор
Е. В. Бодянский, д-р техн. наук, профессор
А. И. Бых, д-р физ.-мат. наук, профессор
Т. К. Винцюк, д-р техн. наук, профессор
З. В. Дударь, канд. техн. наук, профессор
Н. С. Лесная, канд. техн. наук, профессор
В. П. Машталир, д-р техн. наук, профессор
А. Г. Нерух, д-р физ.-мат. наук, профессор
А. В. Палагин, акад. НАНУ, д-р техн. наук, профессор
Е. П. Путятин, д-р техн. наук, профессор
О. Г. Руденко, д-р техн. наук, профессор
А. Г. Руткае, д-р физ.-мат. наук, профессор
И. Б. Сироджа, д-р техн. наук, профессор
С. В. Смеляков, д-р физ.-мат. наук, профессор
А. Д. Тевяшев, д-р техн. наук, профессор
С. Ю. Шабанов-Кушнарченко, д-р техн. наук, профессор
В. А. Широков, чл.-корр. НАНУ,
д-р техн. наук, профессор



© Харьковский национальный университет радиоэлектроники, 2006
© ООО «Компания СМІТ», 2006

БИОНИКА ИНТЕЛЛЕКТА

ИНФОРМАЦИЯ, ЯЗЫК, ИНТЕЛЛЕКТ

№ 1 (64)

2006

СОДЕРЖАНИЕ

| | |
|--|-----|
| <i>Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П.</i> О бионике интеллекта | 3 |
| <i>Тевяшев А. Д., Кобьлинский К. В.</i> Интеллектуальная система обнаружения криминальных отборов из продуктопроводов | 6 |
| <i>Машталир В. П., Шляхов В. В.</i> Индуцированная согласованность отношений в задачах грануляции информации | 19 |
| <i>Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П.</i> Модели цветового зрения | 27 |
| <i>Бодянский Е. В., Винокурова Е. А., Слипченко А. Н., Шило А. В.</i> Идентификация фаз сна в реальном времени на основе гибридных нейронных сетей | 42 |
| <i>Ерохин А. Л., Турута А. П.</i> Идентификация нештатных ситуаций в информационных сетях | 46 |
| <i>Сироджа И. Б., Фролова Г. А.</i> Автоматизация принятия решений при разработке межцеховых технологических маршрутов на приборостроительных предприятиях | 56 |
| <i>Алисейко З. А., Булкин В. И., Канищева О. В., Шаронова Н. В.</i> Использование алгебры предикатов и предикатных операций для формализации декларативной и процедурной составляющих знаний | 59 |
| <i>Четвериков Г. Г., Дмитриева М. А.</i> Обоснование выбора и принципы построения многозначных обратимых логических элементов | 64 |
| <i>Левыкин В. М., Неофитная Т. М.</i> Разработка правил пополнения иерархической базы знаний | 68 |
| <i>Любченко Т. П.</i> Морфологічна модель словозміни флективної мови та електронний граматичний словник | 72 |
| <i>Бондаренко М. Ф., Работягов А. В., Щепковский С. В.</i> Проблемы распознавания речевых образов спектральными методами в области прикладной лингвистики | 78 |
| <i>Волченко Е. В.</i> Модифицированный метод потенциальных функций | 86 |
| <i>Тесленко Н. А., Чурюмова И. Г.</i> Нечеткая кластеризация массивов биомедицинских данных в условиях избыточности информации | 92 |
| <i>Гороховатский В. А., Кузьмин С. В.</i> Исследование детекторов точек интереса при построении локальных признаков изображений | 96 |
| <i>Белоус Н. В., Высоцкая Е. В., Жемчужкина Т. В., Ковалев А. Н., Кобзарь Г. А., Порван А. П.</i> Об одном подходе к распознаванию электрокардиографического сигнала в реальном масштабе времени | 101 |
| <i>Кулишова Н. Е.</i> Интерполяция коэффициентов отражения красок с помощью радиально-базисной искусственной нейронной сети | 105 |
| Рефераты | 109 |
| Сведения об авторах | 117 |
| Правила оформления рукописей для авторов научно-технического журнала «Бионика интеллекта» | 118 |

УДК 519.7

М. Ф. Бондаренко, Ю. П. Шабанов-Кушнаренко

О БИОНИКЕ ИНТЕЛЛЕКТА

На наших глазах происходит информатизация всего, что только можно себе представить, и специалисты в области бионики интеллекта находятся в самой гуще событий. Говорят, что совершается вторая научно-техническая революция, а ее результатом является усиление интеллектуальных возможностей людей. Еще раньше началась и продолжается первая научно-техническая революция, в результате которой достигается усиление физических возможностей человека. На одно из первых мест выдвигается задача создания искусственного интеллекта. Наблюдается экспоненциальный рост возможностей средств вычислительной техники. Создается впечатление, что всех нас увлекает нечто такое, что очень похоже на снежную лавину. Кажется, что наступил критический момент в развитии человеческого общества. С одной стороны, перед человечеством открываются заманчивые перспективы дальнейшего развития, но, с другой стороны, возникают серьезные осложнения и угрозы. Быстро нарастают опасения по поводу близости «конца света».

Что происходит, куда мы идем, что с нами будет и что делать, чтобы избежать катастрофы? Ключ к пониманию происходящего, на наш взгляд, может дать бионика интеллекта. Термин «бионика» образован из первой части слова «биология» и второй части слова «техника». Он выражает идею перехода от живых систем к искусственным. Причина происходящего, очевидно, заключается в развитии техники. А если копнуть поглубже, то можно прийти к выводу, что стремление и способность к созданию технических изделий глубоко коренится в самой природе человека. Более того, неистребимое влечение к технике — это характеристическое свойство представителей вида *Homo sapiens*, определяющее все их действия. Похоже на то, что человечество стремится взять свою судьбу в собственные руки, хочет само управлять ходом собственной эволюции. Сказанное попытаемся подкрепить несколькими цитатами.

Первое высказывание — Роджера Шенка, известного специалиста в области искусственного интеллекта. Он пишет: «Искусственный интеллект как область науки — это лишь малая часть грандиозной попытки постичь мышление. Мы считаем, что это основная цель данной области науки и здесь достигнуты немалые успехи. Программы, которые мы пишем, важны как эксперимент, а не как конечный результат. Главный интерес для нас представляет именно интеллект, а не его искусственное происхождение. Если мы достигнем успеха в этом направлении, то проложим путь для создания механических помощников человеку в его повседневных делах и заботах. Но не в этом главное. Самое важное, чего мы тогда добьемся, — более глубокого понимания самих

себя, что, безусловно, гораздо ценнее, чем любая программа» [1, с. 26].

Второе высказывание — известного московского математика Г. Н. Поварова из предисловия к русскому изданию книги Винера «Кибернетика». Он пишет: «Действительно, научно-технический прогресс ставит перед человечеством серьезные проблемы. Стремительное развитие науки и техники возлагает на нас колоссальную ответственность за разумное использование полученного нами могущества. «Кто живет в стеклянном доме, тот не должен бросать камней», — гласит старинная поговорка. Человек стал настолько могущественным, что любое его нерассчитанное движение — с роботами, с атомной энергией, с химией — может иметь тяжелые непредвиденные последствия. Это парадокс могущества. Нельзя забывать, однако, что наука и техника не только возлагают новую ответственность на человека, но и доставляют ему новые средства справиться с нею. Это относится и к роботам. Альтернатива «человек или робот», «опасное развитие искусственного разума или своевременный отказ от него», чем ограничивается большинство авторов, имеет третье, более необычное и, пожалуй, более вероятное решение, если только искусственный разум и искусственная жизнь вообще возможны. Человек, научившийся создавать искусственный разум и искусственную жизнь, не остановится перед коренной переделкой самого себя. Не роботы вместо людей, а новый человек вместо старого! Человек будущего вряд ли останется таким же «натуральным» существом, таким же теплокровным позвоночным, каким он вышел из горнила естественного отбора. Почти наверное, он будет искусственно развивать свой мозг и свое тело, будет по воле лепить и изменять свою физическую оболочку. Ему по силам быть впереди любого возможного робота. Это будет биологическая революция, и если смелые гипотезы оправдаются, она будет означать преобразование всего человеческого существования. Быть может, далекий смысл «безумной» винеровской идеи о передаче человека по телеграфу и есть достижение человеком перевоплощаемости? Позволим себе минуту фантазии: не станет ли тогда человек новым космическим существом, свободным от земных ограничений? Есть ли абсолютная граница могущества и сложности для человека и его творений, абсолютная граница могущества и сложности для саморазвивающихся систем вообще?.. Впрочем, это вопросы для науки будущего, на которые она сумеет ответить лучше нас» [2, с. 26–27].

Какая сила заставляет человека двигаться по пути научно-технического прогресса? Очевидно, та, которая его породила, а именно — процесс эволюции, направляемый генетическим интеллектом. Эволюция

организмов буксует, она не может оторваться от белковых структур, требующих земных (то есть тепличных) условий для своего существования. Половину срока, отведенного судьбой для развития организмов на Земле, эволюция уже израсходовала. Если и дальше так будет продолжаться, то жизнь на Земле рано или поздно погибнет. Нужны революционные меры. И вот генетический разум создает специальный инструмент в виде человека, который призван перевести жизнь на техническую основу. Циолковский пишет: «Земля — колыбель человечества. Но нельзя вечно жить в колыбели». И еще: «Итак, нет конца жизни, конца разуму и совершенствованию человечества. Прогресс его вечен. А если это так, то невозможно сомневаться и в достижении бессмертия. Смело же идите вперед, великие и малые труженики земного рода, и знайте, что ни одна из черт ваших трудов не исчезнет бесследно, но принесет вам в бесконечности великий плод» [3, с. 55]. Свернуть в сторону с этого пути или остановиться на полдороге человечеству уже не удастся. Оно вынуждено будет все быстрее бежать вперед. Человек слишком далеко зашел вперед в своем развитии. Он не сможет, как вирус, миллиарды лет отсиживать в своей экологической нише в неизменном виде. Чтобы выжить, человечество обречено во все ускоряющемся темпе самосовершенствоваться, опираясь на достижения техники.

Выше было изложено общее видение проблемы, которое связывается с идеей бионики интеллекта. Далее пойдет речь о том, как, на наш взгляд, можно было бы продвигаться вперед в решении этой проблемы. Человек — это материальная система, которая является носителем интеллекта. А что такое интеллект? Он выполняет роль передсмотрящего на корабле. Его цель — обеспечить выживание человека. Чтобы выжить, человечеству приходится постоянно совершенствоваться, подключая для этого технику. На человека природой возложена трудная и ответственная миссия: не только выжить самому, но и обеспечить выживание и беспрепятственное развитие разума вообще. Для этого надо уметь предвидеть события и принимать действенные меры по управлению ими. Человек успешно это делает, обращаясь за помощью к своему мозгу. Основу мозга составляет нейронная сеть, реализующая модель внешнего и внутреннего мира человека. Говоря техническим языком, интеллект — это база данных, из которой человек черпает нужные ему сведения. Интеллект — это система отношений. Математики выражают отношения уравнениями. Информацию из этих уравнений они извлекают путем решения уравнений. Мозг — это сетевая система параллельного действия, которая решает логические уравнения.

По современным представлениям, мозг человека содержит до 10^{11} нервных клеток, каждая может иметь до 10^4 нервных волокон. По нервным волокнам движутся импульсы с частотой до 10^3 Гц. Производительность мозга оценивается величиной порядка 10^{15} бит в секунду. Современная же ЭВМ может

иметь в своем процессоре порядка 10^6 решающих элементов. Частота импульсов, передаваемых по отдельным проводникам процессора, составляет около 10^9 бит в секунду. Все решающие элементы процессора, как и клетки мозга, могут работать параллельно. Потенциальная производительность ЭВМ при современном состоянии электронной техники оценивается величиной порядка 10^{15} бит/сек. Полагают, что именно в нынешнем году потенциальная производительность вычислительных машин сравнялась с производительностью мозга человека. Как будет изменяться потенциальная производительность вычислительных машин в дальнейшем? На этот вопрос дает ответ закон Мура, который гласит: производительность ЭВМ растет по экспоненте. На нынешнем этапе потенциальная производительность ЭВМ ежегодно удваивается. Это означает, что через 10 лет она в тысячу раз превзойдет производительность мозга человека, а через 20 лет — в миллион раз. Итак, технические возможности для создания и совершенствования искусственного интеллекта имеются. Путь открыт. Осталось построить высокопроизводительный процессор ЭВМ параллельного действия, работающий по принципу биологической нейронной сети. Если это удастся сделать, на свет появится мозгоподобная ЭВМ.

Что для этого надо сделать? Нам представляется, что в первую очередь следовало бы решить следующие задачи. Процессы, наблюдаемые во внешнем мире, описываются числовыми отношениями. Основные из них выражаются дифференциальными уравнениями в частных производных. Эти уравнения описывают законы природы, по которым совершаются физические процессы во внешнем мире. Решая такие уравнения, люди получают возможность предсказывать события, происходящие во внешнем мире, и управлять ими. Процессы, наблюдаемые человеком в его собственном внутреннем мире, то есть информационные процессы, тоже выражаются уравнениями, но не числовыми, а логическими. Нужен язык для формульной записи логических уравнений. Такой язык дает нам алгебра предикатов. Насколько мы можем судить, альтернативы языку алгебры предикатов нет. Язык алгебры предикатов универсален, на нем можно записать в виде логического уравнения любое отношение.

Необходимы методы решения логических уравнений. Нужны методы представления различных информационных процессов в виде логических уравнений, выражающих законы протекания этих процессов. Решая логические уравнения, можно будет искусственно воспроизводить интеллектуальную деятельность человека. Требуются методы преобразования логических уравнений в структуру процессора ЭВМ параллельного действия, работающего по принципу биологической нейронной сети. Имея мозгоподобную ЭВМ с таким процессором, можно будет с ее помощью осуществить дальнейшее совершенствование интеллектуальных систем.

Выше было изложено, как, на наш взгляд, проблему бионики интеллекта можно развернуть в систему задач, требующих решения. Теперь остается рассказать о том, как мы (то есть авторы статьи и все те, кто участвовал в движении по этому пути) пытались решать эти задачи.

Отправным пунктом в этом деле для нас послужили следующие два достижения. Первое — это осознание того, что субъективный, то есть внутренний, мир человека можно описывать в точных математических терминах столь же надежно и полно, как это принято в физике, которая имеет дело с явлениями в объективном внешнем мире. И второе — это осознание важности понятия отношения. Первое явилось результатом попыток моделировать психофизические процессы, то есть зависимость ощущений человека от воспринимаемых им внешних предметов. Второе родилось в результате попыток моделирования механизмов естественного языка.

В процессе разработки проблемы вызвало большое удивление, когда обнаружилось, что нет алгебры, на языке которой можно было бы описывать логические уравнения, выражающие произвольные отношения. Представьте себе на минуту, что в вашем распоряжении имеется интегральное исчисление, на языке которого можно описывать преобразование одной числовой функции в другую, но вам неизвестна школьная алгебра, с помощью которой вы могли бы записывать формулами сами числовые функции, фигурирующие в подынтегральном выражении. В этом случае вам пришлось бы столкнуться с большими трудностями при практическом использовании аппарата интегрального исчисления.

Подобное положение до недавнего времени существовало в учении об отношениях. Существовал хорошо развитый математический аппарат для формульного выражения действий над отношениями. Это — язык кванторов, разработанный в конце XIX века австрийским логиком Фреге и широко используемый в математике и логике, а также алгебра Кудда, применяемая в реляционных базах данных. Но как формулой записать конкретную таблицу базы данных или какой-нибудь конкретный предикат, стоящий под знаком квантора? Как это ни удивительно, но 30 лет назад обращение к литературным источникам ответа на этот вопрос не давало.

Стало ясно, почему механизм естественного языка, который, как известно, представляет собой сложнейшую систему отношений, не поддавался математическому описанию. Ответ оказался предельно простым: потому что не существовало математического языка, на котором можно было бы записывать логические уравнения, выражающие законы языка. Тем более, не существовало методов решения таких уравнений. Когда это было осознано, уже не составило большого труда создать нужный язык формульного описания отношений, который получил название алгебры предикатов.

Строение алгебры предикатов предельно просто. Она устроена в некотором смысле даже проще, чем

алгебра булевых функций. В то время, как в алгебре булевых функций имеются три основных операции — конъюнкция, дизъюнкция и отрицание, в алгебре предикатов их всего две — конъюнкция и дизъюнкция. Вместе с тем, в алгебре предикатов используются не двоичные переменные, как в алгебре булевых функций, а буквенные — как в многозначной логике. Поэтому алгебра булевых функций поглощается алгеброй предикатов как частный случай. Замечательно то, что формулы алгебры предикатов могут быть без каких-либо проблем схемно реализованы с помощью той же самой электронной техники, которая применяется при построении современных ЭВМ. В то время, как формулами алгебры булевых функций можно выразить структуру только процессора ЭВМ, формулами алгебры предикатов охватываются также устройства ввода и вывода вычислительной машины.

Сравнивая логические схемы, построенные по формулам алгебры булевых функций, с нейронными сетями мозга, обнаруживают лишь отдаленное сходство между ними. Но если провести такое же сравнение нейронных сетей мозга с логическими схемами, порождаемыми формулами алгебры предикатов, то обнаруживается поразительное сходство их строения. Это сходство настолько велико, что открывается возможность, опираясь на него, с высокой степенью уверенности указать функции многих анатомических нейронных структур. Становится почти несомненным, что алгебра предикатов — это именно то средство, которым воспользовалась природа при создании систем естественного интеллекта. Это — главное. Все остальное можно отнести к деталям. Впечатление такое, что путь к искусственному интеллекту открыт. Однако для заметного продвижения вперед необходимо выполнить еще очень большой объем работ, которые все же не кажутся неприступными.

В заключение перечислим основные темы уже выполненных работ в области бионики интеллекта: 1) разработка алгебры предикатов и операций над предикатами, ее применение в базах данных, экспертных системах, нейроподобных сетях; 2) разработка моделей и методов моделирования механизмов зрения, слуха, других органов чувств, узнавания, оценивания, понимания естественного языка, сознания; 3) соединение алгебры предикатов и предикатных операций с теорией категорий в единую алгебροлогическую систему; 4) формальное и абстрактное описание механизма логики; 5) разработка лингвистической алгебры; 6) разработка методов решения логических уравнений и их применение при проектировании мозгоподобных ЭВМ.

Список литературы: 1. Шенк Р., Хантер Л. Познать механизмы мышления // Реальность и прогнозы искусственного интеллекта. — М.: Мир, 1987. — С. 15–26. 2. Винер Н. Кибернетика. — М.: Сов. радио, 1968. — 326 с. 3. Циолковский К. Э. Грезы о земле и небе. — 1895.

Поступила в редколлегию 28.08.2006

УДК [519.95 + 518.5]: 622.692.4

А. Д. Тевяшев, К. В. Кобылинский

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ОБНАРУЖЕНИЯ КРИМИНАЛЬНЫХ ОТБОРОВ ИЗ ПРОДУКТОПРОВОДОВ

1. Введение

Криминальные отборы углеводородного сырья из продуктопроводов (ПП) являются одной из постоянных проблем, стоящих перед трубопроводными компаниями. Криминальные отборы, кроме прямых экономических потерь, на порядки увеличивают и риски экологических катастроф, связанных с разливом углеводородного сырья, взрывами и пожарами.

В настоящее время все трубопроводные компании имеют собственные службы безопасности, обеспечивающие сохранность как самих трубопроводных систем, так и целевых продуктов (ЦП), транспортируемых по этим системам. Каждая компания применяет свои собственные модели оценки риска криминальных отборов и определяет приемлемый для нее уровень риска, в зависимости от которого формируется структура, состав и степень технической оснащённости ее службы безопасности в условиях ограничения как технических, так и финансовых ресурсов. Наличие мобильных и хорошо оснащённых служб безопасности является необходимым, но не достаточным условием для успешной борьбы с криминальными отборами.

Для эффективной работы служб безопасности они должны обеспечиваться оперативной и достоверной информацией о факте обнаружения криминального отбора, его интенсивности, времени начала и местоположении на трассе ПП. Обеспечение этой информацией возможно за счет включения в состав информационно-аналитических систем управления режимами транспорта и распределения ЦП в трубопроводных системах модуля интеллектуальной системы обнаружения криминальных отборов (ИСОКО). Необходимость в ИСОКО, в отличие от систем обнаружения утечек ЦП из ПП, предопределена ее целевым назначением — интеллектуальным противодействием криминальным группам в сокрытии факта, интенсивности, времени и месте отбора ЦП из ПП. Это противодействие обусловлено использованием криминальными группами множества различных стратегий отбора: от простейших — максимального отбора в течение всего времени, необходимого для заполнения имеющейся емкости путем полного открытия вентиля, через который осуществляется отбор на фоне квазистационарного режима, до комбинированных стратегий — последовательного увеличения/уменьшения отборов с малыми приращениями на фоне существенно нестационарных режимов, вызванных включением/выключением насосных агрегатов (НА), переключением резервуаров (Р) на входе/выходе ПП, значительными изменениями оборотов приводов НА на насосных станциях (НС).

В силу указанных обстоятельств разработка ИСОКО является чрезвычайно актуальной.

2. Структура и параметры продуктопровода

Разработанная ИСОКО обеспечивает эффективную работу на ПП, характеризующихся следующими параметрами:

- длина ПП не ограничена;
- структура ПП — древовидная, однониточная;
- количество НС на ПП не ограничено;
- количество НА, установленных на каждой из НС — до 5;
- количество замерных узлов (ЗУ) на ПП не ограничено;
- расстояния между соседними ЗУ — от 10 до 50 км;
- измеряемые параметры в каждом ЗУ — избыточное давление, массовый расход, плотность и температура ЦП;
- цикл измерения физических параметров ЦП в каждом ЗУ — 1 с;
- величина рассинхронизации данных, поступающих из ЗУ, — не более 1 с.

В основу ИСОКО положены интеллектуальные алгоритмы автоматической обработки, анализа и принятия решений по оперативным данным, поступающим из ЗУ ПП. В состав этих алгоритмов входят следующие процедуры:

- проверка результатов измерений физических параметров ЦП в ЗУ на достоверность;
- заполнение пропусков в результатах измерений физических параметров ЦП;
- фильтрация результатов измерений физических параметров ЦП;
- оценивание и компенсация систематических ошибок результатов измерений массового расхода и плотности ЦП;
- проверка гипотезы и принятие решения о квазистационарности режима ПП;
- методы обнаружения утечек:
 - а) статического баланса;
 - б) динамического баланса;
 - в) знаков приращений расходов;
 - г) знаков приращений давлений;
- проверка сложных гипотез и принятие решения о наличии утечек;
- методы определения времени возникновения утечки и локализации ее местоположения:
 - а) отношения правдоподобия;
 - б) разности между временем прихода фронтов волн возмущений в ЗУ по и против направления движения ЦП;
 - в) максимального правдоподобия.

3. Анализ режимов работы продуктопроводов

Анализ фактических режимов работы продуктопроводов позволяет разделить их на два класса: квазистационарные и существенно нестационарные режимы.

На практике стационарные режимы практически не встречаются из-за наличия ряда возмущающих факторов, которые оказывают постоянное влияние на режим работы ПП. Основными такими факторами являются: опорожнение резервуара в начале ПП, наполнение резервуара в конце ПП, переключение резервуаров, поступление в продуктопровод ЦП с различным химическим составом. При перечисленных возмущениях режим работы ПП можно считать квазистационарным, т. к. данные возмущения малы и не оказывают существенного влияния на его режим работы.

Причиной возникновения существенно нестационарных режимов работы продуктопроводов являются пуски и остановки НА на НС, значительные изменения положения регулирующих задвижек и оборотов приводов НА, аварийные утечки и криминальные отборы.

В момент возникновения криминального отбора из продуктопровода происходит резкое возмущение движения ЦП по ПП. При этом в месте отбора происходит изменение значений параметров ЦП: расхода, давления и температуры. Волны изменения расхода и давления распространяются в обе стороны от места отбора со скоростью, примерно равной скорости звука в рассматриваемом ЦП. Причем значение расхода ЦП в ЗУ до места отбора возрастает, а в ЗУ после отбора уменьшается, в то же время значение давления во всех ЗУ падает. Изменение температуры при криминальном отборе повторяет характер изменения расхода — в ЗУ до места отбора температура ЦП возрастает, а в ЗУ после отбора падает, но эти изменения пропорциональны величине изменения скорости потока ЦП, носят несущественный характер и происходят со значительной временной задержкой. Поэтому такие изменения температуры ЦП в ЗУ не могут служить основой для оперативного принятия решений о наличии криминального отбора.

Если величина отбора постоянная, то распространяющиеся по трубопроводу волны возмущения в течение некоторого времени постепенно затухают и устанавливается новый квазистационарный режим с криминальным отбором.

4. Моделирование режимов работы магистрального продуктопровода

Разработанная ИСОКО предназначена для магистральных ПП линейной одноточечной структуры. Предполагается, что в общем случае магистральный ПП включает в себя ряд НС: одну головную — в начале ПП, остальные промежуточные — вдоль трассы ПП. Расстояние между соседними НС может составлять от нескольких десятков до сотни километров. ИСОКО может контролировать режимы

транспорта ЦП по всему ПП в целом или отдельно по каждому участку ПП между соседними ЗУ.

При разработке математических моделей квазистационарных и существенно нестационарных неизотермических режимов транспорта ЦП по магистральному ПП необходимо учитывать влияние следующих факторов:

- неоднородность плотности ЦП вдоль трассы ПП, обусловленную различным химическим составом ЦП, поступающего в ПП из различных резервуаров;
- сжимаемость и расширяемость ЦП, обусловленные влиянием на него давления и температуры;
- изменение геометрических размеров трубопровода (длины и диаметра), обусловленные влиянием давления и температуры ЦП вдоль трассы ПП.

4.1. Математическая модель установившегося потокораспределения в магистральном продуктопроводе

Математическая модель установившегося потокораспределения (УПР) в магистральном ПП включает в себя математические модели структуры ПП и его технологических элементов: насосных агрегатов (НА), участков трубопровода (УТ), регулирующих задвижек (РЗ), отсекающих задвижек (ОЗ) и обратных клапанов (ОК).

Математическая модель структуры ПП. Технологическая структура ПП рассматривается в виде связанного ориентированного графа с множеством V вершин, которым отвечают узлы сети, и множеством E дуг, которым отвечают технологические элементы. К технологическим элементам относятся: насосные агрегаты (множество L), участки трубопровода (множество P), регулирующие задвижки (множество R), отсекающие задвижки (множество Z), обратные клапаны (множество KL). Полученный граф дополняется «нулевой» вершиной и рядом фиктивных дуг, часть из которых соединяют «нулевую» вершину со входами ПП (множество I), а остальные фиктивные дуги соединяют выходы ПП с «нулевой» вершиной (множество O). Перенумеруем все дуги графа сети. Тогда множество номеров всех дуг графа можно представить в виде объединения следующих взаимно непересекающихся множеств:

$$E = I \cup L \cup P \cup R \cup Z \cup KL \cup O.$$

Произведем выбор дерева графа. Множества номеров дуг, соответствующих ветвям дерева, обозначим индексом 1, а соответствующих хордам — индексом 2. В этом случае имеем:

$$E = E_1 \cup E_2.$$

Математическая модель структуры УПР в магистральном ПП. Для выбранного дерева графа математическая модель структуры УПР в магистральном ПП может быть описана системой уравнений, соответствующей первому и второму законам Кирхгофа [1], и может быть представлена в следующем виде:

$$W_i = \sum_{r \in E_2} b_{1ri} W_r, \quad i \in E_1, \quad (4.1)$$

$$\Delta P_r = \sum_{i \in E_1} b_{1ri} \Delta P_i, \quad r \in E_2, \quad (4.2)$$

где $b_{1ri} \in \{-1; 0; 1\}$ ($i \in E_1, r \in E_2$) — элемент матрицы B_1 , с помощью которой описывается граф с выбранным на нем деревом; W_j — массовый расход на элементе $j \in E$; ΔP_j — разность давлений на входе и на выходе элемента $j \in E$, причем

$$\Delta P_j = \Delta P_j(W_j), \quad j \in L \cup P \cup R \cup Z \cup KL. \quad (4.3)$$

Зависимость (4.3) определяется математической моделью УПР на соответствующем технологическом элементе.

Математическая модель насосного агрегата:

$$\Delta P_j(W_j) = a_{0j} + a_{1j} \cdot \left(\frac{W_j}{\rho}\right) + a_{2j} \cdot \left(\frac{W_j}{\rho}\right)^2, \quad j \in L, \quad (4.4)$$

где a_{0j}, a_{1j}, a_{2j} — коэффициенты квадратичной аппроксимации напорной характеристики j -го НА [2].

Математическая модель участка трубопровода:

$$\Delta P_j(W_j) = \frac{8L_j \lambda_j}{8\pi^2 d_j^5 \rho E_j^2} \cdot W_j \cdot |W_j| + \rho g \Delta z_j, \quad j \in P. \quad (4.5)$$

где L_j — длина j -го УТ, м; d_j — внутренний диаметр трубопровода, м; E_j — коэффициент эффективности; λ_j — коэффициент гидравлического трения; Δz_j — разность геодезических отметок начала и конца j -го УТ. Для расчета значения коэффициента λ_j А. Д. Альтшулем была предложена следующая полуэмпирическая формула:

$$\lambda_j = 0,11 \left(\frac{k_{эj}}{d_j} + \frac{68\nu}{v_j d_j} \right)^{0,25}$$

где $k_{эj}$ — эквивалентная шероховатость стенок труб, м; v_j — скорость потока, м/с; ν — кинематическая вязкость ЦП, м²/с.

Математическая модель регулирующей задвижки:

$$\Delta P_j(W_j) = \frac{c_{P3j}}{E_{P3j}^2} \cdot W_j \cdot |W_j|, \quad j \in R, \quad (4.6)$$

где c_{P3j} — эквивалентное гидравлическое сопротивление j -й РЗ, находящейся в положении «открыто»; $E_{P3j} \in [0, 1]$ — степень открытия РЗ, определяемая положением штока РЗ или углом поворота перекрывающего клапана, причем $E_{P3j} = 1$ в положении «открыто», $E_{P3j} \rightarrow 0$ в положении «закрыто».

Математическая модель отсекающей задвижки:

$$\Delta P_j(W_j) = \frac{c_{O3j}}{E_{O3j}^2} \cdot W_j \cdot |W_j|, \quad j \in Z, \quad (4.7)$$

где c_{O3j} — эквивалентное гидравлическое сопротивление j -й ОЗ в открытом состоянии; E_{O3j} — степень открытия ОЗ, причем $E_{O3j} = 1$ в положении «открыто», $E_{O3j} \rightarrow 0$ в положении «закрыто».

Математическая модель УПР на обратном клапане:

$$\Delta P_j(W_j) = \frac{c_{OKj}}{E_{OKj}^2} \cdot W_j \cdot |W_j|, \quad j \in KL, \quad (4.8)$$

где c_{OKj} — эквивалентное гидравлическое сопротивление j -го ОК в открытом состоянии; $E_{OKj}(W_j)$ — степень открытия ОК, причем $E_{OKj}(W_j) = 1$ при $W_j \geq 0$, $E_{OKj}(W_j) \rightarrow 0$ при $W_j < 0$.

Таким образом, математическая модель УПР в магистральном ПП включает в себя две системы алгебраических уравнений для первого и второго законов Кирхгофа (4.1), (4.2) и уравнения математических моделей УПР на технологических элементах ПП (4.4)–(4.8). Моделирование УПР в магистральном ПП заключается в решении системы нелинейных алгебраических уравнений, входящих в математическую модель УПР в магистральном ПП, одним из численных методов.

Это классический подход к моделированию стационарного или установившегося потокораспределения в магистральных ПП. Однако, как уже отмечалось в предыдущем разделе, на практике стационарные режимы практически не встречаются. Поэтому модель УПР в магистральном ПП должна быть обобщена на случай существенно нестационарных неизотермических режимов (ННР). При построении такой модели предполагается, что основными инерционными элементами в магистральном ПП являются участки трубопровода протяженностью более одного километра, а это означает, что инерционностью процессов в НС, включающей в себя участки трубопровода до десятков метров, можно пренебречь. Это позволяет строить математическую модель существенно нестационарных неизотермических режимов (ММСННР) в виде двух взаимосвязанных классов моделей — модели УПР для НС и ММСННР для всех участков ПП между ЗУ.

4.2. Математическая модель существенно нестационарных неизотермических режимов транспорта ЦП по участку магистрального ПП

Пусть $\{X_1, X_2, \dots, X_N\}$ — координаты ЗУ магистрального ПП. Рассмотрим участок ПП между n -м и $(n+1)$ -м ЗУ. Основными предпосылками при построении ММСННР транспорта ЦП по участку ПП являются три закона:

- закон сохранения количества движения [3]:

$$\frac{\partial P}{\partial x} + \frac{\partial(\rho v)}{\partial t} + \frac{\lambda}{2d_0} \rho v^2 + \rho g \frac{dz}{dx} = 0;$$

- закон сохранения массы [3]:

$$\frac{\partial(\rho S)}{\partial t} + \frac{\partial(\rho v S)}{\partial x} = 0;$$

- закон сохранения энергии [4]:

$$\frac{\partial T}{\partial t} + v \frac{\partial T}{\partial x} + \frac{\pi d_0 K}{\rho S C_p} (T - T_{\text{тр}}) - \frac{v}{C_p} \frac{dz}{dx} = 0,$$

где $x \in [X_n, X_{n+1}]$ — пространственная переменная, м; $t \in [t_1, t_2]$ — переменная времени, с; $P = P(x, t)$ — среднее по сечению давление, Па; $\rho = \rho(x, t)$ — средняя по сечению плотность ЦП, кг/м³; $v = v(x, t)$ — средняя по сечению скорость потока, м/с; $T = T(x, t)$ — температура ЦП, °С; $S = S(x, t)$ — площадь поперечного сечения участка трубопровода, м²; $d_0 = d_0(x)$ — внутренний диаметр участка трубопровода, м; λ — коэффициент гидравлического сопротивления, который рассчитывается по формуле Альтшуля

$$\lambda = 0,11 \left(\frac{k_s}{d_0} + \frac{68v}{vd} \right)^{0,25};$$

k_s — эквивалентная шероховатость стенок труб, м; v — кинематическая вязкость ЦП, м²/с; $g = 9,81$ м/с² — ускорение свободного падения; $z = z(x)$ — геодезическая отметка в точке x , м; K — коэффициент теплопередачи от ЦП к грунту, Вт/(м²·°С); C_p — удельная теплоемкость ЦП, Дж/(кг·°С); $T_{гр}$ — температура грунта, °С.

При определении параметров ПП в реальных условиях необходимо учитывать продольные и поперечные деформации участка трубопровода, обусловленные влиянием давления и температуры транспортируемого ЦП. В этом случае площадь внутреннего сечения участка трубопровода в точке x в момент времени t должна рассчитываться по формуле [4]:

$$S = k_S \cdot S_0,$$

$$\text{где } k_S = 1 + \frac{d_0}{E_{Ю}\delta} \cdot (P - P_0) + 2\alpha_L \cdot (T - T_0);$$

$k_S = k_S(x, t)$ — коэффициент поперечного расширения участка трубопровода; $P_0 = 4,013 \cdot 10^5$ Па, $T_0 = 20$ °С — давление и температура в стандартных условиях; $S_0 = S_0(x)$ — площадь внутреннего сечения участка трубопровода в точке x в стандартных условиях, причем $S_0 = \pi d_0^2 / 4$; $E_{Ю} = E_{Ю}(x)$ — модуль Юнга материала стенок участка трубопровода в точке x , Па; $\alpha_L = \alpha_L(x)$ — коэффициент линейного (температурного) расширения материала стенок участка трубопровода в точке x , 1/°С.

ЦП также обладает свойством сжимаемости и расширяемости, которое обусловлено влиянием давления и температуры, что выражается следующей формулой изменения его плотности [4]:

$$\rho = k_p \cdot \rho_0, \quad (4.9)$$

$$\text{где } k_p = 1 + \frac{1}{K_{ЦП}} \cdot (P - P_0) - \xi_{ЦП} \cdot (T - T_0);$$

$k_p = k_p(x, t)$ — коэффициент сжатия ЦП; $\rho_0 = \rho_0(x, t)$ — приведенная (к стандартным условиям) плотность ЦП; $K_{ЦП}$ — модуль упругости ЦП, Па; $\xi_{ЦП}$ — коэффициент объемного (температурного) расширения ЦП, 1/°С.

При магистральном транспорте неоднородных ЦП необходимо учитывать процессы продольного размытия фронтов ЦП с различной плотностью. Уравнение продольного перемешивания ЦП на границе фронтов, записанное для приведенной плотности ЦП в точке x участка трубопровода в момент времени t , имеет следующий вид [4]:

$$\frac{\partial \rho_0}{\partial t} + v \cdot \frac{\partial \rho_0}{\partial x} + K_{пр} \frac{\partial^2 \rho_0}{\partial x^2} = 0,$$

где $K_{пр}$ — эффективный коэффициент продольного перемешивания, значение которого оценивается экспериментально или приближенно рассчитывается по формуле [4]:

$$K_{пр} = 1,785 \cdot \sqrt{\lambda} \cdot v \cdot d_0.$$

С учетом уравнений сжимаемости и расширяемости ЦП, после соответствующих преобразований уравнений, получим ММСНП транспорта ЦП на участке трубопровода в виде следующей системы дифференциальных уравнений в частных производных:

$$\begin{cases} \frac{\partial P}{\partial x} + \frac{\rho_0}{S_0 k_S} \cdot \frac{\partial Q_0}{\partial t} + \frac{0,5\lambda \rho_0 Q_0^2}{d_0 k_p S_0^2 k_S^2 E^2} + \rho_0 k_p g \frac{\partial z}{\partial x} = 0, \\ \frac{\partial P}{\partial t} + \frac{1}{S_0} \cdot \left(\frac{d_0}{E_{Ю}\delta} + \frac{1}{K_{ЦП}} \right)^{-1} \cdot \frac{\partial Q_0}{\partial x} = 0, \\ \frac{\partial T}{\partial t} + \frac{Q_0}{S_0} \cdot \frac{\partial T}{\partial x} + \frac{(T - T_{гр})4K}{\rho_0 k_p d_0 C_p} - \frac{Q_0}{S_0 C_p} \cdot \frac{\partial z}{\partial x} = 0, \\ \frac{\partial \rho_0}{\partial t} + \frac{Q_0}{S_0} \cdot \frac{\partial \rho_0}{\partial x} + 1,785 \sqrt{\lambda} d_0 \cdot \frac{Q_0}{S_0} \cdot \frac{\partial^2 \rho_0}{\partial x^2} = 0, \end{cases} \quad (4.10)$$

$$x \in [X_n, X_{n+1}], t \in [t_1, t_2]$$

где E — коэффициент эффективности участка трубопровода (оценивается по экспериментальным данным); $Q_0 = Q_0(x, t)$ — приведенный к стандартным условиям объемный расход ЦП, причем

$$Q_0 = \frac{W}{\rho_0}. \quad (4.11)$$

Моделирование существенно нестационарных неизоэотермических режимов транспорта ЦП на участке ПП между n -м и $(n + 1)$ -м ЗУ заключается в численном решении краевой задачи, включающей в себя:

- систему дифференциальных уравнений (4.10) относительно неизвестных функций $P(x, t)$, $Q_0(x, t)$, $T(x, t)$ и $\rho_0(x, t)$;
- начальные условия, т. е. профили давления, расхода, температуры и приведенной плотности в начальный момент времени t_1 :

$$\begin{aligned} P(x, t_1) &= P^{(нач)}(x), Q_0(x, t_1) = Q_0^{(нач)}(x), \\ T(x, t_1) &= T^{(нач)}(x), \rho_0(x, t_1) = \rho_0^{(нач)}(x), \end{aligned} \quad (4.12)$$

$$x \in [X_n, X_{n+1}]$$

где $P^{(\text{нач})}(x), Q_0^{(\text{нач})}(x), T^{(\text{нач})}(x), \rho_0^{(\text{нач})}(x)$ — заданные функции;

- граничные условия, состоящие из условий в начальной точке участка:

$$\begin{aligned} P(X_n, t) &= P_n(t), T(X_n, t) = T_n(t), \\ \rho_0(X_n, t) &= \rho_{0n}(t), t \in [t_1, t_2], \end{aligned} \quad (4.13)$$

и в конечной точке участка:

$$Q_0(X_{n+1}, t) = Q_{0n+1}(t), t \in [t_1, t_2], \quad (4.14)$$

где $P_n(t), T_n(t), \rho_{0n}(t), Q_{0n+1}(t)$ — заданные функции.

Для решения краевой задачи (4.10), (4.12)–(4.14) используется конечно-разностная аппроксимация частных производных на основе неявной схемы [5]. Поиск решения полученной системы нелинейных уравнений осуществляется численным методом. Решением задачи являются функции оценок давления $\hat{P}(x, t)$, расхода $\hat{Q}_0(x, t)$, температуры $\hat{T}(x, t)$ и приведенной плотности $\hat{\rho}_0(x, t)$ при $x \in [X_n, X_{n+1}], t \in [t_1, t_2]$.

4.3. Математическая модель квазистационарных неизоэнтальпических режимов (ММКШР) транспорта ЦП по участку магистрального ПП

При квазистационарном режиме работы ПП уравнения математической модели можно упростить за счет того, что производные по времени обращаются в ноль. В этом случае из (4.10) для участка трубопровода постоянного диаметра между n -м ЗУ и $(n+1)$ -м ЗУ для произвольных точек и получим:

$$\begin{cases} P_{x_1} - P_{x_2} = \frac{(x_1 - x_2)\lambda \rho_{0x_1, x_2}^{(\text{cp})} Q_{0x_1}^2}{2E^2 k_{S_{x_1, x_2}}^{(\text{cp})^2} k_{\rho_{x_1, x_2}}^{(\text{cp})^2} d_0 S_0^2} + \\ \quad + \rho_{0x_1, x_2}^{(\text{cp})} k_{\rho_{x_1, x_2}}^{(\text{cp})} g \cdot (z(x_2) - z(x_1)), \\ Q_{0x_1} = Q_{0x_2}, \\ \frac{T_{x_2} - T_{\text{гр}}}{T_{x_1} - T_{\text{гр}}} = \exp\left(\frac{(x_1 - x_2)\pi d_0 K}{\rho_{0x_1, x_2}^{(\text{cp})} k_{\rho_{x_1, x_2}}^{(\text{cp})} C_p Q_{0x_1}}\right), \\ \hat{\rho}_{0x_1, x_2}^{(\text{cp})} = 0,5 \cdot (\rho_{0x_1} + \rho_{0x_2}), \\ \forall x_1, x_2 \in [X_n, X_{n+1}]; x_1 < x_2, \end{cases} \quad (4.15)$$

где $k_{S_{x_1, x_2}}^{(\text{cp})} = \frac{1}{2} \cdot (k_S(P_{x_1}, T_{x_1}) + k_S(P_{x_2}, T_{x_2}))$, $k_{\rho_{x_1, x_2}}^{(\text{cp})} = \frac{1}{2} \cdot (k_{\rho}(P_{x_1}, T_{x_1}) + k_{\rho}(P_{x_2}, T_{x_2}))$ — средние значения соответствующих коэффициентов; $P_x, Q_{0x}, T_x, \rho_{0x}$ — соответственно давление, расход, температура и приведенная плотность в заданной точке x , т. е. x_1 или x_2 .

Запишем (4.15) для n -го и $(n+1)$ -го ЗУ:

$$\begin{cases} P_n - P_{n+1} = \frac{(X_{n+1} - X_n)\lambda \rho_{0n, n+1}^{(\text{cp})} Q_{0n}^2}{2E^2 k_{S_{n, n+1}}^{(\text{cp})^2} k_{\rho_{n, n+1}}^{(\text{cp})^2} d_0 S_0^2} + \\ \quad + \rho_{0n, n+1}^{(\text{cp})} k_{\rho_{n, n+1}}^{(\text{cp})} g \cdot (z(x_2) - z(x_1)), \\ Q_{0n} = Q_{0n+1}, \\ \frac{T_{n+1} - T_{\text{гр}}}{T_n - T_{\text{гр}}} = \exp\left(\frac{(X_{n+1} - X_n)\pi d_0 K}{\rho_{0n, n+1}^{(\text{cp})} k_{\rho_{n, n+1}}^{(\text{cp})} C_p Q_{0n}}\right), \\ \hat{\rho}_{0n, n+1}^{(\text{cp})} = 0,5 \cdot (\rho_{0n} + \rho_{0n+1}), \end{cases} \quad (4.16)$$

где $k_{S_{n, n+1}}^{(\text{cp})} = \frac{1}{2} \cdot (k_S(P_n, T_n) + k_S(P_{n+1}, T_{n+1}))$, $k_{\rho_{n, n+1}}^{(\text{cp})} = \frac{1}{2} \cdot (k_{\rho}(P_n, T_n) + k_{\rho}(P_{n+1}, T_{n+1}))$; $P_n, Q_{0n}, T_n, \rho_{0n}$ — соответственно давление, расход, температура и приведенная плотность ЦП в n -м ЗУ; $P_{n+1}, Q_{0n+1}, T_{n+1}, \rho_{0n+1}$ — соответственно давление, расход, температура и приведенная плотность ЦП в $(n+1)$ -м ЗУ.

4.4. Проверка квазистационарности режима магистрального продуктопровода

Для оценивания степени связи стационарности изменений физических параметров в некоторой точке магистрального ПП достаточно исследовать поведение массового расхода, так как любое существенное изменение давления, плотности или температуры на входах или выходах магистрального ПП приводит к существенному изменению массового расхода.

Для проверки квазистационарности режима работы магистрального ПП будем использовать метод отношения правдоподобия, предназначенный для выявления скачка в поведении процесса [6]:

$$\max_{l \leq r \leq K} S_n^{(r, K)}(\hat{W}_n^0, \hat{v}_n(r)) < \beta_n, \quad (4.17)$$

где

$$\begin{aligned} S_n^{(r, K)}(\hat{W}_n^0, \hat{v}_n) &= v_n \cdot \sum_{k=r}^K \left(\tilde{W}_n(t_k) - \hat{W}_n^0 - \frac{v_n}{2} \right); \\ \hat{v}_n(r) &= \frac{1}{K-r+1} \cdot \sum_{k=r}^K \left(\tilde{W}_n(t_k) - \hat{W}_n^0 \right); \end{aligned}$$

$\tilde{W}_n(t_k)$ — результат измерения массового расхода в n -м ЗУ в момент времени t_k ; \hat{W}_n^0 — оценка среднего значения процесса до начального момента времени t_1 ; β_n — некоторое пороговое значение критерия, численное значение которого получается в результате обучения (настройки) ИСОКО.

Численное значение параметра, при котором выражение (4.17) достигает максимума, соответствует моменту времени t_n прихода фронта волны возмущения на n -й ЗУ. Считаем, что режим является квазистационарным на участке магистрального ПП между n -м и $(n+1)$ -м ЗУ, если условия квазистационарности (4.17) выполняются в n -м и $(n+1)$ -м ЗУ.

5. Оценивание параметров ММСНП и ММКСНП

На практике в режиме on-line или off-line экспериментально оценивается коэффициент эффективности E и коэффициент теплопередачи K для участка ПП между n -м и $(n + 1)$ -м ЗУ. Если в момент времени t выполняется условие квазистационарности в n -м и $(n + 1)$ -м ЗУ (п. 4.4), то при отсутствии отборов ЦП оценки коэффициента эффективности и коэффициента теплопередачи рассчитываются и соответственно, по формулам, вытекающим из (4.16):

$$\hat{E}^2 = \frac{((\hat{Q}_{0n}^t + \hat{Q}_{0n+1}^t) / 2)^2}{\hat{P}_n^t - \hat{P}_{n+1}^t - \hat{\rho}_0^{(cp)} \hat{k}_p^{(cp)} g \cdot (z(X_{n+1}) - z(X_n))} \times \frac{(X_{n+1} - X_n) \lambda \hat{\rho}_0^{(cp)}}{2 \hat{k}_S^{(cp)2} \hat{k}_p^{(cp)} d_0 S_0^2},$$

$$\hat{K} = \frac{\hat{\rho}_0^{(cp)} \hat{k}_p^{(cp)} C_p \cdot ((\hat{Q}_{0n}^t + \hat{Q}_{0n+1}^t) / 2)^2}{(X_{n+1} - X_n) \pi d_0} \cdot \ln \frac{\hat{T}_{n+1}^t - T_{гр}}{\hat{T}_n^t - T_{гр}},$$

где $\hat{P}_n^t, \hat{P}_{n+1}^t$ — результаты измерений давления в момент времени t в n -м и $(n + 1)$ -м ЗУ; $\hat{T}_n^t, \hat{T}_{n+1}^t$ — результаты измерений температуры в момент времени t в n -м и $(n + 1)$ -м ЗУ; $\hat{Q}_{0n}^t, \hat{Q}_{0n+1}^t$ — оценки приведенного объемного расхода в момент времени t в n -м и $(n + 1)$ -м ЗУ, рассчитанные на основе результатов прямых измерений; $\hat{\rho}_0^{(cp)}, \hat{k}_S^{(cp)}, \hat{k}_p^{(cp)}$ — оценки средних значений, соответственно, приведенной плотности, коэффициента сжатия ЦП и коэффициента поперечного расширения трубопровода для данного участка ПП, рассчитанные на основе результатов прямых измерений.

6. Предварительная обработка оперативных данных

6.1. Проверка оперативных данных на достоверность

Анализ на достоверность информации, переданной из ЗУ системой сбора данных в виде флага:

- 0 — измеренное значение достоверно;
- 1 — измеренное значение не достоверно.

Область допустимых значений для каждого измерения:

$$X_{\min} < \tilde{X}_n^t < X_{\max},$$

где \tilde{X}_n^t — измеренное значение с n -го ЗУ в момент времени t ; X_{\min} — минимальное значение, которое может принимать \tilde{X}_n^t ; X_{\max} — максимальное значение, которое может принимать \tilde{X}_n^t .

Скорость изменения измеренного значения:

$$v_{\min} < \left| \tilde{X}_n^t - \tilde{X}_n^{t-1} \right| < v_{\max},$$

где v_{\min} — минимальное значение скорости изменения; v_{\max} — максимальное значение скорости изменения.

6.2. Восстановление пропусков в оперативных данных

Пусть пропуск данных в n -м ЗУ составляет измерений. Предполагаем, что до и после пропуска име-

ется по n значений измеренных данных: до пропуска — $(\tilde{x}_{i-n+1}, \tilde{x}_{i-n}, \dots, \tilde{x}_i)$ и после пропуска — $(\tilde{x}_{i+k+1}, \tilde{x}_{i+k+2}, \dots, \tilde{x}_{i+k+n})$.

Введем обозначения:

$$\bar{x}^+ = \sum_{r=i+k+1}^{i+k+n} \tilde{x}_r, \quad \bar{x}^- = \sum_{r=i-k+1}^i \tilde{x}_r.$$

Тогда каждое из k пропущенных измерений рассчитаем по формуле:

$$\tilde{x}_j = \frac{\bar{x}^+ - \bar{x}^-}{k} \cdot j + \bar{x}^-, \quad j = \overline{1, k}.$$

Восстановление пропуска в оперативных данных в каждом ЗУ производится только в том случае, если величина пропуска не превышает заданной величины k , в противном случае рассматриваемые данные исключаются из анализа до поступления новых достоверных оперативных данных. Численное значение k определяется при обучении (настройке) ИСОКО.

6.3. Фильтрация оперативных данных

Для уменьшения случайной ошибки результатов измерений, все измеренные значения подвергаются процедуре фильтрации:

$$\hat{X}_n^t = \frac{1}{2K_n^t + 1} \cdot \sum_{i=t-K_n^t}^{t+K_n^t} \tilde{X}_n^i,$$

где \hat{X}_n^t — отфильтрованное значение с n -го ЗУ в момент времени t ; \tilde{X}_n^i — измеренное значение с n -го ЗУ в момент времени i ; K_n^t — глубина фильтра для n -го ЗУ в момент времени t . Численное значение K_n^t определяется при обучении (настройке) ИСОКО.

7. Оценивание систематических ошибок результатов измерений массового расхода и плотности ЦП и устранение их влияния

Результаты измерений физических параметров ЦП в ЗУ ПП содержат в себе два вида ошибок измерений: случайные и систематические. Фильтрация оперативных данных снижает уровень случайных ошибок измерений. Для устранения систематических ошибок измерений существует два подхода: калибровка измерительных датчиков и применение математических методов оценивания систематических ошибок.

В каждом ЗУ осуществляются измерения следующих величин: избыточное давление — \tilde{P}_n^t , массовый расход — \tilde{W}_n^t , плотность — $\tilde{\rho}_n^t$, температура — \tilde{T}_n^t , где t — дискретный момент времени, n — номер ЗУ, $n = 1, 2, \dots, N$.

Математическая модель результата измерения параметра Y имеет вид:

$$\tilde{Y} = Y^* + \Delta_Y + \xi_Y,$$

где \tilde{Y} — результат измерения параметра Y ; Y^* — истинное значение параметра Y ; Δ_Y — систематическая ошибка результата измерения параметра Y ; ξ_Y — случайная ошибка результата измерения параметра Y .

Считаем, что случайные ошибки измерений имеют нормальное распределение вероятностей с нулевым математическим ожиданием и дисперсией σ_Y^2 , т. е. $\xi_Y \sim N(0, \sigma_Y^2)$. Дисперсия σ_Y^2 определяется классом точности датчика. Фактическое значение дисперсии σ_Y^2 оценивается экспериментальным путем.

Таким образом, математические модели результатов измерений, соответственно, избыточного давления, массового расхода, плотности и температуры в n -м ЗУ в момент времени t выражаются формулами вида:

$$\begin{aligned} \tilde{P}_n^t &= P_n(t) + \Delta_{P_n} + \xi_{P_n}^t, \quad \xi_{P_n}^t \sim N(0, \sigma_{P_n}^2); \\ \tilde{W}_n^t &= W_n(t) + \Delta_{W_n} + \xi_{W_n}^t, \quad \xi_{W_n}^t \sim N(0, \sigma_{W_n}^2); \\ \tilde{\rho}_n^t &= \rho_n(t) + \Delta_{\rho_n} + \xi_{\rho_n}^t, \quad \xi_{\rho_n}^t \sim N(0, \sigma_{\rho_n}^2); \\ \tilde{T}_n^t &= T_n(t) + \Delta_{T_n} + \xi_{T_n}^t, \quad \xi_{T_n}^t \sim N(0, \sigma_{T_n}^2). \end{aligned}$$

Считаем, что случайные ошибки измерений независимы друг от друга.

В ИСОКО систематические ошибки измерений давления и температуры оказывают малое влияние на результаты вычисления соответствующих оценок и принятия решений, так что ими можно пренебречь. В то же время влияние систематических ошибок измерений массового расхода и плотности ЦП значимо, оно оценивается и устраняется.

Пусть в момент времени t имеет место квазистационарный режим на участке ПП между n -м и $(n+1)$ -м ЗУ. Согласно математической модели квазистационарного неизотермического режима, имеем

$$Q_{0n}(t) = Q_{0n+1}(t). \quad (7.1)$$

Уравнение (7.1) при подстановке в него выражения (4.11) с учетом (4.9) принимает вид:

$$k_{\rho_n}(t) \cdot \frac{W_n(t)}{\rho_n(t)} = k_{\rho_{n+1}}(t) \cdot \frac{W_{n+1}(t)}{\rho_{n+1}(t)}$$

или

$$k_{\rho_n}(t) \cdot W_n(t) \cdot \rho_{n+1}(t) - k_{\rho_{n+1}}(t) \cdot W_{n+1}(t) \cdot \rho_n(t) = 0. \quad (7.2)$$

Предположим, зафиксировано M различных квазистационарных режимов работы ПП, и по каждому интервалу времени, на котором была установлена квазистационарность m -го режима, для каждого n -го ЗУ рассчитаны оценки среднего значения и дисперсии всех видов измерений, которые обозначим соответственно через $\hat{W}_n^{(m)}$, $\hat{\sigma}_{W_n}^{(m)2}$, $\hat{\rho}_n^{(m)}$, $\hat{\sigma}_{\rho_n}^{(m)2}$, $\hat{P}_n^{(m)}$, $\hat{\sigma}_{P_n}^{(m)2}$, $\hat{T}_n^{(m)}$, $\hat{\sigma}_{T_n}^{(m)2}$, $n = 1, 2, \dots, N$, $m = 1, 2, \dots, M$.

Приведем обоснование формальной постановки задачи оценивания систематических смещений измерений массового расхода $\Delta_{W_{n,n+1}}$ и плотности $\Delta_{\rho_{n,n+1}}$, которые выражаются через систематические ошибки следующим образом:

$$\begin{aligned} \Delta_{W_{n,n+1}} &= \Delta_{W_{n+1}} - \Delta_{W_n}, \\ \Delta_{\rho_{n,n+1}} &= \Delta_{\rho_{n+1}} - \Delta_{\rho_n}. \end{aligned}$$

Запишем выражение для расчета величины невязки, соответствующей левой части выражения (7.2) при подстановке в него результатов измерений по m -му зафиксированному квазистационарному режиму, с учетом поправок на систематические смещения измерений массового расхода и плотности в $(n+1)$ -м ЗУ относительно n -го ЗУ:

$$\begin{aligned} \hat{f}^{(m)} &= \hat{k}_{\rho_n}^{(m)} \cdot \hat{W}_n^{(m)} \cdot (\hat{\rho}_{n+1}^{(m)} - \Delta_{\rho_{n,n+1}}) - \\ &- \hat{k}_{\rho_{n+1}}^{(m)} \cdot (\hat{W}_{n+1}^{(m)} - \Delta_{W_{n,n+1}}) \cdot \hat{\rho}_n^{(m)}. \end{aligned} \quad (7.3)$$

Невязка $\hat{f}^{(m)} = \hat{f}^{(m)}(\Delta_{W_{n,n+1}}, \Delta_{\rho_{n,n+1}})$ при фиксированных значениях $\Delta_{W_{n,n+1}}, \Delta_{\rho_{n,n+1}}$ является случайной величиной, так как подставляемые в (7.3) результаты измерений физических параметров ЦП есть случайные величины. Необходимо найти такие значения неизвестных систематических смещений $\Delta_{W_{n,n+1}}, \Delta_{\rho_{n,n+1}}$, при которых отклонения от нуля невязок $\hat{f}^{(m)}$, $m = 1, 2, \dots, M$, будут обусловлены только влиянием случайных ошибок измерений.

По предположению, случайные ошибки измерений имеют нормальное распределение вероятностей. В силу малости случайных ошибок измерений можно считать, что распределение случайной величины $\hat{f}^{(m)}$ близко к нормальному распределению вероятностей:

$$\hat{f}^{(m)} \sim N(0, \sigma_f^{(m)}).$$

Дисперсию невязки $\hat{f}^{(m)}$ с большой степенью точности можно оценить на основе линеаризации зависимости (6.3):

$$\begin{aligned} \hat{\sigma}_f^{(m)2} &= \left(\hat{k}_{\rho_n}^{(m)} \hat{\rho}_{n+1}^{(m)} \hat{\sigma}_{W_n}^{(m)} \right)^2 + \left(\hat{k}_{\rho_{n+1}}^{(m)} \hat{W}_{n+1}^{(m)} \hat{\sigma}_{\rho_n}^{(m)} \right)^2 + \\ &+ \left(\hat{k}_{\rho_{n+1}}^{(m)} \hat{\rho}_n^{(m)} \hat{\sigma}_{W_{n+1}}^{(m)} \right)^2 + \left(\hat{k}_{\rho_n}^{(m)} \hat{W}_n^{(m)} \hat{\sigma}_{\rho_{n+1}}^{(m)} \right)^2. \end{aligned}$$

Оценивание систематических смещений $\Delta_{W_{n,n+1}}$ и $\Delta_{\rho_{n,n+1}}$ выполним, воспользовавшись методом максимального правдоподобия. Для совокупности реализаций нормально распределенной случайной величины $\hat{f}^{(m)}$ по M зафиксированным квазистационарным режимам логарифмическая функция максимального правдоподобия имеет вид:

$$L = \sum_{m=1}^M \frac{1}{\hat{\sigma}_f^{(m)2}} \hat{f}^{(m)2}.$$

В этом случае задача оценивания систематических смещений $\Delta_{W_{n,n+1}}$ и $\Delta_{\rho_{n,n+1}}$ записывается в следующей постановке:

$$\begin{aligned} &I(\Delta_{W_{n,n+1}}, \Delta_{\rho_{n,n+1}}) = \\ &= \sum_{m=1}^M \frac{\hat{f}^{(m)2}(\Delta_{W_{n,n+1}}, \Delta_{\rho_{n,n+1}})}{\hat{\sigma}_f^{(m)2}} \rightarrow \min_{(\Delta_{W_{n,n+1}}, \Delta_{\rho_{n,n+1}}) \in R^2}. \end{aligned} \quad (7.4)$$

В результате решения задачи безусловной минимизации выпуклой функции $L(\Delta_{W_{n,n+1}}, \Delta_{P_{n,n+1}})$ (7.4) одним из численных методов получим оценки систематических смещений $\hat{\Delta}_{W_{n,n+1}}$ и $\hat{\Delta}_{P_{n,n+1}}$.

Оценки систематических ошибок измерений массового расхода и плотности рассчитываются по следующим формулам:

$$\hat{\Delta}_{W_n} = \hat{\Delta}_{W_{n,1}} - \hat{\Delta}_{W_n}^0, \quad \hat{\Delta}_{P_n} = \hat{\Delta}_{P_{n,1}} - \hat{\Delta}_{P_n}^0, \quad (7.5)$$

$$n = 1, 2, \dots, N,$$

где $\hat{\Delta}_{W_n}^0 = \frac{1}{N} \cdot \sum_{n=1}^N \hat{\Delta}_{W_{n,1}}$, $\hat{\Delta}_{P_n}^0 = \frac{1}{N} \cdot \sum_{n=1}^N \hat{\Delta}_{P_{n,1}}$; $\hat{\Delta}_{W_{n,1}} = \sum_{i=2}^n \hat{\Delta}_{W_{i-1,i}}$, $\hat{\Delta}_{P_{n,1}} = \sum_{i=2}^n \hat{\Delta}_{P_{i-1,i}}$, $n = 1, 2, \dots, N$.

Полученные оценки систематических ошибок измерений массового расхода и плотности используются для коррекции результатов измерений в каждом из N ЗУ.

Дисперсии оценок систематических ошибок $\hat{\Delta}_{W_n}$, $\hat{\Delta}_{P_n}$, $n = 1, 2, \dots, N$, могут быть оценены методом Монте-Карло.

8. Методы обнаружения утечек и криминальных отборов

8.1. Метод статического баланса

Метод статического баланса (МСБ) используется каждый раз перед обучением (настройкой) ИСОКО и позволяет оценить:

- систематические ошибки измерения геодезических отметок ЗУ;
- наличие, местоположение и объемы стационарных утечек на каждом из участков продуктопровода между ЗУ.

Суть метода заключается в следующем. После подтверждения квазистационарности режима ПП начинается постепенное (во избежание гидравлических ударов) перекрытие ОЗ на всех выходах ПП при всех работающих НС. После полного перекрытия ОЗ на выходах ПП расходы ЦП на НС становятся равными нулю и НС отключаются при помощи ОЗ.

В таком состоянии ПП находится в течение некоторого интервала времени ΔT_1 , пока не будут завершены все переходные процессы. Критерием окончания переходных процессов является условие:

$$\forall j \in N : \sigma_{P_j}^2 \leq \varepsilon, \quad (8.1)$$

где $N = \{1, 2, \dots, N\}$ — множество ЗУ ПП, $\sigma_{P_j}^2$ — оценка дисперсии давления в i -м ЗУ, $\varepsilon > 0$ — некоторая малая положительная величина, численное значение которой определяется при обучении (настройке) ИСОКО. Начальное значение $\varepsilon = 0.1$.

Далее перекрываются ОЗ на каждом из ЗУ. В этом случае весь ПП оказывается разбитым на участки между ЗУ, закрытыми с обеих сторон ОЗ и находящимися под рабочими давлениями. В таком состоянии продуктопровод выдерживается в те-

чение времени ΔT_2 , необходимого для получения всех необходимых оценок. При отсутствии стационарных утечек на участках продуктопровода и отсутствии перетоков ЦП через ОЗ, расходы по всем ЗУ остаются равными нулю, а давления постоянными:

$$\forall j \in N : W_j = 0 \wedge P_j = \text{const}. \quad (8.2)$$

Равенство (8.2) означает, что при отсутствии утечек из ПП и завершении всех переходных процессов устанавливается равенство абсолютных давлений во всех ЗУ ПП:

$$\forall i, j \in N : P_i = P_j, \quad (8.3)$$

где $P_j = \tilde{P}_j + \tilde{h}_j^\Gamma$ — абсолютное давление в j -м ЗУ; \tilde{P}_j — измеренное значение избыточного давления в j -м ЗУ; \tilde{h}_j^Γ — измеренное значение геодезической отметки j -го ЗУ.

Равенство (8.3) позволяет оценить систематическую ошибку измерения геодезической отметки j -го ЗУ:

$$\Delta \tilde{h}_j^\Gamma = P_j - \tilde{P}_j + \tilde{h}_j^\Gamma \quad (8.4)$$

и учесть ее в дальнейших расчетах.

Если между n -м и $(n + 1)$ -м ЗУ имеется стационарная утечка ЦП через свищи или другие нарушения герметичности участка ПП, то давление на обоих концах этого участка с течением времени будет падать. Скорость падения давления определяется величиной стационарной утечки, а численные значения абсолютных значений величин падения давления на обоих концах ПП позволяют определить ее местоположение.

Уравнение статического баланса по форме совпадает с уравнением динамического баланса, а его использование позволяет оценить величину стационарной утечки на интервале времени $[t_1, t_2]$.

8.2. Метод динамического баланса

Уравнение динамического баланса ЦП на участке ПП между n -м и $(n + 1)$ -м ЗУ для заданного интервала времени $[t_1, t_2]$ имеет вид:

$$M_n^{(3Y)}(t_1, t_2) - M_{n+1}^{(3Y)}(t_1, t_2) + M_{n,n+1}^{(\text{тр})}(t_1) - M_{n,n+1}^{(\text{тр})}(t_2) - M_{n,n+1}^{(\text{yт})}(t_1, t_2) = 0, \quad (8.5)$$

где $M_n^{(3Y)}(t_1, t_2)$, $M_{n+1}^{(3Y)}(t_1, t_2)$ — массы ЦП, прошедшего соответственно через n -й и $(n + 1)$ -й ЗУ за интервал времени $[t_1, t_2]$; $M_{n,n+1}^{(\text{тр})}(t_1)$, $M_{n,n+1}^{(\text{тр})}(t_2)$ — массы ЦП, аккумулированного внутри трубопровода между n -м и $(n + 1)$ -м ЗУ соответственно в моменты времени t_1 и t_2 ; $M_{n,n+1}^{(\text{yт})}(t_1, t_2)$ — масса потерь ЦП на участке ПП между n -м и $(n + 1)$ -м ЗУ за интервал времени $[t_1, t_2]$. Потери ЦП могут возникать за счет технологических отборов, стационарных утечек или криминальных отборов.

Масса ЦП, прошедшего через вход или через выход n -го участка ПП за интервал времени $[t_1, t_2]$, определяется по формуле:

$$M_i^{(3Y)}(t_1, t_2) = \int_{t_1}^{t_2} W_i(t) dt, \quad i = n, n+1 \quad (8.6)$$

Массу ЦП, аккумулированного на n -м участке ПП в момент времени t_j , где $(j = 1, 2)$, необходимо определять с учетом продольных и поперечных деформаций участка трубопровода по следующей формуле:

$$M_{n,n+1}^{(TP)}(t_j) = \int_{X_n}^{X_{n+1}} \rho_0(x, t_j) k_p(x, t_j) S_0(x) k_S(x, t_j) k_L(x, t_j) dx, \quad j = 1, 2, \quad (8.7)$$

где

$$k_L(x, t) = 1 + \alpha_L(x) \cdot (T(x, t) - T_0); \quad (8.8)$$

$k_L(x, t)$ — коэффициент продольного расширения участка трубопровода в точке x в момент времени t .

Результаты измерений физических параметров ЦП, полученные в ЗУ ПП, позволяют рассчитать численные оценки масс $\hat{M}_n^{(3Y)}(t_1, t_2)$, $\hat{M}_{n+1}^{(3Y)}(t_1, t_2)$, $\hat{M}_{n,n+1}^{(TP)}(t_1)$, $\hat{M}_{n,n+1}^{(TP)}(t_2)$, а также СКО данных оценок — $\hat{\sigma}_{M_n^{(3Y)}}(t_1, t_2)$, $\hat{\sigma}_{M_{n+1}^{(3Y)}}(t_1, t_2)$, $\hat{\sigma}_{M_{n,n+1}^{(TP)}}(t_1)$, $\hat{\sigma}_{M_{n,n+1}^{(TP)}}(t_2)$ соответственно.

Оценка дисбаланса и СКО данной оценки рассчитываются по формулам:

$$\hat{M}_{n,n+1}^{(\Delta\delta)}(t_1, t_2) = \hat{M}_n^{(3Y)}(t_1, t_2) - \hat{M}_{n+1}^{(3Y)}(t_1, t_2) + \hat{M}_{n,n+1}^{(TP)}(t_1) - \hat{M}_{n,n+1}^{(TP)}(t_2), \quad (8.9)$$

$$\hat{\sigma}_{M_{n,n+1}^{(\Delta\delta)}}^2(t_1, t_2) = \hat{\sigma}_{M_n^{(3Y)}}^2(t_1, t_2) + \hat{\sigma}_{M_{n+1}^{(3Y)}}^2(t_1, t_2) + \hat{\sigma}_{M_{n,n+1}^{(TP)}}^2(t_1) + \hat{\sigma}_{M_{n,n+1}^{(TP)}}^2(t_2). \quad (8.10)$$

Оценка дисбаланса включает в себя три составляющие:

$$\hat{M}_{n,n+1}^{(\Delta\delta)}(t_1, t_2) = \hat{M}_{n,n+1}^{(YT)}(t_1, t_2) + \Delta_{M_{n,n+1}^{(YT)}}(t_1, t_2) + \xi_{M_{n,n+1}^{(YT)}}(t_1, t_2), \quad (8.11)$$

где $\Delta_{M_{n,n+1}^{(YT)}}(t_1, t_2)$ — систематическая ошибка дисбаланса, обусловленная систематическими ошибками результатов измерений массового расхода и плотности, что выражается формулой

$$\Delta_{M_{n,n+1}^{(YT)}}(t_1, t_2) = (t_2 - t_1) \cdot (\Delta_{W_n} - \Delta_{W_{n+1}}); \quad (8.12)$$

$\xi_{M_{n,n+1}^{(YT)}}(t_1, t_2)$ — случайная ошибка величины дисбаланса, обусловленная суммарным влиянием случайных ошибок результатов измерений физических параметров ЦП в ЗУ, причем

$$\xi_{M_{n,n+1}^{(YT)}}(t_1, t_2) \sim N(0, \sigma_{M_{n,n+1}^{(\Delta\delta)}}^2(t_1, t_2)).$$

Оценка систематической ошибки дисбаланса $\hat{\Delta}_{M_{n,n+1}^{(YT)}}(t_1, t_2)$ может быть рассчитана по формуле

(8.11) при известных оценках систематических ошибок измерений массового расхода $\hat{\Delta}_{W_n}$ и $\hat{\Delta}_{W_{n+1}}$. Если же систематические ошибки измерений массового расхода неизвестны или оценены с недостаточной точностью, то в качестве оценки систематической ошибки дисбаланса $\hat{\Delta}_{M_{n,n+1}^{(YT)}}(t_1, t_2)$ можно выбрать среднее значение дисбаланса в серии экспериментов по расчету баланса на последовательных интервалах времени при отсутствии криминального отбора ЦП.

Критерий отсутствия утечек на участке ПП между n -м и $(n+1)$ -м ЗУ на интервале времени $[t_1, t_2]$ имеет вид:

$$\left| \hat{M}_{n,n+1}^{(\Delta\delta)}(t_1, t_2) - \hat{\Delta}_{M_{n,n+1}^{(YT)}}(t_1, t_2) \right| \leq k \cdot \hat{\sigma}_{M_{n,n+1}^{(\Delta\delta)}}(t_1, t_2), \quad (8.13)$$

где $k > 0$ — пороговое значение критерия, численное значение которого определяется в результате обучения системы.

Если на интервале времени $[t_1, t_2]$ условие (8.13) выполняется, то булевой переменной M_1 присваивается значение, равное нулю, а если не выполняется, то M_1 присваивается значение, равное единице. Если $M_1 = 1$, то это означает, что величина дисбаланса не может быть объяснена чисто случайными совпадениями и на интервале времени $[t_1, t_2]$ появилась утечка или криминальный отбор.

Рассматривая последовательность интервалов $[t_1, t_2], [t_2, t_3], \dots, [t_k, t_{k+1}]$ и сопоставляя ей значения логических переменных P_1, P_2, \dots, P_k , вычисляемые в соответствии с (8.13), можно утверждать, что, если криминальный отбор осуществлялся в течение всей последовательности интервалов от $[t_1, t_2]$ до $[t_n, t_{n+1}]$, то и значение логической переменной $M = M_1 \wedge M_2 \wedge \dots \wedge M_n$ будет равно единице.

Численное значение n зависит от длительности и стратегий осуществления криминального отбора.

Если условие (8.13) выполняется, то оценка массы потерь ЦП в результате криминального отбора на интервале времени оценивается по формуле:

$$\hat{M}_{n,n+1}^{(YT)}(t_1, t_2) = \hat{M}_{n,n+1}^{(\Delta\delta)}(t_1, t_2) + \hat{\Delta}_{M_{n,n+1}^{(YT)}}(t_1, t_2). \quad (8.14)$$

8.3. Метод знаков приращений расходов

Относится к прямым методам обнаружения утечек и криминальных отборов. Основан на том факте, что, если на интервале времени $[t_1, t_2]$ между n -м и $(n+1)$ -м ЗУ осуществляется криминальный отбор с расходом $Q_0^{(YT)}(t)$, то количество ЦП, прошедшего через n -й ЗУ, увеличивается, а через $(n+1)$ -й ЗУ — уменьшается. При этом объем потерь ЦП равен

$$V_0^{(YT)}(t_1, t_2) = \int_{t_1}^{t_2} V_0^{(YT)}(t) dt,$$

а количества ЦП, прошедшего через n -й и $(n + 1)$ -й ЗУ за интервал $[t_1, t_2]$, могут быть оценены по формуле

$$\hat{V}_{0j}^{(3Y)}(t_1, t_2) = \int_{t_1}^{t_2} \hat{Q}_{0j}(t) dt, \quad j = n, n + 1,$$

где \hat{Q}_{0j} — оценка приведенного объемного расхода в j -м ЗУ в момент времени t , полученная на основе прямых измерений параметров ЦП в соответствующем ЗУ.

В квазистационарном режиме с отбором имеем

$$\hat{V}_{0n}^{(3Y)}(t_1, t_2) - \hat{V}_{0n+1}^{(3Y)}(t_1, t_2) \approx V_0^{(Yr)}(t_1, t_2).$$

В нестационарных режимах, вызванных возникновением, изменением интенсивности или окончанием отбора, имеет место следующее неравенство:

$$0 < \hat{V}_{0n}^{(3Y)}(t_1, t_2) - \hat{V}_{0n+1}^{(3Y)}(t_1, t_2) < V_0^{(Yr)}(t_1, t_2).$$

Это свойство позволило построить индикаторную функцию наличия криминального отбора, достаточно устойчивую к уровню помех и характеру режима транспорта ЦП:

$$P_1 = \text{sgn}[\hat{V}_{0n}^{(3Y)}(t_1, t_2) - \hat{V}_{0n+1}^{(3Y)}(t_1, t_2)]. \quad (8.15)$$

Функция $\text{sgn}(x)$ принимает значение 1, если $x > 0$, и значение 0, если $x \leq 0$. При наличии криминального отбора аргумент этой функции в выражении (8.15) не зависит от используемой стратегии отбора и, следовательно, логическая переменная P_1 принимает значение 1, а при его отсутствии $P_1 = 0$.

Рассматривая последовательность интервалов $[t_1, t_2], [t_2, t_3], \dots, [t_k, t_{k+1}]$ и сопоставляя ей значения логической переменной P_1, P_2, \dots, P_k , вычисляемые в соответствии с (8.15), можно утверждать, что, если криминальный отбор осуществлялся в течение всей последовательности интервалов от $[t_1, t_2]$ до $[t_k, t_{k+1}]$, то и значение логической переменной $P = P_1 \wedge P_2 \wedge \dots \wedge P_k$ будет равно единице.

Численные значения k и длительности интервалов $[t_k, t_{k+1}]$ зависят от технических требований к ИСОКО и стратегий отбора, используемых криминальными группами, и определяются в результате обучения (настройки) системы.

8.4. Метод знаков приращений давлений

Относится к косвенным методам обнаружения утечек или криминальных отборов и основан на сравнении результатов расчета параметров газовых потоков на основе ММСННР с результатами измерений этих параметров.

Пусть в момент времени t имеем результаты измерений $\hat{P}_n^t, \hat{W}_n^t, \hat{p}_n^t, \hat{T}_n^t$ с n -го ЗУ и результаты измерений $\hat{P}_{n+1}^t, \hat{W}_{n+1}^t, \hat{p}_{n+1}^t, \hat{T}_{n+1}^t$ с $(n + 1)$ -го ЗУ.

В качестве начальных условий выберем результаты расчета ММСННР в момент времени $t - 1$:

$$\hat{P}(x, t-1), \hat{Q}_0(x, t-1), \hat{T}(x, t-1), \quad x \in [X_n, X_{n+1}].$$

В качестве граничных условий выберем следующие измерения в момент времени t : $\hat{P}_n^t, \hat{Q}_{0n+1}^t, \hat{T}_n^t$, где $\hat{Q}_{0n+1}^t = \hat{W}_{n+1}^t / \hat{p}_{0n+1}^t$.

Приведенный объемный расход ЦП на n -м ЗУ вычисляем, как $\hat{Q}_{0n}^t = \hat{W}_n^t / \hat{p}_{0n}^t$. В результате расчета ММСННР получим следующие оценки:

$$\hat{P}(x, t), \hat{Q}_0(x, t), \hat{T}(x, t), \quad x \in [X_n, X_{n+1}].$$

Вычисляем значения индикаторных функций

$$D_0 = \text{sgn}(\hat{P}(x_{n+1}, t) - \hat{P}_{n+1}^t), \quad (8.16)$$

$$Q_0 = \text{sgn}(\hat{Q}_{0n}^t - \hat{Q}_0(x_n, t)). \quad (8.17)$$

Если на интервале времени $[t_1, t_2]$ между n -м и $(n + 1)$ -м ЗУ осуществляется криминальный отбор, то, начиная с некоторого момента времени $t_0 \in [t_1, t_2]$, аргументы функций (8.16), (8.17) будут положительными и логические переменные D_0 и Q_0 принимают значения, равные 1.

Индикатором наличия криминального отбора служит единичное значение булевой переменной $R_0 = D_0 \wedge Q_0$.

Если криминальный отбор осуществляется в последовательные моменты времени $t_0, t_1, t_2, \dots, t_l$, то и значение булевой переменной $R = R_0 \wedge R_1 \wedge R_2 \wedge \dots \wedge R_l$ будет также равно единице.

9. Процедура принятия решения о наличии криминального отбора

Полученные оценки и факты позволяют построить эффективную процедуру принятия решения о наличии криминального отбора.

Введем гипотезы: H_0 — нет отбора, при альтернативе H_1 — есть отбор.

Рассмотрим условные вероятности:

$P(H_0 / H_0)$ — вероятность правильного решения об отсутствии отбора;

$P(H_1 / H_1)$ — вероятность правильного решения о наличии отбора;

$P(H_0 / H_1)$ — вероятность ошибки второго рода, т. е. вероятность пропуска утечки: принимается гипотеза H_0 (нет отбора), в то время как верна гипотеза H_1 (есть отбор);

$P(H_1 / H_0)$ — вероятность ошибки первого рода, т. е. вероятность ложной тревоги: принимается гипотеза H_1 (есть отбор), в то время как верна гипотеза H_0 (нет отбора).

В ИСОКО в качестве критериев оценки эффективности процедуры принятия решения о наличии/отсутствии криминальных отборов используются группы критериев, включающих: минимум вероятностей ошибок первого и второго рода, т. е. минимум вероятности ложных тревог и минимум вероятности пропуска (необнаружения) криминального отбора; минимум математического ожидания времени обнаружения криминального отбора с момента его возникновения; минимум математического ожидания величины криминального отбора, обнаруживаемого системой; минимум дисперсии

ошибки определения места криминального отбора по трассе ЦП.

Все рассмотренные критерии практически не имеют области «согласия» и «оптимальное решение» всегда находится в области Парето. Математическая постановка задачи принятия решения в ИСОКО и разработка формализованного метода ее решения оказались неоправданными ввиду необходимости введения большого количества допущений и предположений, которые далеко не всегда подтверждаются практикой. Поэтому в ИСОКО используется самонастраивающаяся по параметрам система принятия решений.

Принятие решения о наличии криминального отбора осуществляется в соответствии с правилом

$$KO = R \wedge P \wedge M.$$

Механизм обучения системы и оценивания (адаптации) параметров n, k, l к используемым или гипотетически возможным стратегиям криминального отбора в данной работе не излагается.

Если логическая переменная KO принимает значение, равное единице, то в ИСОКО формируется и выдается оператору звуковой сигнал и специальная форма о возникновении на участке ЦП между n -м и $(n + 1)$ -м ЗУ утечки или криминального отбора.

10. Определение оценок параметров криминального отбора

Под параметрами криминального отбора будем понимать следующие величины: время начала, местоположение, текущая интенсивность, суммарный объем потерь, время окончания.

В ИСОКО реализованы три метода определения параметров криминального отбора:

- грубый метод, основанный на анализе информации о переходных процессах;
- метод, основанный на ММКСНР;
- метод, основанный на ММСННР.

10.1. Определение грубых оценок параметров криминального отбора

Одной из важных характеристик ИСОУ является величина задержки между временем появления и временем обнаружения криминального отбора.

В момент возникновения криминального отбора волны возмущения давления и расхода распространяются в обе стороны со скоростью звука в среде ЦП. Фиксируя моменты времени прихода фронта волны в n -й и $(n + 1)$ -й ЗУ, можно получить грубые оценки местоположения утечки и времени ее начала:

$$X_{YT} = \frac{1}{2} \cdot (c \cdot (t_n - t_{n+1}) + L_{n,n+1}),$$

$$t_{YT} = \frac{1}{2} \cdot \left(t_n + t_{n+1} - \frac{L_{n,n+1}}{c} \right),$$

где c — скорость распространения звука в ЦП (устанавливается экспериментально), t_n — время прихода волны в n -й ЗУ, t_{n+1} — время прихода волны

в $(n + 1)$ -й ЗУ, $L_{n,n+1}$ — длина участка ЦП между n -м и $(n + 1)$ -м ЗУ.

Моменты t_n и t_{n+1} прихода волны возмущения давления и расхода в n -й и $(n + 1)$ -й ЗУ определяем с помощью критерия отношения правдоподобия, описанного в п. 4.4.

10.2. Оценивание параметров криминального отбора на основе ММКСНР

Если в момент времени t выполняются условия квазистационарности режима в n -м и $(n + 1)$ -м ЗУ (п. 4.3), то, исходя из уравнений ММКСНР (4.16), оценку расхода утечки получим по формуле

$$\hat{Q}_{YT}^t = \hat{Q}_n^t - \hat{Q}_{n+1}^t,$$

а оценка местоположения утечки \hat{X}_{YT}^t , вычисляемая в момент времени t , определяется в результате численного решения следующего уравнения:

$$f(X_{YT}^t) =$$

$$= \left((\hat{Q}_{0n}^t)^2 \cdot (X_{YT}^t - X_n) + (\hat{Q}_{0n+1}^t)^2 \cdot (X_{n+1} - X_{YT}^t) \right) \times$$

$$\times \frac{\lambda \hat{\rho}_0^{(cp)}}{2 \hat{E}^2 \hat{k}_S^{(cp)2} \hat{k}_P^{(cp)} d_0 S_0^2} + \hat{\rho}_0^{(cp)} \hat{k}_P^{(cp)} g \times$$

$$\times (z(X_{n+1}) - z(X_n)) - (\tilde{P}_n^t - \tilde{P}_{n+1}^t) = 0, \quad (10.1)$$

где $\tilde{P}_n^t, \tilde{P}_{n+1}^t$ — результаты измерений давления в момент времени t в n -м и $(n + 1)$ -м ЗУ; $\hat{Q}_{0n}^t, \hat{Q}_{0n+1}^t$ — косвенные измерения (оценки) приведенного объемного расхода в момент времени t в n -м и $(n + 1)$ -м ЗУ, полученные на основе результатов прямых измерений $\tilde{W}_n^t, \tilde{p}_n^t; \hat{\rho}_0^{(cp)}, \hat{k}_S^{(cp)}, \hat{k}_P^{(cp)}$ — оценки средних значений соответственно приведенной плотности, коэффициента сжатия ЦП и коэффициента поперечного расширения трубопровода для данного участка ЦП, рассчитанные на основе результатов прямых измерений в ходе оценивания параметров ММСННР и ММКСНР (п. 5).

Уточнение местоположения криминального отбора осуществляется в течение всего времени его существования $t, t + 1, t + 2, t + 3, \dots$ при выполнении условия квазистационарности режима в соответствии с рекуррентным соотношением:

$$\hat{X}_{YT}^{t+n-1} = \hat{X}_{YT}^{t+n-2} - \frac{1}{n[\hat{X}_{YT}^{t+n-2} - \hat{X}_{YT}^{t+n-1}]}, \quad (10.2)$$

$$n = 1, 2, 3, \dots,$$

где \hat{X}_{YT}^{t+n-1} — уточненная оценка местоположения криминального отбора на n -м шаге; \hat{X}_{YT}^{t+n-1} — текущая оценка местоположения криминального отбора, вычисленная в момент времени $t + n - 1$ в результате решения уравнения (10.1).

10.3. Оценивание параметров утечки на основе ММСННР

Если криминальный отбор осуществляется на фоне существенно нестационарного режима, то оценки (10.1), (10.2) могут оказаться неэффективными, т. е. иметь завышенное значение дисперсии. Для повышения точности (уменьшения дисперсии)

оценки местоположения криминального отбора в этом случае используется ММСННР транспорта ЦП в магистральных ПП.

Перед использованием метода должен быть установлен факт наличия криминального отбора и время его возникновения $t_{\text{YT}}^{\text{нач}}$. Функцию интенсивности криминального отбора в момент времени t будем оценивать следующим образом:

$$\hat{Q}_{\text{YT}}^t = \hat{Q}_n^{t+\Delta t_n} - \hat{Q}_{n+1}^{t+\Delta t_{n+1}}, t \in [t_{\text{YT}}^{\text{нач}}, t_{\text{YT}}^{\text{кон}}],$$

где $\Delta t_n = t_n - t_{\text{YT}}^{\text{нач}}$, $\Delta t_{n+1} = t_{n+1} - t_{\text{YT}}^{\text{нач}}$, t_n — время прихода волны в n -й ЗУ, t_{n+1} — время прихода волны в $(n+1)$ -й ЗУ.

Известно, что координата точки места отбора X_{YT} разбивает участок, на котором он происходит, на две части:

- участок трубопровода до местоположения отбора $[X_n, X_{\text{YT}}]$;
- участок трубопровода после местоположения отбора $[X_{\text{YT}}, X_{n+1}]$.

Тогда ММСННР с отбором на участке ПП $[X_n, X_{n+1}]$ включает в себя:

- систему дифференциальных уравнений вида (4.10) для первой части участка при $x \in [X_n, X_{\text{YT}}]$, $t \in [t_{\text{нач}}, t_{\text{кон}}]$;
- систему дифференциальных уравнений вида (4.10) для второй части участка при $x \in [X_{\text{YT}}, X_{n+1}]$, $t \in [t_{\text{нач}}, t_{\text{кон}}]$;
- условия согласования параметров ЦП в точке X_{YT} :

$$P_n(X_{\text{YT}}, t) = P_{n+1}(X_{\text{YT}}, t), T_n(X_{\text{YT}}, t) = T_{n+1}(X_{\text{YT}}, t),$$

$$Q_{0n}(X_{\text{YT}}, t) = Q_{0n+1}(X_{\text{YT}}, t) + Q_{\text{YT}}(t), t \in [t_{\text{нач}}, t_{\text{кон}}];$$

- начальные условия:

$$P(x, t_{\text{нач}}) = \hat{P}^{(\text{нач})}(x), Q_0(x, t_{\text{нач}}) = \hat{Q}_0^{(\text{нач})}(x),$$

$$T(x, t_{\text{нач}}) = \hat{T}^{(\text{нач})}(x), x \in [X_n, X_{n+1}];$$

- граничные условия:

$$P(X_n, t) = \tilde{P}_n^t, T(X_n, t) = \tilde{T}_n^t, Q_0(X_{n+1}, t) = \tilde{Q}_{0n+1}^t,$$

$$t \in [t_{\text{нач}}, t_{\text{кон}}].$$

Уточненную оценку местоположения криминального отбора на участке ПП между n -м и $(n+1)$ -м ЗУ получаем в результате решения следующей оптимизационной задачи:

$$F(X_{\text{YT}}) = \sum_{t_{\text{нач}} \leq t \leq t_{\text{кон}}} \frac{1}{\hat{\sigma}_{Q_{0n}}^2} \left(\hat{Q}_{0n}^t - \hat{Q}_{0n}^t(X_{\text{YT}}) \right)^2 + \frac{1}{\hat{\sigma}_{P_{n+1}}^2} \left(\tilde{P}_{n+1}^t - \hat{P}_{n+1}^t(X_{\text{YT}}) \right)^2 \rightarrow \min_{X_n \leq X_{\text{YT}} \leq X_{n+1}},$$

где $\hat{Q}_{0n}^t(X_{\text{YT}})$ и $\hat{P}_{n+1}^t(X_{\text{YT}})$ — модельные значения параметров ЦП, полученные на основе вышеизложенной ММСННР с криминальными отборами.

11. Оценка эффективности функционирования ИСОКО

Оценка эффективности функционирования ИСОКО была проведена на одном из конденсаторов ДК «Укргазвидобування» НАК «Нафтогаз України».

Результаты работы ИСОКО после обучения на специально спланированных и проведенных натуральных экспериментах с отборами приведены в таблице 1.

Таблица 1

| № | Участок (длина, км) | Оценки параметров отбора | | | Фактич. погрешность оценки места отбора, км |
|---|---------------------|----------------------------------|---------------------------------------|-----------------|---|
| | | Расстояние от начала участка, км | Средняя интенсивность отбора, т/ч (%) | Объем отбора, т | |
| 1 | ЗУ2—ЗУ3 (29,03) | 0,01 | 0,81 (1,2) | 0,33 | 0,01 |
| 2 | ЗУ2—ЗУ3 (29,03) | 0,01 | 0,75 (1,1) | 0,21 | 0,01 |
| 3 | ЗУ2—ЗУ3 (29,03) | 10,20 | 13,50 (19,6) | 1,98 | 0,40 |
| 4 | ЗУ2—ЗУ3 (29,03) | 9,98 | 2,60 (3,8) | 0,26 | 0,18 |
| 5 | ЗУ3—ЗУ4 (35,93) | 1,10 | 1,38 (2,0) | 0,32 | 1,10 |
| 6 | ЗУ3—ЗУ4 (35,93) | 0,10 | 0,54 (0,8) | 0,09 | 0,10 |
| 7 | ЗУ4—ЗУ5 (56,12) | 37,57 | 2,20 (3,2) | 0,15 | 1,23 |
| 8 | ЗУ4—ЗУ5 (56,12) | 40,04 | 0,79 (1,2) | 0,11 | 1,24 |
| 9 | ЗУ2—ЗУ3 (29,03) | 29,04 | 38,20 (61,1) | 4,91 | 0,01 |

Задержка по времени при обнаружении экспериментальных отборов не превышала длительность цикла опроса замерных узлов (2,5 мин).

На рис. 1–4 представлены графики измерений физических параметров стабильного конденсата в смежных ЗУ при различных стратегиях проведения криминальных отборов, которым ИСОКО в состоянии эффективно противодействовать.

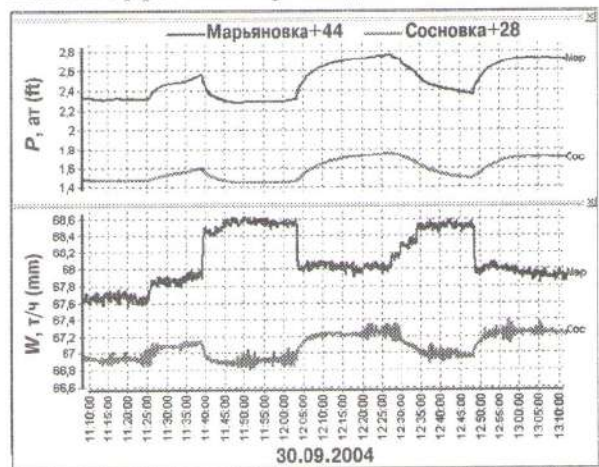


Рис. 1. Графики изменения давления (P) и массового расхода (W) в ближайших ЗУ при отборе с резким и с плавным открытием/закрытием вентили на врезке

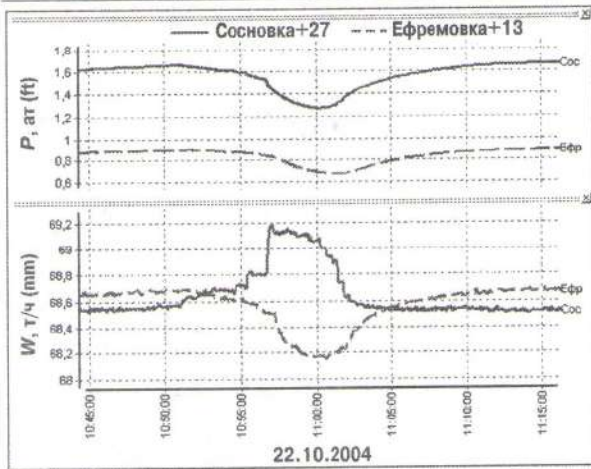


Рис. 2. Графики изменения давления (P) и массового расхода (W) в ближайших ЗУ при кратковременном отборе

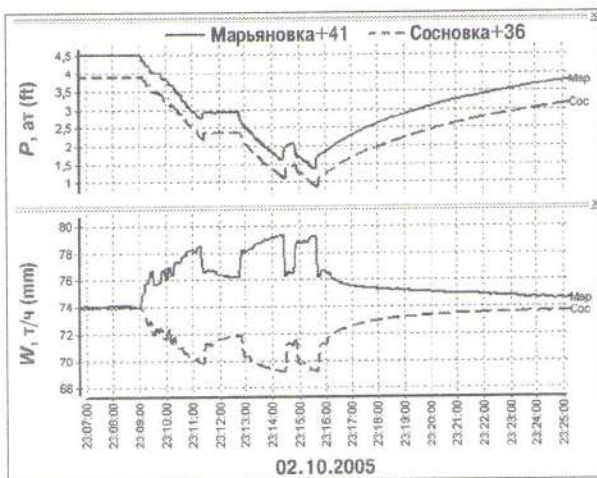


Рис. 3. Графики изменения давления (P) и массового расхода (W) в ближайших ЗУ при отборе с переменной интенсивностью

12. Заключение

Рассмотрена интеллектуальная система обнаружения криминальных отборов из продуктопроводов, в основу которой положены интеллектуальные алгоритмы автоматической обработки, анализа и принятия решений по оперативным данным с замерных узлов продуктопровода в режиме реального времени. Предлагаемые методы обнаружения криминальных отборов и оценивания их параметров базируются на использовании разработанной математической мо-

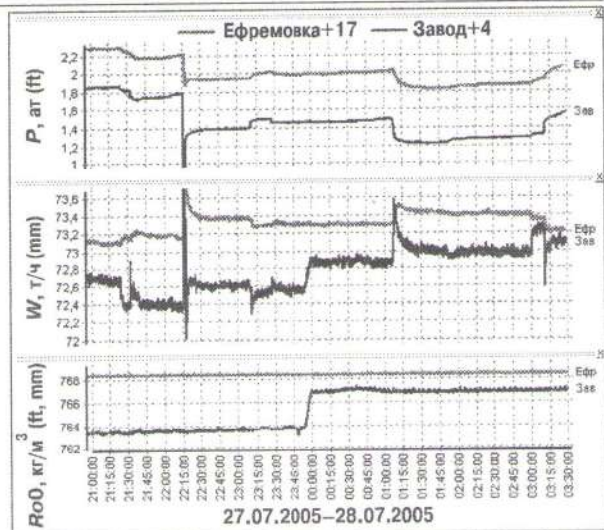


Рис. 4. Графики изменения давления (P), массового расхода (W) и приведенной плотности (RoD) в ближайших ЗУ при длительном отборе малой интенсивности на фоне различных технологических возмущений (изменения положения регулирующей задвижки, плотности поступающего ЦП)

дели квазистационарных и существенно нестационарных неизотермических режимов транспорта жидких продуктов по магистральным продуктопроводам. Анализ результатов работы ИСОКО при проведении экспериментальных отборов позволяет сделать вывод, что метрологические характеристики ИСОКО не только соответствуют, но и превышают заявленные.

Список литературы. 1. Евдокимов А. Г., Дубровский В. В., Тевяшев А. Д. Потокораспределение в инженерных сетях. – М.: Стройиздат, 1979. – 199 с. 2. Карелин В. Я., Минаев А. В. Насосы и насосные станции. – М.: Стройиздат, 1986. – 280 с. 3. Чарный И. А. Неустановившееся движение реальной жидкости в трубах. – М.: Недра, 1975. – 296 с. 4. Лурье М. В. Математическое моделирование процессов трубопроводного транспорта нефти, нефтепродуктов и газа. – М.: Изд-во «Нефть и газ» РГУ нефти и газа им. И. М. Губкина, 2003. – 336 с. 5. Самарский А. А. Теория разностных схем. – М.: Наука, 1989. 6. Обнаружение изменения свойств сигналов и динамических систем: Пер. с англ. / Под ред. М. Бассвиль, А. Банвениста. – М.: Мир, 1989. – 278 с.

Поступила в редколлегию 12.05.2006

УДК 519.6

В. П. Машталир, В. В. Шляхов

ИНДУЦИРОВАННАЯ СОГЛАСОВАННОСТЬ ОТНОШЕНИЙ В ЗАДАЧАХ ГРАНУЛЯЦИИ ИНФОРМАЦИИ

1. Введение

Развитие методов совместной регистрации, обработки и интерпретации в конкретной предметной области разнородных (количественных, качественных, недостаточно определенных, избыточных и т. п.) данных — пожалуй, единственный в настоящее время путь к созданию эффективных систем продукции знаний. В 1997 Л. А. Заде выделил три базовых компонента, лежащих в основе когнитивных подходов: грануляцию, организацию и причинно-следственные отношения. Он так характеризовал их: «Granulation involves decomposition of whole into parts, organization involves integration of parts into whole, and causation involves association of causes and effects» [1]. Первая составляющая (грануляция) — фактически построение разбиений (реже — покрытий) входной информации. Гранулы — группировки данных, получаемые с точки зрения их неразличимости или сходства, огрубляют или, напротив, детализируют с желаемой точностью представление информации [2]. При этом основной интерес представляют внутренние, внешние и контекстные свойства выделяемых множеств. Вторую составляющую можно трактовать как структуризацию классов эквивалентности (или в наиболее общем случае — толерантности), имеющую целью получить целостное представление о всей совокупности данных на различных стратах и их семействах [3]. Наконец, третья составляющая — некоторая продукция знаний, базирующаяся или на заданных правилах вывода, или на функциональных соотношениях вход-выход, или на выявлении внутренних неочевидных сущностных связей [4].

Методы теории грануляции (в оригинале — Granular Computing, далее будем говорить GrC-парадигма или просто GrC) — суть междисциплинарных исследований, до настоящего времени активно развивающихся в информационных системах (прежде всего на основе аппроксимационных множеств — rough sets [2, 5]), вычислительном интеллекте [6], базах данных с контекстным поиском [7], а также интервальном анализе, универсальном шкалировании, кластерных методах и т. д. [3, 4, 8–10]. Следует подчеркнуть, что до настоящего времени не дано даже достаточно строгое определение GrC-парадигмы. Основные идеи, подходы и принципы привлечены в GrC из различных отраслей знаний, поэтому, с одной стороны, не является неожиданностью, что ряд исследований представляют собой лишь перевод известных положений в сферу GrC, с другой — терминология и аксиоматика GrC-парадигмы находятся лишь в состоянии становления. Таким образом, на-

сущной является проблема обобщений, быть может, метаматематических, что создаст предпосылки для конструктивного развития данного направления.

В качестве одного из возможных «языков» GrC-парадигмы можно указать мультиалгебраические системы [11], позволяющие для различных математических структур синтезировать и идентифицировать алгебры, модели и алгебраические системы с носителями, объединяющими семейства множеств различной природы, а в конечном итоге — оперировать гранулами (классами эквивалентности), не различая отдельных представителей классов. Цель данной работы — развитие и обобщение результатов [12, 13] в плане исследования индуцированной согласованности отношений и операций, определяющих мультиалгебраические системы.

2. Постановка задачи

Рассмотрим внутренние эквивалентности, индуцируемые произвольными отношениями. Пусть n -арное отношение $E(x_1, \dots, x_n)$ задано на декартовом произведении $A_1 \times A_2 \times \dots \times A_n$, где A_i — произвольные множества. Зафиксируем k -е место (плечо) среди аргументов отношения E . Если для отношения E для каких-либо x_k и x'_k имеет место тождество

$$E(x_1, \dots, x_k, x_{k+1}, \dots, x_n) \equiv E(x_1, \dots, x'_k, x_{k+1}, \dots, x_n), \quad (1)$$

то очевидно, что элементы x_k и x'_k неразличимы с точки зрения этого отношения. Следовательно, можно считать, что элементы x_k и x'_k находятся в отношении эквивалентности P_k^E (здесь и далее отношения эквивалентности будем представлять предикатами, принимающими значение «1» на элементах одного фактора и «0» в противном случае), порожденном отношением E на k -м месте

$$\begin{aligned} P_k^E(x_k, x'_k) = 1 &\Leftrightarrow E(x_1, \dots, x'_k, x_{k+1}, \dots, x_n) \equiv \\ &\equiv E(x_1, \dots, x_k, x_{k+1}, \dots, x_n) \end{aligned} \quad (2)$$

на декартовом квадрате $A_k \times A_k$. Отношение P_k^E есть эквивалентность, поскольку выполнение свойств рефлексивности, симметричности и транзитивности очевидно. Таким образом, на каждом A_k задается разбиение, порожденное отношением E . Задача работы — поиск условий, при которых произвольные n -арные отношения порождают отношения на классах эквивалентностей, обеспечивающих анализ в рамках GrC-парадигмы.

3. Генезис мультиотношений

Допустим, для каких-либо номеров k и l из набора $\{1, 2, \dots, n\}$ имеет место $A_k = A_l$. Тогда области определений эквивалентностей P_k^E и P_l^E совпадают

и можно говорить о сравнении этих отношений, в том числе и их равенстве. На языке отношения E сравнение выглядит так:

$$\begin{aligned} \text{если } E(x_1, \dots, x_k, \dots, x_n) \equiv E(x_1, \dots, x'_k, \dots, x_n), \text{ то} \\ E(x_1, \dots, x_{l-1}, x_k, x_{l+1}, \dots, x_n) \equiv \\ \equiv E(x_1, \dots, x_{l-1}, x_k, x_{l+1}, \dots, x_n), \end{aligned} \quad (3)$$

т. е. эквивалентность P_k^E реализует более детальное разбиение множества $A_k = A_l$, чем отношение P_l^E . При этом важно понимать, что условие (3) должно выполняться для любых элементов $x_k, x'_k \in A_k$, а не для некоторых пар (в противном случае разбиения будут различны).

Если условие (3) имеет место для любых пар, то P_k^E как отношение, реализующее более детальное разбиение, чем отношение P_l^E , можно считать вложенным и обозначать $P_k^E \subseteq P_l^E$. Если же условие (3) выполняется в обоих направлениях и для любых пар, то эквивалентности P_k^E и P_l^E совпадают. Таким образом, номера мест k и l образуют множество $\{1, 2, \dots, n\}$, на котором исходным отношением E индуцируется функция F , принимающая четыре значения (например, $\{-1, 0, 1, 2\}$)

$$F(k, l) = \begin{cases} -1, \text{ если } A_k = A_l \text{ и } P_k^E \subseteq P_l^E; \\ 0, \text{ если } A_k \neq A_l \text{ или } P_k^E \not\subseteq P_l^E; \\ 1, \text{ если } A_k = A_l \text{ и } P_k^E = P_l^E; \\ 2, \text{ если } A_k = A_l \text{ и } P_k^E \supseteq P_l^E, \end{cases} \quad (4)$$

где « $\not\subseteq$ » означает несравнимость отношений. Зависимость (4) можно представить утверждением 1.

Утверждение 1. Если множества A_k, A_l не равны, то $F(k, l) = 0$, если же $A_k = A_l$, то значение $F(k, l)$ зависит от вида выполнения условия (3): если (3) имеет место в обе стороны, причем для любой пары элементов множества $A_k = A_l$, то $F(k, l) = 1$; если условие (3) выполняется непосредственно, то $F(k, l) = -1$; если же оно выполняется в обратную сторону, то $F(k, l) = 2$; наконец, если условие (3) нарушается, то $F(k, l) = 0$.

Доказательство. Действительно, при условии $A_k \neq A_l$, согласно определению, $F(k, l) = 0$. Если $A_k = A_l$ и для любых двух аргументов $x_k, x'_k \in A_k = A_l$ имеет место условие (3), то из (2) следует, что как только $P_k^E(x_k, x'_k) = 1$, следует и выполнение отношения $P_l^E(x_k, x'_k) = 1$. Это означает, что разбиение, соответствующее эквивалентности P_k^E , более детальное, чем разбиение, соответствующее эквивалентности P_l^E , т. е. $P_k^E \subseteq P_l^E$, и по определению получаем, что $F(k, l) = -1$. Выполнение условия (3) в обратном направлении влечет за собой противоположное вложение эквивалентностей $P_k^E \supseteq P_l^E$, значит, $F(k, l) = 2$. Наконец, если условие (3) выполняется в обе стороны, то в этом случае эквивалентности (и разбиения) совпадают, значит, $F(k, l) = 1$. Утверждение доказано.

Заметим, что если $F(k, l) = 1$, то места k и l для отношения E не отличаются: на этих местах исходным отношением индуцируется одно и то же разбиение, а следовательно, одни и те же эквивалентности, что в дальнейшем будет представлять наибольший интерес.

Определение 1. Произвольное n -арное отношение E , заданное на $A_1 \times A_2 \times \dots \times A_n$, будем называть внутренне (k, l) -согласованным тогда и только тогда, когда $A_k = A_l$ и $P_k^E = P_l^E$.

Отметим, что с точки зрения определения 1 исходное (произвольное) n -арное отношение E порождает эквивалентность V_E на $\{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$

$$V_E(k, l) = 1 \Leftrightarrow A_k = A_l \text{ и } P_k^E = P_l^E, \quad (5)$$

т. е. $V_E(k, l) = 1$ тогда и только тогда, когда отношение E внутренне (k, l) -согласовано (проверка рефлексивности, симметричности и транзитивности отношения V_E не составляет труда). Если это отношение представить в виде матрицы, которую в дальнейшем мы будем называть матрицей внутренней согласованности и обозначать $J(E)$, то она, естественно, будет состоять из 0 и 1, будет симметрична и на главной диагонали всегда содержать 1. В частных случаях, когда $J(E)$ — единичная матрица, то внутренняя согласованность отсутствует, когда же она состоит только из единиц, то все носители совпадают: $A_1 = A_2 = \dots = A_n$ и $P_1^E = P_2^E = \dots = P_n^E = P^E$, т. е. отношение E задано на A^n и порождает на A^2 одну и ту же эквивалентность P^E , классы которой для E неразличимы.

Если вычеркиванием одних и тех же строк и столбцов в $J(E)$ удастся выделить $(r \times r)$ -подматрицу, состоящую из одних 1, то можно говорить о внутренней (k_1, \dots, k_r) -согласованности отношения E в том смысле, что $A_{k_1} = \dots = A_{k_r}$ и $P_{k_1}^E = P_{k_2}^E = \dots = P_{k_r}^E$. Путем перенумерации переменных исходного отношения матрицу внутренней согласованности $J(E)$ всегда можно привести к блочно-диагональному виду

$$J^*(E) = \left(\begin{array}{cccccccccccc} 1 & \dots & 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \vdots & & \vdots & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 1 & \dots & 1 & & & & & & & & & \vdots \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots & 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & & \vdots & 1 & \dots & 1 & & & & & & \vdots \\ \vdots & & \vdots & 0 & \dots & 0 & \ddots & \dots & \dots & \dots & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & 1 & \dots & 1 & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 & & \vdots \end{array} \right) \left. \begin{array}{l} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \right\} \begin{array}{l} r_1 \\ r_2 \\ \vdots \\ r_p \end{array}, \quad (6)$$

где $r_1 + \dots + r_p = n$,

$$\left\{ \begin{array}{l} A_1 = A_2 = \dots = A_{r_1} = B_1; \\ A_{r_1+1} = A_{r_1+2} = \dots = A_{r_1+r_2} = B_2; \\ \dots \\ A_{n-r_p+1} = \dots = A_n = B_p \\ P_1^E = \dots = P_{r_1}^E = L_1^E; \\ P_{r_1+1}^E = \dots = P_{r_1+r_2}^E = L_2^E; \\ \dots \\ P_{n-r_p+1}^E = \dots = P_n^E = L_p^E \end{array} \right\} \quad (7)$$

и отношение E по сути определено на $B_1^{r_1} \times \dots \times B_p^{r_p}$ и порождает p различных эквивалентностей L_i^E , заданных на $B_i \times \dots \times B_i$, $i = 1, 2, \dots, p$. Следует заметить, что это возможно, когда исходное отношение инвариантно относительно произвольной подстановки из n элементов.

Отношение V_E и, соответственно, матрица согласованности несут в себе далеко не всю информацию об исходном отношении E , и главное — о тех эквивалентностях или разбиениях, которые оно индуцирует, что нетрудно увидеть из анализа (4). По сути, эквивалентность V_E не учитывает «детализацию» и вложенность разбиений, которые индуцирует отношение E . Главное, что оно позволяет, — это сравнивать носители на различных плечах сравнения. Однако на практике — в частности, при грануляции информации — вложенность разбиений может играть принципиальную роль. В качестве типичных примеров следует указать кластеризацию данных, сегментацию изображений. Достижение оптимальной в определенном смысле детализации создает предпосылки для правильной и полной автоматической интерпретации. Учитывая это, введем $(n \times n)$ -матрицу общей внутренней согласованности $F(E)$, элементы которой $f_{kl} = F(k, l)$ также определяются равенством (4). Как и матрица внутренней согласованности $J(E)$, она будет содержать 0 и 1 на тех же местах, но с некоторым изменением. Когда в равенстве (4) будут выполняться условия (2) и (3), то 0 и 1 сохранятся, если же будет иметь место условие (1), то будет стоять -1 , а при условии (4), соответственно, 2. При этом матрица $F(E)$, в определенном смысле, является симметрической. Относительно главной диагонали симметрия для 0 и 1 останется обычной, а вот для -1 и 2 она будет такой: симметрично -1 будет стоять 2 и наоборот.

Многообразие ситуаций, возникающих с произвольными отношениями и индуцируемыми ими эквивалентностями, рассмотренными случаями, конечно же, не исчерпывается. Например, можно говорить о двух произвольных отношениях E_1 , заданном на $A_1 \times \dots \times A_n$, и E_2 , заданном на произведении $C_1 \times \dots \times C_m$, где A_i и C_j ($i = 1, \dots, n$, $j = 1, \dots, m$) — наборы произвольных множеств. Допустим, для двух номеров i и j имеет место $A_i = C_j$, тогда, как и ранее, можно говорить о совпадении или вложении эквивалентностей $P_i^{E_1}$, $P_j^{E_2}$. На языке E_1 и E_2 это означает, что для произвольных элементов $x_i, x'_i \in A_i = C_j$ выполняется условие

$$\begin{aligned} E_1(x_1, \dots, x_i, \dots, x_n) &\equiv E_2(x_1, \dots, x'_i, \dots, x_n) \Leftrightarrow \\ E_1(y_1, \dots, y_{j-1}, x_i, y_{j+1}, \dots, y_m) &\equiv \\ \equiv E_2(y_1, \dots, y_{j-1}, x'_i, y_{j+1}, \dots, y_m). \end{aligned} \quad (8)$$

Если имеет место (8), то эквивалентности совпадают, т. е. $P_i^{E_1} = P_j^{E_2}$ и можно говорить о (i, j) -согласованности отношений E_1, E_2 . Если же условие (8) выполняется в одну сторону, например слева направо, то разбиение, соответствующее эквивалентности

$P_i^{E_1}$, более детальное, чем разбиение, соответствующее эквивалентности $P_j^{E_2}$, т. е. $P_i^{E_1} \subseteq P_j^{E_2}$ и наоборот. Аналогично может быть введено отношение V_{E_1, E_2} на произведении $\{1, \dots, n\} \times \{1, \dots, m\}$ и матрица согласованности $J(E_1, E_2)$ размером $n \times m$, которая состоит из 0 и 1 и обладает свойством: на (i, j) месте стоит 1 тогда и только тогда, когда имеет место (i, j) -согласованность отношений E_1 и E_2 .

Заметим, что матрицу согласованности двух отношений формально можно всегда считать квадратной. Действительно, пусть для определенности $n < m$, тогда прямое произведение $A_1 \times \dots \times A_n$ можно дополнить до размерности m прямым произведением пустых множеств, т. е. $A_1 \times A_2 \times \dots \times A_n \times \underbrace{\emptyset \times \dots \times \emptyset}_{m-n}$.

Матрица согласованности в этом случае будет иметь вид

$$J(E_1, E_2) = \left(\begin{array}{c|ccc} \dots & \vdots & \dots & \dots & \dots \\ \dots & \vdots & \dots & \dots & \dots \\ \dots & \vdots & \dots & \dots & \dots \\ \dots & \vdots & \dots & \dots & \dots \\ \dots & \vdots & \dots & \dots & \dots \\ 0 & \vdots & 0 & \dots & 0 \\ \dots & \vdots & \dots & \dots & \dots \\ 0 & \vdots & 0 & \dots & 0 \end{array} \right) \begin{array}{l} \left. \vphantom{\begin{matrix} \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ 0 \\ \dots \\ 0 \end{matrix}} \right\} n \\ \left. \vphantom{\begin{matrix} \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ 0 \\ \dots \\ 0 \end{matrix}} \right\} m-n \end{array} \quad (9)$$

Допустим, задан набор произвольных отношений E_1, E_2, \dots, E_q , каждое из которых имеет свою арность и задано на своем декартовом произведении множеств. Тогда может быть найдена максимальная арность, допустим N , и могут быть построены все матрицы согласованности $J(E_\alpha, E_\beta)$ размера $N \times N$, количество которых будет равно C_q^2 , где C_q^2 — число сочетаний по 2 из q различных элементов. В этих матрицах и в матрицах $J(E_\alpha)$ заложена вся информация: какие отношения и на каких местах индуцируют одни и те же разбиения или неразличимые с точки зрения данного набора отношений классы. Ранее уже отмечалось, что матрицы согласованности не учитывают детализации разбиений или вложенность индуцируемых произвольными отношениями эквивалентностей. В связи с этим может быть введена функция $L(i, j)$ для пары различных отношений аналогичной функции $F(k, l)$, определяемой (4):

$$L(i, j) = \begin{cases} -1, & \text{если } A_k = C_j \text{ и } P_i^{E_1} \subseteq P_j^{E_2}; \\ 0, & \text{если } A_i \neq C_j \text{ или } P_i^{E_1} \not\subseteq P_j^{E_2}; \\ 1, & \text{если } A_i = C_j \text{ и } P_i^{E_1} = P_j^{E_2}; \\ 2, & \text{если } A_i = C_j \text{ и } P_i^{E_1} \supseteq P_j^{E_2}. \end{cases} \quad (10)$$

Как и выше, может быть введена матрица $L(E_1, E_2)$, на (i, j) месте которой будут стоять элементы $l_{ij} = L(i, j)$. Следует отметить, что если переставить аргументы в матрице L , т. е. рассмотреть матрицу $L(E_1, E_2)$, то между ними ни в каком смысле симметрии наблюдаться не будет: на месте (i, j) может наблюдаться любое значение из множества $\{-1, 0, 1, 2\}$. Для матриц согласованности $J(E_1, E_2)$ при перестановке аргументов симметрия в обычном смысле тоже не наблюдается.

Таким образом, произвольное n -арное отношение индуцирует новое n -арное отношение со своим носителем, являющимся произведением классов эквивалентностей, иногда полностью совпадающим с первоначальным, если матрица согласованности единична, или существенно меняющим и носитель, и само отношение. Приведем простой, но достаточно наглядный пример.

Зафиксируем число $p \in \mathbb{N}$ и введем тернарное отношение $E(n_1, n_2, n_3)$ по формуле

$$E(n_1, n_2, n_3) = \begin{cases} 1, & \text{если } \text{res}(n_1 + n_2; p) = \text{res}(n_3; p); \\ 0, & \text{если } \text{res}(n_1 + n_2; p) \neq \text{res}(n_3; p); \end{cases} \quad (11)$$

где $\text{res}(a; b) \in \mathbb{N}$ — $a \div b$ — остаток от целочисленного деления. Если зафиксировать $k = 1$ и рассмотреть эквивалентность P_1^E вида

$$P_1^E(n_1, n'_1) = 1 \Leftrightarrow E(n_1, n_2, n_3) \equiv E(n'_1, n_2, n_3),$$

то она разбивает множество натуральных чисел на классы вычетов по модулю p : $(n_1 + n_2) \equiv n_3 \pmod{p}$. Действительно, если у n_1 и n_2 остатки при делении на число p совпадают, то справедлива цепочка равенств для любых n_2 и n_3

$$\begin{aligned} \text{res}(n_1 + n_2; p) &= \text{res}(ps_1 + r_1 + ps_2 + r_2; p) = \\ \text{res}(r_1 + r_2; p) &= \text{res}(ps'_1 + r_1 + ps_2 + r_2; p) = \\ &= \text{res}(n'_1 + n_2; p), \end{aligned}$$

откуда следует $E(n_1, n_2, n_3) \equiv E(n'_1, n_2, n_3)$ (здесь подразумевалось, что $n_1 = ps_1 + r_1$, $n'_1 = ps'_1 + r_1$, $n_2 = ps_2 + r_2$, а $s_1, s'_1, s_2 \in \mathbb{N}$). Обратное утверждение тоже справедливо: если $E(n_1, n_2, n_3) \equiv E(n'_1, n_2, n_3)$, то остатки при делении на p у n_1 и n'_1 должны быть равны. В противном случае, когда $n_1 = ps_1 + r_1$, $n'_1 = ps'_1 + r_1$ и $r_1 = r'_1$, при $n_2 = 0$, а $n_3 = r_1$ будем иметь цепочку равенств

$$\begin{aligned} \text{res}(n_1 + n_2; p) &= \text{res}(ps_1 + r_1; p) = \\ &= \text{res}(r_1; p) = \text{res}(n_3; p), \end{aligned}$$

т. е. $E(n_1, 0, r_1) = 1$ с одной стороны. С другой стороны,

$$\begin{aligned} \text{res}(n'_1 + n_2; p) &= \text{res}(ps'_1 + r_1; p) = \\ &= \text{res}(r'_1; p) = \text{res}(n_3; p). \end{aligned}$$

Поскольку $r'_1, r_1 \leq p$ и $r'_1 \neq r_1$, то $E(n'_1, 0, r_1) = 1$ что противоречит исходной предпосылке. Из соотношений (11) следует, что $P_1^E = P_2^E = P_3^E$, а матрица внутренней согласованности $J(E)$ и матрица, учитывающая сравнения эквивалентностей $F(E)$, в данном случае имеют вид

$$J(E) = F(E) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Отметим ряд важных обстоятельств. Во-первых, если носителем исходного отношения было множество натуральных чисел \mathbb{N} , то индуцируемая исходным отношением эквивалентность существенно изменила его, трансформировав в конечное множество $P = \{0, 1, \dots, p-1\}$. На новом носителе образовалось но-

вое отношение, которое в дальнейшем будем называть мультиотношением (точное определение дадим позднее) и обозначать E_m . Это отношение осталось тернарным, но заданное не как исходное на \mathbb{N}^3 , а на P^3 . Во-вторых, мультиотношение E_m приобрело новое свойство, которое можно представить бинарной операцией по следующим правилам

$$r_1 \oplus r_2 = r_3 \Leftrightarrow E_m(r_1, r_2, r_3) = 1 \Leftrightarrow E(n_1, n_2, n_3) = 1,$$

где \oplus — операция сложения по модулю p , а переменные — $n_i = ps_i + r_i, i = 1, 2, \dots, p, r_i \in \{0, 1, \dots, p-1\}$. Наконец, если следовать терминологии алгебраических систем [12], тройка $\langle A, F, S \rangle$, где A — произвольное множество (носитель), F — набор отношений и S — набор операций на данном носителе (т. е. в зависимости от арности на нем или на его декартовых степенях), является алгебраической системой. При этом: если $S = \emptyset$, то система называется моделью, если $F = \emptyset$ — алгеброй. Тогда в нашем случае, изначально имея модель $\langle \mathbb{N}, E, \emptyset \rangle$, в силу естественной эквивалентности, заложенной «природой» в отношении E , а не привнесенной «извне» (как это, как правило, происходит при рассмотрении процессов факторизации в традиционных алгебраических структурах), мы перешли к алгебре $\langle P, \emptyset, \oplus \rangle$, носителем которых стали классы эквивалентностей, образующие известную алгебраическую структуру — циклическую абелеву группу порядка p .

Надо понимать, что носитель исходного отношения может представлять собой одно множество, а порожденное им мультиотношение, т. е. отношение на индуцированных классах эквивалентности или толерантности, в качестве носителя может иметь декартово произведение различных множеств. Рассмотрим еще один пример, когда тернарное отношение E с прежним носителем — множеством натуральных чисел — задано равенством

$$E(n_1, n_2, n_3) = \begin{cases} 1, & \text{если } \text{res}(n_1 + n_2; p) = \text{res}(n_3; p); \\ 0, & \text{если } \text{res}(n_1 + n_2; p) \neq \text{res}(n_3; p); \end{cases} \quad (12)$$

где $p_1 \neq p_2$. В этом случае нетрудно заметить, что $P_1^E = P_2^E \neq P_3^E$, т. е. носитель мультиотношения E_m является декартовым произведением множеств $A \times B$, где $A = \{0, 1, \dots, p_1-1\}$, $B = \{0, 1, \dots, p_2-1\}$. Как тернарное отношение E_m задано на $A^2 \times B$ и представляет собой операцию из A^2 в B (но не групповую, как выше). Матрица внутренней согласованности $J(E)$ будет иметь вид

$$J(E) = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Уместно отметить, что при определенных p_1 и p_2 не просто возникает неравенство эквивалентностей и соответственно разбиений, но меняются и степень детализации разбиений, и вложенность эквивалентностей. Например, если $p_1 = 4$, а $p_2 = 2$, то $P_1^E = P_2^E \subseteq P_3^E$, т. к. классы множества \mathbb{N} разбиваются

на четыре класса, соответствующие остаткам $A = \{0, 1, 2, 3\}$ за счет эквивалентности $P_1^E = P_2^E$. Эквивалентность P_3^E разбивает это же множество на два класса: четные и нечетные числа, т. е. $B = \{0, 1\}$. При этом классы $\{0, 2\}$ входят в множество четных чисел (условно говоря, в класс «0» множества B), соответственно классы $\{1, 3\}$ входят в множество нечетных чисел (в класс «1» множества B). Следовательно, матрица $F(E)$ (вложенности разбиений) будет иметь вид

$$F(E) = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ 2 & 2 & 1 \end{pmatrix}.$$

В рассмотренных примерах исходное отношение E индуцировало мультиотношение, а точнее мультиоперацию, т. е. операцию на классах эквивалентностей. Может сложиться впечатление, что подобная ситуация наблюдается всегда, однако это далеко не так. Приведем пример. Рассмотрим бинарное отношение E , определенное табл. 1.

Таблица 1

| | | A | | B | C | | | | | |
|------------|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | a_1 | a_2 | b_1 | c_1 | c_2 | c_3 | c_4 | c_5 | c_6 |
| Π_I | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 2 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Π_{II} | 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 4 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 5 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |

Отношение E задано на декартовом произведении множеств $\{1, 2, 3, 4, 5\} \times \{a_1, a_2, b_1, c_1, c_2, c_3, c_4, c_5, c_6\}$. Нетрудно убедиться, что две эквивалентности P_1^E и P_2^E , индуцируемые исходным отношением, разбивают первое множество $A_1 = \{1, 2, 3, 4, 5\}$ на два класса $\Pi_I = \{1, 2\}$ и $\Pi_{II} = \{3, 4, 5\}$, тогда как второе множество $A_2 = \{a_1, a_2, b_1, c_1, c_2, c_3, c_4, c_5, c_6\}$ разбивается на три класса $A = \{a_1, a_2\}$, $B = \{b_1\}$, $C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$. При этом мультиотношение E_m уже задано на декартовом произведении индуцированных классов эквивалентностей $\{\Pi_I, \Pi_{II}\} \times \{A, B, C\}$ (табл. 2).

Таблица 2

| | A | B | C |
|------------|---|---|---|
| Π_I | 0 | 1 | 1 |
| Π_{II} | 1 | 1 | 0 |

Заметим, что любая n -арная операция может быть представлена в виде $(n+1)$ -арного отношения, когда на $(n+1)$ -м месте в этом отношении стоит единственный элемент носителя. Результат операции, как правило, не выводится за рамки носителя, которому принадлежат аргументы (это характерно для общепринятых алгебраических структур [13]). Однако во многих прикладных задачах (в частности, при факторизации множеств произвольной природы в некоторых признаковых пространствах) под n -арной операцией фактически понимают однозначное отображение аргументов. В данном случае мультиотношение может быть представлено даже

в виде двух отображений из $\{\Pi_I, \Pi_{II}\}$ в $\{A, B, C\}$ и наоборот, которые могут быть представлены в явном виде. Если их обозначить F_{E_m} и $F_{E_m}^{-1}$, то $F_{E_m}(\Pi_I) = \{A, B\}$, $F_{E_m}(\Pi_{II}) = \{A, B\}$ и $F_{E_m}^{-1}(B) = \{1, 2\}$, $F_{E_m}^{-1}(C) = 1$. И в том и в другом варианте никакой однозначности не наблюдается, поэтому говорить об индуцируемости какой-либо операции на классах эквивалентности, а точнее мультиоперации, как в предыдущих двух примерах, мы не можем. Здесь исходное отношение индуцирует в явном виде лишь бинарное мультиотношение.

Как следует из приведенных выше примеров, модель, содержащая одно отношение, может приводить на классах эквивалентностей к различным алгебраическим системам: и к мультимоделям, и к мультиалгебрам. Однако при оперировании одним исходным отношением мы в состоянии прийти только к одному из вышеперечисленных вариантов, т. е. к полноценной алгебраической системе, содержащей и отношения, и операции, мы перейти не сможем. Следует отметить, что в произвольной алгебраической системе, представляющей абстрактную информационную систему (и в традиционных структурах, таких как полугруппа, группа, кольцо, поле, различные типы линейных пространств и т. д.), носитель один и тот же для включенных в систему отношений и операций. Это утверждение, по нашему мнению, надо трактовать в следующем смысле. Если рассматривать любое линейное пространство над каким-либо полем (допустим, векторное над полем действительных чисел), то носителем этой системы можно считать множество векторов или декартово произведение множества векторов и поля действительных чисел (что, на наш взгляд, логичнее). При этом операцию сложения векторов, казалось бы, не действующую элементов поля, следует представлять как линейную комбинацию элементов носителя с единичными коэффициентами. Таким образом, важным становится вопрос: когда два произвольных отношения формируют один носитель? Точнее говоря, этот вопрос будет интересен с точки зрения мультиотношений, индуцируемых исходными отношениями. Здесь уместно обратить внимание на матрицу $J(E)$, определяемую равенством (6), и на равенства (7). Если произвольное n -арное отношение порождает p различных эквивалентностей ($p \leq n$) L_1^E, \dots, L_p^E , то носителем является $B_1 \times B_2 \times \dots \times B_p$, где L_i^E задано на $B_i \times B_i, i = 1, 2, \dots, p$. Если обозначить множество классов эквивалентностей, формирующееся на множестве B_i за счет эквивалентностей L_i^E , через $B_i / L_i^E, i = 1, 2, \dots, p$, то ясно, что мультиотношение E_m будет иметь в качестве носителя декартово произведение смежных классов $B_1 / L_1^E \times B_2 / L_2^E \times \dots \times B_p / L_p^E$.

Приведем, наконец, точное определение мультиотношений.

Определение 2. Отношение E_m , индуцированное произвольным n -арным отношением E , с носителем $B_i / L_i^E \times B_i / L_i^E$ в виде декартова произведения смежных классов, порождаемых «внутренними» эквивалентностями L_i^E , назовем мультиотношением.

4. Условие равенства носителей мультиотношений

Теперь становится очевидным, что, по крайней мере, для совпадения носителей двух различных мультиотношений E_{1m} и E_{2m} , индуцируемых отношениями E_1 и E_2 , количества единичных блоков матрицы внутренней согласованности p_1 матрицы $J(E_1)$ и p_2 матрицы $J(E_2)$ должны совпадать. Рассмотрим далее матрицу согласованности двух отношений $J(E_1, E_2)$ размером $n \times m$, (i, j) -й элемент (обозначим его через l_{ij}) которой определяется условием (8) или равенством

$$l_{ij} = \begin{cases} 1, & \text{если } A_i = C_j \text{ и } P_i^{E_1} = P_j^{E_2} \\ 0, & \text{в противном случае,} \end{cases} \quad (13)$$

где E_1 задано на $A_1 \times \dots \times A_n$, а E_2 — на $C_1 \times \dots \times C_m$. Допустим, эта матрица, состоящая из 0 и 1, путем перестановок строк и столбцов может быть приведена к виду

$$J^*(E_1, E_2) = \begin{matrix} & \beta_1 & \beta_2 & \dots & \beta_i & \dots & \beta_s \\ \left. \begin{matrix} 1 & \dots & 1 & 0 & \dots & 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \dots & 1 & 0 & \dots & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 0 & 1 & \dots & 1 & 0 & \dots & \dots & \dots & 0 \\ & & 0 & 0 & \dots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & 0 & 1 & \dots & 1 & 0 & \dots & 0 \\ & & & & & 0 & 1 & \dots & 1 & 0 & \dots & 0 \\ & & & & & & & & & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & \dots & \dots & 0 & 1 & \dots & 1 \end{matrix} \right\} \begin{matrix} \alpha_1 \\ \alpha_2 \\ \alpha_1 \\ \alpha_s \end{matrix} \end{matrix} \quad (14)$$

Матрица (14) задает бинарное отношение V_{E_1, E_2} на множестве $\{1, 2, \dots, n\} \times \{1, 2, \dots, m\}$.

Следует заметить, что в общем случае это бинарное отношение может быть произвольным, поскольку нет никаких ограничений на отношения E_1 и E_2 . Поэтому в отличие от матриц внутренней согласованности, которые фактически задают отношения эквивалентностей и для которых блочно-диагональный вид достигается всегда, о матрицах $J(E_1, E_2)$ этого сказать нельзя. Например, к блочному виду не приводится матрица

$$J(E_1, E_2) = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

В то же время все иначе, если это отношение является дифункциональным. Напомним, что бинарное отношение дифункционально [14], если для любых $i, i_1, j, j_1 \in \{1, 2, \dots, \max(n, m)\}$ справедлива импликация

$$V_{E_1, E_2}(i, j_1) = 1, V_{E_1, E_2}(i_1, j) = 1 \\ V_{E_1, E_2}(i_1, j) = 1 \Rightarrow V_{E_1, E_2}(i, j) = 1$$

Справедливо следующее утверждение.

Утверждение 2. Пусть X и Y — произвольные конечные множества мощности n и m соответственно. Тогда для любого бинарного отношения R на $X \times Y$ матрица $G(R)$ этого отношения путем перестановок строк и столбцов (фактически — перенумерацией элементов множеств X и Y) приводится к блочно-диагональному виду $G^*(R)$ с размерностями блоков $(\alpha_i \times \beta_i)$, $i = 1, \dots, s$,

$$G^*(R) = \begin{matrix} & \beta_1 & \beta_2 & \dots & \beta_i & \dots & \beta_s \\ \left. \begin{matrix} 1 & \dots & 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \dots & 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & 1 & \dots & 1 & 0 & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & & & & & & 1 & \dots & 1 & 0 & \dots & 0 \\ & & & & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & 1 & \dots & 1 & 0 & \dots & 0 \\ & & & & & & 0 & \dots & 0 & 0 & \dots & 0 \\ & & & & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & \dots & \dots & 0 & \dots & 0 & 0 & \dots & 0 \end{matrix} \right\} \begin{matrix} \alpha_1 \\ \alpha_2 \\ \alpha_s \end{matrix} \end{matrix} \quad (15)$$

лишь при условии, что отношение R дифункционально.

Доказательство. Предварительно дадим интерпретацию свойства дифункциональности при матричном представлении отношений. Заметим, что любое бинарное отношение — это равносильное представление многозначного отображения $f: X \rightarrow Y$, где образом любого $x \in X$ является некоторое подмножество $f(x) \in Y$, а прообразом любого $y \in Y$ является некоторое подмножество $f^{-1}(y) \in X$. В общем случае могут найтись элементы $x \in X$ и $y \in Y$, не имеющие образа и прообраза соответственно: их можно сгруппировать в подмножества X_\emptyset и Y_\emptyset (рис. 1).

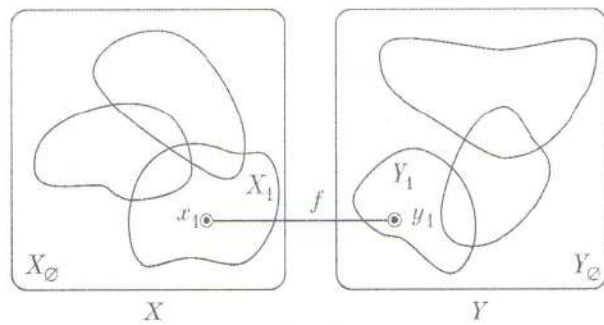


Рис. 1

На рис. 1 образом смежного класса $X_1 \in X$ есть смежный класс $Y_1 \in Y$ при наличии элементов $x_1 \in X$ и $y_1 \in Y$, связанных равенством $f(x_1) = y_1$, т. е. исходным отношением. При этом смежные классы как во множестве образов Y , так и во множестве образов X могут пересекаться, а их число может быть различно. Однако, если отношение дифункционально, различные смежные классы не пересекаются [14], и отображение f осуществляет между ними взаимно однозначное соответствие. Вообще говоря, дифункциональные отношения — это отношения, порожденные взаимно однозначными соответствиями между системами попарно не пересекающихся подмножеств X и Y . В случае, когда множества X и Y конечны, число смежных классов образов и прообразов одинаково, и отображение f может быть схематично представлено в виде рис. 2.

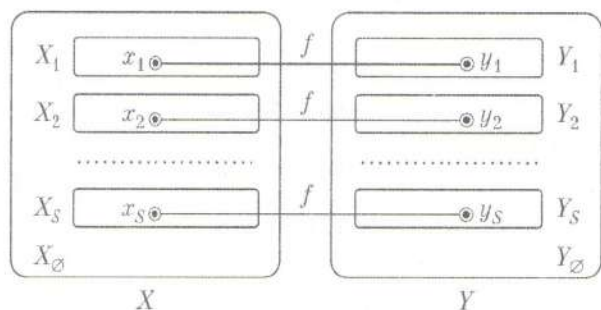


Рис. 2

Следует подчеркнуть, что $\alpha_1 = \text{card } X_1, \dots, \alpha_s = \text{card } X_s, \beta_1 = \text{card } Y_1, \dots, \beta_s = \text{card } Y_s$ и $\text{card } X = \alpha_1 + \dots + \alpha_s + X_\emptyset = n, \text{card } Y = \beta_1 + \dots + \beta_s + Y_\emptyset = m$. Итак, в случае дифункциональности, если в i -й строке матрицы $G(R)$ встречаются две 1 (допустим, в столбцах j и j_1) и в одном из этих столбцов находится еще одна 1 (допустим, в j -столбце и строке i_1), то 1 появится и в j_1 -столбце в той же строке i_1 . Это свойство выполняется и для строк. Сформулируем это свойство как свойство «четырехугольника». Если в вершинах четырехугольника

$$G(R) = \begin{pmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & A & \dots & B & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & C & \dots & D & \dots \\ \dots & \dots & j & \dots & \dots \end{pmatrix} \begin{matrix} i \\ i_1 \\ i_1 \\ i \end{matrix}$$

на любых трех местах появляется 1, то она появится и на месте, соответствующем четвертой вершине. Действительно, пусть i -й строке соответствует элемент x_i, i_1 -й строке — элемент x_{i_1}, j -столбцу — элемент y_j и, наконец, j_1 -столбцу отвечает элемент y_{j_1} , тогда если в трех вершинах, например A, B, C , стоит 1, то $R(x_i, y_j) = R(x_i, y_{j_1}) = R(x_{i_1}, y_j) = 1$. Из дифункциональности отношения R непосредственно следует $R(x_{i_1}, y_{j_1}) = 1$, что соответствует вершине D . Очевидно, что это свойство инвариантно относительно перестановок индексов (выбора вершин). Теперь можно перейти к доказательству утверждения.

Докажем необходимость. Допустим, отношение R дифункционально (оно путем соответствующей нумерации элементов множества X и Y и их классов смежности реализует отображение f). Нетрудно понять, что матрица этого отображения будет иметь вид (15). Действительно, если некоторый блок имеет хотя бы одну «соседнюю» 1 (сверху, снизу, слева, справа), то в силу свойства «четырехугольника», 1 должна появиться на всей «границе» блока в соответствующей строке или столбце. Поясним это утверждение. Выделим блок под номером k в матрице $G^*(R)$ и допустим, что в i -й строке под этим блоком появилась 1. Схематично трансформации можно представить цепочкой импликаций

$$\begin{aligned} G^*(R) &= \left(\begin{array}{ccc} \dots & \overbrace{\dots}^{\beta_k} & \dots \\ \dots & \left(\begin{array}{ccc} 1 & 1 & \dots & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 & \dots & 1 \\ \dots & & & & & \end{array} \right) & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{array} \right) \alpha_k \Rightarrow \\ \Rightarrow G^*(R) &= \left(\begin{array}{ccc} \dots & \overbrace{\dots}^{\beta_k} & \dots \\ \dots & \left(\begin{array}{ccc} 1 & 1 & \dots & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 & \dots & 1 \\ \dots & & & & & \end{array} \right) & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{array} \right) \alpha_k \Rightarrow \dots \Rightarrow \\ \Rightarrow G^*(R) &= \left(\begin{array}{ccc} \dots & \overbrace{\dots}^{\beta_k} & \dots \\ \dots & \left(\begin{array}{ccc} 1 & 1 & \dots & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 & \dots & 1 \\ \dots & & & & & \end{array} \right) & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{array} \right) \alpha_k \end{aligned}$$

Применение свойства «четырехугольника» $\beta_k - 1$ раз (столько же раз, если 1 появилась над блоком, и $\alpha_k - 1$ раз, если она расположена слева или справа от k -го блока) означает расширение блока или, что равносильно, расширение смежного класса образов или прообразов, а это невозможно. Таким образом, полученное противоречие доказывает, что из дифункциональности следует вид (15) матрицы $G^*(R)$. Достаточность проверяется непосредственно. Утверждение доказано.

Предположим теперь, что заданы два отношения E_1 и E_2 , для которых матрица согласованности $J(E_1, E_2)$ приводится к виду (15). Допустим, что она еще не содержит нулевых строк и столбцов. Тогда нетрудно понять, что каждому блоку под номером k в области определения отношения E_1 соответствует декартово произведение $A_k^{\alpha_k}$, а в области определения E_2 декартово произведение $A_k^{\beta_k}$ в силу определения $J(E_1, E_2)$. При этом разбиения, индуцируемые E_1 и E_2 на множестве A_k , совпадают. Таким

образом, для мультиотношений E_{1m} и E_{2m} каждому блоку соответствует один и тот же носитель, а так как и количество блоков и число ненулевых строк (столбцов) одно и то же, то нет фактически «лишних» или различных множеств среди носителей E_1 и E_2 , т. е. у мультиотношений формируется один носитель. В обратном случае, когда у мультиотношений E_{1m} и E_{2m} носители совпадают, вид (15) для матрицы $G^*(R)$ без нулевых строк и столбцов устанавливается непосредственной проверкой. Теперь можно сформулировать теорему.

Теорема (о равенстве носителей мультиотношений). Два произвольных отношения — n -арное E_1 и m -арное E_2 — индуцируют два мультиотношения E_{1m} и E_{2m} с совпадающим носителем тогда и только тогда, когда матрица согласованности этих отношений $J(E_1, E_2)$ путем перестановок строк (столбцов) приводится к блочному виду (15) без нулевых строк и столбцов, что равносильно условию: отношение V_{E_1, E_2} , соответствующее матрице $J(E_1, E_2)$, является дифункциональным.

Доказательство необходимого и достаточного условия приведено непосредственно перед формулировкой теоремы, а их эквивалентность условию дифункциональности отношения V_{E_1, E_2} очевидно следует из утверждения 2.

5. Результаты и перспективы исследований

Грануляция информации позволяет представлять нечто «целое» семействами классов, каждый из которых, в свою очередь, является множеством элементов, неразличимых с точки зрения внутренних (сущностных), внешних (структурных) или контекстных (предметно-ориентированных) свойств. Более того, сами гранулы могут содержать в себе множества вложенных гранул. Возникающие иерархии (частичный порядок на гранулах, индуцированных различными свойствами) часто могут интерпретироваться как индуктивные («снизу вверх») и дедуктивные («сверху вниз») методы вывода. В работе впервые установлены условия, при которых отношения на отдельных уровнях продуцируют новые отношения, не различающие элементы гранул, на более высоких ступенях. Таким образом, синтезированы предпосылки для получения ответов на вопросы: что разделяет уровни иерархии, что объединяет эти уровни и как они взаимодействуют? В этом плане важное значение приобретает «множественное» сравнение, т. е. возможность сопоставлять факторно-множества. Иначе говоря, поиск метрик на разбиениях, в том числе вложенных, представляет определенную перспективу не только для мультиалгебраических систем, но и других подходов в рамках ГС-парадигмы.

Список литературы: 1. Zadeh L. A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic // Fuzzy Sets Systems.— 1997.— Vol. 19.— P. 111–127. 2. Pal S. K., Shankar B. U., Mitra P. Granular computing, rough entropy and object extraction // Pattern Recognition Letters.— 2005.— Vol. 26.— P. 2509–2517. 3. Yao Y. Y. Perspectives of granular computing // Proceedings of 2005 IEEE International Conference on Granular Computing.— 2005.— Vol. 1.— P. 85–90. 4. Yao J. T., Yao Y. Y. Induction of classification rules by granular computing // Rough Sets and Current Trends in Computing / J. J. Alpigini, J. F. Peters, A. Skowron, N. Zhong (Eds.). Lecture Notes in Artificial Intelligence.— Springer-Verlag Berlin Heidelberg.— 2002.— Vol. 2475.— P. 331–338. 5. Yao Y. Y. Information granulation and rough set approximation // International Journal of Intelligent Systems.— 2001.— Vol. 16, No. 1.— P. 87–104. 6. Doherty P., Lukasiewicz W., Szalas A. Information granules for intelligent knowledge structures // Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing / G. Wang, Q. Liu, Y. Yao, A. Skowron (Eds.). Lecture Notes in Artificial Intelligence.— Springer-Verlag Berlin Heidelberg.— 2003.— Vol. 2639.— P. 405–412. 7. Yao Y. Y. Granular computing for data mining // Proceedings of SPIE Conference on Data Mining, Intrusion Detection, Information Assurance and Data Networks Security / B. V. Dasarathy (Ed.).— Kissimmee, Florida, USA.— 2006.— P. 1–12 (624105). 8. Yager R. R. Using granular objects in multi-source data fusion // Rough Sets and Current Trends in Computing / J. J. Alpigini, J. F. Peters, A. Skowron, N. Zhong (Eds.). Lecture Notes in Artificial Intelligence.— Springer-Verlag Berlin Heidelberg.— 2002.— Vol. 2475.— P. 324–330. 9. Lin T. Y. Granular computing (Structures, Representations, and Applications) // Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing / G. Wang, Q. Liu, Y. Yao, A. Skowron (Eds.). Lecture Notes in Artificial Intelligence.— Springer-Verlag Berlin Heidelberg.— 2003.— Vol. 2639.— P. 16–24. 10. Bargiela A., Pedrycz W. Granular computing: an introduction.— Boston, Kluwer Academic Publishers, The Kluwer International Series in Engineering and Computer Science.— 2002.— Vol. 717.— 478 p. 11. Машталир В. П., Шляхов В. В. Свойства мультиалгебраических систем в задачах компаративного распознавания // Кибернетика и системный анализ.— 2003.— № 6.— С. 12–32. 12. Луганский А. М., Машталир В. П., Путькина А. Е., Шляхов В. В. О согласованности отношений в информационных системах // Радиоэлектроника и информатика.— 2004.— № 4.— С. 118–120. 13. Машталир В. П., Шляхов В. В. Необходимые и достаточные условия индуцирования мультиалгебраических систем n -арными отношениями // Интеллектуальные системы принятия решений и прикладные аспекты информационных технологий: М-лы междунар. конф.— Херсон: ХМИ.— 2006.— Т. 1.— С. 136–140. 14. Мальцев А. И. Алгебраические системы.— М.: Наука, 1970.— 392 с.

Поступила в редколлегию 07.06.2006

УДК 519.7

М. Ф. Бондаренко, Ю. П. Шабанов-Кушнарченко

МОДЕЛИ ЦВЕТОВОГО ЗРЕНИЯ

1. Задача моделирования зрительных процессов

Орган зрения человека, являющийся сложнейшей системой приема и обработки информации, давно привлекает внимание исследователей. Многие о принципах работы органа зрения человека уже известно, и эти знания широко используются в светотехнике, технике кино и телевидения, полиграфии и в ряде других областей. Глаз современного человека работает в контакте со многими техническими устройствами. К ним относятся установки телевидения, кино и радиолокации, пульта управления, измерительные приборы, светотехнические устройства, книги, различные знаки (например, дорожные), карты и многое другое. Все эти устройства создаются с учетом свойств человеческого зрения. Так, например, скорость развертки луча в телевизионном изображении, размеры и форма шрифтов книг, частота смены кадров в кино, характер сигнальных огней, размеры и форма шкал, режимы освещения, — все это выбирается в зависимости от свойств человеческого зрения. Чем лучше и полнее мы будем знать эти свойства, тем эффективнее, проще и дешевле будут технические устройства, создаваемые для совместной работы с органом зрения человека.

Данные о принципах работы органа зрения важны также и в связи с тем, что уже в настоящее время для различных технических целей создаются специальные устройства фотоприема информации. Это — различного рода фотоприемные устройства для измерительных приборов и систем автоматики, телевизионные системы и системы радиолокации, устройства фотоприема информации в вычислительные, управляющие и информационные машины. Орган зрения человека также является фотоприемником информации. Будучи неизмеримо более совершенным по сравнению с техническими приборами, орган зрения во многих отношениях может служить образцом при создании новых и усовершенствовании существующих устройств фотоприема информации.

При исследовании органа зрения человека целесообразно рассматривать его как устройство, осуществляющее прием и преобразование сигналов. В настоящей работе рассматривается задача исследования закономерностей преобразования сигналов в органе зрения с целью формулировки этих закономерностей в виде математических моделей. Вопрос об анатомических структурах и физиологических механизмах, обеспечивающих выполнение этих преобразований, не будет затрагиваться. Основой при построении математических моделей работы органа зрения для нас будут служить психофизические реакции зрительного анализатора. При такой постановке задачи для наблюдения доступны лишь входные и выходные сигналы органа зрения, и мате-

матическое моделирование зрительных процессов можно вести по методу, получившему название метода кибернетического «черного ящика». Входными сигналами органа зрения являются зрительные картины, наблюдаемые испытуемым, а выходными — возникающие при этом у испытуемого зрительные ощущения.

Следует отметить, что математическое моделирование зрения человека, основанное на изучении психофизических реакций глаза и фактическом использовании понятия «черного ящика», не является чем-то принципиально новым. Такая работа, хотя и в иных терминах, ведется уже давно, и к настоящему времени в области математического моделирования зрения получен ряд существенных результатов. В этой области успешно работали Ньютоном, Юнг, Максвелл, Грассман, Гельмгольц, Шредингер. Значительный вклад в развитие проблем математического моделирования зрения человека внесли отечественные ученые П. П. Лазарев, А. В. Луизов, В. В. Мешков, Н. Д. Ньюберг, М. М. Бонгард, А. Б. Матвеев, М. С. Смирнов, Д. А. Шкловер и др. Однако, в связи со сложностью законов преобразования информации в органе зрения человека, многие стороны и свойства зрения остаются пока еще не охваченными математическими моделями. Настоящая работа посвящена дальнейшей разработке проблем математического моделирования зрения человека.

Прежде чем приступить к исследованию избранного нами вопроса, рассмотрим, что представляют собой входные и выходные сигналы органа зрения человека и дадим их математическое описание. С этой целью воспользуемся схемой, представленной на рис. 1. На ней изображен эскиз преобразования сигналов в органе зрения человека. Перед глазом 1 расположена зрительная картина 2, заданная в поле зрения. Выделим в поле зрения точку фиксации o и проведем через нее горизонтальную и вертикальную оси. Координатами x и y произвольной точки M поля зрения будем называть соответственно углы $oo'a$ и $oo'b$:

$$x = \angle oo'a, \quad y = \angle oo'b. \quad (1)$$

Точка o' расположена внутри хрусталика глаза. Каждой точке поля зрения соответствует входной сигнал b_λ , характеризующий зрительную картину в этой точке в данный момент времени, а именно спектральная плотность лучистой яркости в функции длины волны или просто спектр излучения $b(\lambda)$ [1, с. 34].

Любая конкретная зрительная картина, заданная в поле зрения, может быть математически описана в виде функции входного сигнала b_λ от координат поля зрения x , y и времени t :

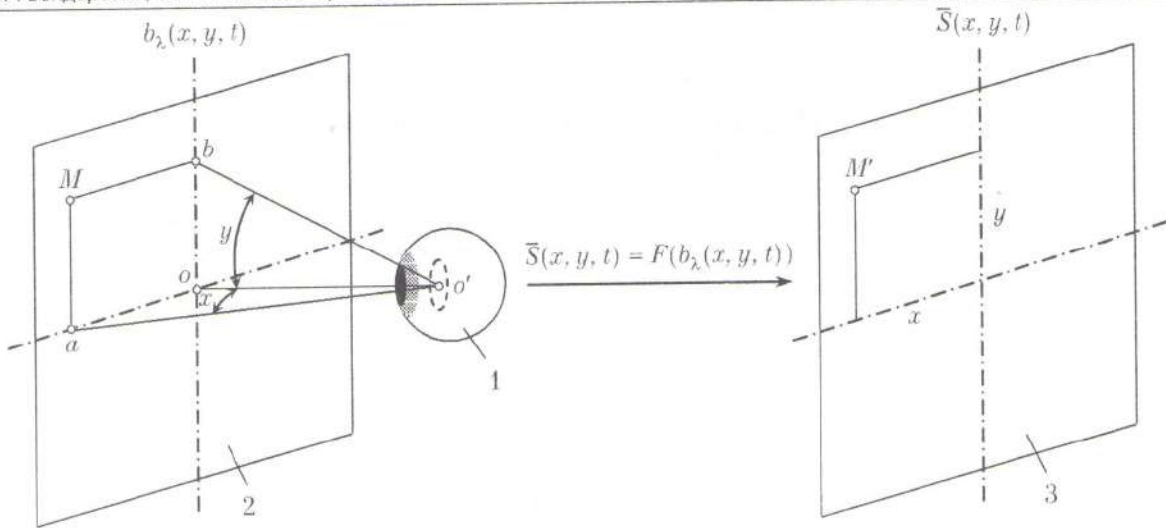


Рис. 1. Схема преобразования сигналов в органе зрения человека

$$b_\lambda = b_\lambda(x, y, t). \quad (2)$$

Таким образом, в принципе, можно математически описать любую зрительную картину, изменяющуюся в пространстве и во времени. В органе зрения человека осуществляется преобразование зрительной картины в зрительное ощущение 3. Зрительное ощущение также задано в двумерном поле. Точке M' зрительного ощущения приписываем координаты x, y соответствующей точки M зрительной картины. Зрительное ощущение в каждой точке поля зрения характеризуется цветом. Считаем, что цвет может быть разложен на три компонента: цветовой тон, насыщенность и светлоту [2, с. 11]. Таким образом, зрительное ощущение в каждой точке поля зрения будем описывать с помощью трехмерного вектора цвета \bar{S} , компонентами которого служат светлота S_1 , насыщенность S_2 и цветовой тон S_3 . Зрительное ощущение в целом, так же как и зрительная картина, может быть описано некоторой зависимостью цвета \bar{S} от координат поля зрения x, y и времени t :

$$\bar{S} = \bar{S}(x, y, t). \quad (3)$$

Мы ввели в рассмотрение входной $b_\lambda(x, y, t)$ и выходной $\bar{S}(x, y, t)$ сигналы органа зрения. Теперь более строго и детально сформулируем задачу исследования. Она состоит в том, чтобы отыскать закон F преобразования входного сигнала органа зрения в выходной:

$$\bar{S}(x, y, t) = F(b_\lambda(x, y, t)). \quad (4)$$

Функция F преобразует бесконечномерную вектор-функцию $b_\lambda(x, y, t)$ трех переменных x, y, t в трехмерную вектор-функцию $\bar{S}(x, y, t)$ тех же переменных. Можно считать, что функция F есть векторный оператор [3, с. 12], который ставит в соответствие вектор-функциям $b_\lambda(x, y, t)$ вектор-функции $\bar{S}(x, y, t)$. Исходя из повседневного опыта зрения, можно заключить, что этот оператор непрерывен [3, с. 18], поскольку всегда при непрерывном изменении зрительной картины $b_\lambda(x, y, t)$ наблюдается непрерывное изменение зрительного ощущения $\bar{S}(x, y, t)$.

Даже не опираясь на специальные эксперименты, а используя только повседневную практику зрения каждого человека, можно заключить, что искомый оператор зрения F не будет очень простым. Напротив, легко прийти к выводу о его большой сложности. Этот оператор должен воспроизводить разнообразные свойства цветового зрения, инерции и иррадиации зрения, адаптации и контраста. В связи с этим имеет смысл общую задачу отыскания оператора F разбить на ряд более простых задач.

Наиболее простой постановка задачи получается, если ограничить класс входных сигналов лишь однородными и стационарными зрительными картинами. Однородной мы называем такую зрительную картину, у которой излучения во всех точках поля зрения одинаковы. В этом случае входной сигнал b_λ не зависит от координат поля зрения x, y , а зависит лишь от времени t , то есть $b_\lambda = b_\lambda(t)$. Стационарной назовем такую зрительную картину, у которой излучения во всех точках не меняются со временем. Иными словами, функция b_λ , описывающая зрительную картину, не зависит от времени, а зависит лишь от координат поля зрения, то есть $b_\lambda = b_\lambda(x, y)$. Функция, описывающая однородную и стационарную зрительную картину, не зависит от координат поля зрения и времени и представляет собой спектр излучения $b_\lambda = b(\lambda)$. Такая картина представляет собой равномерное поле сигналов, не меняющихся с течением времени. Опыт показывает, что таким зрительным картинам соответствует зрительное ощущение, также являющееся однородным и стационарным, то есть не зависящее от координат поля зрения. Зрительное ощущение имеет вид равномерного фона одного цвета, заполняющего все поле зрения. Выходной сигнал представляет собой трехмерный вектор \bar{S} .

Далее можно усложнить задачу, расширив класс входных сигналов за счет введения нестационарных зрительных картин. В этом случае входной сигнал становится функцией времени, то есть $b_\lambda = b_\lambda(t)$. Опыт показывает, что зрительное ощущение также будет представлять собой функцию времени и не

будет зависеть от координат поля зрения, то есть $\bar{S} = \bar{S}(t)$. Зрительное ощущение будет выглядеть в виде равномерного фона, цвет которого меняется со временем. Точно так же возможно изолированное рассмотрение задачи о связи $b_\lambda(x)$ и $\bar{S}(x)$, $b_\lambda(y)$ и $\bar{S}(y)$. Далее могут быть изолированно рассмотрены задачи о связи сигналов $b_\lambda(x, y)$ и $\bar{S}(x, y)$, $b_\lambda(x, t)$ и $\bar{S}(x, t)$, $b_\lambda(y, t)$ и $\bar{S}(y, t)$. Наконец, мы приходим к общей задаче о связи между сигналами $b_\lambda(x, y, t)$ и $\bar{S}(x, y, t)$.

Построение математической модели стационарных и однородных зрительных процессов $\bar{S} = F(b(\lambda))$ должно основываться на фактах так называемых «низшей» и «высшей» метрик цвета. Исследование нестационарных зрительных процессов $\bar{S}(t) = F(b_\lambda(t))$ требует рассмотрения явлений инерции и адаптации зрения. Математическая модель неоднородных зрительных процессов $\bar{S}(x, y) = F(b_\lambda(x, y))$ может быть построена на основе изучения фактов иррадиации зрения и явлений зрительного контраста. Задача моделирования зрительных процессов в общем случае $\bar{S}(x, y, t) = F(b_\lambda(x, y, t))$ приводит к необходимости изучения реакций органа зрения на зрительные картины произвольной сложности. Таким образом, мы приходим к необходимости освоения обширного экспериментального материала. В полном объеме такая работа не может быть выполнена в одном исследовании. В силу этого приходится сузить задачу и ограничиться построением математических моделей для ряда частных случаев функционирования органа зрения.

При построении этих моделей мы придерживались следующего порядка в проведении работ и изложении полученных результатов. Вначале конструируется математическая модель, которая, хотя бы в качественном отношении, воспроизводит ту или иную область фактов зрения. Затем на основе этой модели чисто дедуктивным путем определяются предсказываемые зрительные реакции, которые сравниваются с фактическими реакциями зрительного анализатора, описанными в литературе. Если литературных данных оказывается недостаточно или они вовсе отсутствуют, то проводятся новые необходимые опыты. Модель считается заслуживающей внимания, если хотя бы некоторые из предсказаний, полученных на основе этой модели, совпадают с фактическими реакциями органа зрения. После этого решается задача построения такой новой модели, которая в логическом отношении была бы равносильна следствиям из прежней модели, оправдавшим себя на опыте. Новая модель отличается от первоначальной тем, что в ней устранены те детали, справедливость которых не удалось подтвердить в опыте. В случаях, когда такую задачу удается решить, мы получаем модель, которая логически вытекает из надежно установленных экспериментальных законов, принимаемых в качестве аксиом. Степень достоверности полученной таким путем модели определяется лишь надежностью установления со-

ответствующих экспериментальных законов. Хотя мы и стремились к возможно более полному осуществлению изложенной программы, однако ее удалось реализовать только в ряде частных случаев. Не во всех из этих случаев эта программа была выполнена полностью. Полное осуществление изложенной программы означало бы получение достоверной и исчерпывающей информации об операторе человеческого зрения $\bar{S}(x, y, t) = F(b_\lambda(x, y, t))$.

2. Задача моделирования стационарных и однородных зрительных процессов

Задача математического моделирования стационарных и однородных зрительных процессов сводится к отысканию вида зависимости:

$$\bar{S} = F(b(\lambda)), \quad (5)$$

где $b(\lambda)$ — входной сигнал органа зрения в виде спектра излучения; \bar{S} — выходной сигнал органа зрения в виде трехмерного вектора цвета; F — искомая функциональная зависимость выходного сигнала от входного. Исследования Ньютона [5], Ломоносова [6], Юнга [7], Максвелла [8], Гельмгольца [9] и других авторов привели к построению трехкомпонентной теории цветового зрения, которую можно сформулировать следующим образом. Всевозможные излучения $b(\lambda)$, для которых совпадают тройки чисел B_1, B_2, B_3 , вычисляемые по формулам

$$B_1 = \int_{\lambda_1}^{\lambda_2} b(\lambda)m(\lambda)d\lambda, \quad B_2 = \int_{\lambda_1}^{\lambda_2} b(\lambda)n(\lambda)d\lambda, \quad B_3 = \int_{\lambda_1}^{\lambda_2} b(\lambda)p(\lambda)d\lambda, \quad (6)$$

порождают одинаковые цвета. Вместе с тем, излучения, для которых эти тройки различны, порождают различные цвета. В равенствах (6) $m(\lambda), n(\lambda), p(\lambda)$ обозначают линейно независимые функции (так называемые функции сложения), определяемые для органа зрения экспериментально [10, с. 224]. Числа λ_1 и λ_2 обозначают соответственно наименьшее и наибольшее значения длины волны видимого диапазона спектра электромагнитных колебаний.

Как непосредственно следует из приведенной формулировки трехкомпонентной теории цветового зрения, вектор цвета \bar{S} с компонентами S_1, S_2, S_3 связан некоторой взаимно однозначной вектор-функцией f с вектором \bar{B} , имеющим компоненты B_1, B_2, B_3 :

$$\bar{S} = f(\bar{B}). \quad (7)$$

Конкретный вид зависимости f трехкомпонентной теорией цветового зрения не расшифровывается. Совокупность соотношений (6) и (7) можно рассматривать в качестве математической модели, описывающей вид искомого преобразования сигналов, осуществляемого органом зрения человека. Блок-схема этой модели изображена на рис. 2. В ней блоки $1_1, 1_2, 1_3$ осуществляют вычисление линейных

функционалов B_1, B_2, B_3 по формулам (6). Блок 2 осуществляет некоторое взаимно однозначное преобразование (7) вектора \bar{B} с компонентами B_1, B_2, B_3

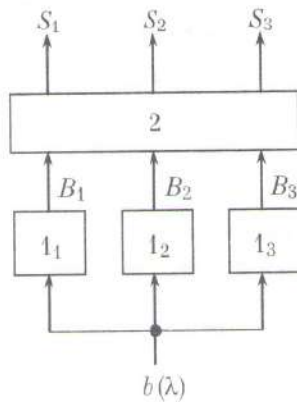


Рис. 2. Математическая модель цветового зрения

в вектор \bar{S} цвета зрительного ощущения с компонентами S_1 — светлотой, S_2 — насыщенностью, S_3 — цветовым тоном.

Возникает вопрос, является ли эта модель всего лишь гипотезой, различные следствия которой подтверждаются в эксперименте, или же ее можно рассматривать как достоверный факт и, следовательно, она может быть логически выведена из прочно установленных

экспериментальных законов, принимаемых в качестве аксиом. Многие авторы вводят эти уравнения прежде, чем рассматриваются экспериментальные факты, подтверждающие их справедливость. В этом случае соотношения (6), (7) фактически фигурируют в качестве гипотезы. Так сделано, например, в книге Мешкова [10, с. 101, 198]. Затем, основываясь на соотношениях (6) и (7), вводят понятие вектора цвета, понимая под ним вектор \bar{B} с компонентами B_1, B_2, B_3 . Далее вводят операции сложения цветов и умножения их на постоянные числа, а также понятие линейной зависимости цветов. Затем с помощью введенных понятий формулируются три закона смешения цветов (законы Грассмана [10]). Иногда эти три закона объединяют в один. Законы Грассмана служат тем основанием, на котором строится затем стройное здание колориметрии. Приведем один из вариантов формулировки законов Грассмана: закон аддитивности — суммы попарно равных цветов также суть равные цвета; закон трехмерности — любые четыре цвета линейно зависимы, однако существуют тройки линейно независимых цветов; закон непрерывности — непрерывному изменению излучения соответствует непрерывное изменение цвета.

В то же время в литературе существует мнение, по-видимому, впервые высказанное в 1920 г. Шредингером в работе [12], что из законов Грассмана в приведенной выше формулировке чисто логически вытекает модель цветового зрения в виде формул (6) и (7). В той же работе Шредингер дал вывод, доказывающий, по его мнению, это положение. Ньюберг в работе [13, с. 158] пишет: «Обычно интегральные выражения цвета выводятся как следствие гипотезы Гельмгольца, но их возможно получить непосредственно из закона Грассмана, не пользуясь никакой гипотезой (курсив Ньюберга. — Авт.). Это положение за недостатком места я оставляю без доказательства, которое можно найти в статье Шредингера». Однако выполненный нами анализ это доказательство по-

казал, что в его основе содержится ошибка «логического круга», делающая вывод неэффективным. Дело в том, что в качестве исходной посылки Шредингер использовал законы Грассмана, сформулированные с привлечением понятий сложения цветов и умножения цвета на постоянные числа.

Однако легко доказать, что законность введения операций сложения цветов и умножения цвета на постоянные числа не может быть строго обоснована до тех пор, пока не будет доказана справедливость модели в виде соотношений (6), (7). Действительно, законность введения операций сложения цветов и умножения цвета на постоянные числа может быть обоснована только в том случае, если под цветом понимать вектор \bar{B} , компоненты B_1, B_2, B_3 которого линейно зависят от излучения $b(\lambda)$, то есть, иными словами, определяются соотношениями (6) (поскольку других линейных зависимостей не существует [14, с. 180]). Однако вектор \bar{B} может рассматриваться в качестве характеристики цвета лишь в том случае, если имеется уверенность в том, что одинаковым векторам \bar{B} соответствуют одинаковые цвета и одинаковым цветам соответствуют одинаковые векторы \bar{B} . Таким образом, мы приходим к необходимости использования зависимости (7).

Следовательно, для строгого обоснования законности введения операций сложения цветов и умножения цвета на постоянные числа необходимо прежде признать справедливой модель цветового зрения в виде соотношений (6) и (7). Поэтому при выводе модели цветового зрения в виде соотношений (6) и (7) мы не имеем права пользоваться законами Грассмана, сформулированными с привлечением понятий сложения цветов и умножения цвета на постоянные числа. Тем не менее, сказанное вовсе не означает, что из законов Грассмана не вытекает модель цветового зрения в виде соотношений (6) и (7). Такое утверждение было бы справедливо лишь в том случае, если бы эти законы было невозможно сформулировать без привлечения понятий сложения цветов и умножения цвета на постоянные числа. Однако, как будет показано ниже, такая формулировка законов Грассмана возможна. Мы также докажем, что из этой формулировки законов Грассмана логически следует модель цветового зрения в виде соотношений (6) и (7).

3. Новая формулировка законов Грассмана

Выше было сказано, что использование понятий сложения цветов и умножения цвета на постоянные числа в неявной форме опирается на гипотетическую модель цветового зрения в виде соотношений (6) и (7). Следовательно, и законы Грассмана, сформулированные с использованием тех же понятий, зависят от этой гипотезы. Но в таком случае уместно поставить вопрос: могут ли законы Грассмана, опирающиеся на гипотезу, называться законами. Для того чтобы восстановить законы Грассмана в своих правах, необходимо их сформулировать без привле-

чения понятий сложения цветов и умножения цвета на постоянные числа. Приступая к решению этой задачи, мы будем пользоваться лишь операциями сложения излучений и умножения излучения на постоянные числа. Законность введения этих операций основана на хорошо изученных свойствах света и не зависит от каких-либо гипотез о виде преобразования сигналов в органе зрения.

Пусть имеются два поля сравнения, причем на одном из них сформировано излучение $b_1(\lambda)$, а на другом $b_2(\lambda)$. При предъявлении этих зрительных картин в органе зрения возникают зрительные ощущения, характеризующиеся соответственно цветами \bar{S}_1 и \bar{S}_2 . Для характеристики условий такого опыта удобно ввести функцию $\beta(\lambda)$, равную разности спектров излучений первого и второго полей сравнения:

$$\beta(\lambda) = b_1(\lambda) - b_2(\lambda). \quad (8)$$

Этой функции мы не приписываем никакого физического смысла. Она вводится лишь затем, чтобы с ее помощью можно было более изящно и кратко сформулировать законы Грассмана. Заметим, что каждой функции $\beta(\lambda)$ соответствует не одна, а бесчисленное множество пар излучений вида $b_1(\lambda) + \beta_0(\lambda)$, $b_2(\lambda) + \beta_0(\lambda)$, где $\beta_0(\lambda)$ — произвольная функция длины волны. Таким образом, прибавляя или вычитая (когда это возможно) на полях сравнения одинаковые излучения, мы не меняем значения функции $\beta(\lambda)$. Важно также отметить, что в силу полной равноправности полей сравнения перемена излучений местами соответствует, по существу, одному и тому же опыту. Так что, если для пары излучений $b_1(\lambda)$, $b_2(\lambda)$ наблюдается равенство цветов полей сравнения, то это равенство будет также наблюдаться и для пары излучений $b_2(\lambda)$, $b_1(\lambda)$. С учетом изложенного сформулируем закон аддитивности.

Закон аддитивности. Если в двух опытах с условиями, характеризуемыми функциями $\beta_1(\lambda)$ и $\beta_2(\lambda)$, наблюдается равенство цветов полей сравнения, то во всевозможных опытах с условиями, характеризуемыми функцией $\beta_1(\lambda) + \beta_2(\lambda)$, также будет наблюдаться равенство цветов полей сравнения.

Важно отметить, что в новой редакции закон аддитивности может быть продемонстрирован на опыте. Это можно сделать следующим образом. Берем две произвольные пары излучений $b'_1(\lambda)$, $b'_2(\lambda)$ и $b''_1(\lambda)$, $b''_2(\lambda)$, цвета которых попарно одинаковы (цвет излучения $b'_1(\lambda)$ совпадает с цветом излучения $b'_2(\lambda)$, а цвет излучения $b''_1(\lambda)$ совпадает с цветом излучения $b''_2(\lambda)$). Определяем для этих пар излучений функции, характеризующие условия опытов: $\beta'(\lambda) = b'_1(\lambda) - b'_2(\lambda)$, $\beta''(\lambda) = b''_1(\lambda) - b''_2(\lambda)$, а также суммарную функцию: $\beta(\lambda) = \beta'(\lambda) + \beta''(\lambda)$. Затем берем произвольную пару излучений $b_1(\lambda)$, $b_2(\lambda)$, разность которых дает функцию $\beta(\lambda)$, то есть $b_1(\lambda) - b_2(\lambda) = \beta(\lambda)$, и убеждаемся на опыте, что для этих излучений также будет наблюдаться равенство цветов полей сравнения.

Как следствие из закона аддитивности вытекает, что одинаковые излучения порождают одинаковые цвета. Действительно, пусть на поля сравнения поданы излучения $b'(\lambda)$ и $b''(\lambda)$, для которых наблюдается равенство цветов полей сравнения. Характеристикой условий этого опыта является функция $\beta_1(\lambda)$, равная $\beta_1(\lambda) = b'(\lambda) - b''(\lambda)$. Поменяем излучения местами. Как было отмечено выше, в силу равноправности полей сравнения по-прежнему будет наблюдаться равенство цветов. Характеристикой условий этого опыта служит функция $\beta_2(\lambda)$, равная $\beta_2(\lambda) = b''(\lambda) - b'(\lambda)$. Согласно закону аддитивности, равенство цветов полей сравнения будет наблюдаться также для всевозможных опытов, характеризуемых функцией: $\beta_1(\lambda) + \beta_2(\lambda) = (b'(\lambda) - b''(\lambda)) + (b''(\lambda) - b'(\lambda)) = 0$. Этой функции соответствуют всевозможные пары равных излучений. Следовательно, одинаковые излучения порождают одинаковые цвета.

Это следствие весьма важно в методологическом отношении. Так как одинаковым излучениям, то есть входным сигналам, соответствуют одинаковые цвета, то есть выходные сигналы органа зрения, то между ними существует причинная связь. Этим обосновывается правомерность постановки задачи моделирования стационарных и однородных зрительных процессов. Предположим, что на полях сравнения сформированы излучения, цвета которых одинаковы. Из закона аддитивности следует, что добавление или вычитание на этих полях одинаковых излучений в результате снова дает равенство цветов. Именно это обстоятельство делает естественным введение функций $\beta(\lambda)$, определяемых равенством (4), для характеристики условий опыта. Сформулируем теперь закон трехмерности.

Закон трехмерности. Существует тройка фиксированных функций $\beta_1(\lambda)$, $\beta_2(\lambda)$, $\beta_3(\lambda)$ такая, что для любой функции $\beta(\lambda)$ можно подыскать, и притом единственным образом, тройку чисел U_1 , U_2 , U_3 таких, что в опытах с условиями, характеризуемыми функцией

$$U_1\beta_1(\lambda) + U_2\beta_2(\lambda) + U_3\beta_3(\lambda) + \beta(\lambda), \quad (9)$$

будет наблюдаться равенство цветов полей сравнения.

Закон трехмерности может быть продемонстрирован на опыте следующим образом. Пусть нам даны три пары специально подобранных излучений $b'_1(\lambda)$, $b''_1(\lambda)$; $b'_2(\lambda)$, $b''_2(\lambda)$; $b'_3(\lambda)$, $b''_3(\lambda)$, разности которых обозначим через $\beta_1(\lambda)$, $\beta_2(\lambda)$, $\beta_3(\lambda)$, то есть $\beta_1(\lambda) = b'_1(\lambda) - b''_1(\lambda)$; $\beta_2(\lambda) = b'_2(\lambda) - b''_2(\lambda)$; $\beta_3(\lambda) = b'_3(\lambda) - b''_3(\lambda)$. Закон трехмерности утверждает, что такие три пары излучений всегда могут быть найдены. Пусть, кроме того, дана произвольная пара излучений $b'(\lambda)$, $b''(\lambda)$, разность которых обозначим через $\beta(\lambda)$, то есть $\beta(\lambda) = b'(\lambda) - b''(\lambda)$. Подадим на поля сравнения излучения $b'(\lambda)$ и $b''(\lambda)$. Будем теперь добиваться достижения равенства цветов полей сравнения. С этой целью разрешается излучения каждой из трех пар

одновременно прибавлять на соответствующие поля сравнения, причем в любом порядке и с умножением на любые равные положительные числа. Кроме того, разрешается прибавлять или вычитать (когда это возможно) на полях сравнения любые равные излучения. Производя такие опыты, убеждаемся, что всегда возможен подбор на полях сравнения линейных комбинаций излучений, обеспечивающих равенство цветов полей сравнения. Такому подбору соответствует функция (9), определяемая как разность полученных излучений на полях сравнения. На опыте также убеждаемся, что всякий раз, когда достигается равенство цветов полей сравнения, в функции разности излучений полей сравнения для такого опыта получаются всегда одни и те же коэффициенты U_1, U_2, U_3 . Закон непрерывности сформулируем следующим образом.

Закон непрерывности. Непрерывному изменению функции $\beta(\lambda)$ соответствует непрерывное изменение чисел U_1, U_2, U_3 .

Непрерывность функции $\beta(\lambda)$ понимается в смысле метрики пространства L суммируемых функций [15, с. 150]. Имеются в виду числа U_1, U_2, U_3 , существование и единственность которых постулируется вторым законом. Демонстрация закона трехмерности может быть осуществлена следующим образом. Берем две произвольные пары излучений $b'_1(\lambda), b''_1(\lambda)$ и $b'_2(\lambda), b''_2(\lambda)$ такие, что соответствующие им функции разности излучений $\beta_1(\lambda) = b'_1(\lambda) - b''_1(\lambda)$ и $\beta_2(\lambda) = b'_2(\lambda) - b''_2(\lambda)$ близки в смысле метрики пространства L . Это значит, что расстояние σ между этими функциями, определяемое формулой

$$\sigma = \int_{\lambda_1}^{\lambda_2} |\beta_1(\lambda) - \beta_2(\lambda)| d\lambda,$$

достаточно мало. На опыте убеждаемся, что соответствующие числа $U'_1 - U''_1, U'_2 - U''_2, U'_3 - U''_3$ всегда оказываются достаточно близкими к нулю.

Сравнивая новую формулировку законов Грассмана с прежней, можно видеть, что теперь в формулировке законов совершенно не участвуют операции сложения цветов и умножения цвета на постоянные числа. О цветах утверждается лишь то, что они равны или не равны между собой. Сознание наблюдателя используется при этом лишь как нулевой прибор, фиксирующий равенство или неравенство цветов двух зрительных ощущений. Важно заметить, что стандартные колориметрические опыты фактически выполняются в точности по той процедуре, которая необходима для демонстрации справедливости законов Грассмана в новой формулировке, поскольку в этих опытах операциям сложения и умножения на постоянные коэффициенты подвергаются именно излучения, а не цвета. Цвета же подвергаются единственной операции, состоящей в установлении их равенства или неравенства [16]. Наконец, необходимо заметить, что за отказ от ис-

пользования операций сложения цветов и умножения цвета на постоянные числа приходится расплачиваться более громоздкой формулировкой законов. Однако новая формулировка законов Грассмана, в отличие от прежней, свободна от каких-либо гипотез.

4. Вывод модели из законов Грассмана

Приступим теперь к выводу математической модели цветового зрения в виде соотношений (6) и (7) из законов Грассмана в новой формулировке. Прежде всего докажем, что совокупность функций $\beta(\lambda)$, являющихся разностью спектров пар всевозможных излучений, образует линейное нормированное пространство $L[\lambda_1, \lambda_2]$ суммируемых функций [17, с. 19]. Для этого достаточно показать, что выполняются следующие три условия: 1) для любой функции $\beta(\lambda)$ может быть введена операция умножения на произвольное вещественное число; 2) может быть введена операция сложения двух любых функций $\beta_1(\lambda)$ и $\beta_2(\lambda)$; 3) любой функции $\beta(\lambda)$ может быть поставлено в соответствие конечное число — норма этой функции, определяемая равенством

$$\|\beta\| = \int_{\lambda_1}^{\lambda_2} |\beta(\lambda)| d\lambda \quad (10)$$

и удовлетворяющая аксиомам нормы [18, с. 53].

Возьмем произвольную функцию $\beta(\lambda)$, которую можно представить как разность спектров $b'(\lambda)$ и $b''(\lambda)$ некоторой пары излучений:

$$\beta(\lambda) = b'(\lambda) - b''(\lambda). \quad (11)$$

Образует функцию $m\beta(\lambda)$, где m — любое неотрицательное число. Эту функцию также можно представить в виде разности спектров пары излучений. В качестве таковых можно использовать излучения со спектрами $mb'(\lambda)$ и $mb''(\lambda)$. Действительно,

$$mb'(\lambda) - mb''(\lambda) = m(b'(\lambda) - b''(\lambda)) = m\beta(\lambda). \quad (12)$$

Пусть теперь m — любое отрицательное число. Функцию $m\beta(\lambda)$ также можно представить в виде разности спектров двух излучений, в качестве которых можно принять $-mb''(\lambda)$ и $-mb'(\lambda)$. Действительно,

$$\begin{aligned} -mb''(\lambda) - (-mb'(\lambda)) &= m(b'(\lambda) - b''(\lambda)) = \\ &= m\beta(\lambda). \end{aligned} \quad (13)$$

Таким образом, если функция $\beta(\lambda)$ представима в виде разности спектров пары излучений, то этим свойством обладает также функция $m\beta(\lambda)$, где m — любое вещественное число. Следовательно, первое условие выполняется.

Пусть имеются две функции $\beta_1(\lambda)$ и $\beta_2(\lambda)$, представимые в виде разности спектров пар излучений $b'_1(\lambda), b''_1(\lambda)$ и $b'_2(\lambda), b''_2(\lambda)$:

$$\beta_1(\lambda) = b'_1(\lambda) - b''_1(\lambda), \quad (14)$$

$$\beta_2(\lambda) = b'_2(\lambda) - b''_2(\lambda). \quad (15)$$

Сумма этих функций $\beta_1(\lambda) + \beta_2(\lambda)$ также представима в виде разности спектров пары излучений,

в качестве таковых можно принять спектры излучений $b'_1(\lambda) + b'_2(\lambda)$ и $b''_1(\lambda) + b''_2(\lambda)$. Действительно,

$$b'_1(\lambda) + b'_2(\lambda) - (b''_1(\lambda) + b''_2(\lambda)) = \\ = b'_1(\lambda) - b''_1(\lambda) + (b'_2(\lambda) + b''_2(\lambda)) = \beta_1(\lambda) + \beta_2(\lambda). \quad (16)$$

Таким образом, если функции $\beta_1(\lambda)$ и $\beta_2(\lambda)$ представимы в виде разности спектров пар излучений, то и их сумма $\beta_1(\lambda) + \beta_2(\lambda)$ также обладает этим свойством. Следовательно, второе условие также выполняется.

Пусть имеется функция $\beta(\lambda)$, представимая в виде (11). Норма этой функции определяется по формуле (10) и равна:

$$\|\beta\| = \int_{\lambda_1}^{\lambda_2} |b'(\lambda) - b''(\lambda)| d\lambda \leq \int_{\lambda_1}^{\lambda_2} b'(\lambda) d\lambda + \int_{\lambda_1}^{\lambda_2} b''(\lambda) d\lambda. \quad (17)$$

Интегралы $\int_{\lambda_1}^{\lambda_2} b'(\lambda) d\lambda$ и $\int_{\lambda_1}^{\lambda_2} b''(\lambda) d\lambda$ представляют собой

в некотором масштабе энергию излучений со спектрами $b'(\lambda)$ и $b''(\lambda)$ и, следовательно, всегда имеют конечные значения. Таким образом,

$$\|\beta\| < \infty. \quad (18)$$

Кроме того, выполняются все аксиомы нормы. Действительно, для любой функции $\beta(\lambda)$ имеем: 1) $\|\beta\| \geq 0$, причем $\|\beta\| = 0$, лишь если $\beta(\lambda) = 0$ в смысле метрики L ; 2) $\|m\beta\| = |m| \cdot \|\beta\|$; 3) аксиома треугольника $\|\beta_1 + \beta_2\| \leq \|\beta_1\| + \|\beta_2\|$ следует из неравенства Минковского для интегралов [3, с. 498]. Этим доказано выполнение третьего условия. Итак, мы доказали, что совокупность функций $\beta(\lambda)$ образует линейное нормированное пространство $L[\lambda_1, \lambda_2]$ суммируемых функций.

Докажем теперь, что числа U_1, U_2, U_3 , введенные законом трехмерности, суть линейные непрерывные функционалы функций $\beta(\lambda)$. Для этого достаточно доказать [3, с. 143], что функционалы

$$U_1 = U_1(\beta(\lambda)), U_2 = U_2(\beta(\lambda)), U_3 = U_3(\beta(\lambda))$$

аддитивны и непрерывны. Пусть $\beta(\lambda)$ есть сумма функций $\beta'(\lambda)$ и $\beta''(\lambda)$:

$$\beta(\lambda) = \beta'(\lambda) + \beta''(\lambda). \quad (19)$$

Согласно закону трехмерности, функции $\beta(\lambda)$ соответствует тройка чисел $U_1(\beta(\lambda)), U_2(\beta(\lambda)), U_3(\beta(\lambda))$ такая, что в опытах с условиями, характеризуемыми функцией

$$U_1(\beta(\lambda))\beta_1(\lambda) + U_2(\beta(\lambda))\beta_2(\lambda) + \\ + U_3(\beta(\lambda))\beta_3(\lambda) + \beta(\lambda), \quad (20)$$

наблюдается равенство цветов полей сравнения.

С другой стороны, согласно закону трехмерности, функциям $\beta'(\lambda)$ и $\beta''(\lambda)$ также соответствуют свои тройки чисел $U_1(\beta'(\lambda)), U_2(\beta'(\lambda)), U_3(\beta'(\lambda))$ и $U_1(\beta''(\lambda)), U_2(\beta''(\lambda)), U_3(\beta''(\lambda))$ такие, что в опытах с условиями, характеризуемыми функциями:

$$U_1(\beta'(\lambda))\beta_1(\lambda) + U_2(\beta'(\lambda))\beta_2(\lambda) + \\ + U_3(\beta'(\lambda))\beta_3(\lambda) + \beta'(\lambda), \quad (21)$$

$$U_1(\beta''(\lambda))\beta_1(\lambda) + U_2(\beta''(\lambda))\beta_2(\lambda) + \\ + U_3(\beta''(\lambda))\beta_3(\lambda) + \beta''(\lambda), \quad (22)$$

наблюдается равенство цветов полей сравнения. Образум сумму функций (21) и (22):

$$(U_1(\beta'(\lambda)) + U_1(\beta''(\lambda)))\beta_1(\lambda) + (U_2(\beta'(\lambda)) + \\ + U_2(\beta''(\lambda)))\beta_2(\lambda) + (U_3(\beta'(\lambda)) + \\ + U_3(\beta''(\lambda)))\beta_3(\lambda) + \beta'(\lambda) + \beta''(\lambda). \quad (23)$$

Согласно закону аддитивности, в опытах с условиями, характеризуемыми функцией (23), имеет место равенство цветов полей сравнения.

В выражениях (20) и (23) функции $\beta(\lambda)$ и $\beta'(\lambda) + \beta''(\lambda)$, согласно (19), совпадают. Следовательно, согласно закону трехмерности в этих выражениях совпадают коэффициенты при функциях $\beta_1(\lambda), \beta_2(\lambda), \beta_3(\lambda)$, то есть

$$U_1(\beta(\lambda)) = U_1(\beta'(\lambda)) + U_1(\beta''(\lambda)), \\ U_2(\beta(\lambda)) = U_2(\beta'(\lambda)) + U_2(\beta''(\lambda)), \quad (24) \\ U_3(\beta(\lambda)) = U_3(\beta'(\lambda)) + U_3(\beta''(\lambda)).$$

Заменив $\beta(\lambda)$ согласно равенству (19), окончательно получим:

$$U_1(\beta'(\lambda) + \beta''(\lambda)) = U_1(\beta'(\lambda)) + U_1(\beta''(\lambda)), \\ U_2(\beta'(\lambda) + \beta''(\lambda)) = U_2(\beta'(\lambda)) + U_2(\beta''(\lambda)), \quad (25) \\ U_3(\beta'(\lambda) + \beta''(\lambda)) = U_3(\beta'(\lambda)) + U_3(\beta''(\lambda)).$$

Равенства (25) означают, что функционалы $U_1(\beta(\lambda)), U_2(\beta(\lambda)), U_3(\beta(\lambda))$ аддитивны. Непрерывность этих функционалов непосредственно вытекает из закона непрерывности. Таким образом, числа U_1, U_2, U_3 суть линейные непрерывные функционалы, определенные на линейном нормированном пространстве $L[\lambda_1, \lambda_2]$ суммируемых функций $\beta(\lambda)$.

При этих условиях, согласно теореме об общем виде линейного функционала [14, с. 180], функционалы U_1, U_2, U_3 могут иметь лишь следующий вид:

$$U_1 = \int_{\lambda_1}^{\lambda_2} b(\lambda)m(\lambda)d\lambda, \\ U_2 = \int_{\lambda_1}^{\lambda_2} b(\lambda)n(\lambda)d\lambda, \quad (26) \\ U_3 = \int_{\lambda_1}^{\lambda_2} b(\lambda)p(\lambda)d\lambda,$$

где $m(\lambda), n(\lambda), p(\lambda)$ — некоторые фиксированные функции (эти функции могут выбираться только из класса существенно ограниченных измеримых функций, включающих в себя как частный случай

класс непрерывных функций [15, с. 99]). Других видов линейных непрерывных функционалов, определенных на линейном нормированном пространстве $L[\lambda_1, \lambda_2]$ суммируемых функций $\beta(\lambda)$, не существует.

Можно показать, что функции $m(\lambda)$, $n(\lambda)$, $p(\lambda)$ линейно независимы. Действительно, в случае их линейной зависимости одну из них (пусть это будет, к примеру, функция $p(\lambda)$) можно было бы выразить в виде линейной комбинации остальных:

$$p(\lambda) = a_1 m(\lambda) + a_2 n(\lambda), \quad (27)$$

где a_1, a_2 — некоторые фиксированные вещественные числа. Подставляя (27) в последнее из равенств (26), имеем:

$$U_3 = a_1 \int_{\lambda_1}^{\lambda_2} \beta(\lambda) m(\lambda) d\lambda + a_2 \int_{\lambda_1}^{\lambda_2} \beta(\lambda) n(\lambda) d\lambda. \quad (28)$$

Используя первые два из равенств (26), перепишем (28) в виде:

$$U_3 = a_1 U_1 + a_2 U_2. \quad (29)$$

Таким образом, для любой функции $\beta(\lambda)$ число U_3 однозначно определяется числами U_1 и U_2 по формуле (29). Если, к примеру, $U_1 = U_2 = 0$, то должно быть также $U_3 = 0$. Однако это не так, поскольку для функции $\beta_3(\lambda)$ $U_1 = U_2 = 0$, но $U_3 = -1$, и этот набор чисел, согласно закону трехмерности, является единственным. Итак, мы пришли к противоречию. Следовательно, функции $m(\lambda)$, $n(\lambda)$, $p(\lambda)$ линейно независимы.

Введем теперь в рассмотрение вектор \bar{B} с компонентами B_1, B_2, B_3 , определяемыми для произвольного излучения $b(\lambda)$ с помощью соотношений (6). Можно доказать, что вектор \bar{B} связан взаимно однозначной зависимостью (7) с вектором цвета \bar{S} . Для доказательства этого утверждения нужно установить, что: 1) излучения, имеющие одинаковые значения интегралов (6), порождают одинаковые цвета; 2) излучения, порождающие одинаковые цвета, имеют одинаковые значения интегралов (6).

Докажем справедливость первого утверждения. Возьмем для этого пару излучений $b'(\lambda)$ и $b''(\lambda)$, имеющих одинаковые значения интегралов (6):

$$B'_1 = B''_1, B'_2 = B''_2, B'_3 = B''_3, \quad (30)$$

где

$$\begin{aligned} B'_1 &= \int_{\lambda_1}^{\lambda_2} b'(\lambda) m(\lambda) d\lambda, \\ B'_2 &= \int_{\lambda_1}^{\lambda_2} b'(\lambda) n(\lambda) d\lambda, \\ B'_3 &= \int_{\lambda_1}^{\lambda_2} b'(\lambda) p(\lambda) d\lambda, \end{aligned} \quad (31)$$

$$\begin{aligned} B''_1 &= \int_{\lambda_1}^{\lambda_2} b''(\lambda) m(\lambda) d\lambda, \\ B''_2 &= \int_{\lambda_1}^{\lambda_2} b''(\lambda) n(\lambda) d\lambda, \\ B''_3 &= \int_{\lambda_1}^{\lambda_2} b''(\lambda) p(\lambda) d\lambda, \end{aligned} \quad (32)$$

Введем в рассмотрение функцию $\beta(\lambda) = b'(\lambda) - b''(\lambda)$ и определим для нее значения интегралов (26):

$$\begin{aligned} U_1 &= \int_{\lambda_1}^{\lambda_2} \beta(\lambda) m(\lambda) d\lambda = \int_{\lambda_1}^{\lambda_2} (b'(\lambda) - b''(\lambda)) m(\lambda) d\lambda, \\ U_2 &= \int_{\lambda_1}^{\lambda_2} \beta(\lambda) n(\lambda) d\lambda = \int_{\lambda_1}^{\lambda_2} (b'(\lambda) - b''(\lambda)) n(\lambda) d\lambda, \\ U_3 &= \int_{\lambda_1}^{\lambda_2} \beta(\lambda) p(\lambda) d\lambda = \int_{\lambda_1}^{\lambda_2} (b'(\lambda) - b''(\lambda)) p(\lambda) d\lambda. \end{aligned} \quad (33)$$

Используя формулы (30)–(32), получаем:

$$\begin{aligned} U_1 &= \int_{\lambda_1}^{\lambda_2} b'(\lambda) m(\lambda) d\lambda - \int_{\lambda_1}^{\lambda_2} b''(\lambda) m(\lambda) d\lambda = B'_1 - B''_1 = 0, \\ U_2 &= \int_{\lambda_1}^{\lambda_2} b'(\lambda) n(\lambda) d\lambda - \int_{\lambda_1}^{\lambda_2} b''(\lambda) n(\lambda) d\lambda = B'_2 - B''_2 = 0, \\ U_3 &= \int_{\lambda_1}^{\lambda_2} b'(\lambda) p(\lambda) d\lambda - \int_{\lambda_1}^{\lambda_2} b''(\lambda) p(\lambda) d\lambda = B'_3 - B''_3 = 0. \end{aligned} \quad (34)$$

Таким образом, для функции $\beta(\lambda) = b'(\lambda) - b''(\lambda)$ числа U_1, U_2, U_3 равны нулю. Согласно закону трехмерности, для любой пары излучений, соответствующей функции $\beta(\lambda)$, имеет место равенство цветов полей сравнения. Одной из таких пар излучений являются излучения $b'(\lambda)$ и $b''(\lambda)$, следовательно, для них также имеет место равенство цветов полей сравнения. Итак, мы приходим к выводу, что излучения, имеющие одинаковые значения интегралов (6), порождают одинаковые цвета.

Докажем теперь справедливость второго утверждения. Пусть излучения $b'(\lambda)$ и $b''(\lambda)$ порождают одинаковые цвета. Тогда для функции $\beta(\lambda) = b'(\lambda) - b''(\lambda)$ имеет место равенство цветов полей сравнения. Следовательно, согласно закону трехмерности, для этой функции значения интегралов (26) U_1, U_2, U_3 равны нулю:

$$\begin{aligned} U_1 &= \int_{\lambda_1}^{\lambda_2} \beta(\lambda) m(\lambda) d\lambda = \int_{\lambda_1}^{\lambda_2} (b'(\lambda) - b''(\lambda)) m(\lambda) d\lambda = 0, \\ U_2 &= \int_{\lambda_1}^{\lambda_2} \beta(\lambda) n(\lambda) d\lambda = \int_{\lambda_1}^{\lambda_2} (b'(\lambda) - b''(\lambda)) n(\lambda) d\lambda = 0, \\ U_3 &= \int_{\lambda_1}^{\lambda_2} \beta(\lambda) p(\lambda) d\lambda = \int_{\lambda_1}^{\lambda_2} (b'(\lambda) - b''(\lambda)) p(\lambda) d\lambda = 0. \end{aligned} \quad (35)$$

Отсюда вытекает, что

$$\begin{aligned} \int_{\lambda_1}^{\lambda_2} (b'(\lambda)m(\lambda)d\lambda &= \int_{\lambda_1}^{\lambda_2} (b''(\lambda)m(\lambda)d\lambda, \\ \int_{\lambda_1}^{\lambda_2} (b'(\lambda)n(\lambda)d\lambda &= \int_{\lambda_1}^{\lambda_2} (b''(\lambda)n(\lambda)d\lambda, \\ \int_{\lambda_1}^{\lambda_2} (b'(\lambda)p(\lambda)d\lambda &= \int_{\lambda_1}^{\lambda_2} (b''(\lambda)p(\lambda)d\lambda. \end{aligned} \quad (36)$$

Используя зависимости (31) и (34), приходим к выводу, что

$$B'_1 = B''_1, B'_2 = B''_2, B'_3 = B''_3. \quad (37)$$

Следовательно, излучения, порождающие одинаковые цвета, имеют одинаковые значения интегралов (6).

Итак, мы доказали, что вектор \bar{B} связан взаимно однозначной зависимостью (7) с вектором цвета \bar{S} . Таким образом, доказано, что из законов Грассмана логически вытекает математическая модель цветового зрения в виде соотношения (6) и (7). Только теперь, используя равенства (6) и (7), мы получаем право рассматривать вектор \bar{B} как вектор цвета без привлечения каких-либо гипотез и можем обоснованно ввести понятия сложения цветов и умножения цвета на постоянные числа, лежащие в основе колориметрии. Важно отметить, что утверждение о трехмерности вектора \bar{S} до сих пор принималось нами без доказательства. Теперь это положение может быть строго доказано: оно вытекает как логическое следствие из факта трехмерности вектора \bar{B} и наличия взаимно-однозначной зависимости (7) вектора \bar{S} от вектора \bar{B} .

5. Вывод законов Грассмана из модели

В предыдущем разделе мы доказали, что из законов Грассмана можно чисто логически вывести математическую модель цветового зрения в виде соотношений (6) и (7). Теперь напрашивается вопрос: нельзя ли из законов Грассмана вывести нечто большее, чем эту модель? Может быть, в законах Грассмана заключена какая-либо дополнительная информация, не нашедшая еще отражения в модели? Здесь мы докажем, что на этот вопрос должен быть дан отрицательный ответ. Законы Грассмана, с одной стороны, и математическая модель цветового зрения в виде равенств (6) и (7), с другой стороны, представляют собой равносильные утверждения. Математическая модель цветового зрения в виде соотношений (6) и (7) есть лишь переформулировка законов Грассмана, равно как и законы Грассмана являются переформулировкой математической модели цветового зрения в виде равенств (6) и (7). Чтобы доказать это утверждение, нам достаточно вывести законы Грассмана в виде логического следствия из математической модели цветового зрения, представленной в виде соотношений (6) и (7).

Выведем сначала из равенств (6) и (7) закон аддитивности. Предположим, что две пары излучений со спектрами $b'_1(\lambda)$, $b'_2(\lambda)$ и $b''_1(\lambda)$, $b''_2(\lambda)$ порождают

на выходе модели попарно одинаковые сигналы \bar{S}'_1 , \bar{S}'_2 и \bar{S}''_1 , \bar{S}''_2 , то есть что

$$\bar{S}'_1 = \bar{S}'_2, \bar{S}''_1 = \bar{S}''_2. \quad (38)$$

Попарные разности спектров излучений обозначим через $\beta'(\lambda)$ и $\beta''(\lambda)$:

$$\beta'(\lambda) = b'_1(\lambda) - b'_2(\lambda), \beta''(\lambda) = b''_1(\lambda) - b''_2(\lambda). \quad (39)$$

Образует функцию $\beta(\lambda)$, равную сумме функций $\beta'(\lambda)$ и $\beta''(\lambda)$:

$$\beta(\lambda) = \beta'(\lambda) + \beta''(\lambda). \quad (40)$$

Требуется доказать, что любая пара излучений со спектрами $b_1(\lambda)$ и $b_2(\lambda)$, разность которых дает функцию

$$b_1(\lambda) - b_2(\lambda) = \beta(\lambda), \quad (41)$$

порождает на выходе модели пару одинаковых сигналов \bar{S}_1 и \bar{S}_2 , то есть что $\bar{S}_1 = \bar{S}_2$. Для доказательства этого утверждения, ввиду наличия взаимно однозначной зависимости (7), достаточно показать, что совпадают векторы \bar{B}_1 и \bar{B}_2 , соответствующие сигналам \bar{S}_1 и \bar{S}_2 , то есть что $\bar{B}_1 = \bar{B}_2$. Обозначим компоненты векторов \bar{B}_1 и \bar{B}_2 соответственно через B_{11} , B_{12} , B_{13} и B_{21} , B_{22} , B_{23} . Совпадение векторов \bar{B}_1 и \bar{B}_2 будет доказано, если мы установим, что:

$$B_{11} - B_{21} = 0, B_{12} - B_{22} = 0, B_{13} - B_{23} = 0. \quad (42)$$

Приступим к доказательству справедливости равенств (42). Согласно равенствам (6), имеем:

$$\begin{aligned} B_{11} - B_{21} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda))m(\lambda)d\lambda, \\ B_{12} - B_{22} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda))n(\lambda)d\lambda, \\ B_{13} - B_{23} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda))p(\lambda)d\lambda. \end{aligned} \quad (43)$$

Заменяя в (43) разность $b_1(\lambda) - b_2(\lambda)$ по формулам (39)–(41), получим следующие выражения:

$$\begin{aligned} B_{11} - B_{21} &= \int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) - b'_2(\lambda))m(\lambda)d\lambda + \\ &+ \int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) - b''_2(\lambda))m(\lambda)d\lambda, \\ B_{12} - B_{22} &= \int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) - b'_2(\lambda))n(\lambda)d\lambda + \\ &+ \int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) - b''_2(\lambda))n(\lambda)d\lambda, \\ B_{13} - B_{23} &= \int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) - b'_2(\lambda))p(\lambda)d\lambda + \\ &+ \int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) - b''_2(\lambda))p(\lambda)d\lambda. \end{aligned} \quad (44)$$

Все интегралы, стоящие в правой части формул (44), равны нулю, поскольку по условию (38) излучения

$b'_1(\lambda), b'_2(\lambda)$ и $b''_1(\lambda), b''_2(\lambda)$ дают попарно одинаковые цвета, а это значит, что и интегралы (6) от них попарно одинаковы. Этим доказывается справедливость формул (42). Вывод закона аддитивности из равенств (6) и (7) завершен.

Переходим к выводу закона трехмерности из соотношений (6) и (7). В качестве тройки фиксированных функций примем функции $m(\lambda), n(\lambda), p(\lambda)$. Нам нужно доказать, что для произвольной функции $\beta(\lambda)$ всегда можно построить функцию:

$$U_1 m(\lambda) + U_2 n(\lambda) + U_3 p(\lambda) + \beta(\lambda), \quad (45)$$

притом с единственно возможным набором чисел U_1, U_2, U_3 , такую, что для любой пары излучений со спектрами $b_1(\lambda), b_2(\lambda)$, разность которых равна функции (45):

$$\begin{aligned} b_1(\lambda) - b_2(\lambda) = \\ = U_1 m(\lambda) + U_2 n(\lambda) + U_3 p(\lambda) + \beta(\lambda), \end{aligned} \quad (46)$$

будут совпадать порожденные этими излучениями выходные сигналы модели \bar{S}_1 и \bar{S}_2 . Для доказательства этого утверждения сначала установим, что если совпадение выходных сигналов модели \bar{S}_1 и \bar{S}_2 возможно, то для каждой функции $\beta(\lambda)$ оно достигается лишь при единственном наборе чисел U_1, U_2, U_3 .

Действительно, пусть выходные сигналы модели \bar{S}_1 и \bar{S}_2 совпадают. Это равносильно попарным равенствам координат B_{11}, B_{12}, B_{13} и B_{21}, B_{22}, B_{23} соответствующих векторов \bar{B}_1 и \bar{B}_2 , которые ввиду (6) можно записать в виде условий:

$$\begin{aligned} B_{11} - B_{21} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda)) m(\lambda) d\lambda = 0, \\ B_{12} - B_{22} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda)) n(\lambda) d\lambda = 0, \\ B_{13} - B_{23} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda)) p(\lambda) d\lambda = 0. \end{aligned} \quad (47)$$

После подстановки по формуле (46) равенства (47) примут вид:

$$\begin{aligned} U_1 \int_{\lambda_1}^{\lambda_2} m(\lambda) m(\lambda) d\lambda + U_2 \int_{\lambda_1}^{\lambda_2} n(\lambda) m(\lambda) d\lambda + \\ + U_3 \int_{\lambda_1}^{\lambda_2} p(\lambda) m(\lambda) d\lambda = - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) m(\lambda) d\lambda, \\ U_1 \int_{\lambda_1}^{\lambda_2} m(\lambda) n(\lambda) d\lambda + U_2 \int_{\lambda_1}^{\lambda_2} n(\lambda) n(\lambda) d\lambda + \\ + U_3 \int_{\lambda_1}^{\lambda_2} p(\lambda) n(\lambda) d\lambda = - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) n(\lambda) d\lambda, \\ U_1 \int_{\lambda_1}^{\lambda_2} m(\lambda) p(\lambda) d\lambda + U_2 \int_{\lambda_1}^{\lambda_2} n(\lambda) p(\lambda) d\lambda + \\ + U_3 \int_{\lambda_1}^{\lambda_2} p(\lambda) p(\lambda) d\lambda = - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) p(\lambda) d\lambda \end{aligned}$$

В силу линейной независимости функций $m(\lambda), n(\lambda), p(\lambda)$, составленный из них определитель Грама [19, с. 226]

$$D = \begin{vmatrix} \int_{\lambda_1}^{\lambda_2} m(\lambda) m(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) m(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) m(\lambda) d\lambda \\ \int_{\lambda_1}^{\lambda_2} m(\lambda) n(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) n(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) n(\lambda) d\lambda \\ \int_{\lambda_1}^{\lambda_2} m(\lambda) p(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) p(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) p(\lambda) d\lambda \end{vmatrix} \quad (49)$$

не равен нулю [20, с. 71]. Вместе с тем, этот определитель совпадает с определителем системы трех линейных уравнений (48) относительно неизвестных U_1, U_2, U_3 . Следовательно, для каждой функции $\beta(\lambda)$ решение системы (48) существует, и оно единственно. Итак, мы доказали, что если совпадение выходных сигналов модели \bar{S}_1 и \bar{S}_2 возможно, то оно достигается лишь при единственном наборе чисел U_1, U_2, U_3 .

Докажем теперь, что при наборе чисел U_1, U_2, U_3 , являющемся решением системы уравнений (48), действительно имеет место совпадение выходных сигналов модели \bar{S}_1 и \bar{S}_2 . С этой целью введем в рассмотрение следующие определители:

$$D_1 = \begin{vmatrix} - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) m(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) m(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) m(\lambda) d\lambda \\ - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) n(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) n(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) n(\lambda) d\lambda \\ - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) p(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) p(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) p(\lambda) d\lambda \end{vmatrix}, \quad (50)$$

$$D_2 = \begin{vmatrix} \int_{\lambda_1}^{\lambda_2} m(\lambda) m(\lambda) d\lambda & - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) m(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) m(\lambda) d\lambda \\ \int_{\lambda_1}^{\lambda_2} m(\lambda) n(\lambda) d\lambda & - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) n(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) n(\lambda) d\lambda \\ \int_{\lambda_1}^{\lambda_2} m(\lambda) p(\lambda) d\lambda & - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) p(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} p(\lambda) p(\lambda) d\lambda \end{vmatrix}. \quad (51)$$

$$D_3 = \begin{vmatrix} \int_{\lambda_1}^{\lambda_2} m(\lambda) m(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) m(\lambda) d\lambda & - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) m(\lambda) d\lambda \\ \int_{\lambda_1}^{\lambda_2} m(\lambda) n(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) n(\lambda) d\lambda & - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) n(\lambda) d\lambda \\ \int_{\lambda_1}^{\lambda_2} m(\lambda) p(\lambda) d\lambda & \int_{\lambda_1}^{\lambda_2} n(\lambda) p(\lambda) d\lambda & - \int_{\lambda_1}^{\lambda_2} \beta(\lambda) p(\lambda) d\lambda \end{vmatrix}. \quad (52)$$

Система уравнений (48) имеет следующее решение:

$$U_1 = \frac{D_1}{D}, \quad U_2 = \frac{D_2}{D}, \quad U_3 = \frac{D_3}{D}. \quad (53)$$

Согласно (6), попарные разности соответствующих координат векторов \bar{B}_1 и \bar{B}_2 равны:

$$\begin{aligned} B_{11} - B_{21} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda)) m(\lambda) d\lambda, \\ B_{12} - B_{22} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda)) n(\lambda) d\lambda, \\ B_{13} - B_{23} &= \int_{\lambda_1}^{\lambda_2} (b_1(\lambda) - b_2(\lambda)) p(\lambda) d\lambda. \end{aligned} \quad (54)$$

Производя в формулах (54) подстановку согласно (46) и (53), получим:

$$\begin{aligned} B_{11} - B_{21} &= \\ &= \int_{\lambda_1}^{\lambda_2} \left(\frac{D_1}{D} m(\lambda) + \frac{D_2}{D} n(\lambda) + \frac{D_3}{D} p(\lambda) + \beta(\lambda) \right) m(\lambda) d\lambda, \\ B_{12} - B_{22} &= \\ &= \int_{\lambda_1}^{\lambda_2} \left(\frac{D_1}{D} m(\lambda) + \frac{D_2}{D} n(\lambda) + \frac{D_3}{D} p(\lambda) + \beta(\lambda) \right) n(\lambda) d\lambda, \\ B_{13} - B_{23} &= \\ &= \int_{\lambda_1}^{\lambda_2} \left(\frac{D_1}{D} m(\lambda) + \frac{D_2}{D} n(\lambda) + \frac{D_3}{D} p(\lambda) + \beta(\lambda) \right) p(\lambda) d\lambda. \end{aligned} \quad (55)$$

Непосредственное вычисление показывает, что выражения, стоящие в правой части равенств (55), равны нулю. Следовательно, соответствующие координаты векторов \bar{B}_1 и \bar{B}_2 совпадают. Поэтому совпадают и сами векторы \bar{B}_1 и \bar{B}_2 . В силу соотношения (7) имеет место также совпадение сигналов \bar{S}_1 и \bar{S}_2 . Этим доказана выводимость закона трехмерности из равенств (6) и (7).

Для доказательства выводимости из соотношений (6) и (7) закона непрерывности достаточно обратить внимание на то, что при непрерывном изменении функции $\beta(\lambda)$ непрерывно изменяются значения свободных членов

$$-\int_{\lambda_1}^{\lambda_2} \beta(\lambda) m(\lambda) d\lambda, \quad -\int_{\lambda_1}^{\lambda_2} \beta(\lambda) n(\lambda) d\lambda, \quad -\int_{\lambda_1}^{\lambda_2} \beta(\lambda) p(\lambda) d\lambda$$

системы (48). Вместе с ними непрерывно изменяется решение системы (48) — числа U_1, U_2, U_3 . Итак, мы доказали выводимость законов Грассмана из соотношений (6) и (7), то есть из математической модели цветового зрения. Вместе с тем, в предыдущем разделе было установлено, что справедливо также обратное утверждение — выводимость модели из законов Грассмана. Следовательно, модель цветового зрения в виде соотношений (6) и (7) и законы Грассмана являются равносильными утверждениями. Поэтому из законов Грассмана нельзя вывести ничего такого, чего бы не содержала в себе модель цветового зрения в виде соотношений (6) и (7).

6. Обобщение модели цветового зрения на случай произвольных зрительных картин

Важно подчеркнуть, что математическая модель цветового зрения в виде соотношений (6) и (7) определена и имеет силу пока лишь для весьма узкого класса входных сигналов, а именно для стационарных и однородных зрительных картин $b(\lambda)$. Однако интересно узнать, можно ли сохранить в модели те же самые преобразования в виде интегралов (6) при допущении произвольных входных сигналов $b_\lambda(x, y, t)$, изменяющихся как в поле зрения, так и во времени. На этот вопрос можно ответить утвердительно, если опереться на приведенный ниже постулат.

Постулат 1. Зрительное ощущение $\bar{S}(x, y, t)$ останется тем же, если в порождающей его зрительной картине $b_\lambda(x, y, t)$ произвольным образом произвести замену излучений на любые метамерные излучения.

Под метамерными [20] понимаются излучения, для которых числа B_1, B_2, B_3 , вычисляемые по формулам (6), совпадают. В формулировке этого постулата использована высказанная Ньюбергом в работе [20] идея о замещении метамерных излучений. Важно заметить, что при экспериментальной проверке этого постулата сознание наблюдателя используется как нулевой прибор. Однако теперь происходит сравнение не двух трехмерных векторов, а объектов неизмеримо более сложной природы — двух вектор-функций трех независимых действительных переменных $\bar{B}_1(x, y, t)$ и $\bar{B}_2(x, y, t)$.

Подтверждается ли этот постулат в психофизическом эксперименте? В литературе отсутствуют какие-либо экспериментальные данные, которые могли бы набросить тень сомнения на справедливость этого постулата. В то же время отсутствуют и специальные исследования по его проверке. В некоторых частных случаях справедливость сформулированного принципа не вызывает сомнений. Так, при временной скачкообразной замене излучения $b'_\lambda(x, y)$ на метамерное излучение $b''_\lambda(x, y)$ сознание не замечает каких-либо временных изменений в зрительном ощущении [23, 24]. Сознание наблюдателя также не обнаруживает границу раздела между двумя метамерными излучениями.

Из сформулированного принципа непосредственно вытекает модель преобразования информации в органе зрения в виде блок-схемы, изображенной на рис. 3.

Математическое описание модели следующее:

$$\begin{aligned} \text{блок } 1_1 & - \\ B_1(x, y, t) &= \int_{\lambda_1}^{\lambda_2} b_\lambda(x, y, t) m(\lambda) d\lambda, \\ \text{блок } 1_2 & - \\ B_2(x, y, t) &= \int_{\lambda_1}^{\lambda_2} b_\lambda(x, y, t) n(\lambda) d\lambda, \\ \text{блок } 1_3 & - \\ B_3(x, y, t) &= \int_{\lambda_1}^{\lambda_2} b_\lambda(x, y, t) p(\lambda) d\lambda, \end{aligned} \quad (56)$$

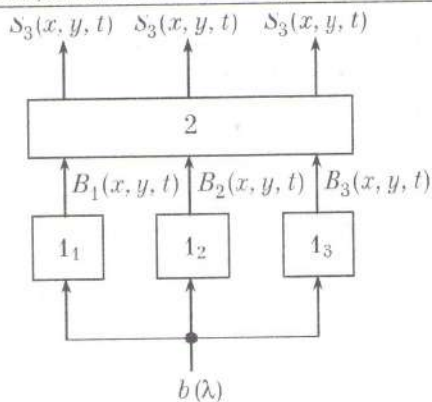


Рис. 3. Обобщенная математическая модель цветового зрения

где $B_1(x, y, t)$, $B_2(x, y, t)$, $B_3(x, y, t)$ — компоненты вектор-функции $\vec{B}(x, y, t)$ — промежуточного сигнала модели; блок 2 —

$$\bar{S}(x, y, t) = F(\vec{B}(x, y, t)), \quad (57)$$

где F — произвольный однозначный оператор; $\bar{S}(x, y, t)$ — вектор-функция зрительного ощущения с компонентами $S_1(x, y, t)$, $S_2(x, y, t)$, $S_3(x, y, t)$.

Важно отметить, что оператор F нельзя полагать взаимно однозначным. Если бы это было так, то любая замена излучений в зрительной картине, кроме той, которая предусмотрена постулатом 1, приводила бы к изменению зрительного ощущения. Однако это заведомо не так. Заменяя стационарное излучение специально подобранной парой нематамерных достаточно быстро мелькающих излучений, можно получить то же самое зрительное ощущение. Такой же эффект можно получить при замещающем нематамерном излучении в виде серии достаточно густых полос (временной и пространственный законы Талбота [25]). Постулат 1 можно дополнить следующим постулатом, обеспечивающим непрерывность оператора F .

Постулат 2. При непрерывном изменении функций $B_1(x, y, t)$, $B_2(x, y, t)$, $B_3(x, y, t)$, определяемых формулами (56), всегда происходит непрерывное изменение соответствующего зрительного ощущения $\bar{S}(x, y, t)$.

Экспериментальная проверка этого постулата возможна ввиду наличия у сознания способности оценивать величину расстояния между двумя вектор-функциями трех переменных $\vec{S}_1(x, y, t)$ и $\vec{S}_2(x, y, t)$, то есть степень различия между двумя зрительными ощущениями. Подтверждение этого постулата на примере простых зрительных картин вселяет уверенность в его справедливости также и в общем виде. Принимая этот постулат, мы имеем право считать оператор F непрерывным.

7. Анализ литературных данных

7.1. Критика вывода Шредингера интегральных соотношений из законов Грассмана

В работе [12] на стр. 411 Шредингер формулирует закон аддитивности в следующем виде: «Одинаково выглядящие излучения дают при сложении снова одинаково выглядящие излучения». Непосредст-

венно за этим Шредингер пишет: «Этот факт — и только он — позволяет нам ... оперировать непосредственно с цветами вместо того, чтобы оперировать с излучениями». Затем Шредингер вводит обозначения для цветов и знак «+» для операции сложения цветов, считая, очевидно, что он имеет для этого все основания. Однако легко показать, что из закона аддитивности еще не следует существование операции сложения цветов.

Сущность только что приведенной формулировки закона аддитивности состоит в следующем. Пусть имеются две пары излучений $b'_1(\lambda)$, $b''_1(\lambda)$ и $b'_2(\lambda)$, $b''_2(\lambda)$, цвета которых попарно одинаковы, то есть что:

$$\bar{S}(b'_1(\lambda)) = \bar{S}(b''_1(\lambda)), \quad \bar{S}(b'_2(\lambda)) = \bar{S}(b''_2(\lambda)). \quad (58)$$

Закон аддитивности утверждает, что цвета суммарных излучений $b'_1(\lambda) + b'_2(\lambda)$, $b''_1(\lambda) + b''_2(\lambda)$ будут также одинаковы, то есть что:

$$\bar{S}(b'_1(\lambda) + b'_2(\lambda)) = \bar{S}(b''_1(\lambda) + b''_2(\lambda)). \quad (59)$$

Однако для строгого введения операции сложения цветов необходима аддитивность цвета, записываемая равенством:

$$\bar{S}(b_1(\lambda) + b_2(\lambda)) = \bar{S}(b_1(\lambda)) + (b_2(\lambda)). \quad (60)$$

В формуле же (60) $b_1(\lambda)$, $b_2(\lambda)$ обозначают два произвольных значения.

Очевидно, что закон аддитивности, с одной стороны, и свойство (60) аддитивности цвета, с другой, представляют собой различные утверждения. Построим пример преобразования излучения в цвет, для которого закон аддитивности будет выполняться, а свойство аддитивности цвета — нет. Допустим, что координаты S_1, S_2, S_3 вектора цвета вычисляются по формулам:

$$\begin{aligned} S_1(b(\lambda)) &= \left(\int_{\lambda_1}^{\lambda_2} b(\lambda)m(\lambda)d\lambda \right)^3, \\ S_2(b(\lambda)) &= \left(\int_{\lambda_1}^{\lambda_2} b(\lambda)n(\lambda)d\lambda \right)^3, \\ S_3(b(\lambda)) &= \left(\int_{\lambda_1}^{\lambda_2} b(\lambda)p(\lambda)d\lambda \right)^3, \end{aligned} \quad (61)$$

где $b(\lambda)$ — излучение, порождающее цвет \bar{S} .

Легко убедиться в том, что принятое преобразование излучения в цвет удовлетворяет закону аддитивности. Действительно, равенства (58) в нашем случае запишутся в виде:

$$\begin{aligned} \left(\int_{\lambda_1}^{\lambda_2} b'_1(\lambda)m(\lambda)d\lambda \right)^3 &= \left(\int_{\lambda_1}^{\lambda_2} b''_1(\lambda)m(\lambda)d\lambda \right)^3, \\ \left(\int_{\lambda_1}^{\lambda_2} b'_1(\lambda)n(\lambda)d\lambda \right)^3 &= \left(\int_{\lambda_1}^{\lambda_2} b''_1(\lambda)n(\lambda)d\lambda \right)^3, \\ \left(\int_{\lambda_1}^{\lambda_2} b'_1(\lambda)p(\lambda)d\lambda \right)^3 &= \left(\int_{\lambda_1}^{\lambda_2} b''_1(\lambda)p(\lambda)d\lambda \right)^3. \end{aligned} \quad (62)$$

$$\begin{aligned} \left(\int_{\lambda_1}^{\lambda_2} b'_2(\lambda) m(\lambda) d\lambda \right)^3 &= \left(\int_{\lambda_1}^{\lambda_2} b''_2(\lambda) m(\lambda) d\lambda \right)^3, \\ \left(\int_{\lambda_1}^{\lambda_2} b'_2(\lambda) n(\lambda) d\lambda \right)^3 &= \left(\int_{\lambda_1}^{\lambda_2} b''_2(\lambda) n(\lambda) d\lambda \right)^3, \\ \left(\int_{\lambda_1}^{\lambda_2} b'_2(\lambda) p(\lambda) d\lambda \right)^3 &= \left(\int_{\lambda_1}^{\lambda_2} b''_2(\lambda) p(\lambda) d\lambda \right)^3. \end{aligned} \quad (63)$$

Извлекая кубический корень из обеих частей каждого из равенств (62), (63), имеем:

$$\begin{aligned} \int_{\lambda_1}^{\lambda_2} b'_1(\lambda) m(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} b''_1(\lambda) m(\lambda) d\lambda, \\ \int_{\lambda_1}^{\lambda_2} b'_1(\lambda) n(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} b''_1(\lambda) n(\lambda) d\lambda, \\ \int_{\lambda_1}^{\lambda_2} b'_1(\lambda) p(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} b''_1(\lambda) p(\lambda) d\lambda; \\ \int_{\lambda_1}^{\lambda_2} b'_2(\lambda) m(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} b''_2(\lambda) m(\lambda) d\lambda, \\ \int_{\lambda_1}^{\lambda_2} b'_2(\lambda) n(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} b''_2(\lambda) n(\lambda) d\lambda, \\ \int_{\lambda_1}^{\lambda_2} b'_2(\lambda) p(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} b''_2(\lambda) p(\lambda) d\lambda. \end{aligned} \quad (64)$$

Складываем левые и правые части соответствующих равенств (64) и (65):

$$\begin{aligned} \int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) + b'_2(\lambda)) m(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) + b''_2(\lambda)) m(\lambda) d\lambda, \\ \int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) + b'_2(\lambda)) n(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) + b''_2(\lambda)) n(\lambda) d\lambda, \\ \int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) + b'_2(\lambda)) p(\lambda) d\lambda &= \int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) + b''_2(\lambda)) p(\lambda) d\lambda. \end{aligned} \quad (66)$$

Наконец, возводим в куб левые и правые части полученных равенств:

$$\begin{aligned} \left(\int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) + b'_2(\lambda)) m(\lambda) d\lambda \right)^3 &= \\ = \left(\int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) + b''_2(\lambda)) m(\lambda) d\lambda \right)^3, \\ \left(\int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) + b'_2(\lambda)) n(\lambda) d\lambda \right)^3 &= \\ = \left(\int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) + b''_2(\lambda)) n(\lambda) d\lambda \right)^3, \end{aligned}$$

$$\begin{aligned} \left(\int_{\lambda_1}^{\lambda_2} (b'_1(\lambda) + b'_2(\lambda)) p(\lambda) d\lambda \right)^3 &= \\ = \left(\int_{\lambda_1}^{\lambda_2} (b''_1(\lambda) + b''_2(\lambda)) p(\lambda) d\lambda \right)^3. \end{aligned}$$

Учитывая (61), имеем:

$$\begin{aligned} \bar{S}_1(b'_1(\lambda) + b'_2(\lambda)) &= \bar{S}_1(b''_1(\lambda) + b''_2(\lambda)), \\ \bar{S}_2(b'_1(\lambda) + b'_2(\lambda)) &= \bar{S}_2(b''_1(\lambda) + b''_2(\lambda)), \\ \bar{S}_3(b'_1(\lambda) + b'_2(\lambda)) &= \bar{S}_3(b''_1(\lambda) + b''_2(\lambda)). \end{aligned}$$

Таким образом, мы приходим к выводу о справедливости равенства (59). Закон аддитивности для построенного примера выполняется.

Проверим теперь, выполняется ли для преобразования (61) свойство (60) аддитивности цвета. Если бы формула (60) была справедлива, то при любых $b_1(\lambda)$ и $b_2(\lambda)$ были бы также справедливы следующие равенства:

$$\begin{aligned} \left(\int_{\lambda_1}^{\lambda_2} (b_1(\lambda) + b_2(\lambda)) m(\lambda) d\lambda \right)^3 &= \\ = \left(\int_{\lambda_1}^{\lambda_2} b_1(\lambda) m(\lambda) d\lambda \right)^3 + \left(\int_{\lambda_1}^{\lambda_2} b_2(\lambda) m(\lambda) d\lambda \right)^3, \\ \left(\int_{\lambda_1}^{\lambda_2} (b_1(\lambda) + b_2(\lambda)) n(\lambda) d\lambda \right)^3 &= \\ = \left(\int_{\lambda_1}^{\lambda_2} b_1(\lambda) n(\lambda) d\lambda \right)^3 + \left(\int_{\lambda_1}^{\lambda_2} b_2(\lambda) n(\lambda) d\lambda \right)^3, \\ \left(\int_{\lambda_1}^{\lambda_2} (b_1(\lambda) + b_2(\lambda)) p(\lambda) d\lambda \right)^3 &= \\ = \left(\int_{\lambda_1}^{\lambda_2} b_1(\lambda) p(\lambda) d\lambda \right)^3 + \left(\int_{\lambda_1}^{\lambda_2} b_2(\lambda) p(\lambda) d\lambda \right)^3. \end{aligned}$$

Однако это заведомо не так. Следовательно, свойство (60) аддитивности цвета для преобразования (61) не выполняется. Построенный пример доказывает, что из закона аддитивности не вытекает аддитивность цвета. Следовательно, закон аддитивности не является достаточным основанием для введения операции сложения цветов.

Продолжая свой вывод, Шредингер, используя введенную им операцию сложения цветов, формулирует закон трехмерности [12, с. 419]: «Имеется линейно независимая тройка цветов. Четыре цвета всегда линейно зависимы». Основываясь на этой формулировке закона трехмерности, Шредингер на с. 436 приходит к интегральным соотношениям (6) для вычисления координат цвета. Таким образом, вывод интегральных соотношений цвета из законов Грассмана, построенный Шредингером, не имеет

доказательной силы, поскольку в нем в значительной мере использована операция сложения цветов, введенная, как уже упоминалось, без достаточного основания.

7.2. К вопросу о формулировке законов Грассмана

Грассман следующим образом сформулировал свои законы. 1) Закон трехмерности: «Для любого излучения можно подобрать одинаково выглядящую смесь белого излучения с некоторым чистым спектральным или же пурпурным излучением. Под пурпурным излучением понимается смесь крайних видимых излучений спектра» [11, с. 78]. 2) Закон непрерывности: «Непрерывному изменению излучения соответствует непрерывное изменение цвета» [11, с. 72]. 3) Закон аддитивности: «Одинаково выглядящие излучения дают при сложении одинаково выглядящие излучения» [11, с. 82].

Очевидно, что формулировка Грассманом закона трехмерности существенно отличается от даваемой нами. Также существенно отличается от нашей формулировка Грассманом закона непрерывности. В то время как в формулировке Грассмана речь идет о непрерывном изменении цвета, в нашей формулировке имеется в виду непрерывное изменение некоторых чисел, а не цвета. Закон аддитивности в формулировке Грассмана содержит меньше информации, чем в нашей формулировке. Для того, чтобы обе формулировки стали равносильными, формулировку Грассмана необходимо дополнить следующим утверждением: «Если к одинаково выглядящим излучениям добавить или вычесть (когда это возможно) равные излучения, то суммарные излучения также будут выглядеть одинаково».

Шредингер, принимая без изменений грассмановские формулировки законов непрерывности и аддитивности, закон трехмерности формулирует следующим образом [12, с. 419]: «Имеется линейно независимая тройка цветов. Четыре цвета всегда линейно зависимы». Недостаток этой формулировки состоит в том, что в ней в значительной мере использовано понятие линейной зависимости цветов, а следовательно, использованы операции сложения цветов и умножения цвета на постоянные числа. Как указывалось выше, это приводит к привлечению гипотетической модели цветового зрения в виде преобразований (6) и (7). Шредингер, кроме трех законов Грассмана, привлекает еще один постулат. Он пишет: «Я напрасно старался обойтись без новых ссылок на опыт и считаю это невозможным» [12, с. 417]. Формулировка дополнительного постулата Шредингера следующая [12, с. 418]: «Не существует двух таких излучений, которые при одинаковом увеличении интенсивности периодически выглядели бы одинаковыми и неодинаковыми». В нашей же работе, как было показано выше, интегральные соотношения цвета (6) выводятся исключительно из трех законов Грассмана без привлечения каких-либо дополнительных постулатов.

Формулировки законов Грассмана, встречающиеся в более поздних работах, как правило, не отличаются или незначительно отличаются от формулировок Шредингера. Нюберг так формулирует закон трехмерности [26, с. 17]: «Для любых четырех цветов всегда существует связывающее их линейное соотношение, причем это соотношение будет единственным, если только три цвета из этих четырех не связаны линейными соотношениями, причем, с другой стороны, можно найти бесчисленное множество троек цветов, которые никаким линейным соотношением не связаны». В другой работе Нюберг формулирует закон трехмерности так: «Для любых четырех цветов (три известных, один неизвестный) всегда можно составить одно и только одно цветовое уравнение одного из указанных выше типов, связывающее четыре данных цвета» [27, с. 53]. В этой же работе мы находим формулировку закона аддитивности (с. 38): «Результат сложения двух или нескольких цветов зависит только от того, каковы эти цвета, но совершенно не зависит от того, из каких спектральных лучей составлен свет, вызывающий тот или иной из складываемых цветов».

В работе [28, с. 159] Нюберг так формулирует закон аддитивности: «Если какие-либо два излучения визуально неразличимы, то после прибавления к обоим любого одинакового излучения полученные новые суммарные излучения также будут визуально неразличимы». В книге Гуревича находим следующую формулировку всех трех законов Грассмана [29, с. 23]. Закон трехмерности: «Любые четыре цвета находятся в линейной зависимости, хотя существует неограниченное число линейно независимых совокупностей из трех цветов». Закон непрерывности: «Непрерывному изменению излучения соответствует также непрерывное изменение цвета». Закон аддитивности: «Цвет смеси зависит только от цветов смешиваемых компонент и не зависит от их спектральных составов».

Наконец, приведем формулировку Нюбергом всех трех законов Грассмана, опубликованную в работе [22, с. 492]. Закон аддитивности: «Цвет суммы двух излучений зависит только от цветов складываемых излучений, но не от их спектрального состава». Закон трехмерности: «Всеякие четыре цвета линейно связаны, но существуют тройки линейно независимых цветов». Закон непрерывности: «При любом непрерывном изменении излучения цвет изменяется непрерывно». Очевидно, что все приведенные формулировки законов Грассмана воспроизводят в точности или с небольшими изменениями формулировки, использованные Шредингером.

В заключение упомянем еще одну формулировку закона аддитивности, опубликованную в работе Шейбнера [30]: «В результате аддитивного смешения излучений в психофизиологической преобразующей системе возникают классы метамерных излучений или цвета. Эти классы являются подмножествами общего множества излучений, для которых

справедливы законы тождества, симметрии и транзитивности». Приведенная формулировка закона аддитивности также отличается от даваемой нами.

8. Заключение

Законы Грассмана, сформулированные с привлечением понятий сложения цветов и умножения цвета на постоянные числа, не могут быть использованы в качестве исходной предпосылки при выводе математической модели цветового зрения, поскольку для строгого обоснования законности введения операций сложения цветов и умножения цвета на постоянные числа необходимо прежде признать справедливой модель цветового зрения. Вывод модели цветового зрения из законов Грассмана, данный Шредингером, не является эффективным, поскольку в нем в качестве исходной посылки использованы законы Грассмана, сформулированные с привлечением понятий сложения цветов и умножения цвета на постоянные числа. Предложена новая формулировка законов Грассмана, в которой не используются понятия сложения цветов и умножения цвета на постоянные числа. Показано, что стандартные колориметрические опыты выполняются в точности по той процедуре, которая необходима для демонстрации справедливости законов Грассмана в новой формулировке. Показано, что математическая модель цветового зрения может быть выведена в виде следствия из новой формулировки законов Грассмана. Построен вывод модели цветового зрения из законов Грассмана в новой формулировке. Показано, что из законов Грассмана в новой формулировке нельзя вывести ничего, кроме математической модели цветового зрения. Построен вывод новой формулировки законов Грассмана из модели цветового зрения. Построена обобщенная математическая модель цветового зрения, реагирующая на произвольные зрительные картины. Сформулированы постулаты, справедливость которых может быть продемонстрирована в эксперименте, обосновывающие обобщенную модель цветового зрения.

Список литературы: 1. Мешков В. В. Основы светотехники: Ч. 1.— М.-Л.: Госэнергоиздат, 1957. 2. Кравков С. В. Цветовое зрение.— М.: Изд-во АН СССР, 1951. 3. Люстерник Л. А., Соболев В. И. Элементы функционального анализа.— М.: Наука, 1965. 4. Shklover D. A. The problem of the equicontrast colorimetric system: Report at Symposium on

visual problems of colours, Teddington, 1957. 5. Ньютон И. Оптика: Изд. 2-е.— М.: Гостехтеориздат, 1954. 6. Ломоносов М. В. Слово о происхождении света, новую теорию о цветах представляющее. Избранные философские произведения.— М.: Госполитиздат, 1950. 7. Young T. Lecture on the theory of light and colours // Phil. Trans. Roy. Soc.— 1802.— Vol. 21. 8. Maxwell J. C. On the theory of compound colours and the relations of the colours of the spectrum // Proc. Roy. Soc.— 1860.— Vol. 10. 9. Helmholtz H. Handbuch der physiologischen Optik.— Hamburg u. Leipzig, 1909—1911. 10. Мешков В. В. Основы светотехники: Ч. 2. Физиологическая оптика и колориметрия.— М.-Л.: Госэнергоиздат, 1961. 11. Grassman H. Zur Theorie der Farbenmischung // Ann. d. Phys. u. Chemie.— 1853.— Bd. 89, № 5. 12. Schrödinger E. Grundlinien einer Theorie der Farbenmetrik im Tagessehen // Ann. d. Phys.— 1920.— Bd. 63. 13. Нюберг Н. Д. Математические основы задачи измерения цвета // Федоров Н. Т. Современное состояние колориметрии.— М.-Л.: Гостехтеориздат, 1933. 14. Шилов Г. Е., Гуревич Б. Л. Интеграл, мера и производная.— М.: Наука, 1967. 15. Натансон И. П. Теория функций вещественной переменной.— М.: Гостехтеориздат, 1957. 16. Майзель С. О., Ратнер Е. С. Цветовые расчеты и измерения.— М.: Госэнергоиздат, 1941. 17. Миллин С. Г. Прямые методы в математической физике.— М.-Л.: Гостехтеориздат, 1950. 18. Канторович Л. В., Акилов Г. П. Функциональный анализ в нормированных пространствах.— М.: Физматгиз, 1959. 19. Гантмахер Ф. Р. Теория матриц: Изд. 2-е.— М.: Физматгиз, 1966. 20. Гельфанд И. М. Лекции по линейной алгебре.— М.-Л.: Гостехтеориздат, 1948. 21. Нюберг Н. Д. Теоретические основы цветной репродукции.— М.: Советская наука, 1947. 22. Нюберг Н. Д. Грассмановы законы // Физический энциклопедический словарь: Т. 1.— М.: Сов. энциклопедия, 1960. 23. Бонгард М. М., Смирнов М. С. Четырехмерность цветового пространства человека // ДАН СССР.— 1956.— Т. 108, № 3. 24. Фридрих Л. Об участии палочкового зрения в работе светоадаптированного глаза человека // Биофизика.— 1957.— Т. 2, вып. 3. 25. Talbot H. F. Experiments on light // Phil. Mag.— 1834.— № 5. 26. Нюберг Н. Д. Измерение цвета и цветовые стандарты.— М.: Стандартизация, 1933. 27. Нюберг Н. Д. Курс цветоведения.— М.-Л.: Гизлегпром, 1932. 28. Нюберг Н. Д. Колориметрические эксперименты как средство исследования цветового зрения и требования к ним // Биофизика.— 1957.— Т. 2, вып. 2. 29. Гуревич М. М. Цвет и его измерение.— М.-Л.: Изд-во АН СССР, 1950. 30. Scheibner H. On colours of the same appearance // Optica acta.— 1966.— Vol. 13, № 3.

Поступила в редакцию 14.03.2006

УДК 519.7:007.52

Е. В. Бодянский, Е. А. Винокурова, А. П. Слипченко, А. В. Шило

ИДЕНТИФИКАЦИЯ ФАЗ СНА В РЕАЛЬНОМ ВРЕМЕНИ НА ОСНОВЕ ГИБРИДНЫХ НЕЙРОННЫХ СЕТЕЙ

1. Введение

В связи с ростом количества заболеваний сердечно-сосудистой системы и сопутствующим ростом расстройств сна в середине 90-х годов прошлого века был развернут Европейский проект SIESTA, одной из целей которого является развитие новых способов описания цикла «сон—бодрствование» человека с высоким временным разрешением, полностью автоматических и свободных от субъективных критериев.

Для создания автоматических анализаторов сна используются системы, построенные на основе амплитудного и частотного анализа в отдельных диапазонах частот, скрытая модель Маркова, использующие один канал электроэнцефалограммы (ЭЭГ), вейвлет-преобразование (для идентификации специфических феноменов сна) и ряд других современных подходов [1–3]. Тем не менее, надежность определения стадий сна составляет 75...80 %, что является недостаточным для корректного анализа и обуславливает необходимость привлечения дополнительных критериев и разработки новых алгоритмов реального времени для более строгой идентификации нарушений цикла «сон — бодрствование». Четкая идентификация последних является чрезвычайно важной как для постановки правильного диагноза, так и для выработки оптимальной стратегии коррекции выявленной патологии [4].

2. Постановка задачи

На содержательном уровне задача идентификации фаз сна/бодрствования может быть разбита на два относительно независимых этапа. На первом этапе необходимо сформировать некоторый набор признаков, характеризующих наблюдаемый сигнал (ЭЭГ) в каждый момент времени. Второй этап будет заключаться собственно в определении класса, к которому относится текущее состояние (бодрствование, медленноволновой сон, парадоксальный сон и т. п.).

На формальном уровне первый этап заключается в идентификации наблюдаемого сигнала. Часто вместо модели системы, генерирующей наблюдаемый временной ряд, рассматривают модель предиктора этого временного ряда, так как эти модели во многих случаях эквивалентны [5]. Таким образом, в качестве набора признаков, характеризующих состояние объекта в текущий момент времени, могут использоваться параметры прогнозирующей модели.

Традиционно для получения прогноза временного ряда используется математическая конструкция, называемая предиктором, которая некоторым образом комбинирует имеющуюся информацию о предыдущих значениях исследуемой последовательности и ошибках прогнозирования. В зависимости от вида

этой комбинации различают линейные и нелинейные предикторы. Линейные модели предикторов (AR, MA, ARMA) получили широкое распространение, однако многие современные приложения (в частности, медико-биологические) явно требуют применения нелинейных подходов. Нелинейные модели предикторов (NAR, NMA, NARMA) являются обобщением соответствующих линейных.

Для того, чтобы иметь возможность отслеживать изменения свойств исследуемого процесса, необходимо проводить настройку параметров модели в реальном масштабе времени по мере поступления новых данных и уменьшать влияние устаревшей информации. Такая постановка приводит к решению следующей оптимизационной задачи:

$$E(k) = \frac{1}{2} \sum_{p=1}^k \alpha^{k-p} (y(p) - \hat{y}(p))^2 \rightarrow \min, \quad (1)$$

где k — текущий момент времени; $y(p)$ — наблюдаемый сигнал в момент времени p ; $\hat{y}(p)$ — прогноз значения наблюдаемого сигнала в момент времени p ; $\alpha \in [0, 1]$ — коэффициент дисконтирования, регулирующий скорость забывания устаревшей информации.

Второй этап, как уже отмечалось, заключается в классификации состояния объекта, которое характеризуется вектором параметров предиктора в каждый момент времени. Как известно, оптимальным в смысле минимизации вероятности ошибки является байесовский классификатор [6, 7]. Формально задача классификации состояния объекта сводится к проверке статистической гипотезы при наличии нескольких альтернатив. При этом предполагается, что известны условные плотности распределения вероятностей вектора признаков ($p(X | \omega_i)$) и априорные вероятности каждого класса $P(\omega_i)$. В этом случае решающее правило будет иметь следующий вид:

$$X \in \omega_i; \\ P(\omega_i | X) > P(\omega_j | X), \quad \forall j \neq i.$$

С учетом теоремы Байеса это выражение может быть переписано следующим образом:

$$X \in \omega_i; \\ P(\omega_i) p(X | \omega_i) > P(\omega_j) p(X | \omega_j), \quad \forall j \neq i. \quad (2)$$

С целью обеспечения возможности более тонкой настройки классификатора, а также в силу того, что не все классы одинаково важны для исследователей, решающее правило (2) может быть модифицировано следующим образом:

$$X \in \omega_i; \\ c_i P(\omega_i) p(X | \omega_i) > c_j P(\omega_j) p(X | \omega_j), \quad \forall j \neq i, \quad (3)$$

где c_i — вес, связанный с i -м классом.

3. Архитектура нейронной сети

Для восстановления нелинейной комбинации входов прогнозирующей модели была применена нейронная сеть [8–11], реализующая отображение вида

$$\hat{y}(k) = \hat{f}(x(k)) = \sum_{i=1}^n \sum_{j=1}^{h_i} w_{ij} \varphi_{ij}(x_i(k)), \quad (4)$$

где w_{ij} — синаптические веса, подлежащие определению; $\varphi_{ij}(\bullet)$ — активационные функции скалярного аргумента. При этом для каждого зафиксированного i функции $\varphi_{ij}(\bullet)$ образуют ортогональную систему. Отметим, что предлагаемая архитектура содержит $h = \sum_{i=1}^n h_i$ настраиваемых параметров, т. е. не подвержена проклятию размерности [8].

Данная архитектура была выбрана в силу того, что:

1) выход сети линейно зависит от настраиваемых параметров, что позволяет использовать оптимальные по быстродействию алгоритмы обучения и решать задачу (4) в реальном времени по мере поступления новых данных [8, 9];

2) данная ИНС обеспечивает возможность восстановления нелинейных отображений [12, 13];

3) поскольку используемые активационные функции образуют ортогональную систему, алгоритм обучения обладает высокой численной устойчивостью [8, 9].

Для классификации временного ряда была разработана модель на основе байесовского подхода.

4. Алгоритм обучения нейронной сети

Для удобства дальнейшего изложения перепишем выражение для выхода нейронной сети в виде:

$$\hat{y}(k+1) = \Phi^T(k) \hat{W}(k),$$

где $\Phi(k) = (\varphi_{11}(k), \dots, \varphi_{1h_1}(k), \dots, \varphi_{nh_n}(k))^T - ((h_1 + \dots + h_n) \times 1)$ — вектор ортогональных функций активации в момент времени k ; $\hat{W}(k) = (\hat{w}_{11}(k), \dots, \hat{w}_{1h_1}(k), \dots, \hat{w}_{nh_n}(k))^T - ((h_1 + \dots + h_n) \times 1)$ — вектор текущих значений синаптических весов.

В силу того, что выход нейронной сети линейно зависит от настраиваемых параметров, для оценки синаптических весов можно применить метод наименьших квадратов

$$\hat{W}(k+1) = (\Phi^T(k) \Phi(k))^{-1} \Phi^T(k) Y(k+1), \quad (5)$$

где $\Phi = (\Phi(0), \Phi(1), \dots, \Phi(k))$, $Y(k) = (y(1), y(2), \dots, y(k))^T$.

Как уже отмечалось, для решения задачи идентификации временного ряда (1) в реальном времени целесообразно применение рекуррентной модификации алгоритма (5), обеспечивающей забывание устаревшей информации (экспоненциально-взвешенный рекуррентный метод наименьших квадратов):

$$\begin{cases} \hat{W}(k+1) = \hat{W}(k) + \\ + \frac{P(k)(y(k+1)\hat{W}^T(k)\Phi(k+1)\Phi(k+1))}{\alpha + \Phi^T(k+1)P(k)\Phi(k+1)}, \\ P(k+1) = \frac{1}{\alpha} \left(P(k) - \frac{P(k)\Phi(k+1)\Phi^T(k+1)P(k)}{\alpha + \Phi^T(k+1)P(k)\Phi(k+1)} \right) \end{cases} \quad (6)$$

где $\alpha \in (0; 1]$ — фактор забывания.

5. Результаты моделирования

Эксперименты были проведены на 7–8-месячных белых крысах линии Вистар массой 180...200 г в летний период времени (июнь — август). Животных содержали в условиях вивария при регулируемом цикле «свет — темнота» (12:12), где они получали стандартный рацион и воду ad libitum. Для контроля за функциональным состоянием животного под тиопенталовым наркозом (смесь тиопентала натрия/оксибутирата натрия) производили стереотаксическую операцию по вживлению двух корковых электродов (область сенсомоторной и зрительной коры, биполярное отведение) и двух электродов для регистрации активности мышц шеи. Опыты начинали не ранее, чем через неделю после операции. По истечении послеоперационного восстановительного периода животных помещали в экспериментальную камеру и через вращающийся токосъемник (не ограничивающий двигательную активность животного) подсоединяли к электроэнцефалографу (BST-1). Усиленные сигналы биоэлектрической активности мозга и мышц через аналогово-цифровой преобразователь (ADDA-14) записывались на жесткий диск персонального компьютера типа Pentium в режиме on-line для визуального мониторинга и сохранения. Последующая обработка проводилась с помощью специально разработанного пакета программ или на М-языке. Первые сутки регистрации рассматривались как адаптационные к условиям регистрации и из анализа исключались. Визуальное выявление стадий сна производили по общепринятым критериям по изменению спектра мощности ЭЭГ, ЭМГ использовалась в качестве вспомогательного параметра.

Эксперименты были проведены в соответствии с «Общими этическими принципами экспериментов на животных», одобренными Первым национальным конгрессом по биоэтике (20 сентября 2001 г., г. Киев, Украина) и согласованными с положениями «Европейской конвенции о защите позвоночных животных, используемых для экспериментальных и других научных целей» (Страсбург, 1985 г.), а также согласно рекомендациям, составляющим один из разделов «Руководства к дотациям и контрактам Национального института здравоохранения США» (1975 г.) — «Принципы использования животных».

Для начальной настройки классификатора часть данных была обработана экспертом, который выделил в них четыре класса: «sws» (медленноволновой сон), «artifact» (нехарактерные участки), «rem» (парадоксальный сон) и «awake» (бодрствование).

Анализ автокорреляционной функции имеющихся данных показал, что для построения прогнозирующей модели достаточно будет взять последние 20 значений (рис. 1). Кроме того, даже визуальный анализ показывает, что автокорреляционные функции, соответствующие различным состояниям, существенно различаются, что свидетельствует о принципиальной возможности их разделения.

В то же время, применение для этих целей самой автокорреляционной функции является плохим решением по целому ряду причин, среди которых стоит отметить следующие:

- для вычисления автокорреляционной функции необходимо наличие достаточно большого объема данных, и такое вычисление не может производиться в реальном времени;
- вычисление автокорреляционной функции на «скользящем» окне приводит к значительному усилению влияния шумов, что в значительной мере усложняет разделение состояний.

Для непосредственной идентификации измеряемого временного ряда использовалась нейронная сеть, на основе которой был синтезирован предиктор на один шаг вперед. Структура входов соответствует модели авторегрессии порядка 24. На каждый вход ИНС было добавлено 5 активационных функций, что дало в сумме 106 настраиваемых параметров.

Обучение ИНС производилось в реальном времени по мере поступления данных с помощью алгоритма (6). Последующие вычислительные эксперименты в ускоренном машинном времени показали, что ПЭВМ с процессором AMD Athlon XP 2000+ и 512 Мб оперативной памяти в состоянии обрабатывать поступающие данные с частотой до 250 Гц.

Настройка параметров c_i модифицированного решающего правила (3) осуществлялась исходя из соображений минимизации ошибки классификации.

Геометрическая интерпретация решающего правила (3) заключается в определении для каждого класса некоторой области Γ_i , при попадании в которую вектора признаков $X(k)$ мы констатируем принадлежность этого вектора классу i . Тогда коэффициенты c_j должны удовлетворять следующей системе неравенств:

$$\forall i, \forall j \neq i: \int_{\Gamma_i} c_i P(\omega_i) p(X | \omega_i) dX > \int_{\Gamma_j} c_j P(\omega_j) p(X | \omega_j) dX. \quad (7)$$

В общем случае эта система содержит $n(n-1)$ неравенств, где n — число различаемых классов.

Вычисление интегралов в неравенствах (7) затруднено в силу того, что нам неизвестен ни аналитический вид областей Γ_i , ни условные плотности

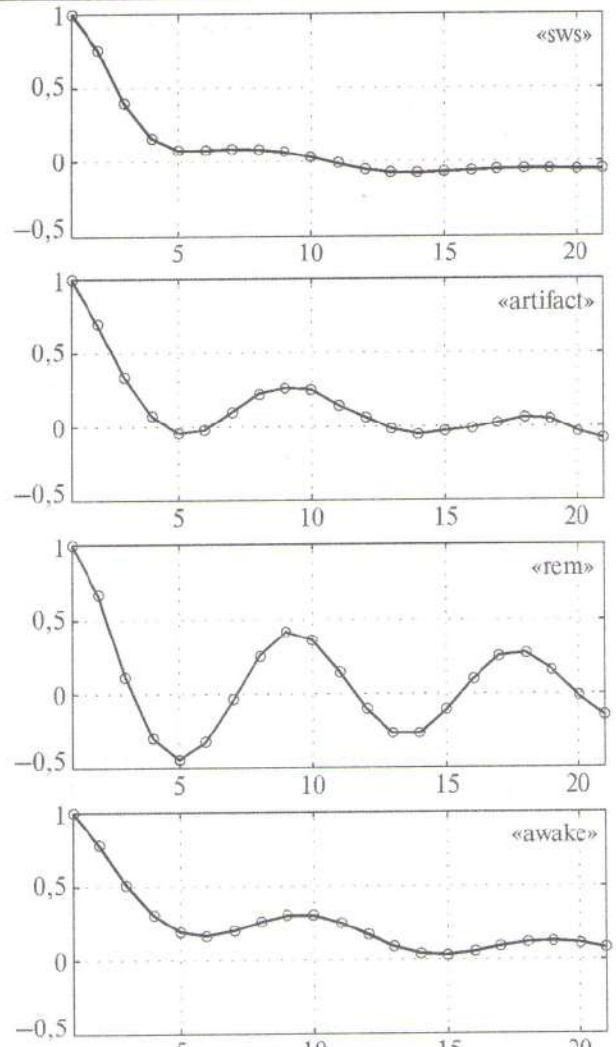


Рис. 1. Усредненная нормированная автокорреляционная функция для каждого состояния

$p(X | \omega_i)$. Очевидно, что при наличии выборки, обработанной экспертом, мы можем перейти к более простой задаче:

$$\forall X \in \omega_i, j \neq i: c_i \sum P(\omega_i) p(X | \omega_i) > c_j \sum P(\omega_j) p(X | \omega_j).$$

Собственно настройка классификатора была произведена на данных, обработанных экспертом. На них была получена относительная частота появления каждого класса, которая затем использовалась в качестве $P(\omega_i)$. На этих же данных проводилось оценивание условных плотностей $p(X | \omega_i)$. Состояния были пронумерованы следующим образом: 1 — «sws», 2 — «artifact», 3 — «rem», 4 — «awake».

Предварительная «ручная» обработка позволила составить систему из 12 неравенств для $c_i, i = 1, 4$. В результате их решения было получено следующее описание области оптимальных значений этих параметров:

$$c_1 > 0, c_2 > 0, c_3 > 0,$$

$$520.345c_1 < c_2 < 964338.c_1,$$

$$4.068c_1 < c_3 < 4.0755c_1,$$

$$3939450c_2 < c_3 < 0.00782844c_2,$$

$$2.92521c_1 < c_4 < 2.92657c_1,$$

$$2833330c_2 < c_4 < 0.00562333c_2,$$

$$0.717458c_3 < c_4 < 0.719219c_3.$$

Анализ полученной области показывает, что она не содержит ни одной точки. Причиной этого является класс № 2 «artifact», для которого не может быть выполнено требование правильной классификации. Исключение этого класса при составлении неравенств позволило получить следующую область для параметров c_1, c_3, c_4 :

$$c_1 > 0, c_3 > 0,$$

$$4.068c_1 < c_3 < 4.0755c_1,$$

$$2.925c_1 < c_4 < 2.926c_1,$$

$$0.717c_3 < c_4 < 0.719c_3.$$

Данная область является открытой и содержит бесконечное число точек. Для классификации временного ряда был использованы следующие значения параметров:

$$c_1 = 1, c_2 = 1, c_3 = 4.0726, c_4 = 2.9261, \quad (8)$$

которые были определены с учетом полученных неравенств исходя из $c_1 = 1$.

С учетом полученных значений классификатора была проведена классификация временного ряда. Качество работы предложенного алгоритма было проверено на 400 000 наблюдений, среди которых 22 фрагмента «sws», 10 фрагментов «rem» и 19 фрагментов «awake» (по мнению эксперта). Параллельно и независимо от эксперта проводилась идентификация временного ряда с последующей автоматической классификацией состояния объекта. Далее результат, полученный с помощью разработанной системы, сравнивался с мнением эксперта. На тестовой выборке ошибка классификации составила 5,0919 %, что является хорошим показателем.

6. Заключение

Проведенные эксперименты подтверждают эффективность предложенного подхода, который позволяет проводить идентификацию фаз сон/бодрствование в автоматическом режиме в реальном масштабе времени.

Работа выполнена в рамках выполнения договора о научном сотрудничестве между Институтом проблем криобиологии и криомедицины НАН Ук-

раины и Харьковским национальным университетом радиоэлектроники.

Дальнейшее развитие данных исследований видится в доработке классифицирующей части алгоритма. Предложенная система все еще не может обойтись без эксперта, проводящего классификацию обучающей выборки. Кроме того, на данном этапе невозможно определить, когда наблюдаемый ряд ведет себя нетипично относительно обучающих примеров, т. е. когда объект приходит в аномальное состояние. Требуется дальнейшего улучшения и процедура настройки параметров классификатора.

Список литературы: 1. *Anderer P., Gruber G., Parapatics S.* An e-health solution for automatic sleep classification according to rechtschaffen and kales: validation study of the somnolyzer 24 x 7 utilizing the siesta database // *Neuropsychobiology.* – 2005. – 51. – P. 115–133. 2. *Jobert M., Degans A.* Wavelets – an improvement in the detection of sleep EEG patterns // *Proc. of 16th Congress of the European Sleep Research Society, Reykjavik, June 3–7, 2002,* P. 168–171. 3. *Clark F.M., Radulovacki M.* An inexpensive sleep-wake state analyzer for the rat // *Physiol. Behav.* – 1988. – 43. – P. 681–683. 4. *Вейн А. М.* Сон человека. Физиология и патология. – М.: Медицина, 1989. – 272 с. 5. *Льюнг Л.* Идентификация систем. Теория для пользователя: Пер. с англ. / Под ред. Я. З. Цыпкина. – М.: Наука, 1991. – 432 с. 6. *Фукунага К.* Введение в статистическую теорию распознавания образов. – М.: Наука, 1979. – 367 с. 7. *Xu R., Wunsch D.* Survey of Clustering Algorithms // *IEEE Trans. on Neural Networks.* – 2005. – 16, No. 3. – P. 645–687. 8. *Bodyanskiy Ye., Kolodyazhnyi V., Slipchenko O.* Artificial neural network with orthogonal activation functions for dynamic system identification / Eds. by Sawodny O., Scharff P. «Synergies between Information Processing and Automation». – Aachen: Shaker Verlag, 2004. – P. 24–30. 9. *Bodyanskiy Ye., Kolodyazhnyi V., Slipchenko O.* Structural and synaptic adaptation in the artificial neural networks with orthogonal activation functions // *Sci. Proc. of Riga Technical University. Comp. Sci., Inf. Technology and Management Sci.* – 2004. – № 20. – P. 69–76. 10. *Бодянский Е. В., Винокурова Е. А.* Адаптивный вейвлет-нейронный предиктор // *Проблемы бионики.* – 2003. – Вып. 58. – С. 10–17. 11. *Bodyanskiy Ye., Lamonova N., Pliss I., Vynokurova O.* An adaptive learning algorithm for a wavelet neural network // *Blackwell Synergy: Expert Systems.* – 2005. – Vol. 22, No. 5 – P. 235–240. 12. *Liying M., Khorasani K.* Constructive Feed-forward Neural Network Using Hermite Polynomial Activation Functions // *IEEE Trans. on Neural Networks.* – 2005. – 16, No. 4. – P. 821–833. 13. *Patra J.C., Kot A.C.* Nonlinear dynamic system identification using Chebyshev functional link artificial neural network // *IEEE Trans. on System, Man, and Cybernetics.* – Part B. – 2002. – 32. – P. 505–511.

Поступила в редколлегию 12.05.2006

УДК 681.518:004:912

А. Л. Ерохин, А. П. Турута

ИДЕНТИФИКАЦИЯ НЕШТАТНЫХ СИТУАЦИЙ В ИНФОРМАЦИОННЫХ СЕТЯХ**1. Введение**

Современный этап развития систем и средств искусственного интеллекта делает актуальной интеллектуализацию информационных сетей. В статье рассматривается разработка метода идентификации нештатных ситуаций в информационных сетях на примере *web*. При современной интенсивности работы сетей обеспечение бесперебойной доставки целевого продукта (информации) становится все более сложной задачей. Можно выделить два подхода в ее решении. Первый связан с аспектами защиты информации и основан на разработке методов и средств обеспечения целостности и конфиденциальности информации в сетях. Второй связан с тем, что в массовых информационных сетях (в частности, *web*), наряду с обеспечением конфиденциальности и целостности, актуальной становится задача обеспечения доступности ресурсов сети. Примером нарушения доступности ресурсов является нештатная ситуация типа «отказ в обслуживании».

Следует заметить, что решение проблем целостности и конфиденциальности информации в сетях находится не только в технической плоскости, но и в юридической. Так, Police and Justice Bill 2006, принятый в Великобритании, к разряду уголовных преступлений относит умышленный вывод из строя любой компьютерной системы [1]. В указанном проекте ответственность за атаку может быть возложена на заказчика атаки, непосредственного исполнителя, владельца бот-сетей, провайдера. Отсутствие в Украине норм права или законопроекта, регулирующего ответственность за совершение атак, направленных на отказ в обслуживании, скорее указывает на то, что такие атаки не фиксируются и не диагностируются, а не на отсутствие проблемы. Поэтому идентификация нештатных ситуаций, связанных с нарушением доступности ресурсов информационных сетей, является актуальной задачей.

2. Анализ предметной области и постановка задач исследования

К причинам нарушений доступности ресурсов относятся: 1) атака типа «отказ в обслуживании» (Denial of service — DoS); 2) распределенная атака типа «отказ в обслуживании» (Distribution denial of service — DDoS); 3) увеличение интенсивности использования ресурсов сети.

Причинами возникновения DoS могут быть: ошибки проектирования системы; большой объем сетевого трафика, заполняющего каналы связи (КС); большое количество пакетов, заполняющих очередь обслуживания и снижающих скорость передачи

данных; повышение вычислительной сложности процесса, занимающего свободные ресурсы системы обслуживания; неправильный расчет нагрузки.

Наряду с умышленными атаками, угрозой для работы сети может стать ее расширение или изменение структуры. После обновления структуры сети может возникать ситуация, когда ресурсов требуется более, чем планировалось. Ущерб от DDoS определяется бесполезным простоем системы. Так, в работе [2] предлагается доход от работы порталов оценивать по формуле $P = ((R_i - C_i) * L_i) - S_i$, где R_i — доход от i -го клиента; C_i — затраты на обслуживание i -го клиента; L_i — время жизни (работы) i -го клиента, выраженное в количестве взаимодействий; S_i — стоимость привлечения клиента.

DDoS существенно влияет на параметр L_i , чем уменьшает доход. Поэтому актуальным является уменьшение времени воздействия атаки, обнаружение и локализация источников атаки, использование алгоритмов восстановления нормального режима работы системы.

Для предотвращения атак типа «отказ в обслуживании» предлагаются различные подходы: «обнаружение сигнатур», анализ статистики, анализ заголовков сетевых пакетов [3]. Для ускорения выполнения запросов, для увеличения полосы пропускания каналов связи дублируются *web*-серверы, устанавливаются дополнительные *proxy*-серверы, обслуживающие запросы в направлении к *web*-серверам, а также дополнительные каналы связи.

До сих пор актуальной является разработка алгоритмов и методов выявления подготовки атак типа «отказ в обслуживании», определение источников атаки и их локализация. Необходимы способы организации систем обслуживания и хранения ресурсов, обеспечивающие эффективное обслуживание запросов. Сетевые атаки и системы защиты от них находятся в постоянном взаимодействии и развитии, поэтому любая система, которая не может вовремя адаптироваться и обучаться, рано или поздно станет уязвима. Для построения надежной системы необходимо разработать механизм самовосстановления и самообучения системы защиты.

Таким образом, для обеспечения непрерывного предоставления информации в информационной сети необходимо решить следующие задачи:

1. Идентификация распределенных атак DoS на основе анализа информации о состоянии каналов связи, *web*-серверов, подконтрольных *proxy*-серверов, маршрутизаторов, результатов распознавания категорий сетевых агентов, а также накопленных

статистических данных. Решение этой задачи состоит из этапов:

1.1. Определение пути самовосстановления системы при идентификации отказа в обслуживании. Система может управлять следующими параметрами: блокировать каналы связи; отказывать в обслуживании отдельным категориям; менять приоритеты доступа и обслуживания.

1.2. Идентификация ситуации, когда система не способна самостоятельно перейти в нормальный режим работы, то есть необходимо внешнее вмешательство лица, принимающего решение (ЛПР). Определение порядка информирования ЛПР, выработка алгоритма внешнего воздействия, сбор и анализ информации, предшествующей отказу и сопровождающей его.

1.3. Построение временных портретов поведения каналов связи, *web*-серверов, промежуточных серверов для выявления аномалий в работе.

С первой задачей связана задача идентификации категории подключившихся сетевых агентов (СА).

2. Идентификация категории СА, запрашивающих файл, на основе: признаков запрашиваемого файла; анализа контента, передаваемого к серверу; анализа параметров, определяющих расположение категории СА; признаков, указывающих на систематичность работы СА; анализа параметров, свидетельствующих о суточных ритмах работы категорий СА. Решение этой задачи состоит из трех этапов:

2.1. При идентификации категории СА необходимо разработать методы выяснения признаков на основе: когнитивного вопроса; анализа области выбора; оценки поведения.

2.2. Для выявления поведения, систематичности биоритмов необходимо решить задачу составления подключений: определение различных подключений одного СА; выявление признаков тематик или групп файлов среди контента серверов; определение точек входа и выхода при просмотре контента файла.

2.3. Построение временных портретов и портретов поведения категорий СА для обнаружения отклонений.

Полученные результаты при идентификации отказа в обслуживании и категорий СА можно использовать для улучшения работы *web*-сервера.

3. Улучшение характеристик обслуживания запросов *web*-серверами на основе: изменения приоритетов КС, обслуживания запросов, выбора ответов для категорий СА. Как функцию цели предполагается использовать: максимальный объем отданной информации; максимальное количество обслуженных СА в пределах Δ -изменений информации; максимальную скорость при обслуживании; минимальную нагрузку на сервер; минимальную нагрузку на канал связи.

4. Прогнозирование изменений нагрузки на *web*-серверы, подконтрольные *proxy*-серверы, маршрутизаторы, каналы связи на основе: анализа исто-

рии нагрузок; мгновенных оценок нагрузки; результатов идентификации категории СА и отказов в обслуживании; анализа соотношения ситуации с «портретами», полученными в предыдущих задачах; анализа обобщенной информации о категориях СА и запрашиваемых файлах.

Полученные результаты прогнозирования, улучшения и идентификации соотносятся с фактически полученными результатами и, в случае необходимости, корректировками «учителя». Для адаптации разработанной системы необходимо решить задачу обучения системы.

5. Задача обучения правилам идентификации категорий СА и отказа в обслуживании, прогнозирования нагрузки и улучшения характеристик работы систем на основе двух подходов: автоматического обучения на ошибках; корректировки «учителя».

Первый подход должен обеспечить обучение системы на основе расхождения прогнозных данных и реально наступивших событий. Автоматическое обучение экономит время человека при настройке системы.

Второй подход используется в таких случаях, когда полученные расхождения не позволяют системе произвести самонастройку. В таком случае система должна предоставить когнитивные ответы «учителю» для внесения необходимых корректировок.

3. Исследование модели работы информационной сети *web*

Общая модель работы глобальной информационной сети *web* рассматривается в [4–7]. Предлагается выделить функции, влияющие на процесс непрерывного предоставления информации (рис. 1).

Представим функциональную модель сети следующими объектами:

1. Сетевые агенты, или клиенты, — объекты сети, которые имеют сетевой идентификатор, подключены линиями связи к компьютерной сети. Основной задачей сетевых агентов является взаимодействие (получение и отправка информации) с серверами.

2. Сервер — объект сети, основной задачей которого является обслуживание запросов клиентов (сетевых агентов). Сервер имеет сетевой идентификатор (IP-адрес), характеризуется системными ресурсами, контентом и способностью обслуживать запросы на статические и динамически формируемые файлы.

3. Доверитель — объект сети, имеющий сетевой идентификатор. Основная задача — обслуживать запросы СА, предоставлять статические файлы или обеспечивать доступ СА к *web*-серверам. Характеризуется системными ресурсами. Доверитель может содержать 0 файлов, тогда он называется маршрутизатором.

В таблице 1 представлены признаки, используемые для разделения функций *web*-серверов и подконтрольных *proxy*-серверов.

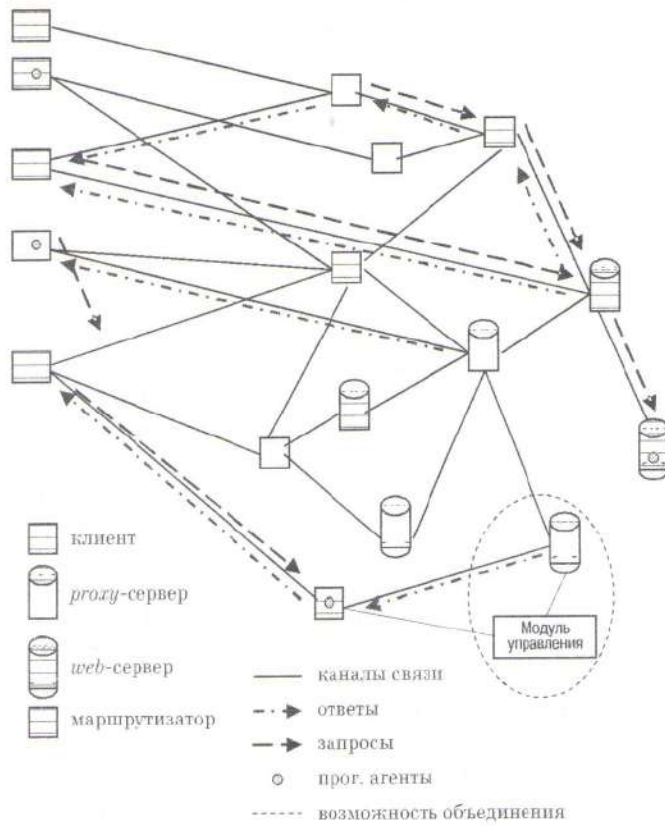


Рис. 1. Модель работы web

Таблица 1

Отличительные признаки web- и proxy-серверов

| Признаки | Web-сервер | Proxy-сервер |
|--|--|---|
| Содержание ответов | Генерирует контент, динамически создает страницы, предоставляет статические файлы. | При первом запросе перезапрашивает файл на web-сервере, далее хранит статические файлы. Динамически формируемые файлы перезапрашивает всегда. |
| Интеллектуальная деятельность | Обработывает параметры запросов, работает с маршрутами, выделяет заданные признаки, генерирует страницы. | Обработывает параметры запросов относительно статических файлов, работает с маршрутами, принимает команды управления. |
| Стоимость простоя, требования к ресурсам | Высокая стоимость, длительна и нежелательна перезагрузка. Больше нуждается в вычислительных ресурсах. | Незначительная стоимость, быстрая перезагрузка. Больше нуждается в ресурсах для хранения и пересылки данных. |
| Приоритеты | Генерация динамических страниц, предоставление постоянных файлов, выделение заданных признаков. | Пересылка пакетов и данных, предоставление статических страниц, выделение заданных признаков. |
| Дополнительные функции | Функция управления. | Функция «предохранителя». |
| Контент | Динамические или статические файлы. | Статические файлы или ∅. |

4. Маршрутизатор — объект, используемый в компьютерных сетях передачи данных, который на основании информации о топологии и определенных правил принимает решение о пересылке пакетов сетевого уровня их получателю. Характеризуется уникальным сетевым идентификатором, нагрузкой и производительностью.

5. Модуль управления — объект, собирающий информацию о других объектах сети, обрабатывающий ее, в случае необходимости может выполнить предписанный алгоритм действий. Работает в составе сервера или отдельно. Может выполнять функции «черного ящика» и реализовывать алгоритм восстановления в случае, если сервер не способен самостоятельно перейти в нормальный режим работы.

6. Каналы связи — объекты, соединяющие другие объекты сети. КС классифицируются по следующим признакам:

6.1. Подконтрольность — контролируемые линии связи (между серверами и маршрутизаторами); неконтролируемые (остальные, во внешней сети).

6.2. Назначение — служебные (связи серверов с управляющим модулем); информационные (используются для получения/отправки информации сетевыми агентами).

6.3. Длительность — маршруты; отдельные участки.

Под линиями связи (ЛС) будем понимать физическое соединение устройств, под каналами связи будем понимать логически объединенное соединение, основанное на ЛС между объектами сети, влияющими на маршрутизацию пакетов. На основе предложенной классификации предлагается выделить следующие виды линий связи: служебные; контролируемые; маршруты; внешние (неконтролируемые); информационные.

Контролируемые ЛС ограничены серверами или маршрутизаторами с обоих концов. Прохождение трафика по такой ЛС контролируется модулем управления.

7. Запросы клиентов — созданные сетевыми агентами специальные требования предоставить файл, передаются по сети по направлению к серверу-обработчику. Запросы характеризуются: адресом отправителя, адресом назначения (обработчиком), фактическим маршрутом, содержанием заголовка запроса, объемом.

8. Ответы — результаты обработки proxy- или web-серверами полученных запросов от клиентов. Ответы на обработанные запросы (ответ может быть и отрицательным, запрос может быть потерян) помещаются в сеть и отправляются сетевому агенту (клиенту). Ответ характеризуется: адресом отправителя (обработчика), адресом получателя, направлением отправки, содержанием, объемом, причиной-запросом, затратами ресурсов на генерацию.

9. Программные агенты — необязательные объекты, используются для логического (программного) улучшения характеристик работы серверов. Предлагается использовать программные агенты двух видов: *web*-ориентированные и локальные. Функции программных агентов: доставка файлов, их частей; нахождение «лучшего» места для обслуживания запроса файла или его части; слежение за перенаправлением СА; учет рекомендаций модуля управления; обеспечение механизма поиска частей файлов, проверки их подлинности и правильности «склеивания» частей в файл.

4. Идентификация сетевых агентов

Сетевые агенты характеризуются идентификатором, содержанием запроса, маршрутом, адресами (отправителя, автономной сети [8, с. 100], оценочным расположением), потоком запросов, текущими характеристиками (например, текущим временем). Дополнительно могут вычисляться соответствующие характеристики, например, соответствие географической зоны, другие признаки. Множество СА может быть разделено на подмножества по следующим критериям: по временному признаку; по географическому; по АС и по адресному пространству; узанным категориям; по программному разбиению.

Пусть N — количество установленных уникальных сетевых агентов, A_i — информация об i -м сетевом агенте, $i = \overline{1, N}$. Каждый i -й сетевой агент характеризуется кортежем $A_i = \{A_{i1}, A_{i2}, \dots, A_{ik}\}$, где k — количество элементов кортежа. Множество A может включать в себя подмножества A^j , где $j = \{1, 2, \dots, 5\}$, а A^1 — множество элементов, объединенных по временному признаку; A^2 — множество элементов, объединенных по территориальному признаку; A^3 — множество элементов, объединенных по адресному признаку; A^4 — множество элементов, объединенных в результате узнавания; A^5 — множество элементов, объединенных как программное объединение. Причем $A^1 \subset A$, $A^2 \subset A$, $A^3 \subset A$, $A^4 \subset A$, $A^5 \subset A$.

Пусть NN — количество значений для обобщенной оценки. Тогда $K_k, k = \overline{1, NN}$ — k -е значение оценки, функция обобщенной оценки будет функцией отнесения объекта A_i к классу K_k , дополнительной оценкой будет подмножество A^j . $A_i \in K_k, A_i \in A^j$, где $i = \overline{1, N}, j = \{1, 2, \dots, 5\}, k = \overline{1, NN}$.

5. Формализация характеристик *web*-сервера

Выделим следующие группы характеристик *web*-сервера:

- 1) характеристики системных ресурсов — память, процессор, жесткий диск, очередь обслуживания запросов, файл страничной подкачки и др.;
- 2) наполнение (содержание) контента — генерируемые файлы, статические файлы, части файлов;
- 3) характеристики сетевых параметров — количество сетевых интерфейсов, трафик, пакеты, очередь, коллизии и ошибки, параметры скорости и эф-

фективности передачи, наличие прямых соединений с серверами и маршрутизаторами;

4) характеристики систем реакции — правила управления ответами, маршрутизация, правила доступа; приоритеты доступа и обработки запросов;

5) (опционально) наличие модуля управления и результаты его работы.

Каждый параметр предлагается учитывать по следующим признакам: «занято»; «свободно»; история; тенденция; расчетные признаки. Пусть M — количество серверов, тогда $B_i, i = \overline{1, M}$ — i -й сервер задан общим количеством признаков P , по группам: P_1, P_2, P_3, P_4, P_5 — границы номеров признаков по группам.

Тогда обозначим характеристики сервера кортежем:

$$B_i = \left\{ \begin{array}{l} B_{i,1}, \dots, B_{i,p1}, B_{i,p1+1}, \dots, B_{i,p2}, B_{i,p2+1}, \dots \\ \dots, B_{i,p3}, B_{i,p3+1}, \dots, B_{i,p4}, B_{i,p4+1}, \dots, B_{i,p5} \end{array} \right\},$$

где $i = \overline{1, M}$; $\{B_{i,1}, \dots, B_{i,p1}\}$ — характеристики системных ресурсов; $\{B_{i,p1+1}, \dots, B_{i,p2}\}$ — содержание контента сервера, принадлежит множеству G , $\{B_{i,p1+1}, \dots, B_{i,p2}\} \subset G$; $B_{i,p2+1}, \dots, B_{i,p3}$ — характеристики сетевых параметров, содержат перечень доступных линий связи; $B_{i,p3+1}, \dots, B_{i,p4}$ — описание правил системных реакций; $B_{i,p4+1}, \dots, B_{i,p5}$ — результаты работы модуля управления.

Каждый параметр $B_{i,j}, i = \overline{1, M}, j = \overline{1, P}$ задается кортежем из 5 параметров $\{b1, b2, b3, b4, b5\}$, что позволяет более полно характеризовать состояние признака («занято», «свободно», история, тенденция, расчетный признак), значение -1 указывает на то, что такой параметр не используется.

Для получения обобщенной характеристики состояния *web*-сервера предлагается использовать K -значную P -местную функцию оценки состояния. В качестве исходных данных использованы P значений характеристик:

$$f_s(x_1, x_2, \dots, x_p) \in \{0, 1, \dots, K-1\}.$$

Например, состояние *web*-сервера может быть оценено как «простой», «работает», «перегружен», «не отвечает». В таком случае $K=4$, а значения функции по порядку обозначают предложенные состояния.

6. Формализация объектов *proxy*-серверов

Рассмотрим характеристики объектов *proxy*-серверов:

1) характеристики сети — сетевые интерфейсы, трафик, пакеты, направление (принятые, отправленные, пересланные данные);

2) характеристика системы — процессор, память, дисковое пространство;

3) контент — статические файлы, части файлов. Если доверитель без файлов — то он только включился. Количество файлов увеличивается, если поступают запросы на новые файлы, уменьшается — если истекает время хранения старых файлов;

4) характеристики систем управления — функции защиты, распределение приоритетов, маршруты.

Пусть M_c — количество доверителей, тогда C_i — идентификатор i -го доверителя, $i = \overline{1, M_c}$. Каждый доверитель задается P_c параметров, тогда

$$C_i = \{C_{i,1}, \dots, C_{i,P_c1}, \dots, C_{i,P_c}\},$$

где $\{C_{i,1}, \dots, C_{i,P_c1}\}$ — характеристики сети; $\{C_{i,P_c1+1}, \dots, C_{i,P_c2}\}$ — характеристики системы; $\{C_{i,P_c2+1}, \dots, C_{i,P_c3}\}$ — подмножество контента, $\{C_{i,P_c2+1}, \dots, C_{i,P_c3}\} \subset G$; $\{C_{i,P_c3+1}, \dots, C_{i,P_c}\}$ — характеристики систем управления.

Обобщенную оценку работы доверителя предлагается реализовать k -значной P -местной функцией:

$$f_s(x_1, x_2, \dots, x_{P_c}) \in \{0, 1, \dots, k-1\}.$$

7. Формализация характеристик маршрутизаторов

Пусть R^* — множество всех маршрутизаторов, а R — подмножество множества R^* ; R задает множество подконтрольных маршрутизаторов. R_l — l -й маршрутизатор, характеризуется кортежем

$$R_l = \{R_l^1, R_l^2, R_l^3, R_l^4, R_l^5\},$$

где R_l^1 — множество интерфейсов и их сетевых идентификаторов l -го маршрутизатора, $R_l^1 \subset X$; R_l^2 — множество подключенных линий связи, $R_l^2 \subset U$; R_l^3 — правила маршрутизации, заданные в стандартном формате [8, с. 201]; R_l^4 — правила фильтрации; R_l^5 — данные о нагрузке и производительности (переданные пакеты, обработанный трафик, возникающие коллизии и их структура по направлениям).

Обобщенную оценку работы маршрутизатора предлагается реализовать k -значной 5-местной функцией:

$$f_s(x_1, x_2, \dots, x_5) \in \{0, 1, \dots, k-1\}.$$

8. Формализация каналов связи

Каналы связи характеризуются:

- 1) идентификатором КС;
- 2) идентификаторами оконечных устройств;
- 3) вхождениями в группы в зависимости от следующих критериев:

- подконтрольность — контролируемые (между *web*-серверами и доверителями) и внешние (неконтролируемые);
 - назначение — информационные и служебные (между доверителями, к управляющему блоку);
 - длительность — маршруты и отдельные сегменты;
- 4) оценками производительности — трафиком (структура, в обоих направлениях), пакетами (структура, в обоих направлениях), скоростью (в обоих направлениях, номинальная, реальная, расчетная), простоем, коллизиям и ошибкам;

5) соединениями — в процессе установки, установленным (в работе), закрытым, оценками производительности для каждого работающего соединения;

6) приоритетами и оценками эффективности;

7) уровнями и протоколами.

Рассмотрим уровни с точки зрения характеристик производительности. Это физический (характеризуется топологией, размером кадра, номинальной скоростью), канальный, транспортный (характеризуется заголовками, реальной скоростью данных), прикладной (характеризуется избыточностью). Для объективной оценки эффективности предлагается учитывать данные, полученные на физическом, транспортном и прикладном уровнях. В соответствии с моделью OSI [9, с. 66], каждый уровень добавляет служебную информацию. Необходимо учитывать, что увеличение такой информации ведет к уменьшению реальной скорости и снижению эффективности работы ЛС.

С учетом того, что исследуемая проблема НС возникла в глобальных и локальных сетях, основанных на стеке протоколов TCP/IP [10, с. 498], то используются следующие характеристики и возможности данного протокола.

Для описания линий связи необходимо ввести множество известных (зарегистрированных или всех) хостов. Пусть X — множество хостов сети. Такое множество счетное и конечное, количество элементов 2^{32} , каждый из них задается IP-адресом. Отметим, что один хост может иметь два адреса, а два хоста не могут иметь один адрес. Множества *proxy* и *web*-серверов, маршрутизаторов включаются в множество хостов ($R^* \subset X, B \subset X, C \subset X$). Элементы множества X будем считать вершинами. Обозначим множеством U множество линий связи. Элемент множества U соединяет два элемента множества X и является дугой графа $GL(U, X)$. Канал связи мы рассматриваем как единицу соединения, на всем промежутке которой не изменяются правила маршрутизации.

Тогда $GL^1(U, X)$ — подконтрольный граф — подграф графа $GL(U, X)$, который включает в себя вершины множеств B, C и R , а также соединяющие их дуги.

Аналогично выделим: $GL^2(U, X)$ — информационный граф, подграф содержит дуги информационного назначения; $GL^3(U, X)$ — подграф маршрутов.

Указанные подграфы в случае необходимости раскрашивают исходный в различные цвета.

Каждая дуга U_i характеризуется кортежем параметров $U_i = \{U_{i,1}, U_{i,2}, U_{i,3}, U_{i,4}, U_{i,5}\}$, где параметр $U_{i,1}$ указывает на вхождение i -й КС в подграфы GL^j , $j = 1, 3$; $U_{i,2}$ содержит оценки производительности, в том числе вычисляемые:

1) фактическую скорость $V_r = D/t$, где D — объем полезной информации, переданной по КС за расчетное время t ;

2) эффективную скорость $V_\Delta = V_n * \frac{Q+14}{Q+64}$, где V_n — номинальная и минимальная среди всех ЛС

скорость работы физического уровня КС, Q — размер кадра.

Соотношение расчетных параметров предлагается использовать для определения скважности, коллизии, простоя системы или неэффективности работы.

Параметр $U_{i,3}$ содержит информацию о протоколах и уровнях, может быть задан кортежем значений для каждого протокола: $U_{i,3} = \{P_1, P_2, \dots, P_{proto1}, P_{proto1+1}, \dots, P_{protoN}\}$.

Параметр $U_{i,4}$ характеризует данные о соединениях, а параметр $U_{i,5}$ — данные о приоритетах и оценках эффективности.

Учитывая разнородность данных, различные точки сбора и неодновременность возникновения таких данных, оценить состояние ЛС в каждый момент времени не всегда представляется возможным. Предлагается ввести обобщенную оценку на основе необходимых известных данных. Такая оценка должна содержать информацию о том, в каком направлении можно увеличивать нагрузку на ЛС, а в каком стоит сокращать. Будем использовать пятиместную k -значную функцию

$$f_{ls}(U_{i1}, U_{i2}, U_{i3}, U_{i4}, U_{i5}) \in \{0, 1, \dots, k-1\}.$$

Модуль управления может быть реализован как программно, так и аппаратно. При программной реализации модуль управления располагается в *web*-сервере, а при аппаратной — отдельно. Модуль выполняет следующие функции:

1. Прием и обработку входных параметров от подконтрольных объектов сети: k -значные входы от объектов сети; корректировки человека-оператора, в случае необходимости.

2. Накопление данных: запись информации на случай выхода из строя сервера в результате атаки или перегрузки для последующего анализа причин выхода из строя; сбор и обработку статистической информации.

3. Адаптацию интеллектуальных систем поддержки принятия решения: обучение ИС; отбор эталонных объектов для обучения.

4. Непосредственно идентификацию и распознавание ситуаций, объектов, категорий.

5. Реализацию внешних воздействий: изменение (улучшение) характеристик работы серверов при обслуживании запросов СА; алгоритм восстановления системы — в случае, когда система самостоятельно не может перейти в нормальный режим работы; визуализацию заданных параметров.

В качестве исходных данных модуль управления получает оценки *web*- и *proxy*-серверов, оценку маршрутизатора, оценку ЛС.

Пусть D — множество значений $D = \{0, 1, \dots, k-1\}$, тогда x_1, x_2, x_3, \dots — места входных параметров, где $x_1 \dots x_m$ — оценки *web*-сервера; $x_{m+1} \dots x_{m+mc}$ — оценки *proxy*-сервера; $x_{m+mc+1} \dots x_{m+mc+mr+u}$ — оценки маршрутизатора; — оценка ЛС, где m, mc, mr, u — количество *web*-, *proxy*-серверов, routes, ЛС соответственно.

Зададим предикат узнавания ситуации, когда система не способна без внешнего воздействия перейти в нормальный режим работы:

$$P(x_1^{aj1}, \dots, x_m^{ajm}, \dots, x_{m+mc}^{aj_{m+mc}}, \dots, x_{m+mc+mr+u}^{aj_{m+mc+mr+u}}),$$

где $a_j \in D$.

Для возможности дальнейшего решения запишем предикатное уравнение узнавания ситуации в виде СДНФ [11, с. 24]:

$$P = \vee x_1^{aj1} \wedge \dots \wedge x_m^{ajm} \wedge \dots \wedge x_{m+mc}^{aj_{m+mc}} \wedge \dots \wedge x_{m+mc+mr+u}^{aj_{m+mc+mr+u}}.$$

9. Идентификация контента сервера

Сервер характеризуется содержанием и протоколами доступа. Протоколы доступа зависят от подключенных КС. Содержание (контент) включает файлы и переходы (ссылки файлов друг на друга). Контент представлен иерархическим деревом или ориентированным графом (рис. 2).

Пусть F — множество файлов, тогда F_i — i -й файл содержания. Обозначим Url — множество переходов между файлами, тогда $Url_{i,j}$ — переход от F_i к файлу F_j .

Получаем $G(F, Url)$ — ориентированный граф, который содержит множество файлов и переходов между ними. Контент (содержание) i -го *web*-сервера опишем как $G^{Bi,j}(F, Url)$, где $i = 1, M, j = p_1, p_2$, а k -го доверителя — как $G^{Ck,l}(F, Url), k = 1, M_c, j = p_c 1, p_c 2$.

Рассмотрим характеристики файлов:

- 1) идентификатор;
- 2) контрольная сумма для проверки целостности;
- 3) размер;
- 4) делимость на части идентификаторами частей;
- 5) вхождение в контенты доверителей и *web*-серверов, а также их расположение;

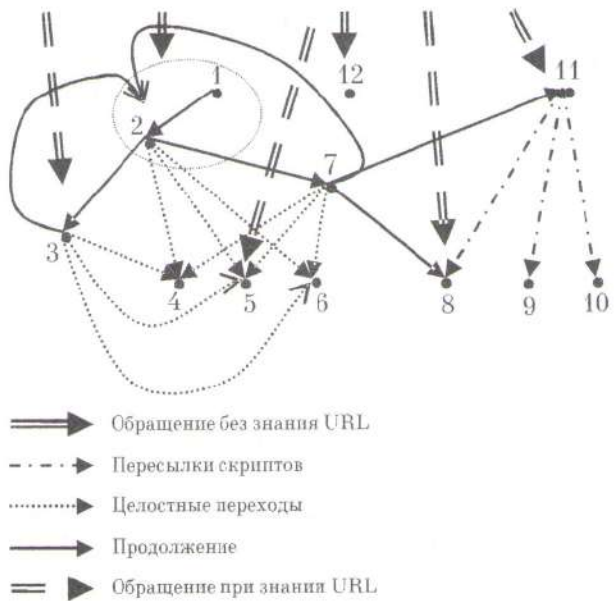


Рис. 2. Схема представления контента сайта:

- 1 — «/», 2 — index.html, 3 — news.html, 4 — style.css,
- 5 — images/*, 6 — function.js, 7 — download.php, 8 — file.zip,
- 9 — file.part1, 10 — file.part2, 11 — script.php, 12 — robot.txt

б) затрата ресурсов на обслуживание и обработку запросов;

7) частота вызова (популярность среди сетевых агентов).

Выделим группы файлов по критериям доступности и возвращаемого контента:

1) доступные по общим правилам («/», robot.txt, index.html);

2) доступные извне при знании идентификатора (обмен ссылками, баннерами; переходы из каталогов, поисковиков, избранного);

3) конечные служебные файлы общего пользования (элементы дизайна CSS, images/*; скрипты Js);

4) скрипты, не возвращающие данных, но обеспечивающие прохождение процедуры (когнитивный вопрос, авторизация, переход к загрузке файла);

5) файлы и части файлов для загрузки.

По информационному назначению выделим следующие группы файлов: страницы с содержанием (страницы с текстовой и графической полезной информацией); элементы дизайна (графические файлы, стили); скрипты (проверка форм, обработка взаимодействия, коммуникации); файлы с неоформленной информацией (RSS, XML); оформленные страницы.

По требованиям к обработке выделим: статические файлы (создаются до выдачи их клиенту); динамические файлы (создаются каждый раз при ответе клиенту); генерируемые файлы (меняются при изменении существенных параметров и, как правило, возвращаются клиенту как статические).

Такое разделение позволяет оценить требовательность обработки файлов к конкретным видам ресурсов сервера. Будем вести учет ресурсов, необходимых для генерации файлов: время формирования и отгрузки; затраты CPU, RAM; объем трафика, количество соединений с БД. Эти параметры следует использовать при распределении файлов и частей на web-серверах и доверителях; при принятии решения о предоставлении доступа сетевым агентам. Файлы будем группировать:

1) автоматически — путем исследования запроса, путем исследования истории запросов и ответов либо по атрибутам файлов и признакам их содержания;

2) в ручном режиме, в случае если автоматическое распределение не может выделить группы.

Пусть F_i — i -й файл графа G , характеризуется кортежем из M параметров $F_i = \{F_{i,1}, F_{i,2}, \dots, F_{i,M}\}$, где $F_{i,1}$ — идентификатор файла; $F_{i,2}$ — количество частей, 0 — файл на части не делится; $F_{i,3}$ — контрольная сумма для проверки целостности файлов; $F_{i,4}$ — размер файла; $F_{i,5}$ — расположение на B_k, C_j серверах, $k = 1, M, j = 1, M_c$; $F_{i,6}$ — затраты ресурсов на обработку файла; $F_{i,7}$ — частота обращения к файлу; $F_{i,8}$ — идентификаторы частей файлов и контрольные суммы к ним.

С целью улучшения характеристик обслуживания запросов предлагается распределять статические файлы и части файлов в контенте (содержании) различных proxy- и web-серверов. При запросе файла оцениваются потенциально возможные точки его обслуживания, оценивается скорость передачи клиенту с зеркал и принимается решение о разбиении обслуживания целого файла на части, о выборе точки обслуживания.

При разбиении файлов на части предлагается разработать механизм поиска файла и его частей по контенту файла. Причем признаки поиска должны характеризовать как файл в целом, так и отдельные его части.

Также необходимо разработать механизм подтверждения целостности загруженных ресурсов (всего файла, каждой части файла). Предпосылки для разбиения файла на части:

1) размер превышает заданный порог ($F_{i,1} > \Delta$);

2) статический файл. В противном случае не определенным может быть процесс докачки частей и их склейки;

3) наличие частей на различных серверах $F_{i,5} \subset (B \cup C)$;

4) предварительная оценка скоростей указывает на целесообразность распараллеливания процесса;

5) наличие простоя отдельных ресурсов на определенных серверах.

Переходы между файлами характеризуются:

1) направлением: переходы направлены; в некоторых случаях переходы одновременно могут в разных направлениях соединять два файла;

2) вероятностью (первым был вызван файл А, затем файл В): вероятность того, что после файла А будет вызван файл В; вероятность того, что файл В был вызван по причине вызова файла А;

3) интервалом времени между вызовом А и вызовом В (время чтения);

4) историей вызовов файлов;

5) возможностью параллельности и последовательности вызова.

Разделим переходы на пять типов:

1. Переход к сайту без знания конкретной страницы (по умолчанию).

2. Запрос страницы со знанием URL.

3. Последовательный/параллельный просмотр страниц.

4. Пересылки скриптами и переход к закачке файла.

5. Доагрузка элементов целой страницы.

Для дальнейшего анализа будем выделять: точку входа; точку выхода; последовательность, цепочку; повторы и возвращения; перебор файлов и случайный просмотр.

Анализ переходов позволяет выделить следующие группы файлов: а) страница; б) тематический обход; в) файлы общего пользования.

Переход $Url_{i,j}$ характеризуется кортежем из N параметров: $Url_{i,j} = \{Url_{i,j,1}, Url_{i,j,2}, \dots, Url_{i,j,n}\}$, где

$Url_{i,j,1}$ — вероятность перехода; $Url_{i,j,2}$ — условная вероятность перехода (байесова); $Url_{i,j,3}$ — частота использования в день; $Url_{i,j,4}$ — частота использования перехода на сессию; $Url_{i,j,5}$ — среднее время срабатывания перехода.

Процесс коммуникации сетевых агентов и серверов по средствам запросов и ответов разделим на три этапа (см. рис. 3).

1. **Запрос.** СА создает отдельный запрос и передает его серверу на обслуживание. Запрос характеризуется параметрами, указывающими на источник, передаваемый контент, и ресурсами, необходимыми для обработки запроса.

2. **Сессия.** В процессе коммуникации СА создает несколько запросов, и эти запросы идентифицированы от одного СА. Сессия характеризуется такими параметрами, как порядок просмотра, темп работы, группы ресурсов, время и интенсивность работы, время прочтения информации.

Если запрос — элементарная часть коммуникации, то объединенные в группу подряд идущие запросы образуют сессию. Сессия — элементарная часть построения образов и сбора информации.

Обозначим S — множество сессий, тогда S_l — l -я сессия, которая характеризуется N_c -местным вектором, $S_l = \{S_l^0, S_l^1, \dots, S_l^{N_c^1}, S_l^{N_c^1+1}, \dots, S_l^{N_c^2}, \dots, S_l^{N_c^3}\}$, с тремя группами параметров, по N_c^1, N_c^2, N_c^3 в каждой:

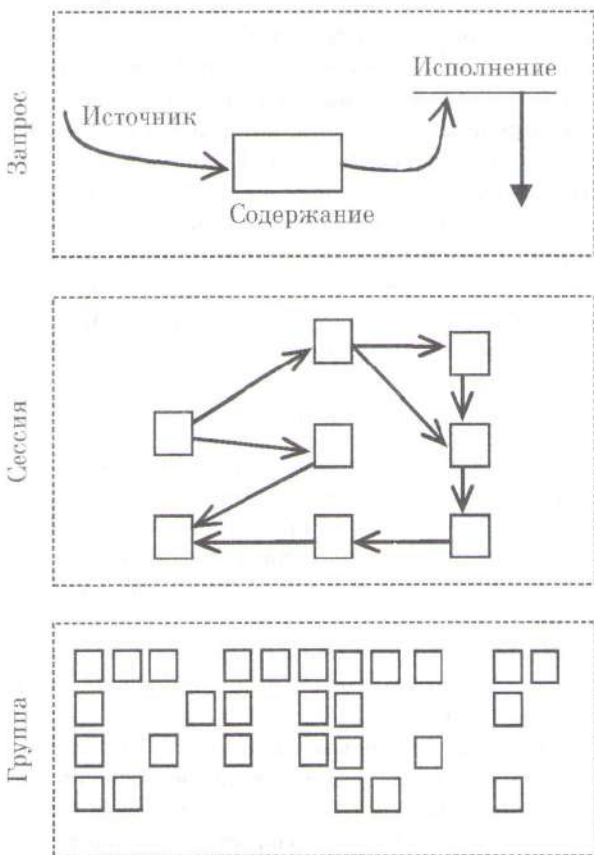


Рис. 3. Процесс коммуникации средствами запросов

1) S_l^0 — множество запросов, образовавших сессию, $S_l^0 \subset RQ, S_l^0 = \{RQ_1^{S_l}, RQ_2^{S_l}, \dots, RQ_{M_s}^{S_l}\}$. $N_c^1 = 5$ — количество признаков, образующих сессию, RQ — множество всех запросов. S_l^1 задает *phpsession*, существует на протяжении сессии, $S_l^4 = RQ_{i,2}^2 \text{ xor } RQ_{i,2}^3$; S_l^2 задает *cookie*, может существовать до, после и во время сессии. При необходимости может меняться сервером во время сессии, $S_l^2 = RQ_{i,2}^2$; S_l^3 — адрес отправителя, определен на протяжении всей сессии, стабильность зависит от способа подключения СА, $S_l^3 = RQ_{i,1}^1$; S_l^4 задает файл, с которого был сделан вызов, имеет смысл только для одного перехода (*refer*), $S_l^4 \subset RQ_{i,2}^4$; S_l^5 — данные формы, могут определять различную информацию — авторизацию, активные действия пользователя, признаки, загрузки файлов и сообщений. Данные формы постоянны для одного перехода, $S_l^5 \in RQ_{i,2}^5$; для всех $RQ_i \subset S_l^0, i = 1, M_{si}$.

2) $N_c^2 = 5$ — количество признаков, которые вычисляются при объединении запросов в сессию. S_l^6 — точка входа, $S_l^6 \approx Url_{0,RQ_{i,2}^1}, Url_{0,RQ_{i,2}^1} \in G(F, Url)$, $RQ_i \subset S_l^0$. S_l^7 — точка выхода, $S_l^7 \approx Url_{RQ_{i,2}^1,0}$, $Url_{RQ_{i,2}^1,0} \in G(F, Url), \max(RQ_{j,5}) \forall RQ_j \in S_l^0$. Сессия

характеризуется максимальной паузой между запросами (TR). Как только это время превышено, следующий запрос автоматически образует новую сессию. Последняя загруженная страница в старой сессии считается точкой выхода.

$$TR > (\text{Реальное время}) - RQ_{j,5}$$

S_l^8 — порядок обхода страниц, $S_l^8 \subset GL(F, Url)$, задан кортежем $S_l^8 = \{RQ_1, RQ_2, \dots, RQ_p\}$, где RQ_1 — точка входа, RQ_p — точка выхода, $RQ_{p,1}^1 = RQ_p$. Этот параметр позволяет определить количество и объем просмотренных страниц. S_l^9 — темп просмотра, время чтения страниц и время срабатывания переходов.

Паузы при просмотре страниц, срабатывании переходов могут являться следствием прочтения предварительно открытых страниц. S_l^{10} — содержание ответов на творческие вопросы и авторизация.

3) $N_c^3 = 5$ — количество обобщенных признаков, определяющих сессию и вычисленных во время нее. S_l^{11} — множество целых страниц с подчиненными файлами, с оценкой статуса страницы. Выделение целой страницы позволяет оценивать статус документа и автоматически распределять вероятности вызовов и приоритеты предоставления следующих файлов в графе $GL(F, Url)$.

Статус страницы определяет следующие параметры: а) приложение активно/неактивно = читается/не

читается; б) документ закрыт; в) страница сохранена, добавлена в избранное; г) загружается страница/часть; д) время прочтения.

S_i^{12} отражает предварительную группировку тематик, порядок основного просмотра. S_i^{13} — влияние темпа просмотра на загрузку файлов с серверов. S_i^{14} — определение нагрузки на сервера, обслуживающие запросы. S_i^{15} — идентификация предустановок СА. Например, загружаются ли изображения, работают ли скрипты, язык отображения, использование объектов, используются ли стили и др.

3. Группы сессий. Перед модулем управления ставится задача объединения различных сессий в группу для отдельного СА или производить группировку по иным критериям. Группы могут характеризовать одного сетевого агента, однотипных СА, временные (суточные, недельные и др.) ритмы сетевых агентов, территориальную, тематическую группу СА и др. Рассмотрим исходные данные, которые содержит запрос.

1. Параметры источника.
2. Параметры контента, сопровождающего запрос.
3. Ресурсы, требуемые на обработку запроса.
4. Параметры, характеризующие работу на протяжении сессии.
5. Параметры, позволяющие группировать запросы.

Обозначим множеством RQ — множеством запросов на получение файлов. Тогда RQ_i — i -й запрос, характеризуется вектором параметров:

- $RQ_i = \{A_j^{RQ_i}, RQ_{i,1}, RQ_{i,2}, RQ_{i,3}\}$, где $RQ_{i,1}$ — кортеж параметров источника;
- $RQ_{i,1} = \{RQ_{i,1}^1, RQ_{i,1}^2, RQ_{i,1}^3, RQ_{i,1}^4, RQ_{i,1}^5\}$, где $RQ_{i,1}^1$ определяет адрес источника соединения, $RQ_{i,1}^1 \in X$; по $RQ_{i,1}^1$ определяется вхождение СА в адресное подмножество, $A_j^{RQ_i} \in A^3$; $RQ_{i,1}^2$ — адреса *proxy*-серверов и скрытые за ними клиенты, $RQ_{i,1}^2 \in X$; $RQ_{i,1}^3$ — маршрут следования пакетов i -го запроса, $RQ_{i,1}^3 \in GL(U, X)$. Маршрут следования характеризуется адресом источника, адресом назначения, изменчивостью маршрута в пределах передачи пакетов для запроса, адресами промежуточных маршрутизаторов [12, с. 294]. Адреса источника соединения могут совпадать с адресом источника маршрута, а могут и отличаться, так как первый может быть скрыт за *proxy*-сервером. Маршрут — последовательность сетевых адресов (IP), через которые были переданы пакеты запросов от сетевого агента к серверу обслуживания; $RQ_{i,1}^4$ — номер автономной системы источника, получается путем запроса в специализированных базах данных [13]; $RQ_{i,1}^5$ — определяет предположительное географическое расположение источника (такую информацию предоставляет сервис Whois).

В дальнейшем этот параметр предлагается использовать: для оценки соответствия географической зоны и регистратора; для группировки и установление суточного времени.

Отдельные запросы предлагается группировать с использованием: адреса источников — для одной сессии; АС и географических признаков — для группировок.

Для группирования не подходят (могут произвольно меняться) маршруты и адреса источников между сессиями;

- $RQ_{i,2}$ — контент (содержание) запроса, задается вектором параметров $RQ_{i,2} = \{RQ_{i,2}^1, RQ_{i,2}^2, RQ_{i,2}^3, RQ_{i,2}^4, RQ_{i,2}^5, RQ_{i,2}^6\}$. Полное описание параметров запроса подробно изложено в спецификации. Выделим признаки, существенные для поставленной задачи.

$RQ_{i,2}^1$ обозначает URI [14, 15] запроса и в общем виде имеет следующий формат: <схема>; <идентификатор_источника/> <путь/> <имя_файла>.

<схема> — определяет протокол и *web*-сервис, который будет обрабатывать запрос;

<идентификатор_источника/> — адрес сервера, $\in (B \cup C)$;

<имя_файла> — идентификатор файла из контента сервера $\in (B \cup C)$, <имя_файла> $\in F$.

Файл, с которого был инициирован запрос, обозначим F_j , а запрашиваемый файл обозначим F_i , тогда переходом будем считать $Url_{i,j}$. Следовательно, запрос имеет отображение в графе $G(F, Url)$. Если файл F_j не определен, то будем считать файл F_i точкой входа и задавать как $Url_{i,0} = \langle \text{путь} \rangle \langle \text{имя_файла} \rangle$.

$RQ_{i,2}^2$ — необязательный параметр и задает информационные поля форм, которые могут быть переданы методами GET, POST. Таким методом могут быть переданы файлы, параметры для авторизации сетевых агентов как клиентов системы.

$RQ_{i,2}^3$ — необязательный параметр, задает данные, которые сохраняются произвольное (или специально определенное) время на стороне клиента и направляются при инициировании запроса (например, *cookie*).

$RQ_{i,2}^4$ — необязательный параметр, задает F_j (*refer*).

$RQ_{i,2}^5$ — необязательный параметр, задает параметры авторизации. Этот параметр может использоваться после того, как на предыдущем шаге ($RQ_{i-\Delta}$) прошел этап авторизации с параметрами $RQ_{i-\Delta}^2$, где Δ — количество пакетов, прошедших после авторизации.

$RQ_{i,2}^6$ — необязательный параметр, обозначает другие параметры (язык, сжатие, поддерживаемые форматы файлов).

Предлагается оценивать контент запроса с целью выявления языка запроса и ключевых слов для дальнейшей проверки соответствия географической зоны, АС, адреса источника и построения портрета.

Предлагается оценивать параметры *cookie* ($RQ_{i,2}^3$), *phpsession* ($RQ_{i,2}^3, RQ_{i,2}^2$), полей в форме ($RQ_{i,2}^2$) перехода для оценки параметров работы сетевого агента.

Для дальнейшего использования предлагается применять:

- *cookie* ($RQ_{i,2}^3$) — для сессии и для группы;
- *phpsession* ($RQ_{i,2}^3, RQ_{i,2}^2$) — только для сессии;
- поля в форме ($RQ_{i,2}^2$) — для одного перехода;
- *refer* ($RQ_{i,2}^4 \in F$) — для одного перехода;
- авторизацию ($RQ_{i,2}^2, RQ_{i,2}^3$) — часть сессии и для групп.

• $RQ_{i,3}$ — определяет предположительные необходимые ресурсы сервера и задается кортежем параметров $RQ_{i,3} = \{RQ_{i,3}^1, RQ_{i,3}^2, RQ_{i,3}^3, RQ_{i,3}^4, RQ_{i,3}^5\}$. Здесь $RQ_{i,3}^{1,k}$ — предположительные системные ресурсы, $k = \{cpu, ram, hdd, net\}$, $RQ_{i,3}^{1,k} \approx F_{i,6}$; $RQ_{i,3}^{2,p}$ — ресурсы обслуживания, $p = \{sess, connect, httpd_proc\}$; $RQ_{i,3}^3$ — предположительный (среднестатистический) объем ответа $RQ_{i,3}^3 \approx F_{i,4}$; $RQ_{i,3}^4$ — предположительное (среднестатистическое) время генерации ответа; $RQ_{i,3}^5$ — расчетная нагрузка на сеть в связи с необходимостью доставить ответ сетевому агенту.

Из характеристик файла $F_i \in G(F, Url)$ можно получить вероятности вызова следующих файлов, таким образом можно оценить последующую (цепную) нагрузку на сервер.

- $RQ_{i,4}$ — задает сетевого агента, который инициировал запрос, $RQ_{i,4} = A_j \in A$, где $j = \overline{1, N}$.
- $RQ_{i,5}$ — задает время поступления i -го запроса на обслуживание сервером.

10. Выводы

Разработаны и исследованы формальные модели компонент информационной сети *web* при нештатной ситуации типа «отказ в обслуживании».

Предложен метод идентификации нештатных ситуаций в информационных сетях *web*, предназначенный для обнаружения распределенных атак на *web*-серверы на основе распознавания категорий сетевых агентов.

Разработан метод выбора объектов для обучения во время функционирования системы. Предложены критерии оценивания относимости и чистоты объектов.

Список литературы: 1. Кошкина Э. В Великобритании будут сажать в тюрьму за организацию DoS-атак. (<http://security.compulenta.ru/294763/>). 2. Черняк Л. Порталы и жизненные циклы (Portal Lifecycle Management) // Открытые системы. — 2002. — № 2 (<http://citforum.ru/consulting/portal/life/>). 3. Амиев А. А., Поселдов А. В. Обнаружение распределенных атак в сети Интернет // Научно-методич. конф. «Интернет и современное общество» (<http://ims2002.nw.ru/02-r2f03.html>). 4. Вишневецкий В. М. Теоретические основы проектирования компьютерных сетей. — М.: Техносфера, 2003. — 512 с. 5. Олійник В. Ф. Основи теорії систем зв'язку: Математичні моделі телекомунікаційних систем. — К.: Техніка, 2000. — 152 с. 6. Кульгин М. В. Компьютерные сети. Практика построения. Для профессионалов. 2-е изд. — СПб.: Питер, 2003. — 462 с. 7. Хелеби С., Мак-Ферсон Д. Принципы маршрутизации в Internet. 2-е изд.: Пер. с англ. — М.: Изд. дом «Вильямс», 2001. — 448 с. 8. Сергеев А. П. Офисные локальные сети: Самоучитель. — М.: Изд. дом «Вильямс», 2003. — 320 с. 9. Олифер В. Г., Олифер Н. А. Компьютерные сети. Принципы, технологии, протоколы. — СПб.: Питер, 2001. — 672 с. 10. Таненбаум Э. Компьютерные сети. 4-е изд. — СПб.: Питер, 2003. — 992 с. 11. Шабанов-Кушнарченко Ю. П. Теория интеллекта. Математические средства. — Харьков: Вища школа, 1984. — 144 с. 12. Колисниченко Д. Н. Сделай сам компьютерную сеть. Монтаж, настройка, обслуживание. — СПб.: Наука и техника, 2004. — 400 с. 13. RIPE Network coordination center, <http://www.ripe.net> (11.09.2006). 14. <http://www.ietf.org/rfc/rfc3986.txt> (27.5.2006). 15. <http://ru.wikipedia.org/wiki/Uri> (10.04.2006).

Поступила в редколлегию 01.09.2006

УДК 004.89

И. Б. Спроджа, Г. А. Фролова

АВТОМАТИЗАЦИЯ ПРИНЯТИЯ РЕШЕНИЙ ПРИ РАЗРАБОТКЕ МЕЖЦЕХОВЫХ ТЕХНОЛОГИЧЕСКИХ МАРШРУТОВ НА ПРИБОРОСТРОИТЕЛЬНЫХ ПРЕДПРИЯТИЯХ

1. Введение

Эффективное функционирование украинских промышленных предприятий в условиях экономической нестабильности, колебаний спроса и ассортимента сопровождается ростом сложности процесса управления. Основопологающим фактором конкурентоспособности наукоемкой продукции и прибыльности мелкосерийного и единичного производства является внедрение автоматизированных систем управления производством (АСУП) [1–3]. На рынке специализированного программного обеспечения существует целый ряд АСУП, которые обеспечивают управление, контроль финансовых и материальных ресурсов предприятия («Omega Production», «Stalker», «Монолит», «Фобос», «Лопман» и другие) [4]. Однако на мелкосерийных приборостроительных предприятиях [5, 6] критическими являются потери из-за отсутствия контроля и управления в процессе технологической подготовки производства (ТПП), поскольку имеет место постоянное увеличение сложности проектных и конструкторских работ, связанных как с созданием изделия, так и с ТПП, что приводит к увеличению сроков разработки и дополнительным финансовым затратам. Технолог анализирует большой объем информации, а также учитывает многокритериальность задачи принятия технологических решений (при разработке межцеховых технологических маршрутов, проектировании оснастки и т. д.) на основании опыта работы и профессиональной интуиции [7, 8]. Следовательно, проблематична задача по формализации исходной информации на этапе ТПП. Исходя из этого, по-прежнему актуальна задача создания программных средств для поддержки принятия решений при ТПП с помощью использования информационных технологий, ориентированных на знания. Предлагается использовать методы инженерии квантов знаний [13, 14] для решения слабоструктурированных задач принятия решений в системах автоматизации разработки межцеховых технологических маршрутов.

Цель данной работы состоит в представлении слабоструктурированных задач, возникающих в процессе ТПП, и знаний технолога с помощью квантовых моделей для построения автоматизированной знаниеориентированной системы поддержки принятия производственных решений.

2. Формулирование проблемы

На основании обзора и анализа публикаций [1–10] сделан вывод, что в настоящее время одним

из перспективных направлений обеспечения конкурентоспособности предприятия является повышение эффективности ТПП выпускаемых изделий. Целью ТПП [2, 3] является оптимальное по срокам и ресурсам обеспечение технологической готовности производства к изготовлению изделий в соответствии с требованиями заказчика или рынка данного класса изделий.

Результаты работ [2–5, 6, 9] показывают, что существующие компьютерные системы технологической подготовки производства обеспечивают автоматизацию большей части рутинных действий технолога при ТПП. При этом за человеком остается творческая часть — принятие решений. Специальные знания, опыт и интуиция помогают технологу решать плохо формализованные (слабоструктурированные) задачи [11–14]. Ярким примером слабоструктурированной задачи процесса разработки ТПП является разработка межцеховых технологических маршрутов (расцеховка). Расцеховка устанавливает последовательность прохождения заготовок, деталей, сборочных единиц по подразделениям цехов. От решения этой задачи зависит оптимизация работ и, в конечном итоге, сокращение временных и финансовых затрат при производстве наукоемкой продукции.

Рассмотрим сложности, возникающие при разработке межцеховых технологических маршрутов. Они обусловлены многономенклатурностью заказов, требующей обработки большого количества исходных данных: комплекта чертежей на новое изделие; срока запуска изделия в производство; организационно-технических условий, предусматривающих различные способы получения заготовок и учитывающих возможности приобретения комплектующих изделий, а также оборудования и оснастки на других предприятиях.

Кроме того, технолог должен учитывать множество разнотипных параметров: конфигурацию, количество, стоимость материалов исходных заготовок; тип и время выполнения операций для изготовления деталей и сборочных единиц; наличие, исправность, изношенность соответствующего оборудования и др.

В настоящее время отсутствуют строго формализованные алгоритмы разработки межцеховых технологических маршрутов; четко не определены критерии и производственные характеристики, влияющие на принимаемое решение. В таких условиях не пригоден к использованию никакой из существующих

ющих формализованных методов автоматизации принятия решений.

В данной работе ставится задача разработки моделей и методов инженерии знаний для обеспечения интеллектуальной поддержки принятия технологических решений при разработке межцеховых технологических маршрутов.

3. Подход к решению проблемы

При нетрадиционном моделировании сложных объектов используют модели и методы инженерии знаний [11, 12], основанные на семантических сетях, фреймовых, продукционных, формальных логических моделях (исчисление высказываний и исчисление предикатов), квантовых моделях [13, 14].

Особенности рассматриваемого процесса ТПП требуют реализации в модели свойства активности знаний, возможности отражения в базе знаний структурных отношений объектов, оперирования нечеткими знаниями, возможность учитывать различные неопределенности, использование многоуровневых представлений (данные, модели, метамодели и т. д.). Метод разноуровневых алгоритмических квантов знаний (РАКЗ-метод), разработанный профессором И. Б. Сироджей, наиболее полно удовлетворяет этому набору требований. РАКЗ-метод обеспечивает строгую формализацию понятия «знание» на основе алгоритмических структур, допускающих алгебраические и логические операции в рамках заданной семантики (смысловой нагрузки).

Созданная общая методология квантового моделирования и алгоритмизации знаниеориентированного принятия решений в условиях неопределенности [13, 14] базируется на построении введенного обобщенного класса $M = \{M_t, M_\pi, M_v\}$ различных РАКЗ-моделей (*t*РАКЗ, *π*РАКЗ, *v*РАКЗ), удовлетворяющих специфическим ограничениям в подклассах: M_t — точных tk-знаний, M_π — приближенных π k-знаний и M_v — вероятностных vk-знаний.

Общность методики обусловлена общим принципом автоматического векторно-матричного квантования информации с доменным представлением характеристик объектов принятия решений, единой структурой пространства РАКЗ-моделей с возможностью их аналитического конечно-предикатного представления, а также общей индуктивно-дедуктивной логической схемой вывода искомых решений, опираясь на базу квантов знаний (БКЗ), иллюстрируемой рис. 1.

Квант знаний 1-го уровня представляет собой осмысленную, поименованную структуру данных об ОПР в виде конечного доменизированного вектора, домены которого разделены двоеточием «:» и отвечают признакам (разнотипным характеристикам) объекта, а компоненты доменов — значениям признаков; *i*-я компонента *j*-го домена должна содержать «1», если наблюдается *i*-е значение *j*-го признака объекта принятия решений (ОПР), в противном случае *i*-я компонента равна «0».

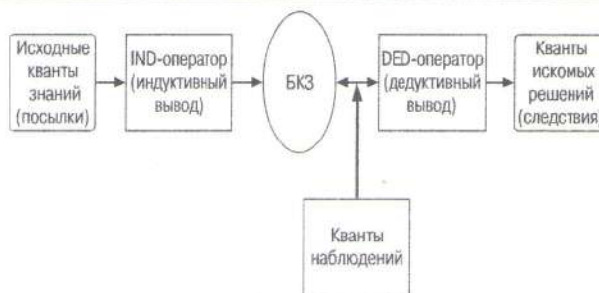


Рис. 1. Общая схема причинно-следственного вывода решений на основе РАКЗ-метода

В соответствии с семантикой (смысловой нагрузкой) данного кванта, его домены логически связаны конъюнкцией, а соответствующие компоненты в домене — дизъюнкцией. В семантику входят также указание шкал измерения значений всех признаков, их количество, семантический код (т. е. имя, тип и уровень), структура и назначение кванта. Если каждый домен кванта 1-го уровня содержит строго по одной «1» в каждом домене, то он называется элементарным, в противном случае — интервальным векторным квантом.

Пусть, например, некоторые сведения о факте или информация о функциональной связи между признаками ОПР $y \in \Omega_j$, который обладает *n* разнотипными признаками $x_j, j = \overline{1, n}$, значения которых $\alpha_{k_j}^j \in \tilde{X}^j, k_j = \overline{1, r_j}$ с соответствующими показателями достоверности (ПД) $p(\alpha_{k_j}^j) \in [0, 1]$, имеют вид:

$$x_1 - \tilde{X}^1 = \{ \alpha_1^1 | p(\alpha_1^1), \alpha_2^1 | p(\alpha_2^1), \emptyset \},$$

$$x_2 - \tilde{X}^2 = \{ \alpha_1^2 | p(\alpha_1^2), \alpha_2^2 | p(\alpha_2^2), \alpha_3^2 | p(\alpha_3^2), \alpha_4^2 | p(\alpha_4^2), \emptyset \},$$

$$x_3 - \tilde{X}^3 = \{ \alpha_1^3 | p(\alpha_1^3), \alpha_2^3 | p(\alpha_2^3), \alpha_3^3 | p(\alpha_3^3), \emptyset \},$$

где x_1 — материал {металл, резина}; x_2 — операция {сварка, слесарно-сборочная, грунтовочная, прессовочная}; x_3 — цех {цех 1, цех 2, цех 5}.

1) Если материал — металл И операция — сварка ИЛИ операция — слесарно-сборочная с соответствующими ПД p_1^1, p_1^2 и p_2^2 , ТО цех — цех 1 с вероятностью p_1^3 ;

2) Если материал — резина И операция — прессовочная с соответствующими ПД p_2^1, p_4^2 , ТО цех — цех 2 с вероятностью p_2^3 ;

3) Если операция — грунтовочная с соответствующим ПД p_3^2 , ТО цех — цех 5 с вероятностью p_3^3 .

При этом существует алгоритм $A(\rightarrow C)$, который обеспечивает вычисление вероятности $p(\rightarrow C)$, как правило, по формуле $p(\rightarrow C) = \max\{p(Y_1), p(Y_2), \dots, p(Y_m)\}$ [13, 14]. Алгоритм $A(\rightarrow C)$ для вычисления $p(\rightarrow C)$ синтезируют в данном случае по специальной методике, либо значение $p(\rightarrow C)$ указывает эксперт.

Решение частных задач по РАКЗ-методу осуществляют в такой последовательности [13, 14]:

1) провести анализ задачи, подготовить исходную информацию;

2) на основании найденных информативных признаков выбрать посыльные (входные) и заключительные (целевые) признаки;

3) сформировать таблицу эмпирических данных (ТЭД), содержащую примеры принятия технологических решений в конкретных производственных условиях относительно соответствующего целевого признака;

4) определить допустимый ранг искомых импlicative закономерностей;

5) с помощью экспертов сформировать соответствующие сценарные примеры обучающих знаний, на которых обучить логическую сеть вероятных рассуждений (ЛСВР);

6) выполнив автоматическое квантование и оптимизацию ЛСВР, получить квантовую сеть вывода решений, используемую одновременно в качестве базы квантов знаний и механизма вывода решений.

Сценарий интеллектуальной поддержки принятия решений технологом при разработке межцеховых технологических маршрутов в ТПП состоит из следующих этапов:

1. На основании данных, полученных от конструкторов, выполнить раскрытие состава изделия (детали сборочных единиц (ДСЕ), покупные комплектующие изделия, стандартные изделия) и внести операции по изготовлению ДСЕ с помощью заранее созданной базы знаний (БЗ).

2. Осуществить автоматический контроль за последовательностью выполнения операций посредством сформированного множества упорядоченных знаний для поддержки решений при проектировании межцеховых технологических маршрутов.

3. Для каждой операции технологического маршрута (ТМ) выбрать тип необходимого оборудования с помощью заранее созданной БЗ.

4. Предоставить технологу информацию из базы данных (БД) об оборудовании выбранного типа, технологические параметры которого удовлетворяют требуемым условиям.

5. Предоставить технологу информацию из БД мониторинга текущего состояния оборудования с прогнозом его работы без отказа и о наличии рабочих требуемой квалификации для выполнения задания.

6. Для каждой операции ТМ на основании БЗ осуществить автоматический выбор последовательности прохождения ДСЕ по подразделениям предприятия.

7. Автоматическая генерация технологической документации.

Очевидно, что в данном сценарии первый, второй, третий, шестой этапы интеллектуальной поддержки принятия решений являются плохо формализованными.

4. Заключение

В работе поставлена и решена задача представления процесса разработки межцеховых технологических маршрутов в ТПП, извлечения, формализации технологических знаний с помощью построения квантовых моделей для эффективного манипулирования имеющимися знаниями и логического вывода решений.

Использование интеллектуальной системы для разработки межцеховых технологических маршрутов повысит качество проектов и позволит сократить общее время разработки ТПП. Экономическая эффективность интеллектуализации ТПП определяется сокращением в несколько раз сроков проектирования, повышением качества решений, снижением себестоимости изготовления документации.

Список литературы: 1. Шербаков П. П. Автоматизация технологического проектирования: Учеб. пособ. Барнаул: Изд-во АлтГТУ, 2002. – 434 с. 2. Яблочников Е. И. Автоматизация технологической подготовки производства в приборостроении: Учеб. пособ. — СПб: СПбГИТМО(ТУ), 2002. – 92 с. 3. Падерин А. Н. Автоматизация задач технологической подготовки производства // Информационный листок № 995-96. – Екатеринбург: ЦНТИ, 1996. – 4 с. 4. Вальков В. М., Вершин В. Е. АСУ технологическими процессами: 3-е изд., перераб. в доп. – Л.: Политехника, 1991. – 269 с. 5. Милаев В. А., Фаткин А. А., Рулева Т. В. Автоматизация процесса управления в условиях многономенклатурного мелкосерийного производства // Автоматизация и современные технологии. – 2002. – № 9. – С. 36–45. 6. Милаев В. А., Фаткин А. А., Рулева Т. В. Информационные технологии в условиях мелкосерийного производства // Автоматизация и современные технологии. – 2002. – № 2. – С. 16–21. 7. Киричук Е. П. Развитие v-РАКЗ метода принятия решений при выборе оборудования в ТПП в условиях неопределенности // Открытые информационные и компьютерные интегрированные технологии. – 2004. – Вып. 25. – С. 100–104. 8. Гордиенко Л. А. Интеллектуальная технология поддержки принятия решений в технологической подготовке производства на основе квантового подхода в инженерии знаний // Открытые информационные компьютерные интегрированные технологии: Сб. науч. трудов. – Х.: ХАИ, 2004. – № 22. – С. 185–189. 9. Милаев В. А., Фаткин А. А., Соколов А. Н. Схематическое представление заказа в системе оперативного управления многономенклатурным мелкосерийным производством // Автоматизация и современные технологии. – 2003. – № 11. – С. 28–37. 10. Норенко И. П., Кузьмик П. К. Информационная поддержка наукоемких изделий. CALS-технологии. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2002. – 320 с. 11. Десятков В. В. Системы искусственного интеллекта: Учеб. пособ. для вузов. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2001. – 352 с. 12. Алиев Р. А., Абдикеев Н. М., Шарназаров М. М. Производственные системы с искусственным интеллектом. – М.: Радио и связь, 1990. – 264 с. 13. Сироджа И. Б., Петренко Т. Ю. Метод разноуровневых алгоритмических квантов знаний для принятия производственных решений при недостатке или нечеткости данных. – Киев: Наук. думка, 2000. – 247 с. 14. Сироджа И. Б. Квантовые модели и методы искусственного интеллекта для принятия решений и управления. – Киев: Наук. думка, 2002. – 420 с.

Поступила в редколлегию 25.08.2006

УДК 681.82:519.81

З. А. Алисейко, В. И. Булкин, О. В. Канищева, Н. В. Шаронова

ИСПОЛЬЗОВАНИЕ АЛГЕБРЫ ПРЕДИКАТОВ И ПРЕДИКАТНЫХ ОПЕРАЦИЙ ДЛЯ ФОРМАЛИЗАЦИИ ДЕКЛАРАТИВНОЙ И ПРОЦЕДУРНОЙ СОСТАВЛЯЮЩИХ ЗНАНИЙ

1. Введение

Дальнейшее развитие информатизации и компьютеризации немислимо без создания новых и усовершенствования существующих интеллектуальных информационных систем, как на базе традиционных компьютеров, так и на базе разрабатываемых в настоящее время прогрессивных компьютеров параллельной обработки данных. Ядром любой интеллектуальной информационной системы является база знаний. Формализацией знаний и наполнением ими базы знаний занимаются так называемые инженеры по знаниям. Их задачей является преобразование знаний, представленных в естественном для человека виде, в вид, удобный для хранения и обработки их в машинной базе знаний. Человек в своей практической и познавательной деятельности использует однородные и комплексные формы представления знаний.

2. Постановка задачи

В данной статье представлены возможности использования предикатов и предикатных операций для формализации декларативной и процедурной составляющих знаний. Рассматриваются современные формы представления знаний, проводится их сравнительный анализ, а также подробно рассматриваются возможности предикатов и предикатных операций для представления знаний.

3. Формы представления знаний

Множество однородных форм представления знаний можно формально определить в следующем виде:

$$M_1 = \{t, s, v, g\},$$

где t — текстовая форма представления знаний (книги, газеты, журналы и т. д.); s — аудиальная форма представления знаний (речь, звуки, музыка); v — визуальная форма представления знаний (анимация, балет, пантомима, жесты, мимика); g — графическая форма представления знаний (картины, рисунки, схемы, графики и т. д.) [1].

Множества комплексных форм представления знаний задаются на декартовых степенях множества M_1 . Декартов квадрат M_1^2 множества M_1 задает множество бинарных комплексных форм представления знаний. Декартов куб M_1^3 множества M_1 задает множество тернарных комплексных форм представления знаний и т. д. Очевидно, что представление знаний с помощью комплексных форм является более эффективным и полным по сравнению с однородными.

Однако наиболее удобными для человека и широко распространенными являются текстовая и речевая формы представления знаний на естественном языке, как в составе однородных, так и в составе комплексных форм. Поскольку современные компьютеры не «понимают» естественного языка, для машинного представления знаний используются специально разработанные модели представления знаний — логические, продукционные, фреймвые и сетевые МПЗ. Каждая из этих моделей имеет свои преимущества и недостатки, однако главным недостатком всех этих моделей является то, что они используют разные формальные средства для представления знаний. Так, *логические модели* для представления знаний используют кортеж вида:

$$M = \langle T, P, A, B \rangle,$$

где T — множество объектов предметной области; P — множество синтаксических правил для построения синтаксически правильных конструкций; A — множество аксиом предметной области; B — множество правил вывода. Основой базы знаний в логических системах является множество A аксиом предметной области. Все остальные знания выводятся из аксиом на основе правил вывода B .

Сетевые модели характеризуются набором:

$$H = \langle I, C_1, C_2, \dots, C_n, \Gamma \rangle,$$

где I — множество информационных единиц; C_1, C_2, \dots, C_n — множество типов связей между информационными единицами; Γ — отображение, устанавливающее связи между информационными единицами из заданного множества типов связей. Выделяются *классифицирующие сети*, *функциональные сети* и *сценарии*. Классифицирующие сети используют отношения структуризации. Функциональные сети вычисляют значения одних информационных единиц по значениям других и называются *вычислительными моделями*. Сценарии описывают каузальные отношения, отношения типов «орудие — действие», «средство — результат». Если в сети используются различные типы связей, то она называется *семантической сетью*.

Продукционные модели — это гибрид сетевых и логических моделей. Декларативная составляющая знаний в них представляется в виде сети, а процедурная — в виде множества правил вывода, которые называются *продукциями*. В результате применения правил вывода сеть может трансформироваться. Появляется *вывод на знаниях* вместо логического вывода. Процедурная составляющая знаний здесь явно

выделена и описывается иными средствами, чем декларативная составляющая.

Фреймовые модели используют жесткую структуру информационных единиц, представляемых в виде протофреймов, которые выглядят следующим образом:

(Имя фрейма:

Слот 1 (Значение слота 1)

Слот 2 (Значение слота 2)

Слот N (Значение слота N)).

В качестве значений слотов может использоваться все что угодно, в том числе и другие фреймы. Таким образом обеспечивается создание иерархических структур. Если в качестве значений слотов подставить конкретные данные, то получим экземпляр фрейма. Некоторые исследователи не выделяют фреймовые модели в отдельный вид, а рассматривают их в общем контексте с сетевыми моделями.

Сам термин «знания» в компьютерной литературе определяется весьма нестрого. Большинство авторов определяют знания как данные, имеющие некоторые особенности. Среди особенностей знаний выделяются различные наборы. В работе [2] утверждается, что знания отличаются от данных интерпретируемостью, наличием классифицирующих связей и ситуативных отношений. В работе [3] выделяются следующие особенности знаний по сравнению с данными: внутренняя интерпретируемость, структурированность, связность, семантическая метрика, активность. Еще одно определение знаний гласит следующее: знания — это данные, которые необходимо передать по каналу связи, чтобы на приемной стороне получить сообщение, пусть не во всех деталях, но вполне понятно его общее содержание [4]. Более конкретным и строгим является следующее определение: знания — это данные об объектах, отношениях и процессах [2]. Последнее определение является наиболее близким к формальному определению понятия знаний, которое предлагает теория интеллекта. Знание о факте — это отношение, выраженное некоторым высказыванием. Факт — это действительное состояние всех интересующих нас мест некоторого предметного пространства. Знание не такое определенное понятие, как факт. Оно лишь ограничивает множество возможных состояний мест предметного пространства [5].

Под предметным пространством в данном случае понимается декартово произведение $S = A_1 \times A_2 \times \dots \times A_m$ произвольных непустых не обязательно различных множеств A_1, A_2, \dots, A_m , которые являются подмножествами некоторого универсума предметов U . Множества A_1, A_2, \dots, A_m являются координатными осями пространства S . Число m называется размерностью пространства S . На данном пространстве задается множество предметных переменных $V = \{x_1, x_2, \dots, x_m\}$. Множество V называется универ-

сумом переменных пространства S . Значениями переменной x_i ($i = \overline{1, m}$) являются элементы множества A_i ($x_1 \in A_1, x_2 \in A_2, \dots, x_m \in A_m$). Множества A_1, A_2, \dots, A_m являются областями задания переменных x_1, x_2, \dots, x_m . Каждой переменной x_i ($i = \overline{1, m}$) ставится в соответствие фиксированная область задания A_i . Переменные x_1, x_2, \dots, x_m содержательно интерпретируются как места пространства S , а их значения (предметы) — как состояния этих мест. Любое подмножество P пространства S называется отношением, заданным на пространстве S . Отношение имеет размерность m . Все отношения, заданные на пространстве S , являются однотипными. Тип отношения задается множеством предметных переменных V и набором A_1, A_2, \dots, A_m областей их задания. Над однотипными отношениями можно выполнять операции объединения, пересечения и дополнения. Объединение отношений P и Q — это такое отношение $P \cup Q$, которому соответствуют все наборы отношения P и все наборы отношения Q . Пересечение отношений P и Q — это такое отношение $P \cap Q$, которому соответствуют все наборы как отношения P , так и отношения Q . Дополнением отношения P называется отношение $\sim P = \tilde{P}$, состоящее из тех наборов предметного пространства S , которые не принадлежат отношению P . Операциям над отношениями соответствуют операции над высказываниями. Объединению отношений $P_A \cup Q_B$ соответствует высказывание $A \vee B$, называемое дизъюнкцией высказываний A и B . Оно образуется с помощью союза «или», соединяющего высказывания A и B . Пересечению отношений $P_A \cap Q_B$ соответствует высказывание $A \wedge B$, называемое конъюнкцией высказываний A и B . Оно образуется соединением высказываний A и B с помощью союза «и». Дополнению \tilde{P}_A отношения P_A соответствует высказывание $\neg A = \bar{A}$, которое называется отрицанием высказывания A . Оно образуется присоединением к исходному высказыванию частицы «не».

Отношения задаются различными способами: в явном виде как множество упорядоченных наборов предметов, в табличном виде, в виде графов и графиков. Однако ни один из этих способов не позволяет представлять отношения в формульном виде. Как известно, язык формул является наиболее удобным и выразительным средством для присвоения имен объектам и описания их характеристик и связей с другими объектами. Кроме того, формульному описанию поддается только один вид отношений — функциональные отношения. Бинарное отношение R , заданное на множествах X и Y , называется функциональным, если из xRy_1 и xRy_2 следует, что $y_1 = y_2$, поэтому можно сказать, что для каждого значения $x \in X$ существует единственное значение $y \in Y$. Любое n -местное отношение R_n ,

заданное на множествах $X_1, X_2, \dots, X_{n-1}, Y$, называется функциональным, если из того, что $(x_1, x_2, \dots, x_{n-1}, y_1) \in R_n$ и $(x_1, x_2, \dots, x_{n-1}, y_2) \in R_n$ следует, что $y_1 = y_2$. Это значит, что для каждого фиксированного набора $(x_1, x_2, \dots, x_{n-1})$ существует единственное значение y такое, что $(x_1, x_2, \dots, x_{n-1}, y) \in R_n$. Отношения общего вида, которые не являются функциональными, можно отнести к классу многозначных функций, когда для каждого набора предметов $(x_1, x_2, \dots, x_{n-1})$, где $x_1 \in X_1, x_2 \in X_2, \dots, x_{n-1} \in X_{n-1}$, может существовать несколько значений $y_1, y_2, \dots, y_k \in Y$ таких, что $y_1 \neq y_2 \neq \dots \neq y_k$. Описать такие отношения с помощью формул не представляется возможным, поэтому воспользуемся методом перевода, когда решение одной задачи заменяется решением другой взаимно однозначно связанной с ней задачи. Для этого заданному отношению во взаимно однозначное соответствие ставится высказывание, а высказыванию — функция, которая принимает значение из множества $B = \{0, 1\}$, где 0 и 1 — булевы элементы «ложь» и «истина» соответственно. Такие функции называются предикатами. Предикат — это любая функция $P(x_1, x_2, \dots, x_m)$, заданная на предметном пространстве $S = A_1 \times A_2 \times \dots \times A_m$, которая отображает пространство S в множество $B = \{0, 1\}$, где 0 и 1 — булевы элементы.

Формально операцию замены отношения предикатом можно записать следующим образом:

$$P(x_1, x_2, \dots, x_m) = \begin{cases} 1, & \text{если } x_1, x_2, \dots, x_m \in P; \\ 0, & \text{если } x_1, x_2, \dots, x_m \notin P. \end{cases}$$

Символом $B = \{0, 1\}$ в данном случае обозначается отношение, а символом P — предикат. Обратный переход от предиката к отношению формально записывается так:

$$\begin{aligned} &\text{если } P(x_1, x_2, \dots, x_m) = 1, \text{ то } x_1, x_2, \dots, x_m \in P; \\ &\text{если } P(x_1, x_2, \dots, x_m) = 0, \text{ то } x_1, x_2, \dots, x_m \notin P. \end{aligned}$$

Переход от произвольного отношения P к предикату P и обратно можно представить в виде схемы (рис. 1).

Таким образом, любому отношению P во взаимно однозначное соответствие ставится предикат, который на языке алгебры предикатов можно записать в формульном виде, используя базисные предикаты 0, 1 и предикат узнавания предмета x_i^a , а также базисные операции конъюнкции \wedge и дизъюнкции \vee . Предикат узнавания предмета a , находящегося на месте x_i , определяется следующей формулой:

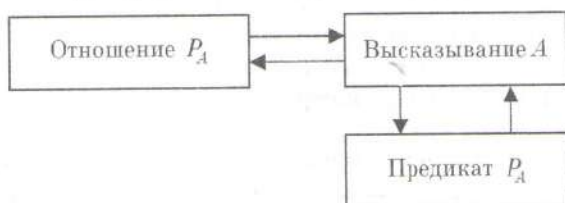


Рис. 1. Схема перехода от произвольного отношения к предикату

$$x_i^a = \begin{cases} 1, & \text{если } x_i = a; \\ 0, & \text{если } x_i \neq a. \end{cases}$$

На языке формул алгебры предикатов очень просто записать любые отношения. Отдельный факт, представленный набором (a_1, a_2, \dots, a_m) состояний a_1, a_2, \dots, a_m мест x_1, x_2, \dots, x_m , можно выразить высказыванием « $x_1 = a_1$ и $x_2 = a_2$ и ... и $x_m = a_m$ », которому соответствует формула алгебры предикатов $x_1^{a_1} \wedge x_2^{a_2} \wedge \dots \wedge x_m^{a_m}$. Формулы такого вида называются конъюнктами единицы предиката. Произвольное знание о факте можно представить отношением $\{(a_{11}, a_{21}, \dots, a_{m1}), (a_{12}, a_{22}, \dots, a_{m2}), \dots, (a_{1r}, a_{2r}, \dots, a_{mr})\}$, где r — число наборов в отношении. Это знание выражается высказыванием:

$$\begin{aligned} &x_1 = a_{11} \text{ и } x_1 = a_{21} \text{ и } \dots \text{ и } x_1 = a_{m1} \text{ или} \\ &x_1 = a_{12} \text{ и } x_1 = a_{22} \text{ и } \dots \text{ и } x_1 = a_{m2} \text{ или } \dots \\ &\text{или } x_1 = a_{1r} \text{ и } x_1 = a_{2r} \text{ и } \dots \text{ и } x_1 = a_{mr}. \end{aligned}$$

На языке алгебры предикатов это знание запишется в виде формулы:

$$\begin{aligned} &x_1^{a_{11}} \wedge x_2^{a_{21}} \wedge \dots \wedge x_m^{a_{m1}} \vee x_1^{a_{12}} \wedge x_2^{a_{22}} \wedge \dots \wedge x_m^{a_{m2}} \vee \\ &\vee \dots \vee x_1^{a_{1r}} \wedge x_2^{a_{2r}} \wedge \dots \wedge x_m^{a_{mr}}, \end{aligned}$$

которая называется совершенной дизъюнктивной нормальной формой (СДНФ) предиката.

Поскольку при переходе от отношений к предикатам в виде промежуточного звена используются высказывания, то операциям над высказываниями будут соответствовать операции над предикатами. Высказыванию $A \vee B$, называемому дизъюнкцией высказываний A и B , соответствует предикат

$$\begin{aligned} &(P_A \vee P_B)(x_1, x_2, \dots, x_m) = \\ &= P_A(x_1, x_2, \dots, x_m) \vee P_B(x_1, x_2, \dots, x_m). \end{aligned}$$

Высказыванию $A \wedge B$, называемому конъюнкцией высказываний A и B , соответствует предикат

$$\begin{aligned} &(P_A \wedge P_B)(x_1, x_2, \dots, x_m) = \\ &= P_A(x_1, x_2, \dots, x_m) \wedge P_B(x_1, x_2, \dots, x_m). \end{aligned}$$

Высказыванию $\neg A = \bar{A}$, которое называется отрицанием высказывания A соответствует предикат

$$(\neg P_A)(x_1, x_2, \dots, x_m) = \neg(P_A(x_1, x_2, \dots, x_m)).$$

Символы \vee, \wedge, \neg , стоящие слева от знака равенства, обозначают операции над предикатами, а те же символы, стоящие справа от знака равенства, — операции над значениями предикатов. Итак, мы убедились, что кроме предикатов существуют еще и операции над предикатами, а значит, должна существовать некоторая алгебра операций над предикатами.

Алгебра предикатов описывает только знания о фактах. Она формирует лишь то, что называется высказыванием, суждением. Алгебра операций над предикатами, или алгебра предикатных операций, должна «оживлять» эти знания. Она должна формализовать операции над знаниями, представленными в виде отношений на некотором предметном пространстве S . Алгебра предикатов описывает

декларативную составляющую знаний, а алгебра предикатных операций — процедурную составляющую знаний. Алгебра предикатных операций служит для осуществления вывода на знаниях, получения новых знаний на основе множества аксиом или первичных знаний, хранящихся в базе знаний интеллектуальной информационной системы. Машина должна не только иметь какие-то знания, но и обладать способностью оперировать ими, получать новые знания.

Приведем общие сведения об алгебрах предикатных операций. Пусть U — универсум предметов теории T , x_1, x_2, \dots, x_m — предметные переменные теории T , M — множество предикатов $P(x_1, x_2, \dots, x_m) = 0$, заданных на пространстве U^m . Множество M называется универсумом предикатов. Переменные X_1, X_2, \dots, X_n , заданные на множестве M , называются предикатными переменными. Их значениями служат предикаты, заданные на U^m . Множество M^n — это пространство предикатов размерностью n . Элементы множества M^n называются предикатными векторами. Любая функция $F(X_1, X_2, \dots, X_n) = Y$, отображающая M^n в M ($F: M^n \rightarrow M$), называется предикатной операцией. Пусть R — множество предикатных операций. *Алгеброй предикатных операций над R* называется любая алгебра на носителе R . Существует несколько алгебр предикатных операций на носителе R .

Булевой алгеброй предикатных операций называется любая алгебра предикатных операций с базисом операций отрицания X , дизъюнкции $X \vee Y$, конъюнкции $X \wedge Y$ для любых предикатных операций X и Y . Различают несколько булевых алгебр предикатных операций в зависимости от того, какие базисные элементы выбираются для каждой конкретной алгебры предикатных операций.

Алгеброй предикатных операций с константами и переменными называется булева алгебра предикатных операций с базисными элементами, состоящими из всевозможных тождественных и константных предикатных операций. Тождественной предикатной операцией по переменной X_j ($j = \overline{1, n}$) называется операция $F(X_1, X_2, \dots, X_j, \dots, X_n) = X_j$ при любых $X_1, X_2, \dots, X_j, \dots, X_n \in M$. Эта операция называется операцией выбора аргумента. Существует n таких операций. Константной предикатной операцией называется операция $F(X_1, X_2, \dots, X_j, \dots, X_n) = P$ при любых $X_1, X_2, \dots, X_j, \dots, X_n \in M$, где P — предикат на множестве M . Каждому предикату $P \in M$ соответствует своя константная предикатная операция. Всего существует $|M|$ константных предикатных операций, где $|M|$ — мощность множества M . Алгебра предикатных операций с константами и переменными при заданных m, n и универсуме предметов U , содержащем более одного элемента, неполна. Средствами этой алгебры невозможно выразить операцию

$$F(X) = \begin{cases} 1, & \text{если } X = P; \\ 0, & \text{если } X \neq P. \end{cases}$$

С этой операцией X_j^P алгебра предикатных операций с константами и переменными становится полной.

Дизъюнктивно-конъюнктивной алгеброй предикатных операций называется такая алгебра предикатных операций, базисными операциями которой являются операции дизъюнкции $X \vee Y$, конъюнкции $X \wedge Y$ для любых предикатных операций X и Y , а базисными элементами являются всевозможные константы $P \in M$ и предикаты узнавания предикатов X_j^P ($j = \overline{1, n}$), $P \in M$. При любых m, n и U дизъюнктивно-конъюнктивная алгебра предикатных операций полна. Любая предикатная операция в ней выражается формулой:

$$F(X_1, X_2, \dots, X_n) = \vee F(P_1, P_2, \dots, P_n) X_1^{P_1} X_2^{P_2} \dots X_n^{P_n}, \\ P_1, P_2, \dots, P_n \in M,$$

и называется совершенной дизъюнктивной нормальной формой (СДНФ) предикатной операции.

Фундаментальной алгеброй предикатных операций называется любая алгебра, у которой базисными операциями являются операции дизъюнкции $X \vee Y$, конъюнкции $X \wedge Y$ для любых предикатных операций X и Y , а базисными элементами являются предикаты 0 и 1 , всевозможные предикаты узнавания предмета x_i^a ($j = \overline{1, n}$; $a, x_i \in A_i$) и всевозможные предикаты узнавания предиката X_j^P ($j = \overline{1, n}$, $P, X_j \in B_j$). При любом носителе N алгебра предикатных операций полна. Дизъюнктивно-конъюнктивная алгебра предикатов на носителе M является подалгеброй алгебры предикатных операций на носителе N . Алгебра предикатных операций вместе с ее подалгеброй, являющейся алгеброй предикатов, называется *алгеброй предикатов и предикатных операций*.

Алгебра предикатов и предикатных операций в теории интеллекта может использоваться для описания законов поведения испытуемого, для описания баз данных и баз знаний, а также для описания структуры предложений и текста. Текст — это формула, записанная на языке некоторой лингвистической алгебры предикатных операций. Наиболее удобным для человека было бы представление знаний в базах знаний в виде текстов естественного языка. В этом случае отпадает необходимость в инженерах по знаниям, которые занимаются формализацией знаний, представленных на естественных языках. В высказывании на естественном языке зафиксированы отношения на трех уровнях:

1. Отношения, заданные на множестве букв естественного языка, которым соответствуют предикаты-слова.

2. Отношения, заданные на множестве слов естественного языка, задающие синтаксис языка, которым соответствуют операции над словами-предикатами.

3. Отношения, заданные на множестве семантических значений текстовых единиц. Они включают:
- отношения, заданные на множестве семантических значений морфем, которым соответствуют семантические предикаты слов;
 - отношения, заданные на множестве семантических значений слов, которым соответствуют операции над семантическими предикатами слов — алгебра семантических операций;
 - отношения, заданные на множестве семантических значений предложений, которым соответствуют семантические операции над семантическими операциями.

В качестве иллюстрации можно привести небольшой пример. Пусть высказывание на естественном языке выглядит следующим образом: «Я пишу статью». Слова, входящие в состав предложения, можно рассматривать как факты, а предложение — как знание о фактах. Первый факт называет имя субъекта «Я». Второй факт называет имя действия «писать». Третий факт говорит о том, что имя объекта «статья». На языке лингвистической алгебры операций предложение записывается следующим образом:

$$F(X_1, X_2, X_3) = X_1 \wedge X_2 \wedge X_3,$$

где X_1, X_2, X_3 — предикаты, соответствующие словам рассматриваемого предложения. На языке алгебры предикатов предикаты-слова выражаются в виде следующих формул: $X_1 = x_{11}^j$, $X_2 = x_{21}^n x_{22}^u x_{23}^u x_{24}^y$, $X_3 = x_{31}^c x_{31}^m x_{31}^a x_{31}^m x_{31}^b x_{31}^o$. Операция \wedge , соединяющая слова-предикаты в предложении, получила название «конъюнкция». Отношения, зафиксированные в данном предложении, соответствующие первым двум уровням, записываются в формульном виде следующим образом:

$$x_{11}^j \wedge x_{21}^n x_{22}^u x_{23}^u x_{24}^y \wedge x_{31}^c x_{31}^m x_{31}^a x_{31}^m x_{31}^b x_{31}^o = 1. \quad (1)$$

Полученное лингвистическое уравнение можно решить программным путем с помощью соответствующего решателя лингвистических уравнений, либо представить в виде лингвистической переключательной цепи, которую можно реализовать аппаратно, используя разработанные в последнее время микроэлектронной промышленностью программируемые логические интегральные микросхемы

(ПЛИС). Переключательная цепь, соответствующая лингвистическому уравнению (1), представлена на рис. 2.

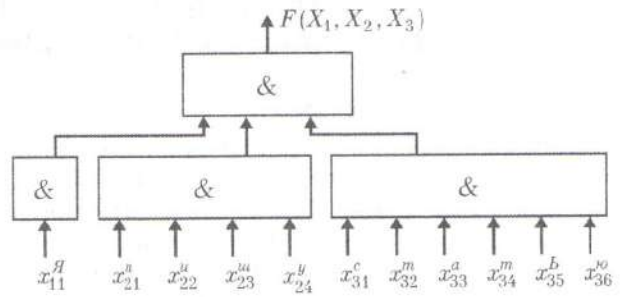


Рис. 2. Переключательная цепь

4. Основные результаты и выводы

Выводы. Используя алгебру предикатов и предикатных операций, можно создать интегрированную модель представления знаний (ИМПЗ), основанную на традиционных моделях представления знаний, а также на моделях представления знаний на естественном языке. В интеллектуальных информационных системах эти знания можно обрабатывать с помощью метода компараторной идентификации, а также современных методов Text Mining, которые позволяют эффективнее получать и обрабатывать информацию. Алгебрологические уравнения теории интеллекта используются для разработки лингвистических сопроцессоров в составе современных компьютеров. Кроме того, полученные предикатные уравнения и переключательные цепи становятся основой для создания компьютерных систем параллельной обработки данных.

Список литературы: 1. Перспективы развития вычислительной техники: В 11 кн.: Справ. пособие / Под ред. Ю. М. Смирнова. Кн. 1: Информационные семантические системы / Н. М. Соломатин. — М.: Высшая школа, 1989. — 127 с. 2. Кокорева Л. В., Перевозчикова О. Л., Ющенко К. Л. Диалоговые системы и представление знаний. — Киев: Наук. думка, 1992. — 448 с. 3. Искусственный интеллект: В 3-х кн. Кн. 2. Модели и методы: Справочник / Под ред. Д. А. Поспелова. — М.: Радио и связь, 1990. — 304 с. 4. Рафаэл Б. Думающий компьютер. — М.: Мир, 1979. — 400 с. 5. Бондаренко М. Ф., Дударь З. В., Ефимова И. А. и др. О мозгоподобных ЭВМ // Радиоэлектроника и информатика. — 2004. — № 2. С.

Поступила в редколлегию 06.04.2006

УДК 519.62

Г. Г. Четвериков, М. А. Дмитриева

ОБОСНОВАНИЕ ВЫБОРА И ПРИНЦИПЫ ПОСТРОЕНИЯ МНОГОЗНАЧНЫХ ОБРАТИМЫХ ЛОГИЧЕСКИХ ЭЛЕМЕНТОВ

1. Введение

Одной из основных тенденций современного развития вычислительной техники является увеличение интеграции в микроэлектронике, вызвавшее в результате переоценку стоимостей программных и аппаратных средств. Конкретными проявлениями этой тенденции являются: увеличение доли аппаратной реализации стандартных элементов программного обеспечения; построение однородных вычислительных систем с программируемой структурой; интенсивное внедрение микропрограммирования; применение табличной арифметики и групповых вычислений; разработка многозначных структур с высокой степенью распараллеливания [1–3].

В результате это привело к переработке и хранению информации на основе больших интегральных схемах (БИС), применение которых в дальнейшем будет сдерживаться технологическими факторами: ограничением количества выводов с кристалла и приближением размеров элементов к физическим пределам. Технологические ограничения и дальнейший рост объемов информации поставили задачу поиска методов ее уплотнения на качественно новом уровне. Существует мнение, что переход от двузначного представления логических функций к k -значному ($k > 2$) — один из возможных путей увеличения плотности упаковки информации. В связи с этим сейчас ведутся исследования принципов реализации многозначных БИС. В будущем это позволит использовать многозначные БИС в сочетании с существующими БИС. Многозначными называют структуры средств обработки данных, которые построены на базе многозначных логических элементов с соответствующими связями [1, 3].

Первые многозначные элементы появились в начале 60-х годов прошлого века, что вывело k -значную логику из области теоретических исследований: стали развиваться методы синтеза функций k -значной логики. Анализ этих методов (в основном, это методы минимизации) и опыт проектирования схем в k -значном алфавите показывают, что их эффективность зависит от выбора аналитического представления функций и количества одноместных операций в базе. Несмотря на то, что эти методы являются обобщением точных методов булевой алгебры, они имеют ряд особенностей, связанных с функциональными возможностями k -значных алгебр. И наконец, эти методы не рассчитаны на реализацию в двузначной схемотехнике: на современном этапе существования и развития средств вычислительной техники наблюдается кризис ее архитектуры и принципов

функционирования процессора Неймана с использованием исключительно двузначного кодирования.

Очевидно, что на многозначных БИС было бы целесообразно перенести способы функциональной и структурной организации двузначных БИС, основанные на принципах программируемости структуры и конструктивной однородности. Наиболее интересным классом БИС, реализующим эти принципы, являются матричные БИС — программируемые логические матрицы (ПЛМ).

2. Принципы построения многозначных обратимых неоднородных логических элементов

Математической основой построения обратимых переключательных цепей первого и второго рода для многозначных неоднородных кодов на базе предикатно-гибридной логики является алгебра конечных предикатов (АКП). Такие цепи используются при создании устройств распознавания образов, интеллектуальных устройств, в робототехнике, а также при создании систем для автоматической обработки текстовой информации. Другой областью применения таких цепей являются универсальные многозначные функциональные преобразователи пространственного типа. Однако создание многозначных структур и связанные с этим исследования проводились и проводятся, в основном, в направлении реализации цифровых вычислительных структур, машин и систем, использующих многозначное кодирование.

В последнее время предложены принципы построения многозначных элементов (МЭ), ориентированные на использование достижений современной интегральной технологии. Наибольшее распространение получили МЭ, построенные по принципу базиса с преобразованием информационных признаков сигналов. Один класс таких элементов с промежуточным преобразованием информационного признака динамического характера (т. е. каждому символу алфавита ставится в соответствие некоторая непериодическая последовательность) имеет ограниченное применение. Другой класс — с пространственным промежуточным представлением информации (каждому символу алфавита ставится в соответствие особое состояние в одной из точек некоторого пространства) — в большей мере отвечает требованиям вычислительной техники и современных автоматизированных информационных и интеллектуальных систем. При этом в большинстве работ, посвященных разработке методов синтеза цифровых устройств, авторы используют математические аппараты булевой алгебры и многозначной

логики. Однако специфика исследования, в качестве которого выступают математические модели естественного языка, указывает на нецелесообразность применения таких аппаратов. Это следует из того, что алгебра логики оперирует с двоичными символами, а аппарат многозначной логики, хотя и направлен на обработку непосредственно буквенной информации, не устраивает нас в силу своей направленности на описание функций, а не отношений. Однако именно отношения являются основным объектом описания естественного языка и его фрагментов.

В качестве основного инструментария математического описания конечных объектов выступает понятие конечного алфавитного оператора (КАО). При этом данное понятие служит хорошим средством для математического описания деятельности интеллекта, то есть тех реакций интеллекта, которые наблюдаются объективно. Правильный выбор же исходной и промежуточных форм представления КАО, с одной стороны, позволяет нам решить задачу построения элементов и устройств с предельным структурным быстродействием, с другой стороны, вооружает нас единообразным и универсальным приемом решения уравнений АКП. Известны различные способы аналитического представления алфавитных операторов. Их целесообразно разбить на два класса: методы явного и неявного задания операторов.

Рассмотрим алфавитный оператор вида $y_1 y_2 \dots y_n = g(x_1 x_2 \dots x_m)$, который можно представить в виде канонического уравнения:

$$g(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = 1, \quad (1)$$

при этом

$$g(x_1, \dots, x_m, y_1, \dots, y_n) = \begin{cases} 1, & \text{если } y_1 \dots y_n = G(x_1 \dots x_m) \\ 0, & \text{если } y_1 \dots y_n \neq G(x_1 \dots x_m) \end{cases} \quad (2)$$

Это уравнение есть реакция оператора G на входное слова длиной m ; результат — выходное слово длиной n . В случае вычисления значения отдельных букв выходного слова $y_1 y_2 \dots y_n$ требуется задать конечные алфавитные операторы $G_1 G_2 \dots G_n$, формирующие отдельные буквы $y_1 y_2 \dots y_n$ выходного слова:

$$\begin{aligned} y_1 &= G_1(x_1, x_2, \dots, x_m), \\ y_2 &= G_2(x_1, x_2, \dots, x_m), \\ &\dots \\ y_n &= G_n(x_1, x_2, \dots, x_m). \end{aligned} \quad (3)$$

При этом алфавитные операторы G_i ($i = \overline{1, n}$) можно записать также в виде уравнений типа (1):

$$\begin{aligned} q_1(x_1, x_2, \dots, x_m, y_1) &= 1, \\ &\dots \\ q_n(x_1, x_2, \dots, x_m, y_1) &= 1. \end{aligned} \quad (4)$$

Таким образом, очевидна связь между формулами (1) и (4), которая может быть представлена в виде:

$$\begin{aligned} &g(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = \\ &g_1(x_1, x_2, \dots, x_m, y_1) \wedge g_2(x_1, x_2, \dots, x_m, y_2) \wedge \dots \\ &\dots \wedge g_n(x_1, x_2, \dots, x_m, y_n). \end{aligned} \quad (5)$$

Ограничив область изменения буквенной переменной y значениями $a_1 a_2 \dots a_k$, свяжем ее уравнением:

$$y^{a_1} \vee y^{a_2} \vee \dots \vee y^{a_k} = 1. \quad (6)$$

С учетом такого ограничения любой конечный алфавитный оператор вида $y = Q(x_1 x_2 \dots x_m)$, заданный уравнением

$$q(x_1 x_2 \dots x_m, y) = 1, \quad (7)$$

можно записать следующей системой уравнений:

$$\begin{aligned} q(x_1, x_2, \dots, x_m, a_1) &= y^{a_1}, \\ q(x_1, x_2, \dots, x_m, a_2) &= y^{a_2}, \\ &\dots \\ q(x_1, x_2, \dots, x_m, a_k) &= y^{a_k}. \end{aligned} \quad (8)$$

В силу однозначности алфавитного оператора Q корень уравнения (7) a_i ($1 \leq i \leq k$) — единственный.

Тогда очевидно, что уравнение (7) и система уравнений (8) — равносильны. Таким образом, алфавитный оператор Q можно легко представить в виде системы уравнений:

$$\begin{aligned} q_1(x_1, x_2, \dots, x_m, a_{11}) &= y_1^{a_{11}}, \\ q_1(x_1, x_2, \dots, x_m, a_{12}) &= y_1^{a_{12}}, \\ &\dots \\ q_1(x_1, x_2, \dots, x_m, a_{1k_1}) &= y_1^{a_{1k_1}}, \\ q_2(x_1, x_2, \dots, x_m, a_{22}) &= y_2^{a_{22}}, \\ &\dots \\ q_1(x_1, x_2, \dots, x_m, a_{2k_2}) &= y_2^{a_{2k_2}}, \\ &\dots \\ q_n(x_1, x_2, \dots, x_m, a_{n2}) &= y_n^{a_{n2}}. \end{aligned} \quad (9)$$

Здесь области изменения переменных y_i ($i = \overline{1, n}$) ограничены множествами букв $M_i = \{a_{i1}, a_{i2}, \dots, a_{ik_i}\}$, где k_i — количество букв в множестве M_i . Система (9) задает алфавитный оператор G в явной форме и допускает непосредственное вычисление букв входного слова в отличие от формы (1), требующей решения соответствующих уравнений. Задание алфавитного оператора в форме (9) служит отправным моментом при построении многозначных обратимых логических элементов, методы синтеза которых будут подробнее рассмотрены ниже.

Таким образом, для построения быстродействующих многозначных структур и элементов, реализующих конечные отношения, целесообразно использовать предикатно-гибридную логику, математической основой которой является АКП. Последняя обладает возможностью широкого распараллеливания обрабатываемой информации, включая ее

значность ($k \geq 2$) и неоднородность на уровне отдельных обратимых логических элементов, сохраняя при этом однородность на уровне создаваемых структур. Перспективность такого подхода основана на возможности использования современной интегральной технологии изготовления СБИС. Рассмотренные принципы и предложенные методы синтеза многозначных обратимых неоднородных логических элементов расширяют функциональные возможности и повышают их надежность.

3. Модификация методов построения многозначных обратимых неоднородных логических элементов

Существенным отличием переключательных цепей второго рода является их обратимость «по частям». В данном случае схема многозначного обратимого неоднородного ЛЭ как бы повторяет структуру исходного многоместного отношения. В качестве аналога служит структура комбинационной схемы, построенной для исходного многоместного k -значного отношения. При этом все логические элементы в такой комбинационной схеме заменяем аналогичными по своему функциональному назначению многозначными обратимыми неоднородными логическими элементами первого рода, способы построения которых были рассмотрены в разделе 2. Рассмотрим алгоритм построения МОН ЛЭ второго рода на несложном примере.

Пусть задано многоместное отношение вида:

$$y = ab \vee cd. \tag{10}$$

Область определения переменных a, b, c и d ограничиваем следующими законами истинности:

$$\begin{aligned} a^0 \vee a^1 \vee a^2 &= 1, \\ b^0 \vee b^1 \vee b^2 \vee b^3 &= 1, \\ c^0 \vee c^1 \vee c^2 &= 1, \\ d^0 \vee d^1 \vee d^2 \vee d^3 \vee d^4 &= 1. \end{aligned} \tag{11}$$

Значность переменной y :

$$y(a, b, c, d) = \max(g(a, b), f(c, d)),$$

где функции g и f определяются как $g(a, b) = \min(a, b)$, $f(c, d) = \min(c, d)$ согласно логике их задания. Другими словами, исходное отношение может быть установлено в виде эквивалентной ему системы предикатных уравнений:

$$\begin{cases} y = g \vee f; \\ g = a \cdot b; \\ f = c \cdot d. \end{cases} \tag{12}$$

Модифицированная таблица истинности второго уравнения будет представлена ниже, в системе (13). Таблицы для первого, второго и третьего уравнений системы (12) представлены ниже (табл. 1–3).

Таблица 1

Модифицированная таблица истинности $g(a, b)$

| $a \setminus b$ | 0 | 1 | 2 | 3 | 4 |
|-----------------|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 1 | 1 |
| 2 | 0 | 1 | 2 | 2 | 2 |

Таблица 2

Модифицированная таблица истинности $f(c, d)$

| $c \setminus d$ | 0 | 1 | 2 | 3 |
|-----------------|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 1 |
| 2 | 0 | 1 | 2 | 2 |

Таблица 3

Модифицированная таблица истинности $y(g, f)$

| $g \setminus f$ | 0 | 1 | 2 |
|-----------------|---|---|---|
| 0 | 0 | 1 | 2 |
| 1 | 1 | 1 | 2 |
| 2 | 2 | 2 | 2 |

Эти таблицы более компактны по сравнению с традиционными таблицами истинности и являются отправным моментом при построении МОН ЛЭ второго рода. Обратимая переключательная цепь второго рода для исходного отношения (10) представлена на рис. 1.

Алгоритм функционирования такого устройства базируется на трех обратимых ЛЭ первого рода: обратимых конъюнкторах и обратимом дизъюнкторе (рис. 1).

Существенным отличием является наличие на входе устройства промежуточных переменных g и f . Однако, с учетом установленной интерпретации и k -фазного представления информации в таких устройствах, на все входные шины данных переменных (количество шин соответствует значности переменной) следует подать логическую единицу. Тем самым допускаем возможность любого значения для данной переменной из вычисляемой области ее определения по заданному отношению (10) или эквивалентной ему системы логических уравнений (12) с учетом значности исходных переменных, заданных усеченными законами истинности (11).

Для нашего случая логика работы первого, второго и третьего МОН ЛЭ определяется табл. 1, 2 и 3 соответственно.

Соединение логических элементов И и ИЛИ в первом МОН ЛЭ, например, полностью задается следующей системой предикатных уравнений:

$$\begin{aligned} g_{\text{ВЫХ}}^0 &= a^0 b^0 g^0 \vee a^1 b^0 g^0 \vee a^2 b^0 g^0 \vee \\ &\vee a^0 b^1 g^0 \vee a^0 b^2 g^0 \vee a^0 b^3 g^0; \\ g_{\text{ВЫХ}}^1 &= a^1 b^1 g^1 \vee a^2 b^1 g^1 \vee a^1 b^2 g^1 \vee a^1 b^3 g^1; \\ g_{\text{ВЫХ}}^2 &= a^2 b^2 g^2 \vee a^2 b^3 g^2; \\ a_{\text{ВЫХ}}^0 &= a^0 b^0 g^0 \vee a^0 b^1 g^0 \vee a^0 b^2 g^0 \vee a^0 b^3 g^0; \\ a_{\text{ВЫХ}}^1 &= a^1 b^0 g^0 \vee a^1 b^1 g^1 \vee a^1 b^2 g^1 \vee a^1 b^3 g^1; \\ a_{\text{ВЫХ}}^2 &= a^2 b^0 g^0 \vee a^2 b^1 g^1 \vee a^2 b^2 g^2 \vee a^2 b^3 g^2; \\ b_{\text{ВЫХ}}^0 &= a^0 b^0 g^0 \vee a^1 b^0 g^0 \vee a^2 b^0 g^0; \end{aligned} \tag{13}$$

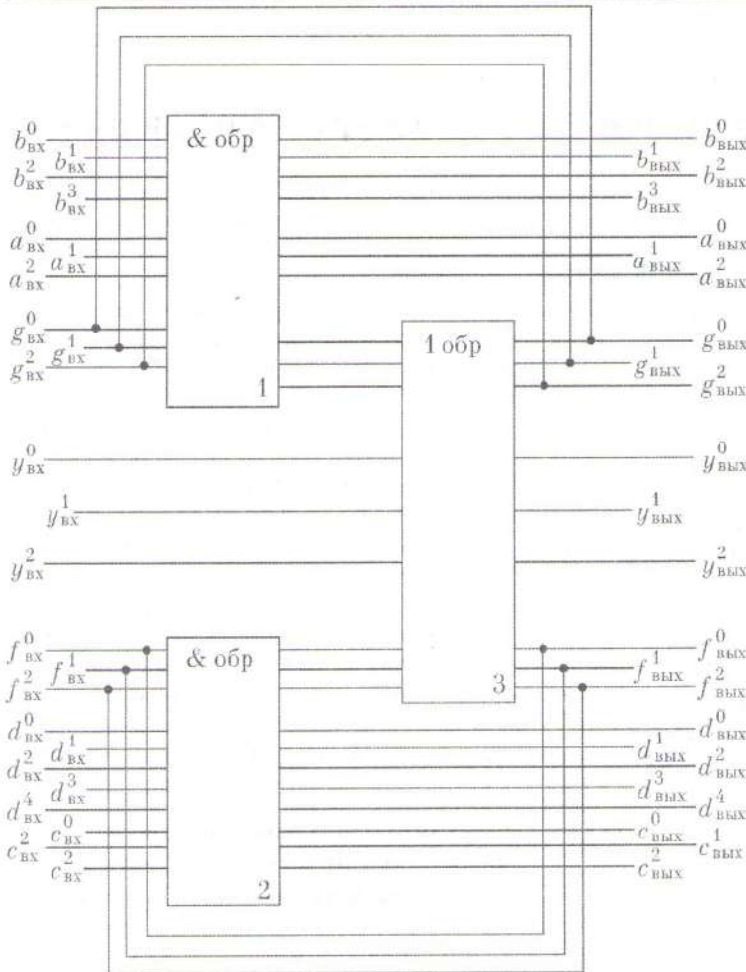


Рис.1. Обратимая переключательная цепь (МОН ЛЭ) второго рода, реализующая логическое уравнение вида $y = ab \vee cd$.

$$b_{\text{ВЫХ}}^1 = a^0 b^1 g^0 \vee a^1 b^1 g^1 \vee a^2 b^1 g^1;$$

$$b_{\text{ВЫХ}}^2 = a^0 b^2 g^0 \vee a^1 b^2 g^1 \vee a^2 b^2 g^2;$$

$$b_{\text{ВЫХ}}^3 = a^0 b^3 g^0 \vee a^1 b^3 g^1 \vee a^2 b^3 g^2.$$

Здесь и далее для краткости записи индекс, обозначающий входное значение той или иной переменной (в правой части уравнения), будем опускать.

Аналогичные системы предикатных уравнений, задающие соединения логических элементов И и ИЛИ для второго и третьего МОН ЛЭ, представленные на рис. 1, могут быть получены из соответствующих табл. 1, 2.

При этом функциональные схемы рассматриваемых МОН ЛЭ первого рода строятся таким образом, чтобы получить предельное структурное быстродействие, подобное тому, которое характерно для линейных дешифраторов и равно обратной величине средней задержки распространения сигнала типового элемента И (ИЛИ) для соответствующих типов интегральных микросхем. Первый ярус содержит полный набор элементов И на три входа, соответствующий

получению всех конституент единицы исходной таблицы истинности. В данные конституенты входят все входные переменные рассматриваемого отношения. Второй ярус предназначен для получения дизъюнкции тех или иных конституент единицы. Выход каждого такого элемента ИЛИ является выходным значением того или иного значения соответствующей переменной.

4. Выводы

В данной статье предложена модификация традиционных таблиц истинности для многозначных неоднородных кодов, связанных предикатными уравнениями, позволяющая представлять эти таблицы в компактной форме.

В статье метод синтеза обратимых переключательных цепей первого рода распространен на случай неоднородных кодов, заданных предикатными уравнениями.

Обоснована целесообразность перехода к переключательным цепям второго рода при числе переменных в исходном предикатном уравнении более 8, при значности переменных, не превышающей 6, что является характерным для задач морфологического анализа и синтеза.

Впервые предложены методы синтеза и осуществлен анализ базовых многозначных неоднородных логических элементов второго рода. Показана возможность эффективного построения переключательных цепей за счет варьирования базовых модулей [3, 4].

Были выявлены особенности и проанализирован ряд характерных свойств (регенерации, частичной регенерации, выявления противоречий) переключательных цепей прямого и обратного действия [3–5].

Список литературы: 1. Раков М. А. Реализация многозначных структур автоматки. – К.: Наук. думка, 1976. – 350 с. 2. Самофалов К. Г., Корнейчук В. П., Романкевич А. М., Тарасенко В. П. Цифровые многозначные элементы и структуры. – Киев: Вища школа, 1974. – 168 с. 3. Четвериков Г. Г. Многозначные структуры (анализ, сравнение, синтез, обобщение): Ч. 1. – К.: ИСМО, 1997. – 192 с. 4. Бондаренко М. Ф., Коноплянко З. Д., Четвериков Г. Г. Основы теории побудови багатозначних структур та кодування в системах штучного інтелекту. – Харків: Фактор-друк, 2003. – 336 с. 5. Бондаренко М. Ф., Коноплянко З. Д., Четвериков Г. Г. Основы теории синтезу над швидкодіючих структур мовних систем штучного інтелекту. – К.: ІЗМН, 1997. – 264 с.

Поступила в редакцию 14.03.2006

УДК 519.7

В. М. Левыкин, Т. М. Неофитная

РАЗРАБОТКА ПРАВИЛ ПОПОЛНЕНИЯ ИЕРАРХИЧЕСКОЙ БАЗЫ ЗНАНИЙ**1. Введение**

Модель концептуализации предметной области, используемая в базе знаний автоматизированной системы, может иметь слишком много релевантных деталей, что затрудняет решение отдельных функциональных задач пользователей, т. е. возникает проблема необозримости модели концептуализации (МК) предметной области (ПрО). При этом высокая степень детализации МК ПрО может оказаться несогласованной с необходимой точностью решения и необоснованными дополнительными затратами. В этом случае эффективной является декомпозиция знаний о ПрО в виде иерархии абстракций, которая позволяет распределить знания о ПрО по различным уровням детализации.

Представление семантики предметной области в виде иерархии абстракций является фундаментальным принципом объектно-ориентированной технологии разработки баз знаний [1]. При этом смысл абстракции состоит в извлечении существенных деталей, опуская при этом несущественные, и в раскрытии деталей на других уровнях детализации. На высших уровнях иерархии абстракций используются наименее детальные концептуальные представления, отражающие только самые общие знания о ПрО. На следующих уровнях степень подробности возрастает, причем рассматриваются только отдельные сущности ПрО, а не ПрО в целом.

Рассмотрим основные преимущества поддержки иерархии абстракций в базе знаний автоматизированной системы [1, 2].

1. Появляется возможность игнорировать многие релевантные детали знаний о ПрО на более высоком уровне абстракции, что часто облегчает поиск решения задачи.

2. В базе знаний могут быть эффективно интегрированы абстракции, имеющие отношение к различным пользователям системы.

3. Наблюдается повышенная стабильность базы знаний по отношению к различным эволюционным изменениям.

Выделяют две основных формы абстракции: агрегацию и обобщение [2]. Агрегация — это такая абстракция, которая превращает связь между концептами в некоторый агрегированный (составной) концепт. Обобщение — это абстракция, превращающая класс концептов в родовую концепт. Таким образом, совокупность моделей концептуализации ПрО можно стратифицировать по различным уровням агрегации с помощью введения специального отношения агрегации, а также по различным уровням обобщенности с помощью введения специального отношения обобщения.

Существование отношения обобщения между МК ПрО означает, что одна модель является обобще-

нием другой модели. Существование отношения агрегации между МК ПрО означает, что одна модель является компонентом другой модели. Комбинация отношений обобщения и агрегации приводит к сложной иерархической структуре базы знаний о ПрО, которая имеет определенные семантические закономерности. Данные закономерности необходимо учитывать при разработке процедуры создания баз знаний в диалоге с пользователем. Поэтому целью данной работы является выявление и учет семантических закономерностей иерархических структур систем знаний о ПрО. Для этого в данной работе решаются следующие задачи: моделирование отношений обобщения и агрегации МК ПрО; выявление семантических закономерностей иерархических структур систем знаний о ПрО; разработка правил пополнения иерархической базы знаний в процессе ее разработки с учетом выявленных метазакономерностей иерархических структур систем знаний о ПрО. Данные правила позволят поддержать семантическую корректность иерархической базы знаний о ПрО.

2. Семантические закономерности иерархических структур систем знаний

База знаний автоматизированной системы разрабатывается таким образом, чтобы иерархическая структура отношений функций классов объектов ПрО, описываемых МК ПрО, предопределяла иерархическую структуру отношений между МК ПрО. Пусть F — множество функций классов объектов, описываемых МК ПрО. На данном множестве естественно определить отношения обобщения и агрегации, которые обозначим «*is-a*» и «*part-of*». Отношения агрегации и обобщения, заданные на множестве функций, являются отношениями порядка, т. е. данные отношения рефлексивны, антисимметричны и транзитивны.

Определим отображение $Func$, которое каждой МК ПрО ставит в соответствие функциональное название описываемого данной моделью класса объектов ПрО. Запись $Func(B) \text{ is-a } Func(A)$ означает, что функция класса объектов, описываемого моделью концептуализации A , является обобщением функции класса объектов, описываемого моделью концептуализации B . Запись $Func(A) \text{ part-of } Func(X)$ означает, что функция класса объектов, описываемого моделью концептуализации A , является частью функции класса объектов, описываемого моделью концептуализации X .

Если функция класса объектов, описываемого моделью концептуализации A , является обобщением функции класса объектов, описываемого моделью концептуализации B , то модель концептуализации A является обобщением модели концептуализации B . Если функция класса объектов, описываемого моделью концептуализации A , является частью функ-

ции класса объектов, описываемого моделью концептуализации X , то модель концептуализации A является частью модели концептуализации X .

Отношения обобщения и агрегации между МК ПрО эффективно описываются с помощью морфизмов. Морфизм из подчиненной МК ПрО в данную МК ПрО назовем морфизмом агрегации. Наиболее строгим определением морфизма агрегации является его определение как обладающего свойствами инъективного гомоморфизма и корреспонденции одновременно. Морфизм из данной МК ПрО в ее обобщенную МК ПрО назовем морфизмом наследования. В общем случае морфизм наследования является сюръективной корреспонденцией. Математические определения гомоморфизма и корреспонденции, заданных между МК ПрО, даны в работе [3].

На основании данных о морфизмах агрегации и морфизмах наследования могут быть получены некоторые другие морфизмы между МК ПрО. Действительно, морфизмы наследования и морфизмы агрегации в самом общем случае являются корреспонденциями. Тогда существуют произведения морфизмов наследования и морфизмов агрегации, называемые далее морфизмами связи, т. е. произведение корреспонденций является корреспонденцией. Для определения морфизмов связи воспользуемся следующими правилами.

Пусть о некоторой модели концептуализации A известно, что она является компонентом модели концептуализации X , т. е. существует морфизм агрегации $r: A \rightarrow X$. Кроме того, известно, что модель концептуализации A является обобщением модели концептуализации B , т. е. существует морфизм наследования $d: B \rightarrow A$. Тогда между моделями концептуализации B и X существует морфизм связи $k: B \rightarrow X$, при котором выполняется $k = dr$ (рис. 1, а).

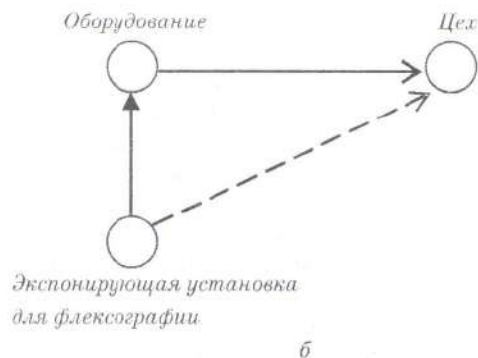
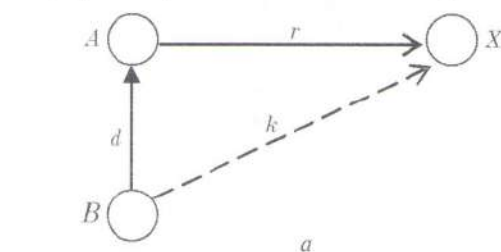


Рис. 1. Определение морфизма связи:

« \longrightarrow » — морфизм наследования;
« \longrightarrow » — морфизм агрегации; « \dashrightarrow » — морфизм связи

Для примера рассмотрим фрагмент базы знаний, касающийся области производства флексографских печатных форм (ФПФ) (рис. 1, б). Пусть в базе знаний определены следующие МК ПрО: «Оборудование», «Экспонирующая установка для флексографии», «Цех». Известно, что МК ПрО «Оборудование» является компонентом МК ПрО «Цех», т. е. существует морфизм агрегации из МК ПрО «Оборудование» в МК ПрО «Цех». Кроме того, МК ПрО «Оборудование» является обобщением МК ПрО «Экспонирующая установка для флексографии», т. е. существует морфизм наследования из МК ПрО «Экспонирующая установка для флексографии» в МК ПрО «Оборудование». Тогда существует морфизм связи из МК ПрО «Экспонирующая установка для флексографии» в МК ПрО «Цех», свидетельствующий о существовании отношения структурного сходства между данными моделями. Семантика данного отношения заключается в том, что экспонирующая установка для флексографии находится в цехе, поскольку оборудование находится в цехе.

Пусть о некоторой модели концептуализации B известно, что она является компонентом модели концептуализации Y , т. е. существует морфизм агрегации $w: B \rightarrow Y$. Кроме того, известно, что модель концептуализации B является обобщением модели концептуализации X , т. е. существует морфизм наследования $h: X \rightarrow B$. Тогда между моделями концептуализации B и X существует морфизм связи $k: B \rightarrow X$, при котором выполняется $k = wh$ (рис. 2, а).

Для примера рассмотрим фрагмент базы знаний, касающийся области производства ФПФ (рис. 2, б). Пусть в базе знаний определены следующие МК ПрО: «Экспонирующая установка для флексографии», «Цех», «Цех производства ФПФ». Известно, что МК ПрО «Экспонирующая установка для флексографии» является компонентом МК ПрО «Цех производства

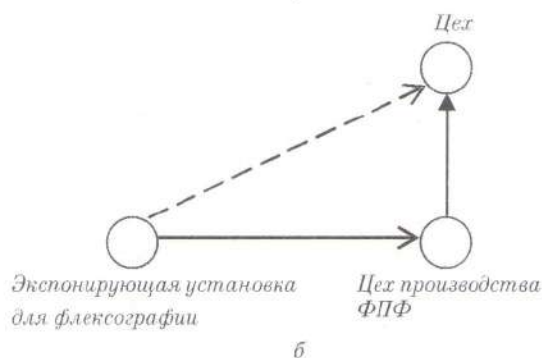
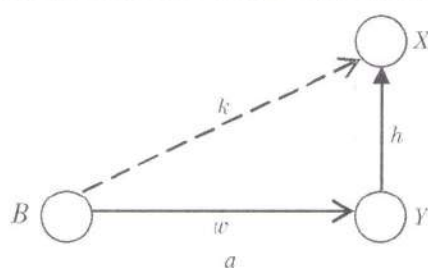


Рис. 2. Определение морфизма связи:

« \longrightarrow » — морфизм наследования;
« \longrightarrow » — морфизм агрегации; « \dashrightarrow » — морфизм связи

ФПФ», т. е. существует морфизм агрегации из МК ПрО «Экспонирующая установка для флексографии» в МК ПрО «Цех производства ФПФ». А МК ПрО «Цех производства ФПФ» является обобщением МК ПрО «Цех производства ФПФ», т. е. существует морфизм наследования из МК ПрО «Цех производства ФПФ» в МК ПрО «Цех». Тогда существует морфизм связи из МК ПрО «Экспонирующая установка для флексографии» в МК ПрО «Цех», который говорит о существовании отношения структурного сходства между данными моделями. Семантика данного отношения заключается в том, что экспонирующая установка для флексографии находится в цехе, поскольку она находится в цехе производства ФПФ.

Рассмотренные примеры проясняют семантику отношений структурного сходства, соответствующих морфизмам связи между МК ПрО. Из данных примеров следует, что, с одной стороны, экспонирующая установка для флексографии находится в цехе, поскольку оборудование находится в цехе, с другой стороны, экспонирующая установка для флексографии находится в цехе, поскольку она находится в цехе производства ФПФ (рис. 3, б). Таким образом, получаем коммутативную диаграмму $dr = k = wh$, где $d: B \rightarrow A$ — морфизм наследования, $r: A \rightarrow X$ — морфизм агрегации, $k: B \rightarrow X$ — морфизм связи, $w: B \rightarrow Y$ — морфизм агрегации, $h: Y \rightarrow X$ — морфизм наследования (рис. 3, а).

Для поддержки коммутативных диаграмм вида $dr = wh$ требуется в рамках итеративной процедуры создания баз знаний в диалоге с пользователем выполнять специальные правила пополнения иерархической базы знаний ПрО. При этом необходимо рассмотреть две стратегии проектирования базы знаний — нисходящую с последовательным уточнением МК ПрО сверху вниз и восходящую с последовательным обобщением МК ПрО снизу вверх.

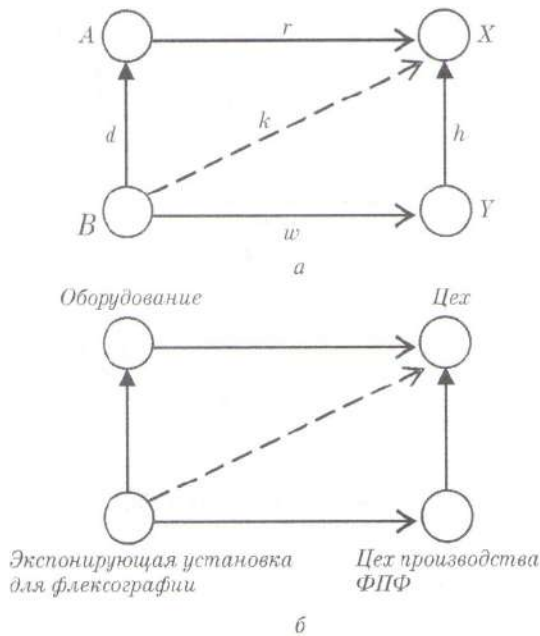


Рис. 3. Фрагмент иерархической структуры знаний:
 $\leftarrow \rightarrow$ — морфизм наследования;
 \rightarrow — морфизм агрегации; \dashrightarrow — морфизм связи

3. Правила пополнения иерархической базы знаний ПрО

в случае нисходящей стратегии проектирования

Пусть известно, что некоторая модель концептуализации A является компонентом модели концептуализации X , т. е. существует морфизм агрегации $r: A \rightarrow X$. Предположим, что требуется уточнить знания о ПрО, представляемые моделью концептуализации X . Для этого от концепции X была унаследована новая концепция Y , т. е. должен существовать морфизм наследования $h: Y \rightarrow X$. В этом случае в качестве компонента модели концептуализации Y вместо модели концептуализации A может быть использована только такая модель концептуализации B , при котором существует морфизм наследования $d: B \rightarrow A$, т. е. модель концептуализации B является унаследованной от модели концептуализации A .

Рассмотрим пример (рис. 3, б). Пусть в базе знаний определены следующие МК ПрО: «Оборудование», «Экспонирующая установка для флексографии», «Цех». Известно, что МК ПрО «Оборудование» является компонентом МК ПрО «Цех», а МК ПрО «Экспонирующая установка для флексографии» является унаследованной от МК ПрО «Оборудование». Предположим, что требуется уточнить знания о ПрО, представляемые моделью концептуализации «Цех». Для этого от МК ПрО «Цех» была унаследована новая МК ПрО «Цех производства ФПФ». Цех производства ФПФ содержит промышленное оборудование; точнее, оборудованием, которое содержится в цехе производства ФПФ, в том числе является экспонирующая установка для флексографии [4]. Поэтому вместо МК ПрО «Оборудование» в качестве компонента МК ПрО «Цех производства ФПФ» используется МК ПрО «Экспонирующая установка для флексографии».

Дополнительный анализ показал, что приведенное выше правило не отражает в полной мере семантические закономерности структуры системы знаний о предметной области. Оказывается, что выбор модели концептуализации B в качестве компонента модели концептуализации Y не является случайным. Осуществляя этот выбор, необходимо учитывать функциональное назначение классов объектов, описываемых моделями концептуализации ПрО. Точнее, модель концептуализации B может быть выбрана в качестве компонента модели концептуализации Y , если функция класса объектов, описываемого моделью концептуализации B , является частью функции класса объектов, описываемого моделью концептуализации Y , т. е. $Func(B) \text{ part-of } Func(Y)$. Кроме того, при определении модели концептуализации Y необходимо, чтобы функция класса объектов, описываемого моделью концептуализации X , являлась обобщением функции класса объектов, описываемого моделью концептуализации Y , т. е. $Func(Y) \text{ is-a } Func(X)$.

Сформулированное правило носит эвристический характер и может учитываться только в том случае, если выполняется предыдущее правило. Уже на ранее рассмотренном примере (рис. 3, б) видно, что

не случайно МК ПрО «Экспонирующая установка для флексографии» была выбрана в качестве компонента МК ПрО «Цех производства ФПФ». Функциональным назначением класса объектов, описываемого МК ПрО «Цех производства ФПФ», является производство ФПФ из флексографских фотополимеризующихся пластин. Функциональным назначением класса объектов, описываемого МК ПрО «Экспонирующая установка для флексографии», является экспонирование флексографских фотополимеризующихся пластин, что является составной частью производства ФПФ из флексографских фотополимеризующихся пластин.

Таким образом, разработаны правила пополнения иерархической базы знаний ПрО в случае нисходящей стратегии проектирования базы знаний. Данные правила, по сути, позволяют предсказать морфизмы агрегации между МК ПрО, если определены функции классов объектов, описываемых данными моделями.

4. Правила пополнения иерархической базы знаний ПрО в случае восходящей стратегии проектирования

Пусть известно, что некоторая модель концептуализации B является компонентом модели концептуализации Y , т. е. существует морфизм агрегации $w: B \rightarrow Y$. Предположим, что требуется обобщить знания о ПрО, представляемые моделью концептуализации Y . Для этого была определена новая концепция X , такая, что должен существовать морфизм наследования $h: Y \rightarrow X$. В этом случае в качестве компонента модели концептуализации X вместо модели концептуализации B может быть использована только такая модель концептуализации A , для которой существует морфизм наследования $d: B \rightarrow A$, т. е. модель концептуализации A является обобщением модели концептуализации B .

Рассмотрим пример (рис. 3, б). Пусть в базе знаний определены следующие МК ПрО: «Экспонирующая установка для флексографии», «Цех производства ФПФ», «Оборудование». Известно, что МК ПрО «Экспонирующая установка для флексографии» является компонентом МК ПрО «Цех производства ФПФ». Кроме того, МК ПрО «Оборудование» является обобщением МК ПрО «Экспонирующая установка для флексографии». Предположим, что требуется обобщить знания о ПрО, представляемые моделью концептуализации «Цех производства ФПФ». Для этого была определена МК ПрО «Цех», которая является обобщением МК ПрО «Цех производства ФПФ». Цех не обязательно содержит экспонирующую установку для флексографии, но обязательно содержит промышленное оборудование. Поэтому вместо МК ПрО «Экспонирующая установка для флексографии» в качестве компонента МК ПрО «Цех» используется МК ПрО «Оборудование».

Однако приведенное выше правило не отражает в полной мере семантические закономерности структуры системы знаний о предметной области. Выбор модели концептуализации A в качестве ком-

понента модели концептуализации X не является случайным. Осуществляя этот выбор, необходимо учитывать функциональное назначение классов объектов, описываемых моделями концептуализации ПрО. Точнее, модель концептуализации A может быть выбрана в качестве компонента модели концептуализации X , если функция класса объектов, описываемого моделью концептуализации A , является частью функции класса объектов, описываемого моделью концептуализации X , т. е. $Func(A)$ *part-of* $Func(X)$. Кроме того, при определении модели концептуализации X необходимо, чтобы функция класса объектов, описываемого моделью концептуализации X , являлась обобщением функции класса объектов, описываемого моделью концептуализации Y , т. е. $Func(Y)$ *is-a* $Func(X)$.

Сформулированное правило является эвристическим и может учитываться только в том случае, если выполняется предыдущее правило. На ранее рассмотренном примере (рис. 3, б) видно, что не случайно МК ПрО «Оборудование» была выбрана в качестве компонента МК ПрО «Цех». Функциональным назначением класса объектов, описываемого МК ПрО «Цех», является производство промышленной продукции. Функциональным назначением класса объектов, описываемого МК ПрО «Оборудование», является выполнение технологических операций, что является составной частью производства промышленной продукции.

Таким образом, разработаны правила пополнения иерархической базы знаний ПрО в случае восходящей стратегии проектирования базы знаний. Данные правила, как и предыдущие, позволяют предсказать морфизмы агрегации между МК ПрО, если определены функции классов объектов, описываемых данными моделями.

5. Выводы

В данной работе рассмотрены семантические закономерности иерархических структур систем знаний ПрО и обоснована необходимость их учета для получения качественных баз знаний автоматизированных систем. На этой основе предложены правила пополнения иерархической базы знаний автоматизированной системы. Использование данных правил в рамках процедуры создания базы знаний в диалоге с пользователем позволит проводить проверку корректности базы знаний и упростит процесс ее уточнения.

Список литературы: 1. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. – СПб: Питер, 2000. – 384 с. 2. Smith J., Smith D. Database Abstractions: Aggregations and Generalization // ACM Transactions on Database Systems. – 1977. – Vol. 2, № 2. – P. 105–133. 3. Неофитная Т. М. Разработка категорной модели базы знаний интеллектуальной системы принятия решений // Проблемы бюрократии. – 2003. – № 58. – С. 108–115. 4. Киппхан Г. Энциклопедия по печатным средствам информации. Технологии и способы производства: Пер. с нем. – М.: МГУП, 2003. – 1280 с.

Поступила в редакцию 14.08.2006

УДК 658.012.56

Т. П. Любченко

МОРФОЛОГІЧНА МОДЕЛЬ СЛОВОЗМІНИ ФЛЕКТИВНОЇ МОВИ ТА ЕЛЕКТРОННИЙ ГРАМАТИЧНИЙ СЛОВНИК

1. Вступ

У системах автоматичної обробки тексту (АОТ), зокрема в системах машинного перекладу (МП), типовими є дві операції над словами: лематизація, тобто редукція текстового слова до його словникової форми, та синтез необхідної текстової словоформи від її початкової (словникової) форми. Вирішення зазначених завдань для мов флективного типу, до яких належить і російська мова, спрямовує до необхідності комп'ютерно-лінгвістичного дослідження системи словозміни, побудови парадигматичної класифікації лексики, створення бази даних, що підтримує дану систему словозміни, розробки алгоритмів формування як повної словозмінної парадигми слова, так і конкретної його текстової форми. Необхідність реалізації такої програми на репрезентативному мовному масиві (хоча б у обсязі 150 тисяч словникових одиниць) робить розв'язання цих завдань виключно складним у «ручному» режимі й вимагає розробки спеціальної комп'ютерної технології.

Для російської мови цю проблему в ряді відомих систем МП і АОТ успішно вирішено, однак скористатися цими результатами майже неможливо, оскільки відомі нам програмні продукти, поширені на ринку, є закритими. Викладене й спонукало нас до самостійної розробки технології автоматизованої побудови словозмінної парадигми російської мови, створення електронного граматичного (парадигматичного) словника, який став би інструментом морфологічного аналізу російських текстів у системі МП, що розробляється в Українському мовно-інформаційному фонді Національної академії наук України (УМІФ НАНУ).

Для побудови граматичного словника визначальним фактором є наявність формальної моделі словозміни, що означає встановлення та формалізацію лінгвістичних критеріїв, згідно з якими вся множина слів мови розбивається на певні підмножини, взаємний перетин яких є порожнім і всередині якого словозміна відбувається за однаковими правилами. Підмножини слів з такими властивостями називаються *словозмінними парадигматичними класами*.

Моделювання розподілу множини слів мови на словозмінні парадигматичні класи відбувається у декілька етапів. На першому визначається поняття *парадигматичного типу*, в означенні якого принципову роль відіграють поняття граматичної категорії, граматичного значення та граматичної форми [1].

2. Словозмінні парадигматичні типи та словозмінні параметри російської мови

Уведемо позначення. Нехай L — фіксована мова (флективна)¹, W — множина слів мови L ;

P_j ($j = 1, 2, \dots, p$) — граматичні класи², p — кількість граматичних класів;

$W(P_j)$ — множина слів мови L , яка належить до граматичного класу P_j ;

T_i ($i = 1, 2, \dots, N$) — парадигматичні (морфологічні) типи, N — кількість парадигматичних (морфологічних) типів;

$W(T_i)$ — множина слів мови L , яка належить до типу T_i ;

$\Omega(T_i)$ — множина граматичних значень, що відповідають типу T_i .

За ознакою приналежності до певної частини мови та за додатковими ознаками, які є класифікуючими (не словозмінними) в межах певної частини мови, множину слів W розподіляємо на підмножини, котрі називатимемо *граматичними класами*, таким чином.

Іменники за значенням граматичної категорії «рід» (яка в межах цієї частини мови є класифікуючою ознакою) розподіляються на три граматичні класи: іменники чоловічого роду, іменники жіночого та іменники середнього роду; множинні іменники утворюють окремий граматичний клас [2]. Таким чином, іменники складають чотири граматичні класи (позначатимемо їх відповідно P_1, P_2, P_3, P_4).

Дієслова за значенням граматичної категорії «вид» (яка не є словозмінною ознакою і розглядається нами як класифікуюча) розподіляються на такі три граматичні класи: дієслова доконаного виду, дієслова недоконаного виду та двовидові дієслова [2].

Усі інші слова з множини W , які не є іменниками або дієсловами, віднесені до своїх граматичних класів за ознакою приналежності до конкретної частини мови (тобто у цьому випадку поняття «граматичний клас» збігається з поняттям «частина мови»).

Таким чином, у російській мові нами виділено такі граматичні класи: іменники чоловічого роду (P_1), іменники жіночого роду (P_2), іменники середнього роду (P_3), множинні іменники (P_4), ад'єктиви (прикметники+порядкові числівники) (P_5), дієслова доконаного виду (P_6), дієслова недоконаного виду (P_7), двовидові дієслова (P_8), дієприкметники (P_9),

¹Наводиться на прикладі російської мови.

²Грамматичний клас — аналог частини мови.

займенники (займенники-іменники) (P_{10}), займенники-прикметники (P_{11}), числівники кількісні (P_{12}), прислівники (P_{13}), вигуки (P_{14}), сполучники (P_{15}), частки (P_{16}), прийменники (P_{17}), предикативи (P_{18}), скорочення (P_{19}).

Отже, $W = \bigcup_{j=1}^{19} W(P_j)$. Вважатимемо омонімію знятою, а омонімі промаркованими. Тоді

$$W(P_{j_1}) \cap W(P_{j_2}) = \emptyset \text{ при } j_1 \neq j_2, j_1, j_2 = 1, 2, \dots, 19.$$

За словозмінними категоріями, що визначають словозмінну парадигму конкретних слів (сукупність граматичних значень та відповідних граматичних форм), вводимо такі *парадигматичні типи*.

Парадигматичний тип, що характеризується граматичними формами, які визначаються граматичними значеннями словозмінних категорій «число» та «відмінок», називатимемо *субстантивним парадигматичним типом*³:

$$W(T_1) \equiv W^S = \{w_1^S, w_2^S, \dots, w_{12}^S\}. \quad (1)$$

При цьому граматичні форми $w_i^S = w_i^S(n, k)$ визначаються множиною граматичних значень $\Omega(T_1) = \{\omega_i^S\}$, елементами якої є значення граматичних категорій «число» та «відмінок» (тобто граматичні форми визначаються парами граматичних значень (число, відмінок)):

$$\omega_i^S = \{n_1, k_i\}, \omega_{i+6}^S = \{n_2, k_i\}, i = 1, 2, \dots, 6, \quad (2)$$

де n_1 — одинна, n_2 — множина; k_i — значення відмінків: k_1 — називний, k_2 — родовий, k_3 — давальний, k_4 — знахідний, k_5 — орудний, k_6 — місцевий.

До субстантивного парадигматичного типу належать усі іменники та займенники-іменники (займенники, які є заміниками іменників — особові займенники *я, ты, он, мы, вы, они* та такі як *кто, что, нечто, нечто*).

Зазначимо, що в кожній з граматичних форм конкретна лексема може мати одну або декілька словоформ чи не мати жодної словоформи. За відсутності реалізації лексеми в певній граматичній формі словозмінну парадигму вважатимемо дефектною. Прикладом подібних випадків є відсутність форм множини в іменників *singularia tantum*, або відсутність форм однини в іменників *pluralia tantum*. Для урахування факту відсутності словоформ для деяких граматичних значень у формулу опису парадигма-

тичного типу (1) введемо параметр *def*, який називатимемо *параметром дефектності*:

$$W(T_1) \equiv W^S = \{w_1^S, w_2^S, \dots, w_{12}^S, def\}, \quad (3)$$

$$\Omega(T_1) = \{\omega_1^S, \omega_2^S, \dots, \omega_{12}^S\},$$

що вказує на номери (числа) граматичних значень, для яких відповідні словоформи відсутні в повній парадигмі; якщо дефектності немає, то за визначенням покладаємо $def = 0$.

Парадигматичний тип, що характеризується граматичними формами, які визначаються множиною граматичних значень $\Omega(T_2) = \{\omega_i^A\}$, елементами якої є значення граматичних (словозмінних) категорій «рід», «число» та «відмінок», будемо називати *ад'єктивним парадигматичним типом*⁵:

$$W(T_2) \equiv W^A = \{w_1^A, w_2^A, \dots, w_{28}^A, def\}, \quad (4)$$

$$\Omega(T_2) = \{\omega_1^A, \omega_2^A, \dots, \omega_{28}^A\},$$

де граматичні форми w_i^A визначаються для граматичних значень ω_i^A , поданих трійками (рід, число, відмінок). Для форм однини:

$$\omega_1^A = \{g_1, n_1, k_i\}, \omega_{i+6}^A = \{g_2, n_1, k_i\}, \quad (5)$$

$$\omega_{i+12}^A = \{g_3, n_1, k_i\}, i = 1, 2, \dots, 6,$$

а для форм множини значення роду нерелевантне:

$$\omega_{i+18}^A = \{n_2, k_i\}, i = 1, 2, \dots, 6. \quad (6)$$

У формулах (4)–(6) g_1 — чоловічий рід, g_2 — жіночий рід, g_3 — середній рід, n_1 — одинна, n_2 — множина; k_i — значення відмінків: k_1 — називний, k_2 — родовий, k_3 — давальний, k_4 — знахідний, k_5 — орудний, k_6 — місцевий.

Короткі граматичні форми $w_{25}^A, w_{26}^A, w_{27}^A, w_{28}^A$ існують тільки для називного відмінку і визначаються категоріями роду та числа:

$$\omega_{25}^A = \{g_1, n_1, k_1\}, \omega_{26}^A = \{g_2, n_1, k_1\}, \quad (7)$$

$$\omega_{27}^A = \{g_3, n_1, k_1\}, \omega_{28}^A = \{n_2, k_1\}.$$

До ад'єктивного парадигматичного типу належать прикметники, займенники-прикметники, порядкові числівники, дієприкметники, кількісний числівник «один».

Для врахування можливої дефектності парадигми у деяких прикметників до формули (4) введено параметр дефектності *def*, який має той самий зміст, що і в (3).

Словозміна дієслів характеризується граматичними формами, що визначаються граматичними значеннями категорій «стан», «час», «число», «особа».

³Для російської мови субстантивний парадигматичний тип характеризується 12 граматичними формами: 6 відмінків у однині + 6 відмінків у множині. Для української мови субстантивний парадигматичний тип характеризується 14 граматичними формами внаслідок того, що відмінків в українській граматиці 7, а не 6, як в російській.

⁴Грамматична форма залежить від значень категорій «число» та «відмінок».

⁵Ад'єктивний парадигматичний тип російської мови характеризується 28 граматичними значеннями: 3 значення роду × 6 відмінків однини + 6 відмінків множини + 4 коротких форми (ч. р., ж. р., с. р. та множ.). В українській мові ад'єктивний парадигматичний тип характеризується 24 граматичними значеннями, які визначають повні словозмінні форми (коротких — немає).

«спосіб», «рід» (категорія роду релевантна тільки для минулого часу).

Зазначені категорії можуть набувати таких значень:

стан = {активний, пасивний} ($z = \{z_1, z_2\}$);
 час = {теперішній, минулий, майбутній} ($t = \{t_1, t_2, t_3\}$);

число = {однина, множина} ($n = \{n_1, n_2\}$);

особа = {перша, друга, третя} ($l = \{l_1, l_2, l_3\}$);

спосіб = {дійсний, умовний, наказовий} ($h = \{h_1, h_2, h_3\}$);

рід = {чоловічий, жіночий, середній} ($g = \{g_1, g_2, g_3\}$).

Отже, дієслівний парадигматичний тип описується формулою:

$$W(T_3) \equiv W^V = \{w_0^V, w_1^V, w_2^V, \dots, w_{45}^V, def\}, \quad (8)$$

$$\Omega(T_3) = \{\omega_0^V, \omega_1^V, \omega_2^V, \dots, \omega_{45}^V\},$$

де w_0 — інфінітив дієслова (збігається з реєстровим словом); $w_1^V, w_2^V, \dots, w_6^V$ — представляють граматичні форми активного стану дійсного способу теперішнього часу; граматичні форми w_i^V ($i = 1, 2, \dots, 6$) визначаються п'ятіркою категорій (стан, спосіб, час, число, особа)⁶ при фіксованих значеннях стану ($z = z_1$ — активний стан), способу ($h = h_1$ — дійсний спосіб) і часу ($t = t_1$ — теперішній час):

$$\omega_i^V = \{z_1, h_1, t_1, n_1, l_1\}, \quad \omega_{i+3}^V = \{z_1, h_1, t_1, n_2, l_1\}, \quad (9)$$

$$i = 1, 2, 3;$$

$w_7^V, w_8^V, \dots, w_{10}^V$ — граматичні форми активного стану дійсного способу минулого часу, які визначаються п'ятіркою категорій (стан, спосіб, час, число, рід)⁷ при фіксованих значеннях стану ($z = z_1$), способу ($h = h_1$) і часу ($t = t_2$ — минулий час):

$$\omega_{i+6}^V = \{z_1, h_1, t_2, n_1, g_1\}, \quad i = 1, 2, 3; \quad (10)$$

$$\omega_{i+9}^V = \{z_1, h_1, t_2, n_2\}; \quad (11)$$

$w_{11}^V, w_{12}^V, \dots, w_{16}^V$ — граматичні форми активного стану дійсного способу майбутнього часу, що визначаються п'ятірками категорій (стан, спосіб, час, число, особа) при фіксованих значеннях стану ($z = z_1$), способу ($h = h_1$) і часу ($t = t_3$ — майбутній час):

$$\omega_{i+10}^V = \{z_1, h_1, t_3, n_1, l_1\}, \quad (12)$$

$$\omega_{i+12}^V = \{z_1, h_1, t_3, n_2, l_1\}, \quad i = 1, 2, 3;$$

w_{17}^V, w_{18}^V — граматичні форми наказового способу, які визначаються четвірками категорій (стан, спосіб, число, особа) при фіксованих значеннях стану ($z = z_1$), способу ($h = h_3$ — наказовий спосіб) та особи ($l = l_2$ — друга особа):

⁶ Категорія роду не є релевантною для форм теперішнього і майбутнього часу.

⁷ Для форм минулого часу категорія особи не є релевантною.

$$\omega_{i+16}^V = \{z_1, h_3, t_2, n_i, l_2\}, \quad i = 1, 2; \quad (13)$$

w_{19}^V, w_{20}^V — дієприслівникові граматичні форми, які визначаються значеннями пар категорій (стан, час) при фіксованих значеннях стану ($z = z_1$) (дієприслівники активного стану теперішнього та минулого часу):

$$\omega_{i+18}^V = \{z_1, t_i\}, \quad i = 1, 2; \quad (14)$$

$w_{21}^V, w_{22}^V, w_{23}^V, w_{24}^V$ — дієприкметникові граматичні форми дієслова, які визначаються категоріями (стан, час) (дієприкметники активного та пасивного стану теперішнього та минулого часу):

$$\omega_{i+20}^V = \{z_1, t_i\}, \quad \omega_{i+22}^V = \{z_2, t_i\}, \quad i = 1, 2; \quad (15)$$

w_{25}^V — інфінітив пасивної форми дієслова;

$w_{26}^V, w_{27}^V, \dots, w_{31}^V$ — граматичні форми пасивного стану теперішнього часу; значення w_i^V ($i = 26, 27, \dots, 31$) визначаються п'ятірками категорій (стан, спосіб, час, число, особа) при фіксованих значеннях стану ($z = z_2$ — пасивний стан), способу ($h = h_1$ — дійсний спосіб) та часу ($t = t_1$ — теперішній час):

$$\omega_{i+25}^V = \{z_2, h_1, t_1, n_1, l_1\}, \quad (16)$$

$$\omega_{i+28}^V = \{z_2, h_1, t_1, n_2, l_1\}, \quad i = 1, 2, 3;$$

$w_{32}^V, w_{33}^V, \dots, w_{35}^V$ — граматичні форми, що визначаються п'ятірками категорій (стан, спосіб, час, число, рід) при фіксованих значеннях стану ($z = z_2$), способу ($h = h_1$) і часу ($t = t_2$ — минулий час):

$$\omega_{i+31}^V = \{z_2, h_1, t_2, n_1, g_i\}, \quad i = 1, 2, 3; \quad (17)$$

$$\omega_{i+35}^V = \{z_2, h_1, t_2, n_2\}; \quad (18)$$

$w_{36}^V, w_{37}^V, \dots, w_{41}^V$ — граматичні форми пасивного стану майбутнього часу; значення w_i^V ($i = 36, 37, \dots, 41$) визначаються п'ятірками категорій (стан, спосіб, час, число, особа) при фіксованих значеннях стану ($z = z_2$ — пасивний стан), способу ($h = h_1$ — дійсний спосіб) і часу ($t = t_3$ — майбутній час):

$$\omega_{i+35}^V = \{z_2, h_1, t_3, n_1, l_1\}, \quad (19)$$

$$\omega_{i+38}^V = \{z_2, h_1, t_3, n_2, l_1\}, \quad i = 1, 2, 3;$$

w_{42}^V, w_{43}^V — граматичні форми пасивного стану наказового способу, що визначаються четвірками категорій (стан, спосіб, число, особа) при фіксованих значеннях стану ($z = z_2$), способу ($h = h_3$ — наказовий спосіб) та особи ($l = l_2$ — 2-га особа):

$$\omega_{i+41}^V = \{z_2, h_3, n_i, l_2\}, \quad i = 1, 2; \quad (20)$$

w_{44}^V, w_{45}^V — відповідно дієприслівникова та дієприкметникова граматичні форми пасивного стану дійсного способу минулого часу (визначаються категоріями стан, спосіб та час: $z = z_2, h = h_1, t = t_2$):

$$\omega_{44}^V = \{z_2, h_1, t_2\}, \quad \omega_{45}^V = \{z_2, h_1, t_2\}; \quad (21)$$

def — параметр дефектності.

Парадигматичні типи російської мови

Формули (9)–(21) описують усі основні можливі граматичні форми синтетичних дієслівних форм, які можуть бути властиві парадигмі дієслова (і які зазвичай залучаються розробниками до словозмінної парадигми дієслова). Не включено аналітичні форми, зокрема форми умовного способу, а також зворотні форми дієслова. Зворотне дієслово розглядається нами як самостійне і змінне відповідно до парадигматичного типу, що описується формулами (8)–(21). Для парадигматичного типу дієслів можливі випадки дефектності декількох видів. Усі вони описуються параметром *def*, який вказує номери граматичних форм, для яких відсутні варіанти словоформ; *def* = 0, якщо дефектність відсутня. Наведемо деякі особливі види дефектності дієслівної парадигми в російській мові:

| Парадигматичний тип | Граматичні класи | Словозмінні граматичні категорії | Кількість граматичних значень у повній парадигмі |
|--|---|--------------------------------------|--|
| Субстантивний | Іменники, займенники-іменники | число, відмінок | 12 |
| Ад'єктивний | Прикметники, займенники-прикметники, порядкові числівники, дієприкметники | рід, число, відмінок | 28 |
| Дієслівний | Дієслова доконаного виду, дієслова недоконаного виду, двовидові дієслова | стан, час, число, особа, спосіб, рід | 46 |
| Парадигматичний тип кількісних числівників | Кількісні числівники | відмінок | 6 |
| «Нульовий» парадигматичний тип — незмінювані слова | Прислівники, вигук, сполучники, частки, прийменники, предикативи | — | 1 |

відсутність синтетичних форм майбутнього часу у дієслів недоконаного виду, за винятком дієслова *быть*; відсутність форм теперішнього часу, а також дієприкметників активного та пасивного станів теперішнього часу у дієслів доконаного виду; відсутність багатьох форм безособових дієслів тощо.

Парадигматичний тип, що характеризується шістьма граматичними формами, які визначаються категорією відмінка, притаманний кількісним числівникам (крім числівника *один*). Такий парадигматичний тип називатимемо *парадигматичним типом числівників*:

$$W(T_4) \equiv W^C = \{w_1^C, w_2^C, \dots, w_6^C\}, \quad (21)$$

$$\Omega(T_4) = \{\omega_1^C, \omega_2^C, \dots, \omega_6^C\}.$$

Усі незмінні слова російської мови можуть бути віднесені до одного парадигматичного типу — вони мають єдину форму подання у мові, а саме ту, яку подано в реєстровій частині словника. До незмінних слів належать прислівники, сполучники, прийменники, вигук, частки, предикативні слова.

Таблиця 1 ілюструє парадигматичні типи російської мови і відношення (відповідність) між парадигматичними типами, граматичними класами та словозмінними категоріями.

3. Словозмінні парадигматичні класи та відношення парадигматизації

Усередині граматичних класів виділяємо парадигматичні класи.

Дамо формальне визначення парадигматичного класу. Довільна лексема *x* (з урахуванням її словозмінних варіантів) може бути подана у вигляді комбінації незмінної та змінної складових:

$$x = c(x) * f(x), \quad (23)$$

де *c(x)* — частина лексеми *x*, яка в процесі словозміни залишається незмінною (квазіоснова), *f(x)* — її змінна складова (квазіфлексія), * — конкатенація.

Змінна та незмінна складові можуть мати як нульову довжину, так і являти собою всю лексему. Наприклад, у парадигмах іменників із суплетивними формами множини (*человек, человека, ... , люди, людей, ...*) незмінна частина дорівнює нулю, а змінна частина представлена всіма словоформами. У парадигмах незмінних слів, навпаки, нулю дорівнює змінна частина.

Повна словозмінна парадигма [*x*] слова *x*, що належить до граматичного класу (парадигматичного типу T_i), має вигляд:

$$\pi(x) = c(x) * \{f_i(x)\}, \quad (24)$$

де $f_i(x)$, $i = 0, 1, 2, \dots, n(T_i)$ — змінні частини слова (квазіфлексії) у відповідних граматичних формах; причому в деяких із них може існувати більше однієї словоформи. Для означення даного факту введемо параметр кратності граматичної форми $v(w_i(x))$, який задається цілим числом, рівним кількості можливих форм лексеми *x* у граматичній формі w_i . У загальному випадку:

$$f_i(x) = \bigcup_{l=0}^{v(w_i(x))} f_{il}, \quad (25)$$

$l = l(i) = 0, 1, 2, \dots$ — індекс кількості словоформ у граматичній формі (залежить від номера граматичної форми *i*); $f_0(x)$ — квазіфлексія початкової форми, яка для іменника конкретного роду відповідає словоформі називного відмінка однини, для

дієслова — його інфінітиву, для прикметника — словоформі чоловічого роду називного відмінка однини тощо; $n(T_i)$ — кількість граматичних форм у парадигматичному типі T_i .

Покладемо

$$F = \bigcup_{x \in W} (\{f_0(x)\}, \{f_{1l}(x)\}, \dots, \{f_{n(T_i)l}(x)\}) \equiv \{f_{jl}^1, f_{jl}^2, \dots, f_{jl}^{N_i}\}, \quad (26)$$

$$j = 0, 1, 2, \dots, n(T_i), l = l(w_j) = 0, 1, 2, \dots$$

Тоді

$$F = \bigcup_{k=1}^{N_i} [F]^k, \quad (27)$$

де $[F]^k = \{f^k\} = \{f_{jl}^k, j = 0, 1, \dots, n(T_i)\}$, $N_i = N(T_i)$, $l = l(w_j)$.

Таким чином, кожна множина $[F]^k$ складається з квазіфлексій слів, які мають у всіх своїх граматичних формах $w_1, w_2, \dots, w_{n(T_i)}$ (парадигматичного типу T_i) однакові змінні складові.

Оскільки $[F]^k$ побудовані таким чином, що до них увійшли унікальні набори квазіфлексій, тобто $[F]^i \neq [F]^j$ при $i \neq j$ ($i, j = 1, 2, \dots, N_i$), то для кожного граматичного класу P_i (парадигматичного типу T_i) можна побудувати відношення π_i на декартовому добутку $P_i \times P_i$, яке визначається так:

$$\forall x^1, x^2 \in P_i, x^1 \pi_i x^2 : x^1 = c(x^1) * f^k, \\ x^2 = c(x^2) * f^k, f^k \in [F]^k. \quad (28)$$

Це відношення є відношенням еквівалентності, оскільки воно, очевидно, є рефлексивним, симетричним та транзитивним. Назвемо його *відношенням парадигматизації*.

Фактор-множина P_i / π_i є множиною парадигматичних класів граматичного класу P_i (парадигматичного типу T_i). Очевидно, що різні словозмінні парадигматичні класи не перетинаються. Отже P_i є об'єднанням парадигматичних класів: $P = \bigcup_{j=1}^n \Pi_j$. До

одного парадигматичного класу входять тільки ті слова, які мають однакові набори квазіфлексій для всіх граматичних форм, а відрізняються один від одного лише незмінною складовою $c(x)$. Зрозуміло також, що слова з одного класу еквівалентності, визначеного в такий спосіб, мають і однакові правила словозміни.

Таким чином, для кожного з граматичних класів (парадигматичних типів T_i) будується розбиття на множини слів, що не перетинаються, і які є парадигматичними класами, всередині кожного з яких діють єдині правила словозміни. (Для мов флективного типу це означає однаковість флексій граматичних форм та збіг характеру чергування в основі.)

4. Оператор парадигматизації

Для автоматичної побудови повної парадигми за вихідною (початковою) формою x_0 визначається оператор парадигматизації

$$H: x \rightarrow [x] = c(x) * \{f_0(x), f_1(x), \dots, f_n(x)\} \equiv \{c(x) * f_0(x), c(x) * f_1(x), \dots, c(x) * f_n(x)\}, \quad (29)$$

для якого визначається відношенням $\pi(x^1, x^2)$.

Оператор повної парадигматизації (який діє на множині лексем W) визначається за формулою:

$$H = \sum_{i=1}^4 H_i \cdot \delta(x; T_i), \quad (30)$$

де

$$\delta(x; T_i) = \begin{cases} 1, & x \in W_{T_i}; \\ 0, & x \notin W_{T_i}. \end{cases} \quad (31)$$

H_i — оператор парадигматизації, який діє на множині лексем відповідного парадигматичного типу $W(T_i)$. На множині лексем кожного з парадигматичних типів діє свій оператор парадигматизації, оскільки кожен із парадигматичних типів характеризується своїм комплексом значень граматичних категорій:

$$H_1: x_0 \rightarrow [x] \forall x \in W(T^S), \quad (32)$$

$$H_2: x_0 \rightarrow [x] \forall x \in W(T^A), \quad (33)$$

$$H_3: x_0 \rightarrow [x] \forall x \in W(T^V), \quad (34)$$

$$H_4: x_0 \rightarrow [x] \forall x \in W(T^C), \quad (35)$$

$$\forall x \in W_{\Pi_k} \subset W(T^S) H_1^k: x_0 \rightarrow c(x) * [F]_1^k, \\ H_1 = \sum_{k=1}^{Cnt(T^S)} H_1^k \cdot \delta(x, W_{\Pi_k}) \quad (36)$$

де H_1 — оператор парадигматизації, який діє на $W(T^S)$; $Cnt(T^S)$ — кількість парадигматичних класів у множині $W(T^S)$; $[F]_1^k$ — множина наборів квазіфлексій слів, які належать до парадигматичного типу $W(T^S)$ (кількість елементів цієї множини дорівнює кількості парадигматичних класів $Cnt(T^S)$). Функція

$$\delta(x; W_{\Pi_k}) = \begin{cases} 1, & x \in W_{\Pi_k}; \\ 0, & x \notin W_{\Pi_k}. \end{cases} \quad (37)$$

Аналогічно для інших парадигматичних типів:

$$\forall x \in W_{\Pi_k} \subset W(T^A) H_2^k: x_0 \rightarrow c(x) * [F]_2^k, \\ H_2 = \sum_{k=1}^{Cnt(T^A)} H_2^k \cdot \delta(x, W_{\Pi_k}). \quad (38)$$

$$\forall x \in W_{\Pi_k} \subset W(T^V) H_3^k: x_0 \rightarrow c(x) * [F]_3^k, \\ H_3 = \sum_{k=1}^{Cnt(T^V)} H_3^k \cdot \delta(x, W_{\Pi_k}). \quad (39)$$

$$\forall x \in W_{\Pi_k} \subset W(T^C) H_4^k: x_0 \rightarrow c(x) * [F]_4^k, \\ H_4 = \sum_{k=1}^{Cnt(T^C)} H_4^k \cdot \delta(x, W_{\Pi_k}). \quad (40)$$

де H_2, H_3, H_4 — оператори парадигматизації, які діють відповідно на $W(T^A)$, $W(T^V)$ та $W(T^C)$; $Cnt(T^A)$ — кількість парадигматичних класів ад'єктивного парадигматичного типу (в $W(T^A)$), $Cnt(T^V)$ — кількість парадигматичних класів у $W(T^V)$, а $Cnt(T^C)$ — кількість парадигматичних класів у $W(T^C)$; $[F]_2^k, [F]_3^k, [F]_4^k$ — множини наборів квазіфлексій слів, які належать до відповідного парадигматичного типу (T^A, T^V або T^C).

Таким чином, для кожного із класів $W(T^N)$, $W(T^A)$, $W(T^V)$ та $W(T^C)$ оператор парадигматизації визначається незалежно. На множині $W(T^0)$ немає необхідності визначати цей оператор через те, що для незмінюваних слів $[x] \equiv x_0$.

Оператор H відображає лексему x на її повну парадигму $[x]$, і його реалізовано за допомогою словника квазіфлексій і набору алгоритмів побудови повних словозмінних парадигм для російської лексики. За допомогою парадигматичного словника довільній лексемі приписується її словозмінний тип. Далі з використанням набору алгоритмів побудови повних словозмінних парадигм здійснюється граматична ідентифікація лексеми x . Після цього лексема набуває представлення (23).

Алгоритмічна реалізація оператора H^{-1} здійснює процес лематизації, тобто зведення довільної словоформи до її вихідної канонічної форми.

Викладена вище морфологічна модель складає концептуальну основу для комп'ютерного моделювання та реалізації функції парадигматичних відношень.

5. Висновки

У роботі подано модель словозмінної системи російської мови, запропоновано формальне визначення поняття парадигматичного типу і парадигматичного класу, розроблено словозмінну класифікацію російської лексики, яка (класифікація) придатна та зручна для використання її в електронному словнику, побудовано і програмно реалізовано оператор парадигматизації, котрий однозначно ставить у відповідність кожному російському слову його парадигма-

тичний клас. Це дозволяє одержувати повні словозмінні парадигми для всіх повнозначних змінюваних частин мови російської мови.

Наведені результати апробовано і верифіковано на масиві російської лексики обсягом близько 170 тисяч лексем. Одержано 1590 парадигматичних класів, серед них 526 класів іменників без урахування власних назв, 633 — з урахуванням власних назв, 792 класи дієслів, 97 класів ад'єктивів, 4 класи дієприкметників, 24 класи кількісних числівників та 39 парадигматичних класів займенників.

Створено електронний граматичний словник російської мови [4] (який є аналогом паперового «Грамматического словаря русского языка» А. А. Зализняка [2]).

Передбачено розробку морфологічних баз даних для інших мов, залучених до системи МП, яка розробляється в УМІФ НАНУ. Ці дослідження вже виконуються для англійської, німецької, іспанської мов. Принципи моделювання системи словозміни російської мови знаходять застосування й для мов іншої будови (які відрізняються від мов флективного типу). Звичайно, кожна мова має свої особливості, урахування котрих спонукає до відповідних змін у структурі даних, а також розробки нових алгоритмів і програм. Паралельно зі створенням ЛБД для згаданих мов буде виконуватися розробка алгоритмів та програмних модулів морфологічного (морфо-синтаксичного) аналізу текстів, написаних відповідними мовами.

Список літератури: 1. Лингвистический энциклопедический словарь / Под ред. В. Н. Ярцева. — М., 1990. — 685 с. 2. Зализняк А. А. Грамматический словарь русского языка: Словоизменение. — М.: Русский язык, 1978. — 878 с. 3. Широков В. А. Элементы лексикографии. — К.: Довіра, 2005. — 304 с. 4. Грязнухина Т. А., Любченко Т. П., Рабулец А. Г. Электронная версия грамматического словаря русского языка (А. А. Зализняк) как инструмент автоматического морфологического анализа русского текста // Докл. научн. конф. «Корпусная лингвистика и лингвистические базы данных», Санкт-Петербург, март 2002 г. — С. 63–70.

Поступила в редколлегию 21.05.2006

УДК 004.934

М. Ф. Бондаренко, А. В. Работягов, С. В. Щепковский

ПРОБЛЕМЫ РАСПОЗНАВАНИЯ РЕЧЕВЫХ ОБРАЗОВ СПЕКТРАЛЬНЫМИ МЕТОДАМИ В ОБЛАСТИ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

1. Введение

Как известно, современное направление в области речевых технологий к решению основных задач — *кто говорит* (идентификация человека), *что говорит* (например, распознавание речевых команд искусственной технической системой, кодирование / декодирование речевого сообщения) и *как говорит* (установление эмоционального состояния человека) — характеризуется в целом как *спектральное*. Однако это направление имеет некоторые существенные проблемы.

2. Положения

акустической теории речеобразования

Фундаментальные положения акустической спектральной теории речи были сформулированы в XIX ст. выдающимся немецким физиком Г. Гельмгольцем и до сегодняшнего дня остаются неизменными. В работах известных ученых-лингвистов, начиная от И. А. Бодуэна де Куртенэ, А. И. Томсона, Л. В. Щербы (XIX — начало XX вв.) и заканчивая Г. Фантом, Дж. Фланаганом, Л. Р. Зиндером, Л. А. Чистович, Л. В. Бондарко, М. А. Сапожковым, Р. К. Потаповой и др. (вторая половина XX — начало XXI вв.), мы видим, что для объяснения речевых процессов и количественного описания речевых сигналов используется их спектральное представление. На сегодняшний день главное положение акустической теории речеобразования¹ формулируется следующим образом. «Речевой сигнал возникает в результате воздействия одного или нескольких источников звука на систему резонаторов, образуемых воздушными полостями речевого тракта. Если обозначить через $S(f)$ амплитудно-частотный спектр колебаний, создаваемых источником звука, через $T(f)$ — передаточную функцию резонаторной системы речевого тракта, то амплитудно-частотный спектр результирующих звуковых колебаний $P(f)$ может быть представлен равенством:

$$P(f) = S(f) \cdot T(f).$$

Как показывает приведенная формула, частотная фильтрация состоит в том, что амплитуда каждой из частотных составляющих источника звука умножается на значение передаточной функции

тракта на той же частоте. Свойства источников звука и резонаторной системы тракта не являются неизменными. Произнесение отдельных звуков и, тем более, звуковых последовательностей представляет собой сложный динамический процесс, характеристики которого меняются во времени. Поэтому в приведенную выше формулу включается параметр времени:

$$P(f, t) = S(f, t) \cdot T(f, t).$$

...В результате получается сложный периодический сигнал со спектром гармонической структуры, задаваемой голосовым источником» [1, с. 102–103].

Как известно, цель спектрального анализа состоит в «разложении» сложного суммарного колебания на отдельные составляющие элементы и измерении свойств этих элементов, представленных в виде количественных *спектральных* характеристик. Спектральные речевые характеристики получаются в результате разложения «речевой» функции в ряд Фурье. В первую очередь к спектральным характеристикам относятся мгновенный и текущие спектры, частоты основного тона, первой, второй и третьей формант, динамика частоты основного тона, которые определяются на участках языковых и речевых единиц: звуке, слого, слове, словосочетании, предложении [2]. Используемые для целей обработки речи экспериментальные методы и математический аппарат (дискретное преобразование Фурье, теории вероятностей, статистики и др.) достаточно глубоко продуманы и детально изложены в научно-технической литературе.

В работах М. А. Сапожкова, Н. Г. Загоруйко и Р. К. Потаповой, Т. К. Винцюка [3–6] изложены общие подходы и методы анализа и описания речевых сигналов, имеющих в своей основе частотную структуру речевого сигнала, т. е. спектр².

Данные методы позволяют аппроксимировать речевой сигнал с различной степенью точности в зависимости от представления спектра. Они применяются

¹«Задача акустической теории речеобразования состоит в том, чтобы выявить и количественно описать аэродинамические и акустические процессы, которые происходят в речевом тракте при артикуляции. Понимание этих процессов создает возможность обратных заключений: от акустики к артикуляционной картине» [1, с. 100].

²Спектр (лат. spectrum — «представление», «образ»). При спектральном представлении акустический сигнал представляется в виде «наложения» большого числа гармоник. Разложение сигнала в спектр обычно проводится с помощью быстрого преобразования Фурье (БПФ), реализованного в большинстве компьютерных звуковых редакторов и специальных программ обработки речи. Исследователи речи чаще всего используют представление речевого сигнала в виде трех- или двумерных сонограмм (спектрограмм). В первом случае по осям координат откладываются время, частота и спектральная плотность (энергия) частотной составляющей, а на двумерной сонограмме ось энергии заменяется интенсивностью цвета в плоскости время–частота.

в различных модификациях в большинстве работ по анализу и распознаванию речи как один из основных уровней переработки сигнала для сегментации, формирования дифференциальных признаков (например, признаков фонем), определения места и способа образования звуков, переходных и временных характеристик. К этим методам относятся:

1) ортогональные методы, которые позволяют описывать мгновенный спектр речевого сигнала последовательностью коэффициентов Фурье;

2) корреляционные методы, использующие автокорреляционную функцию;

3) спектрально-полосные методы, которые проводят разложение сигнала в определенных полосах частот (по применяемому математическому аппарату относят к ортогональным методам);

4) формантные методы, где в качестве характеристик речевого сигнала рассматриваются параметры формант³ и антиформант⁴ — частота, амплитуда, ширина полосы, скорость изменения и др.;

5) метод ро-параметра, основанный на измерении плотности нулевых переходов речевого сигнала в соответствующей формантной области;

6) дискриминантный метод (является вариантом метода ро-параметра);

7) метод моментов спектра (по существу близок к методу нулевых переходов);

8) метод анализа через синтез (формантные параметры попарно сравниваются в компараторах с параметрами, выделяемыми основным анализатором);

9) метод выделения формантных частот путем последовательного исключения формант из речевого сигнала;

10) метод инверсных фильтров;

11) полюсно-нулевой метод;

12) кепстральный⁵ метод (для определения периода основного тона);

13) статистические методы (варианты: метод линейного предсказания, автокорреляционный и ковариационный. Например, анализ речевого сигнала на основе линейного предсказания заключается в разложении его спектра на две составляющие: сглаженного спектра, представленного в виде модели спектра, характеристика которого содержит только полюсы, и спектра функции возбуждения, содержащего информацию о погрешности предсказания;

14) векторный метод;

15) метод кратковременных и быстрых преобразований Фурье;

16) метод преобразования Адамара (Уолша);

17) метод кодирования с предсказанием;

18) методы, базирующиеся на статистических характеристиках ритмики и темпа речи (количество фонетических слов, звуков в секунду, распределение длительности звуковых сегментов, речевых пауз);

19) отдельную группу образуют методы, в которых характеристику голосового источника применяют как компоненту свертки;

20) иерархические композиционные модели, основанные на теории оптимальных решений — динамическом программировании.

Например, на базе перечисленных методов в конкретных устройствах реализованы: бинарные (дихотомические) системы в двоичной метрике; системы на основе клипированной речи и энергетического подхода, где определяются логарифмы энергии в различных полосах частот (т. е. спектр мощности: средний и мгновенный спектр, кросскорреляции спектральных компонентов); системы с пространственным представлением речи, в которых описание параметров происходит на базе нечетких множеств; динамические портреты речевого сигнала; параметры модели речеобразующего тракта; системы функциональной обработки, использующие параметры основного тона голоса — среднее значение, дисперсию, статистические моменты, изрезанность мелодического контура, микро- и макровариация; параметры модели линейного предсказания; гомоморфная обработка речевых сигналов.

В частности, в криминалистических подразделениях правоохранительных органов применяемые на практике компьютерные системы при проведении судебно-акустических (фоноскопических) экспертиз для идентификации человека по признакам речи в качестве основных детерминирующих признаков используют спектральные характеристики, которые получают, в основном, при помощи ортогональных, формантных и кепстральных методов.

3. Проблемы распознавания речевых образов

Следует указать на то важное обстоятельство, что эффективность экспериментально-фонетического исследования при спектральном отображении речевых сигналов зависит от соблюдения определенных требований, предъявляемых к речевому материалу и передаточным каналам записи. Эти требования относятся, главным образом, к сопоставимости (равнозначности):

1) объемов речевого материала;

2) фрагментов (образцов) речи для непосредственного исследования;

3) каналов записи (например, сопоставимость телефонных каналов или технических звукозаписывающих устройств);

4) эмоционального состояния говорящего;

5) манеры произношения говорящего.

³Форманта — частота максимума спектральной энергии, область концентрации энергии в спектре звука речи [3, с. 561].

⁴Антиформанта — частота минимального значения спектральной энергии в огибающей спектра звука речи [3, с. 561].

⁵Кепстр(ум) — «перевернутый» спектр, обратное дискретное преобразование Фурье (преобразование Фурье от логарифма модуля спектра) [3, с. 555].

Но каким образом проводить исследование, если такие требования не соблюдены? Как проводить сравнение образцов речи, если, например, на фоноскопическую экспертизу поступил речевой материал, представленный лишь одной фразой длительностью 3 секунды? На практике же при проведении фоноскопических экспертиз современными спектральными методами должны учитываться, в частности, следующие ограничения:

1) ограничение по длительности — минимальная длительность «чистой речи», как правило, должна составлять не менее 10 мин;

2) ограничение по спектру — речевой сигнал теряет часть полезной информации вследствие ограниченной частотной пропускной способности совокупного передаточного канала записи, предел которой в области верхних частот составляет 3...6 кГц для большинства диктофонов среднего класса (высококачественная звукозаписывающая аппаратура не учитывается, поскольку запись большинства фонограмм производится именно на диктофонах среднего класса) и 3 кГц — для телефонных линий. Это, естественно, отражается на спектре речевого сигнала и ограничивает «идентификационные» возможности спектральных методов.

Разрешить указанное противоречие возможно лишь при помощи иных, не использующих спектральную идеологию методов исследования речи. Это во-первых.

Во-вторых, спектральное отображение является всего лишь одним из возможных вариантов описаний речевых сигналов. Как отмечают специалисты в области науковедения, правила и методы эмпирической интерпретации теории могут развиваться бесконечно, так как любая теория не имеет предела в своем развитии. С позиции гносеологии, формы описания объекта познания бесконечны, поскольку сам объективный мир изменяется и развивается. Исходя из этого обстоятельства, правомерно сделать вывод: как существующая акустическая теория речеобразования, так и разработанные на ее основе спектральные методы исследования речи не могут претендовать на исчерпывающее и тем более абсолютное описание.

В-третьих, чрезмерно широкое применение математики, статистики (например, статистического анализа временных рядов) так или иначе «служит уточненным способом маскировки недостатка знаний о предмете исследования». Профессор У. Сиберг утверждает, что «применение статистики оправдано лишь после того, как решена основная задача классификации или идентификации параметров» [7, с. 133].

Действительно, никто не будет отрицать тот факт, что реальный речевой сигнал на выходе резонаторной системы речевого тракта — это сложный акустический сигнал. Данное обстоятельство логически оправдывает применение надежного, хорошо прове-

ренного математического аппарата для дальнейшего исследования речевых сигналов. В качестве такого аппарата применяется преобразование Фурье для анализа временных рядов⁶ (речевой сигнал как раз и представляется в таком виде). С его помощью находят значения частот и амплитуд гармонических составляющих сигнала. И решение, как кажется на первый взгляд, найдено. Однако, как отмечает известный специалист в области распознавания речи Дж. Р. Доддингтон, на этом пути возникают «определенные трудности в использовании столь привлекательных спектрографических различий (в частности, для распознавания человека по параметрам речи). ... Действительная трудность заключается в разном звучании и спектрографическом представлении одного и того же голоса в разное время» [8].

В отношении указанных выше спектральных методов следует отметить, что все они используют так называемые глобальные свойства объекта, к которым, несомненно, относится спектр. Как отмечается в книге «Распознавание образов» [7], глобальные операторы (признаки) оказались фактически бесполезными для решения некоторых задач распознавания образов. «Глобальные операторы «рассматривают» не деревья, а лес в целом; они, возможно, способны отличить лиственный лес от хвойного, однако не смогут обнаружить яблоню, затерявшуюся в березовой роще» [7, с. 269]. В противовес глобальным признакам в книге отдается предпочтение локальным признакам как наиболее эффективным для решения прикладных задач.

К тому же, исследование нестационарных сигналов при помощи преобразования Фурье обладает рядом существенных недостатков. Крупный специалист в области вычислительной математики (теории сплайнов, теории приближения функций, теории вэйвлетов⁷) и теории обработки сигналов, профессор Чарльз К. Чуи (США) отмечает: «Формула

$$f(\omega) = \int_{-\infty}^{\infty} e^{-it\omega} f(t) dt$$

преобразования Фурье в таком виде неудобна для практических задач» [9, с. 28]. «Неудобство» формулы состоит в следующем.

⁶ Физически преобразование Фурье $S(f)$ представляет собой распределение интенсивности сигнала по частоте, т.е. является функцией плотности.

⁷ Теория вэйвлетов (англ. wavelet — «малые волны»), или как их иногда называют в русской литературе *всплесков*, имеет истоки в таких классических областях математики, как теория функций вещественного переменного, теория ортогональных рядов, преобразование Фурье и др. интегральные преобразования, теория функций комплексного переменного, функциональный анализ. Теория вэйвлетов неразрывно связана с развитием прикладных областей современной науки: цифровой обработки сигналов и изображений, теории фильтрации и кодирования, теории сплайнов, дискретных и быстрых преобразований. Наиболее бурное развитие теории вэйвлетов приходится на 80–90-е годы XX века [9].

1) Чтобы извлечь спектральную информацию $f(\omega)$ об аналоговом сигнале $f(t)$ по этой формуле, «следует использовать бесконечные интегралы времени: иметь информацию о прошлом и будущем сигнала, чтобы вычислить спектр для одной частоты ω ».

2) «Формула не отражает эволюцию частот со временем. Что действительно необходимо — это определить интервалы времени, которые дают спектральную информацию о любой нужной частотной области (или диапазоне частот). Кроме того, так как частота сигнала обратно пропорциональна длительности его периода, то в случае высокочастотной спектральной информации временной интервал может быть взят относительно малым для обеспечения нужной точности, а в случае низкочастотной спектральной информации такой временной интервал должен быть взят относительно большим. Другими словами, важно иметь гибкое частотно-временное окно, которое автоматически сжимается в окрестностях высоких частотных центров и расширяется у низких частотных центров» [9, с. 28].

3) «Для изучения спектрального поведения аналогового сигнала необходимо полное знание сигнала во временной области, включая и будущую информацию. Вдобавок, если сигнал меняется в малой окрестности некоторого момента времени, то это влияет на весь спектр. Действительно, в крайнем случае преобразование Фурье от дельта-распределения $\delta(t - t_0)$ с носителем в единственной точке t_0 есть $e^{-it_0\omega}$, которое покрывает всю частотную область. Следовательно, во многих приложениях, таких как анализ нестационарных сигналов и обработка сигнала в реальном времени, применение одной формулы преобразования Фурье весьма неадекватно» [9, с. 91].

В этом аспекте также показателен вывод специалиста в области распознавания речи В. Н. Сорокина (Институт проблем передачи информации РАН) в отношении существующих методов, которые применяются непосредственно на этапах обработки первичных данных. В. Н. Сорокин отмечает, что системы распознавания речи, основанные на использовании метода скрытых марковских моделей достигли предела своих возможностей и, как следствие, все еще (!) не удовлетворяют большинству практических применений [10].

В-четвертых, оправдание и «весомые» аргументы в пользу применения спектральных характеристик звукового сигнала многие ученые пытаются найти в *спектральном* объяснении некоторых нейрофизиологических процессов, происходящих в так называемом *слуховом пути*. Выделим, в частности, следующее объяснение процессов передачи полезной информации на нейрональном уровне. «Центральный отдел слуховой системы имеет сложное нейроанатомическое строение. Образующие его нейроны

организованы в несколько уровней, постепенно приближающихся к слуховой коре мозга, которая является конечной точкой так называемого классического *слухового пути*. Связь между уровнями в основном последовательная: нейроны данного уровня, получая информацию от предыдущего, передают ее после определенных преобразований следующему уровню. На всем пути сохраняются узкая настройка нервных клеток на определенные частоты, свойственная нейронам слухового нерва, и связь этих частот с пространственным расположением нейронов и их группировок в пределах своего уровня. Это значит, что каждая группа нейронов обрабатывает информацию в ограниченной частотной полосе, а анализ спектральной информации на всем пути ее прохождения к мозгу происходит в отдельных «сквозных» частотных каналах, в совокупности представляющих частотный диапазон человеческого слуха» [1, с. 226]. «Частотные» свойства периферической слуховой системы интересны как для специалистов, изучающих слух, так и для специалистов, исследующих восприятие речи, поскольку «в общем виде известно, что полезная информация, обеспечивающая восприятие речевого сигнала, заключена в его динамическом спектре» [11, с. 171]. Это подтверждается тем фактом, что реакция импульсной активности отдельных нейронов на воздействие какого-либо тонального сигнала действительно имеет место [12]. Однако является ли это действительно веским основанием для однозначного утверждения, что «нейроны обрабатывают именно спектральную информацию» и «анализируя спектр, можно объяснить (насколько это возможно) механизмы, например, восприятия речи? По мнению специалистов в области нейрофизиологии — нет. И этому есть серьезные обоснования. Механизм различения качества в слуховом органе действует «совсем по-иному» [12]. Общеизвестно, что различимых по высоте тонов больше тысячи. «Вдоль основной мембраны улитки лежат тысячи концевых органов, и на первый взгляд кажется, как и считал Гельмгольц, что каждый из них может представлять узкую полосу спектра слышимых звуков. Тем не менее оказалось, что элементы слухового нерва вовсе не обладают узкой настройкой, а что каждый рецептор обладает довольно широкой полосой реактивности, которая находит на полосу его соседа и занимает большой участок слышимых частот. ...Таким образом, надо отказаться от попыток объяснить всякое различие качества однородной теоретической схемой... В нервном коде один и тот же элемент не играет разных ролей в зависимости от всей конstellации. Вместо этого используются элементы разного рода, и каждый имеет свою ограниченную область в спектре воспринимаемых качеств раздражителя. ...Всякий стимул... может активировать много афферентных элементов

с разными рецептивными характеристиками. И полное описание значения стимула для организма осуществляется относительным распределением возбуждения... Такого рода организация получила название *частотный профиль популяции*, или *структура распределения по волокнам*» [12, с. 47].

К сказанному добавим следующее. Проведенное нами небольшое исследование показало, что слуховое восприятие двух речевых сигналов, имеющих одинаковый спектр, но разную пространственную ориентацию, совершенно различно. В исследовании диктор произносил слово *вокзал*. Речевой сигнал записывался в цифровом виде в компьютер. Этот сигнал назначался нами как исходный. Из исходного сигнала путем компьютерной «инверсии» нами формировался второй сигнал. В результате второй сигнал имел вид «перевернутого наоборот»⁸ первого исходного сигнала, т. е. конец первого исходного сигнала становился началом второго сигнала. Оба сигнала хранились в цифровом виде в памяти компьютера. При прослушивании этих двух сигналов их аудитивное восприятие было совершенно различным: исходный сигнал звучал как *вокзал*, а смысл второго речевого сигнала был полностью разрушен. Однако в процессе определения акустических характеристик двух сигналов выяснилось, что их спектры одинаковы. Одним из объяснений этому явлению может стать представление речевого сигнала как некоторого события [12], определяемого не наличием некоторой совокупности частотных составляющих, а последовательностью некоторых акустических элементарных микрособытий (микроколебаний), где на первое место выдвигаются пространственно-временные и структурные представления [13].

Аналогичных представлений придерживается А. В. Бару. В своей работе «Слуховые центры и опознание звуковых сигналов» он пишет: «Предполагается, что человек имеет специализированные механизмы обработки акустического речевого сигнала, отличающиеся от процессов, при помощи которых

⁸В исследованиях Н. П. Дукельского такой способ называется *инверсальным* или *обратным* прослушиванием. Инверсальное прослушивание, т. е. прослушивание в обратном порядке ранее записанных на пленке слов или слогов, следует отличать от *инверсии*, т. е. произнесения в обратном порядке слов или слогов. При инверсальном прослушивании оказывается нарушенной последовательность расположения отрезков речевого потока, в то время как инверсия предполагает автоматическую замену звуков и, соответственно, положений. В отличие от обычного, нормального восприятия речевого потока, при инверсальном прослушивании «большинство звуков теряет в большей или меньшей степени естественный характер звучания, а именно — свою ясность, однородность, слитность, громкость, нормальную длительность» [16, стр. 48]. При этом необходимо выделить психологические особенности восприятия. «По сравнению с нормальным прослушиванием при инверсальном прослушивании период времени, необходимый для восприятия слога, значительно возрастает, что свидетельствует о функциональном нарушении протекания нервного процесса» [16, стр. 52].

обрабатываются неречевые акустические сигналы. ... В слуховом анализаторе должны быть некоторые наборы детекторов формы и скорости измерения сигнала» [14].

В-пятых, одним из основных препятствий, стоящих на пути распознавания речевых образов, является несостоятельность существующих в настоящее время акустико-фонетических признаков. Например, в современных компьютерных системах автоматического распознавания и понимания слитной речи, ориентированных на дикторонезависимый интерфейс (например, в системе *Via Voice*), все еще наиболее уязвимым является отсутствие базы параметрических данных в области просодии⁹ того или иного языка [3, с. 528].

Как известно, универсальная акустическая классификация (известная также как теория различительных признаков [15]) является, по существу, попыткой описания звуковых контрастов на основе акустических характеристик. Ее в 1952 г. предложили, а затем дополняли и уточняли Р. Якобсон, Г. Фант и М. Халле. В этой классификации выделены основные акустические признаки, на основе которых формируются важнейшие звуковые противопоставления¹⁰. Признаковая база акустической классификации включает 12 бинарных различительных признаков, набор которых считается достаточным для описания смыслозначительных звуковых контрастов, возможных в конкретных языках. Каждый признак представляет собой противопоставление между двумя относительными проявлениями одного и того же акустического свойства, т. е. своего рода шкалу выраженности определенного акустического качества. Звуки, у которых степень проявления акустического свойства больше некоторого порога, имеют положительное значение соответствующего признака-классификатора, в противном случае — отрицательное. Отличительная особенность этой системы признаков состоит в том, что для описания согласных и гласных предлагается один и тот же набор признаков, в соответствии с которым звук речи классифицируется [1, с. 296–303; 15] как:

- 1) вокальный / невокальный (вокальные звуки, в отличие от невокальных, акустически характеризуются четко выраженной формантной структурой);
- 2) консонантный / неконсонантный (консонантные звуки, в отличие от неконсонантных, имеют более низкий общий уровень интенсивности);

⁹Просодия — «система фонетических средств (высотных, силовых, временных), реализующихся в речи на всех языковых уровнях. Вышеуказанные фонетические средства соответствуют основным физическим характеристикам: частоте основного тона, интенсивности и длительности» [3, с. 558].

¹⁰Например, спектрограммы, иллюстрирующие большинство звуковых контрастов, полно представлены в «Общей фонетике» С. В. Кодзасова и О. Ф. Кривновой [1].

3) прерванный / непрерывный (к прерванным относятся согласные, у которых акустически выражен интервал отсутствия или сильного ослабления звуковой энергии в полосе частот выше основного тона, после которого следует «взрыв» или резкое изменение формантной картины);

4) глоттализированный / неглоттализированный (глоттализированные согласные характеризуются резким включением интенсивного источника шума);

5) резкий / нерезкий (в основе признака лежат относительные различия в интенсивности, длительности и степени упорядоченности фрикативного шума, т. е. своего рода шкала шумности. Резкие согласные, в отличие от нерезких звуков, имеют интенсивный и длительный шум);

6) звонкий / глухой (звонкие звуки, в отличие от глухих, произносятся с участием голосового источника);

7) компактный / диффузный (компактные звуки, в отличие от диффузных, характеризуются большей концентрацией энергии в относительно узкой серединной (у диффузных — краевой) части частотного диапазона спектра, большей интенсивностью и длительностью);

8) низкий / высокий (к низким (низкотоновальным) относятся звуки, у которых энергия сосредоточена в более низких частотах, чем у высоких (высокотоновальных));

9) бемольный / простой (у бемольных звуков, в отличие от простых, частоты всех формант понижаются и энергия высокочастотной части спектра ослабляется);

10) дизный / простой (у дизных вокальных звуков увеличиваются частоты нижних формант и возрастает интенсивность по сравнению с простыми (недизными) звуками);

11) напряженный / ненапряженный (напряженные звуки, в отличие от ненапряженных, характеризуются большей длительностью, большей интенсивностью, более отчетливым и богатым спектром);

12) носовой / ртовый (у носовых звуков, образующихся с участием носового резонатора, в отличие от ртовых, в спектре появляются устойчивые, мало изменяющиеся форманты назализации (200...300 Гц)).

В-шестых, на сегодняшний день среди специалистов в области лингвистики нет единого, однозначного мнения по поводу того, какую же структурную единицу (участок, элемент) речевого сигнала на акустико-фонетическом (физическом) уровне принимать за минимально-информационный *сегмент* для надежного распознавания.

Изучение любого речевого сигнала «предполагает предварительную сегментацию этого сообщения с выделением ключевых сегментов и их признаков, отношения между ними, их структурной организации с целью дальнейшего распознавания и понима-

ния последнего. При этом встает вопрос определения самих единиц сегментации речевого высказывания, методики, критериев их обнаружения, специфики и формы их взаимодействия» [3, с. 269].

Для подтверждения к сказанному достаточно перечислить такие структурные единицы, как фонемы, аллофоны, транземы (или дифоны), слоги, звуки и т. п. Эти единицы принимают за минимальный участок речевого сигнала и обозначают такими понятиями, как «сегмент», «минимальный сегмент». Большинство исследователей относит понятие «минимальный сегмент» либо к слогу, либо к звуку [3], либо к фонеме [1], либо к некоторой области речевого сигнала, длительностью порядка 10...30 мс, либо к некоторому гибкому частотно-временному окну [9]. В работе [13] в качестве «минимального сегмента» принят так называемый «элементарный сегмент гласного звука речи».

Так, например, Р. К. Потапова к элементам микросегментации относит сегменты следующего порядка: интразвуковые (межзвуковые переходные процессы, смычка, фрикация, эксплозия и т. д.), звуковые, интерзвуковые (сочетания двух соседних звуков) и слоговые. Она пишет: «Для решения проблемы сегментации звучащей речи большое значение имеет обращение к слогу... Сегментация может проводиться в два этапа: на слоги, а затем на звуки, их составляющие, в результате чего уточняются границы между слогами» [3, с. 300]. Л. В. Бондарко определяет слог как «минимальный сегмент речевой цепи». С. В. Кодзасов и О. Ф. Кривнова указывают, что «звуковая информация об означаемом языкового знака не может быть сведена только к его фонемному составу, поскольку существуют звуковые явления, сферой реализации которых служат фонемные цепочки: слоги или слова. Такие явления относят к *супрасегментным* (от лат. *supra* — «над») звуковым средствам языка, в отличие от фонем — *сегментных* единиц» [1, с. 26–27]. Н. И. Дукельский в работе «Принципы сегментации речевого потока» отмечает, что «единственным критерием для отделения в речевом потоке кратчайшего функционально значимого отрезка... является... изменение вида источника речеобразования. Такого рода кратчайшие функционально значимые отрезки речевого потока будем называть сегментами» [16, с. 21]. На представленных в работе Н. И. Дукельского осциллограммах речевого потока «сегменты» выглядят в виде некоторых речевых участков, соответствующих определенному звуку речи, как согласному, так и гласному. Границами, определяющими расположение «сегмента», являются границы между двумя соседними звуками.

Однако, по нашему мнению, назначая объект своего исследования в соответствии с указанными выше представлениями, многие ученые сталкиваются

с трудностями принципиального характера. Они заключаются в том, что первичные признаки, которые определяются на рассмотренных выше минимальных сегментах, теряют свою информативную способность. А ведь, как известно, от правильного выбора минимального сегмента и, следовательно, первичных признаков зависит успех распознавания в целом. Решение любой прикладной задачи в области речевых технологий неразрывно связано с построением формальной системы в базисе теории распознавания образов, где одним из основных этапов процесса распознавания является этап формирования первичных *информативных* признаков.

Поэтому, если рассматривать в качестве минимального сегмента:

1) слог, звук или «сегмент» Дукельского, то это не отвечает объективной (физической) реальности, так как сами слог, звук или «сегмент» Дукельского «построены» из еще более простых составных структурных элементов (как будет показано ниже);

2) фонеме¹¹, то данная лингвистическая категория не может рассматриваться на акустическом уровне, поскольку это идеальные семиотические элементы, идеальные семиотические ценности, не заключающие в себе ничего физического, это «абстрактные единицы» [3, с. 85];

3) некоторую автоматически выделяемую область (области) речевого сигнала фиксированной или переменной (так называемые временные окна¹²) длительности, что вообще характерно для исследований речи спектральными методами, то в этом случае разрушается целостность объекта (как физическая, так и функциональная) и, как следствие, его информативность.

¹¹ «Основными конкретными единицами фонетического уровня являются звуки, а основными абстрактными единицами — фонемы и морфемы» [3, с. 122]. В устной речи фонема представлена в виде своих оттенков, вариантов, аллофонов, один из которых меньше зависит от окружающих звуков, другие — больше. Первые называются *основными* аллофонами, вторые — *комбинаторными* и *позиционными* [3, с. 88].

¹² Оценка спектральных параметров речи производится с использованием окон частотного взвешивания. Существует множество различных временных окон, наиболее известны из них: окно Хемминга, косинусное окно, окно Гаусса, прямоугольное окно, окно Блэкмана, «функция-окно» Габора, гибкое частотно-временное окно интегрального вейвлет-преобразования и др. Наибольший интерес в исследовании речи спектральными методами представляет: 1) окно Гаусса, т. к. оно определено на всей числовой оси, содержит только один максимум в амплитудном спектре и не содержит ни одного артефакта, а также легко применимо при разного рода аналитических вычислениях; 2) «функция-окно» Габора, параметры которого используются для перемещения окна с целью покрытия всей временной области для получения локальной информации о преобразовании Фурье сигнала [9]; 3) гибкое частотно-временное окно интегрального вейвлет-преобразования, которое автоматически сжимается в окрестностях высоких частотных центров и расширяется у низких частотных центров [9].

Как видим, вследствие указанных причин первичные признаки, формируемые в границах приведенных вариантов минимального сегмента речи, лишены своей качественной информативной определенности.

Следует сказать, что особенно остро проблема определения материальных (физических) границ, например между звуками, «внутри» самого звука, встает каждый раз перед теми, кто пытается описать речевые явления в терминах точных наук. Об отсутствии четких акустических признаков границ звука известный шведский фонетист Г. Фант пишет: «В результате подобной чисто акустической сегментации может быть получено некоторое число минимальных звуковых единиц, имеющих размер, равный размеру звука речи или меньший... Число таких последовательных во времени звуковых единиц, как правило, больше числа символов фонетической или фонематической транскрипции» [17, с. 35].

Целесообразно привести примеры акустических признаков сегментации речевого сигнала, которые так же, как и в случае идентификации человека при помощи современных методов, носят, в основном, спектральный характер. К ним относятся [3, с. 271]: наличие — отсутствие частоты основного тона; скачкообразное повышение частоты основного тона на переходе от согласного к гласному; скачкообразное понижение частоты основного тона на переходе от гласного к согласному; наличие / отсутствие шума; локализация полос шума на шкале частот; интенсивность полос шума; крутизна нарастания шума; длительность шума; наличие низко-, высокочастотной энергии; длительность сегмента.

В-седьмых, на сегодняшний день недостаточно глубоко, по нашему мнению, продуман вопрос, связанный с методами последующей обработки первичных акустических признаков.

Например, методы фоноскопической идентификации человека, в основном, используют так называемую «кодирующую» концепцию сравнения образцов речи. К этой группе относятся методы, формирующие для цели идентификации: 1) «кодированную страницу», 2) «параметрические коды»¹³, 3) «отпечатки голосов»¹⁴ и 4) «эталонные» дикторов¹⁵. Несмотря на различия в названиях, суть указанных методов одинакова. Обычно полная процедура идентификации, в которой диктору присваива-

¹³ Галляшина Е. Речь под микроскопом. (Интернет-статья), 1999 г.

¹⁴ Т. н. Система оперативной верификации и идентификации голоса (СОВИГ); разработчики — Д. Н. Коновалов, д.ф.-м.н., и А. Г. Бояров; система была представлена на прошедшей 2 ноября 2000 г. конференции «Информационная безопасность компьютерных систем».

¹⁵ Галунов В. И. Верификация и идентификация говорящего. (Интернет-статья), 2000 г.

ется определенного рода код (кодовая комбинация), такова. Анализ речевого сигнала начинается с перевода его в цифровую форму. Производится сегментация сигнала на отдельные элементы. Затем цифровой сигнал обрабатывается с помощью определенных алгоритмов (спектрального анализа, линейного предсказания, кепстральной обработки и др.). В результате получается параметрическое описание сегментов речевого сигнала в виде вектора первичных параметров. Следующий этап — это сравнение с имеющимися эталонными описаниями зарегистрированного числа дикторов в базе данных компьютера. Механизм сравнения реализуется при помощи метода динамического программирования, скрытых марковских моделей (в основном, для распознавания слитной речи), искусственных нейронных сетей или комбинаций указанных методов. По существу же задача компьютера — распознать человека по параметрам речевого сигнала — не отличается от задачи, решаемой экспертом-криминалистом. Задача компьютера заключается в том, чтобы сравнить некоторый код неизвестного диктора с эталонным кодом заявленного диктора (при верификации) или с эталонным параметрическим «кодовым» описанием каждого из конечного числа зарегистрированных дикторов (при идентификации).

Если сравнение при верификации показывает приемлемую «близость», которая вычисляется по определенным критериям, то система считает диктора «своим», а если значение близости превышает некий порог, то диктор объявляется «чужим». При идентификации компьютер, сравнивая «спектральные коды», выбирает наиболее «близкий» код диктора из числа кодов, имеющихся в его памяти. При этом следует подчеркнуть, что попытки применения на практике рассмотренной «кодовой» концепции привели к ряду критических замечаний (Е. Галяшина, В. И. Галунов и др.).

4. Заключение

Таким образом, анализ современного состояния исследований в области распознавания речевых образов показал ряд проблем существующей речевой теории. Одно из решений этих проблем мы видим в развитии структурного направления в области прикладной лингвистики [13].

Список литературы: 1. *Кодзасов С. В., Кривцова О. Ф.* Общая фонетика: Учебник. — М.: Рос. гос. гуманитар. ун-т, 2001. — 592 с. 2. *Головин Б. Н.* Введение в языкознание: Учебн. пособие. — М.: Высшая школа, 1983. — 232 с. 3. *Поталова Р. К.* Речь: коммуникация, информация, кибернетика: Учебн. пособие. — М.: Едиториал УРСС, 2003. — 568 с. 4. Распознавание слуховых образов / Под ред. Н. Г. Загоруйко. — Новосибирск: Наука, Сибирск. отд-ние, 1970. — 388 с. 5. *Сапожков М. А.* Речевой сигнал в кибернетике и связи. — М.: Связьиздат, 1963. — 452 с. 6. *Витцок Т. К.* Анализ, распознавание и интерпретация речевых сигналов. — Киев: Наук. думка, 1987. — 262 с. 7. Распознавание образов. Исследование живых и автоматических распознающих систем: Пер. с англ. — М.: Мир, 1970. — 288 с. 8. *Доддингтон Дж. Р.* Распознавание дикторов: Идентификация людей по голосу // ТИИЭР. — 1985. — Т. 73, № 11. — С. 129–137. 9. *Чуи Ч.* Введение в вэйлеты: Пер. с англ. — М.: Мир, 2001. — 412 с. 10. Современные речевые технологии: Сб. трудов IX сессии Российского акустического общества. — М.: ГЕОС, 1999. — 166 с. 11. *Чистович Л. А., Венцов А. В., Гранстрем М. Б. и др.* Физиология речи. Восприятие речи человеком. — Л.: Наука, Ленингр. отд-ние, 1976. — 386 с. 12. *Сомьен Дж.* Кодирование сенсорной информации: Пер. с англ. — М.: Мир, 1975. — 414 с. 13. *Бондаренко М. Ф., Дрюченко А. Я., Шабанов-Кушнарченко Ю. П.* Гласные звуки в теории и эксперименте. — Харьков: ХНУРЭ, 2002. — 348 с. 14. *Бару А. В.* Слуховые центры и опознание звуковых сигналов. — Л.: Наука, Ленингр. отд-ние, 1978. — 192 с. 15. *Якобсон Р., Фант Г., Халле М.* Введение в анализ речи. Различительные признаки и их корреляты // Новое в лингвистике. Вып. II. — М.: Изд-во АН СССР, 1962. — С. 57–68. 16. *Дукельский Н. И.* Принципы сегментации речевого потока. — Л.: Ленингр. отд-ние изд-ва АН СССР, 1962. — 140 с. 17. *Фант Г.* Акустическая теория речеобразования: Пер. с англ. / Под ред. В. С. Григорьева. — М.: Наука, 1964. — 284 с.

Поступила в редакцию 18.02.2006

Е. В. Волченко

МОДИФИЦИРОВАННЫЙ МЕТОД ПОТЕНЦИАЛЬНЫХ ФУНКЦИЙ**1. Введение**

Задача выбора решающего правила классификации является одним из пяти этапов построения системы автоматического распознавания [1]. Она состоит в нахождении (построении) численного выражения или алгоритма, результатом выполнения которого будет номер класса, к которому будет отнесен распознаваемый объект. В зависимости от объема априорной информации, доступной разработчику системы на этапе ее проектирования, системы распознавания делятся на три вида, для каждого из которых существуют свои особенности выбора решающего правила классификации [2]:

- системы без обучения (в качестве решающего правила используется критерий Байеса и его модификации);
- системы с учителем (для заданной классифицированной обучающей выборки необходимо найти решающее правило итерационно в процессе обучения);
- самообучающиеся системы (для заданной неклассифицированной выборки необходимо выделить классы объектов и построить алгоритм классификации во время самообучения).

Как известно, каждый объект в системе распознавания описывается набором признаков, количество которых может быть достаточно велико. Если каждому признаку в многомерном пространстве поставить в соответствие координатную ось, то каждому объекту будет соответствовать точка в полученном пространстве признаков. Объекты, имеющие то или иное сходство, на основе которого они относятся к одному классу в пространстве признаков, локализуются в определенные области. Всем распознаваемым классам в пространстве признаков соответствует несколько областей, поэтому решить задачу распознавания на основе обучающей выборки означает установить поверхность (гиперплоскость), которая разделяет эти области [3].

Для решения задачи распознавания в обучающихся системах существует несколько групп алгоритмов, отличающихся способом построения решающей функции (правила) [4]. К первой группе относятся методы, с помощью которых строится некоторый функционал на основе концепции допустимых преобразований. Основной проблемой таких методов являются слабые адаптивные способности. Вторая группа методов характеризуется тем, что решающее правило находится рекуррентно в процессе обучения. Различие между алгоритмами данной группы состоит в выборе аппроксимирующей функции, вида экстремизируемого функционала и способа экстремизации этого функционала. Наиболее из-

вестными алгоритмами этой группы являются метод секущих гиперплоскостей, метод потенциальных функций и метод группового учета аргументов. К третьей группе относятся системы перцептронного типа и методы лингвистического распознавания. Однако эти методы отличаются низкой помехоустойчивостью и требуют наличия значительной априорной информации о распознаваемых объектах.

В данной статье рассматривается метод потенциальных функций, предложенный в [5], как один из наиболее часто используемых при построении сложных систем и к которому могут быть сведены многие алгоритмы обучения, используемые, в том числе, и в нейронных сетях.

При решении практических задач довольно часто приходится иметь дело с большими выборками объектов, описываемых несколькими десятками признаков. Это приводит к существенным временным и ресурсным затратам при хранении и обработке таких выборок. Поэтому очевидна потребность обработки обучающей выборки с целью сокращения ее длины, не уменьшая при этом качества обучения и распознавания [6].

В [7] предлагается два алгоритма уменьшения размера обучающей выборки. Алгоритм STOLP включает в новую выборку только «точки опоры», расстояние до которых от объектов «своего» класса меньше расстояния от объектов «чужого» класса. Существенным недостатком алгоритма является его комбинаторная сложность. При выполнении распознавания по «точкам опоры» предлагается использовать метод ближайшего соседа [8], который рассматривает только один ближайший объект каждого из классов, что может приводить к ошибкам классификации. Алгоритм ДРЭТ (метод «дробящихся эталонов») основан на идее покрытия всего обучающего множества объектов в признаковом пространстве сферами минимального радиуса. Каждая сфера покрывает только объекты одного класса, а ее радиус подбирается во время работы алгоритма. При проведении распознавания объект будет отнесен к тому классу, расстояние до центра ближайшей сферы которого минимально.

В [9] предлагается алгоритм формирования обучающей выборки по алгоритму, сходному с алгоритмом STOLP. Выбор «узловых точек» осуществляется в результате анализа матрицы расстояний.

На наш взгляд, общим недостатком методов, приведенных в [7] и [9], является исключение из обучающей выборки некоторой части объектов без сохранения значений их признаков для построения решающей функции.

Авторы статьи [10] предлагают устранить этот недостаток посредством нахождения центра сосредоточения объектов каждого из классов и замены всех объектов класса на один центральный объект. При проведении такой замены существенно уменьшаются временные и ресурсные характеристики алгоритмов обучения и распознавания, однако при пересечении классов в признаковом пространстве эффективность алгоритма существенно снижается. Процесс распознавания в алгоритме также сводится к нахождению класса, расстояние до центрального объекта которого минимально.

В данной статье предлагается иной подход к построению решающего правила методом потенциальных функций. Он заключается в том, что группа близкорасположенных объектов одного класса заменяется одним мета-объектом. Из мета-объектов формируется новая обучающая выборка, которая является исходными данными для модифицированного метода потенциальных функций. Предложенный подход позволяет существенно уменьшить время обучения и распознавания, сократить объем памяти для хранения обучающей выборки, однако при этом сохранить эффективность распознавания.

2. Постановка задачи

Пусть задана некоторая обучающая классифицированная выборка объектов $\bar{W} = \{\bar{W}_1, \bar{W}_2, \dots, \bar{W}_s\}$, каждый из которых относится к классу V_1 или V_2 . Каждый объект задан набором своих признаков $\bar{W}_i = \{w_{i1}, w_{i2}, \dots, w_{im}\}$, где w_{ij} — j -й признак i -го объекта обучающей выборки. Необходимо найти такую разделяющую функцию $y = f(W)$, которая по своему знаку определяет две области в n -мерном пространстве признаков:

$$\text{sign } y = \text{sign } f(W) = \begin{cases} +1, & \text{если } \bar{W}' \in V_1; \\ -1, & \text{если } \bar{W}' \in V_2, \end{cases}$$

где \bar{W}' — распознаваемый объект.

Задача нахождения такой гиперплоскости методом потенциальных функций сводится к рекуррентному построению функции вида

$$U(\bar{W}', W_k) = \sum_{m=1}^k U(\bar{W}', \bar{W}_m),$$

где W_k — множество из k объектов обучающей выборки, включенных в потенциальную функцию; $U(\bar{W}', \bar{W}_m) = \exp(-\alpha \cdot d^2)$, $d^2 = \|\bar{W}' - \bar{W}_m\|^2$; α — параметр крутизны потенциальной функции, выбираемый эвристически.

Для оценки качества распознавания построенной потенциальной функции [5] введем меру $N(U(\bar{W}', W_k))$, равную количеству неверно классифицированных объектов обучающей выборки потенциальной функцией $U(\bar{W}', W_k)$.

В качестве эталонного значения N_0 меры $N(U(\bar{W}', W_k))$ для дальнейшей оценки произволь-

ной потенциальной функции примем количество неверно классифицированных объектов функцией, включающей в себя все объекты обучающей выборки $N(U(\bar{W}', W_k))$.

Очевидно, что время выполнения классификации прямо пропорционально числу k . Поэтому при решении задачи нахождения оптимальной разделяющей функции необходимо минимизировать это число без потери качества распознавания, т. е. должны выполняться следующие условия:

$$N(U(\bar{W}', W_k)) = N_0; k \rightarrow \min. \quad (1)$$

3. Описание метода построения мета-выборки и алгоритма построения решающей функции

Для простоты геометрического представления объектов обучающей выборки в виде точек, расположенных в пространстве признаков Π , без потери сущности рассматриваемой задачи примем, что каждый объект обучающей выборки описывается двумя некоррелированными признаками w_1 и w_2 , а система распознавания имеет только два класса в составе алфавита.

При изучении возможных вариантов расположения объектов двух классов в признаковом пространстве были выделены следующие ситуации:

1) классы в признаковом пространстве линейно разделимы, что позволяет использовать в качестве решающего правила одну плоскость (рис. 1);

2) классы линейно неразделимы, однако при разделении их некоторой плоскостью количество неверно классифицированных объектов обучающей выборки достаточно мало (рис. 2);

3) классы в признаковом пространстве существенно пересекаются, и среди объектов одного класса имеются области, в которых присутствуют только неверно классифицируемые объекты другого класса (рис. 3);

4) классы в признаковом пространстве существенно пересекаются, нет областей, в которых присутствуют только неверно классифицируемые объекты одного из классов (рис. 4).

Анализируя все возможные варианты расположения объектов в признаковом пространстве (рис. 1–4), приходим к выводу, что для всех ситуаций характерно

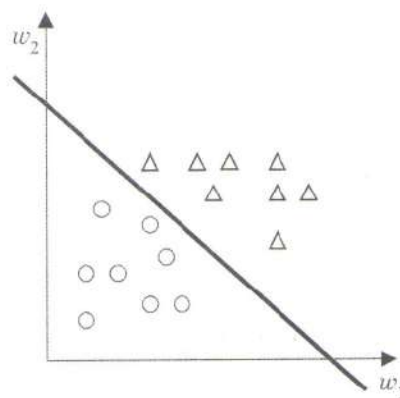


Рис. 1

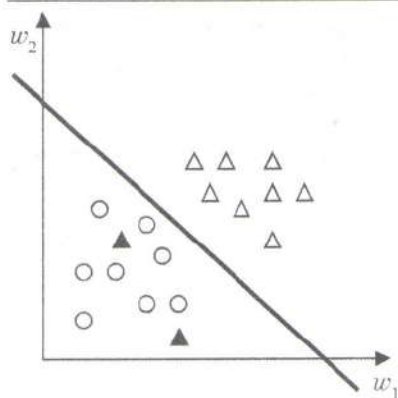


Рис. 2

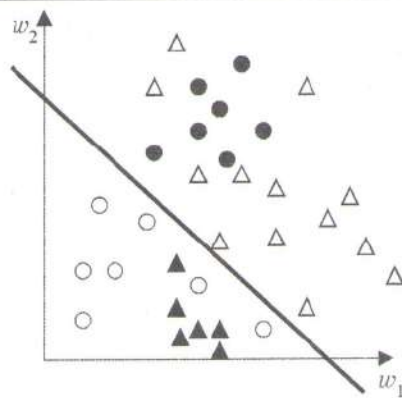


Рис. 3

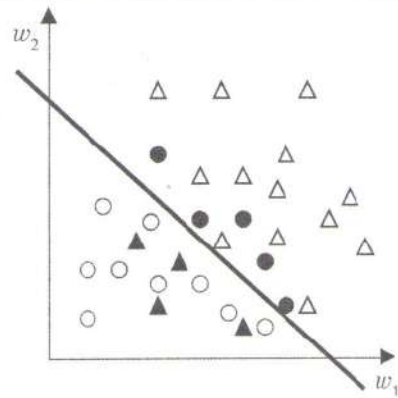


Рис. 4

наличие областей, в которых находятся объекты только одного из классов. Если предположить, что построена решающая функция, которая верно классифицирует все объекты обучающей выборки, то объекты такой области будут классифицироваться одинаково. Очевидно, что в таком случае в обучающей выборке может быть оставлен только один из этих объектов. Однако такой подход не позволит учитывать количество объектов в каждой из групп и их расположение относительно других групп, что может привести к существенному изменению решающей функции, построенной по усеченной выборке по сравнению с исходной выборкой. В связи с этим нами предлагается метод объединения близкорасположенных объектов в один мета-объект и модифицированный метод потенциальных функций, позволяющий учитывать количество объектов, объединенных в мета-объект.

Он состоит в выполнении следующих шагов.

1. Рассчитаем расстояния между всеми парами объектов двух классов

$$r_{ij} = \|\bar{W}_i - \bar{W}_j\|, \quad (2)$$

где $\bar{W}_i \in V_1, i = 1, |V_1|; \bar{W}_j \in V_2, j = 1, |V_2|; |V_i|$ — мощность класса V_i .

2. Выбираем объект f одного из классов, для которого сумма расстояний до всех объектов другого класса, рассчитанная по формуле (2), максимальна:

$$f = \arg \max_{j=1, |V_1|} \sum_{j=1}^{|V_2|} r_{ij}.$$

3. Находим объект s другого класса, до которого расстояние от f минимально:

$$s = \arg \min_{j=1, |V_1|} r_{fs}.$$

4. Выбираем все объекты того же класса, что и f , расстояние до которых от f меньше, чем до s (рис. 5), и помещаем их во множество W_f

$$W_f = \{\bar{W}_i | \bar{W}_i \in W_f (r_{fi} < r_{si})\}.$$

Отметим, что $\bar{W}_f \in W_f$.

5. Заменяем W_f мета-объектом \overline{MW}_f . При этом будем говорить, что объект \bar{W}_f является включен-

ным в мета-объект \overline{MW}_f , если он принадлежит множеству W_f . Значения признаков этого мета-объекта рассчитаем как средние значения признаков всех объектов, в него включенных:

$$w_{f1} = \frac{\sum_{i=1}^p w_{i1}}{p}; w_{f2} = \frac{\sum_{i=1}^p w_{i2}}{p},$$

где $p = |W_f|$ — вес мета-объекта \overline{MW}_f .

Таким образом получен мета-объект \overline{MW}_f со значениями признаков $\{w_{f1}, w_{f2}\}$.

6. Удалим из исходной обучающей выборки все объекты, включенные в созданный мета-объект \overline{MW}_f .

7. Выполняем пп. 2–5 до тех пор, пока в исходной обучающей выборке не останется ни одного объекта. Сформированную новую обучающую выборку мета-объектов назовем мета-выборкой.

Из описания метода следует, что:

- 1) один мета-объект может содержать от 1 до объектов исходной обучающей выборки;
- 2) два мета-объекта не могут содержать один и тот же объект исходной выборки;
- 3) мета-объект будет относиться к тому же классу, что и все объекты, в него включенные.

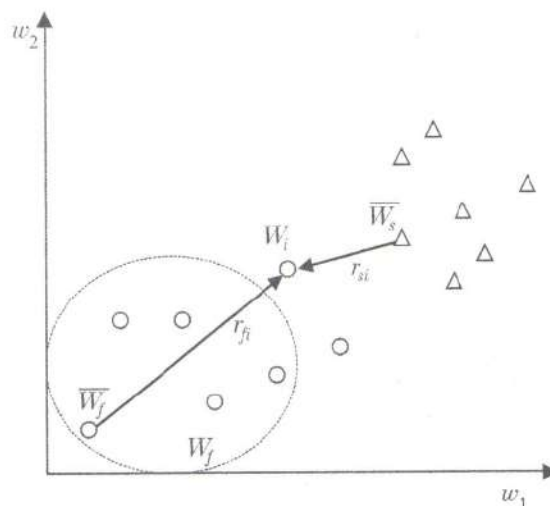


Рис. 5

Для нахождения разделяющей функции по мета-выборке будем использовать метод потенциальных функций. Для учета количества объектов, включенных в каждый из мета-объектов, выполним модификацию этого метода. С этой целью заменим вектор принадлежности объектов потенциальной функции к одному из классов

$$\alpha = \{\alpha_1, \dots, \alpha_j, \dots, \alpha_k\},$$

$$\text{где } \alpha_j = \begin{cases} +1, & \text{если } \overline{W_j} \in V_1; \\ -1, & \text{если } \overline{W_j} \in V_2. \end{cases}$$

на вектор

$$\alpha' = \{\alpha'_1, \dots, \alpha'_j, \dots, \alpha'_{mk}\},$$

$$\text{где } \alpha'_j = \begin{cases} +p, & \text{если } \overline{MW_j} \in V_1; \\ -p, & \text{если } \overline{MW_j} \in V_2. \end{cases}$$

Введение вектора α' позволяет учитывать совместные значения потенциалов всех объектов, включенных в мета-объекты. Заметим, что если объектам обучающей выборки приписаны веса m , то полагаем, что

$$\alpha'_j = \begin{cases} +\sum_{l=1}^p m_l, & \text{если } \overline{MW_j} \in V_1; \\ -\sum_{l=1}^p m_l, & \text{если } \overline{MW_j} \in V_2. \end{cases}$$

4. Результаты экспериментов

Предложенный выше подход к построению решающего правила модифицированным методом потенциальных функций по мета-выборке был реализован программно. Для оценки его эффективности использовалась обучающая выборка, значения признаков объектов которой распределены по нормальному закону, достаточно распространенному в прикладных задачах.

При проведении экспериментов оценивались длина потенциальной функции k и количество неверно классифицированных объектов исходной обучающей выборки $N(U(\overline{W}', W_k))$, описанные в условиях (1).

В каждом из экспериментов были получены три решающие функции:

1) $U(\overline{W}', W_k)$, построенная простым методом потенциальных функций по исходной обучающей выборке;

2) $U(\overline{W}', MW_{ms})$, в которую включены все мета-объекты, где ms — количество мета-объектов;

3) $U(\overline{W}', MW_{mk})$, построенная модифицированным методом потенциальных функций по мета-выборке, где mk — количество объектов, включенных в потенциальную функцию, построенную модифицированным методом по мета-выборке.

При проведении испытаний использовались обучающие выборки длиной 200...1000 объектов, так как, согласно [11], дальнейшее увеличение объема выборки не улучшает качества получаемого решаю-

щего правила. При оценке решающих функций для различной площади пересечения классов рассматривались ситуации, когда классы пересекаются не более чем на 70%, что соответствует данным, используемым в реальных системах. Для непересекающихся классов рассматривалось две ситуации:

- классы расположены близко друг к другу;
- классы существенно удалены друг от друга.

На рис. 6 приведены исходная обучающая выборка (а) и полученная по ней мета-выборка (б). Анализ результатов объединения объектов в мета-объекты показал, что вес мета-объекта тем больше, чем дальше этот мета-объект находится от границы между классами (области пересечения классов). Было получено, что в зависимости от размера исходной обучающей выборки и площади пересечения классов использование мета-объектов позволяет уменьшить обучающую выборку в 5...300 раз.

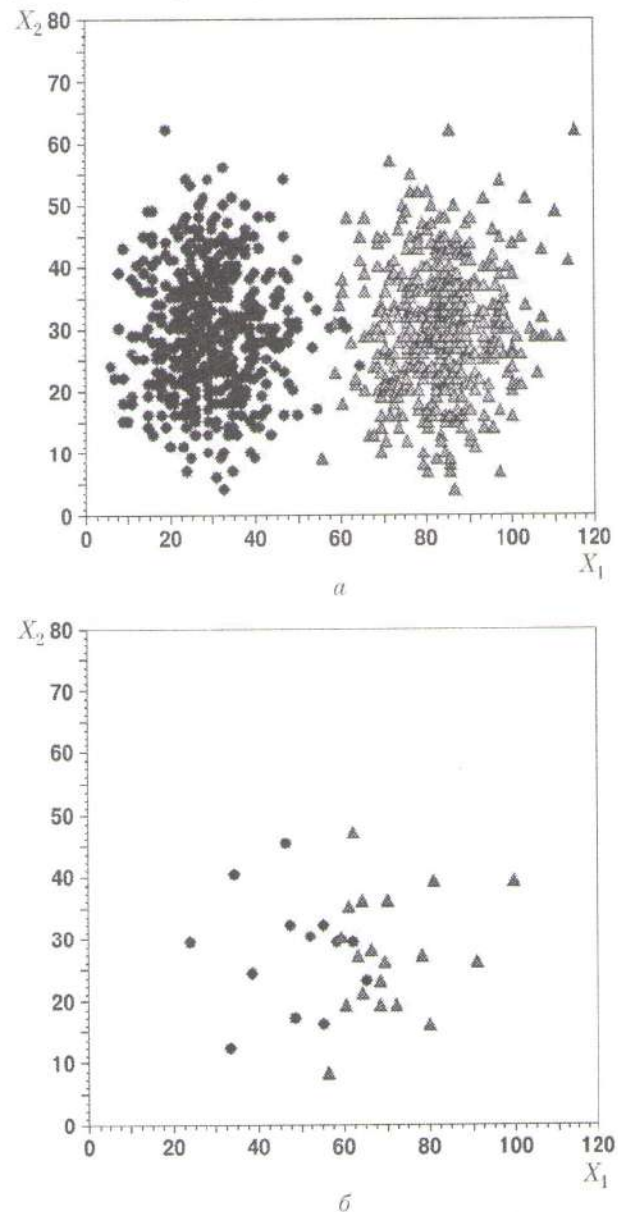


Рис. 6. Исходная обучающая выборка (а) и построенная по ней мета-выборка (б)

В таблицах 1–3 приведена зависимость длины решающей функции и количества неверно классифицированных объектов обучающей выборки от ее длины для близкорасположенных, но не пересекающихся классов. В таблицах 4–6 приведена зависимость длины решающей функции и количества неверно классифицированных объектов обучающей выборки от площади пересечения классов в признаковом пространстве при размере обучающей выборки 1000 объектов (по 500 объектов каждого из классов). Все значения, приведенные в таблицах, являются средними по 20 проведенным экспериментам.

Анализ приведенных в таблицах 1–6 результатов экспериментов показывает, что для любого размера обучающей выборки и площади пересечения классов длина решающей функции $U(\bar{W}', MW_{mk})$ в среднем на 30 % меньше длины $U(\bar{W}', W_k)$. При этом количество неверных классификаций объектов исходной обучающей выборки $N(U(\bar{W}', MW_{mk}))$ на 10 % меньше $N(U(\bar{W}', W_k))$. Сравнительный анализ $U(\bar{W}', MW_{ms})$ с $U(\bar{W}', W_k)$ и $U(\bar{W}', MW_{mk})$ показывает, что:

- длина $U(\bar{W}', MW_{ms})$ в 2 и 2,5 раза больше длины $U(\bar{W}', W_k)$ и $U(\bar{W}', MW_{mk})$ соответственно;
- $N(U(\bar{W}', MW_{ms}))$ в среднем на 40 % и 30 % меньше $N(U(\bar{W}', MW_{mk}))$ и $N(U(\bar{W}', W_k))$ соответственно.

Таблица 1

| Размер обучающей выборки | Метод потенциальных функций, реализованный по исходной выборке | |
|--------------------------|--|-----------------------|
| | k | $N(U(\bar{W}', W_k))$ |
| 200 | 3,62 | 0,55 |
| 400 | 4,91 | 0,85 |
| 600 | 5,57 | 1,12 |
| 800 | 6,34 | 1,18 |
| 1000 | 7,79 | 1,25 |

Таблица 2

| Размер обучающей выборки | Метод потенциальных функций, реализованный по мета-выборке | |
|--------------------------|--|---------------------------|
| | mk | $N(U(\bar{W}', MW_{mk}))$ |
| 200 | 2,45 | 0,47 |
| 400 | 2,65 | 0,81 |
| 600 | 3,73 | 1,07 |
| 800 | 4,51 | 1,08 |
| 1000 | 4,75 | 1,1 |

Таблица 3

| Размер обучающей выборки | Использование в качестве потенциальной функции всех объектов мета-выборки | |
|--------------------------|---|---------------------------|
| | ms | $N(U(\bar{W}', MW_{ms}))$ |
| 200 | 6,3 | 0 |
| 400 | 7,3 | 0 |
| 600 | 9,5 | 0 |
| 800 | 11,7 | 0 |
| 1000 | 16,2 | 0 |

Таблица 4

| Площадь пересечения классов, % | Метод потенциальных функций, реализованный по исходной выборке | |
|--------------------------------|--|-----------------------|
| | k | $N(U(\bar{W}', W_k))$ |
| Классы существенно обособлены | 3,1 | 0,5 |
| 0 | 7,8 | 1,25 |
| 10 | 13,6 | 3,4 |
| 20 | 23,4 | 5,75 |
| 30 | 45,6 | 13,5 |
| 40 | 74,2 | 30,8 |
| 50 | 130,1 | 55,4 |
| 60 | 205,7 | 89,8 |
| 70 | 303,3 | 161,6 |

Таблица 5

| Площадь пересечения классов, % | Метод потенциальных функций, реализованный по мета-выборке | |
|--------------------------------|--|---------------------------|
| | mk | $N(U(\bar{W}', MW_{mk}))$ |
| Классы существенно обособлены | 2,1 | 0,1 |
| 0 | 4,75 | 1,1 |
| 10 | 8,9 | 3,1 |
| 20 | 15,5 | 5,3 |
| 30 | 34,3 | 12,3 |
| 40 | 59,3 | 26,6 |
| 50 | 104,2 | 47,6 |
| 60 | 175,3 | 87,7 |
| 70 | 263,5 | 140,4 |

Таблица 6

| Площадь пересечения классов, % | Использование в качестве потенциальной функции всех объектов мета-выборки | |
|--------------------------------|---|---------------------------|
| | ms | $N(U(\bar{W}', MW_{ms}))$ |
| Классы существенно обособлены | 6,7 | 0 |
| 0 | 16,2 | 0 |
| 10 | 32 | 0,35 |
| 20 | 49,1 | 0,85 |
| 30 | 95,1 | 4,7 |
| 40 | 159,7 | 14,1 |
| 50 | 283,8 | 32,6 |
| 60 | 434,1 | 65,3 |
| 70 | 589,7 | 128,1 |

Также при проведении экспериментальных исследований были рассчитаны среднеквадратический разброс σ_k длины мета-выборки и среднеквадратический разброс $\sigma_{N(U(\bar{W}', MW_{mk}))}$ количества неверно классифицированных объектов потенциальной функцией, построенной по мета-выборке при различном порядке следования объектов в обучающей выборке (порядок следования объектов в обучающей мета-выборке изменялся посредством «взбалтывания»). Результаты этого исследования для исходных обучающих выборок различных размеров

приведены в таблицах 7 и 8. Было получено, что вне зависимости от длины обучающей выборки $\sigma_{N(U(W', MW_{mk}))}$ больше $\sigma_{N(U(W', W_k))}$, т. е. модифицированный метод потенциальных функций более чувствителен к порядку следования мета-объектов в мета-выборке.

Таблица 7

| Площадь пересечения классов, % | Метод потенциальных функций, реализованный по исходной выборке | |
|--------------------------------|--|--------------------------|
| | σ_k | $\sigma_{N(U(W', W_k))}$ |
| Классы существенно обособлены | 0,8 | 0,44 |
| 0 | 1,67 | 3,76 |
| 10 | 1,62 | 2,87 |
| 20 | 2,37 | 4,28 |
| 30 | 3,04 | 4,35 |
| 40 | 4,36 | 9,72 |
| 50 | 5,8 | 15,8 |

Таблица 8

| Площадь пересечения классов, % | Метод потенциальных функций, реализованный по мета-выборке | |
|--------------------------------|--|------------------------------|
| | σ_k | $\sigma_{N(U(W', MW_{mk}))}$ |
| Классы существенно обособлены | 0 | 1,4 |
| 0 | 0,93 | 5,97 |
| 10 | 1,26 | 10,33 |
| 20 | 1,85 | 18,75 |
| 30 | 2,61 | 23,25 |
| 40 | 4,11 | 27,19 |
| 50 | 5,09 | 28,56 |

5. Выводы

В настоящей работе предложен модифицированный метод потенциальных функций для нахождения решающего правила обучающейся системы распознавания. В качестве обучающей выборки для реализации этого метода было предложено исполь-

зовать выборку мета-объектов, методика формирования которой также описана в данной статье. Проведенные эксперименты показали преимущества предложенного подхода по сравнению с простым методом потенциальных функций.

Автор благодарен Л. А. Белозерскому и И. С. Грунскому за ряд полезных замечаний, способствовавших улучшению статьи.

Список литературы: 1. Белозерский Л. А. Введение в теорию автоматического распознавания. – Киев: Наук. думка, 2005. – 434 с. 2. Горелик А. Л., Скрипкин В. А. Методы распознавания: Учеб. пособ. для вузов. – М.: Высшая школа, 1977. – 222 с. 3. Букатова Л. И. Эволюционное моделирование: идеи, основы теории, приложения. – М.: Знание, 1981. – 64 с. (Новое в науке и технике. Сер. «Математика, кибернетика»; № 10). 4. Васильев В. И. Распознающие системы: Справочник. – Киев: Наук. думка, 1983. – 422 с. 5. Айзерман М. А., Браверманн Э. М., Розноэр Л. И. Метод потенциальных функций в теории обучения машин. – М.: Наука, 1970. – 384 с. 6. Белозерский Л. А. Совместимость обучаемого классификатора и систем автоматического распознавания // Искусственный интеллект. – 2003. – № 4. – С. 325–344. 7. Загоруйко И. Г. Прикладные методы анализа знаний и данных. – Новосибирск: Изд-во Ин-та математики, 1999. – 270 с. 8. Дуда Р., Харт П. Распознавание образов и анализ сцен. – М.: Мир, 1976. 9. Дубровин В. И., Субботин С. А., Кривенко В. И., Евченко Л. Н. Сокращение объема данных в задачах распознавания и диагностики // Тр. VIII Всероссийской конференции «Нейрокомпьютеры и их применение» (Москва, 21–22 марта 2002 г.) / Под редакцией проф. А. И. Галушкина. – М.: Ин-т проблем управления им. В. А. Трапезникова РАН, 2002. – С. 954–963. 10. Дубровин В. И., Корецкий Н. Х., Субботин С. А. Модифицированный метод потенциальных функций // Складные системы и процессы. – 2002. – № 1. – С. 12–19. 11. Жук А. В. Обучаемые классификаторы в статистических испытаниях // Искусственный интеллект. – 2005. – № 1. – С. 168–179.

Поступила в редколлегию 23.04.2006

УДК [615.471:616-071]:681.513

Н. А. Тесленко, И. Г. Чурюмова

НЕЧЕТКАЯ КЛАСТЕРИЗАЦИЯ МАССИВОВ БИМЕДИЦИНСКИХ ДАННЫХ В УСЛОВИЯХ ИЗБЫТОЧНОСТИ ИНФОРМАЦИИ

1. Введение

В последнее время пристальное внимание общественности уделяется проблеме здоровья студентов и других контингентов длительно учащейся молодежи [1, 2, 3]. Молодые люди в возрасте 16–25 лет приходят в вузы во время так называемого пика активного динамического развития. Функционально формирующиеся мышечные группы, в зависимости от адекватной физической нагрузки, оказывают прямое влияние на все внутренние органы и системы организма, управляя их работой и синхронизируя их функции [4]. Но двигательная активность студента резко ограничена как во времени, так и в пространстве. По характеру учебы он находится в состоянии постоянной гиподинамии, создаются условия для формирования, а затем развития статического мышечного перенапряжения [5]. Этот термин емко отображает функциональную недогруженность одних мышечных групп (например, мышц ног) и перегруженность других (например, мышц спины). Прежде всего это находит отражение в асинхронности мышечных сокращений и негативном влиянии на работу внутренних органов и особенно сердца.

Находясь ежедневно приблизительно в одном и том же положении, недостаточно нагруженные в физическом смысле, например, мышцы ног теряют свой тонус из-за уменьшения притока крови к ним, кровеносные сосуды сужаются, что ухудшает микроциркуляцию мышечных волокон, в расслабленных мышцах кровоток затрудняется и уменьшается. Растет общее периферическое сопротивление сосудов, что пагубно отражается на работе сердца как насоса, вынужденного с большим усилием работать, проталкивая кровь через мышечные группы. В то же время мышцы спины, находясь в тонусе, а затем переходя в состояние гипертонуса, нуждаются в усиленной работе сердца [6, 7]. Это находит отражение в центральной и периферической нервной регуляции просвета кровеносных сосудов. Сужение и расширение кровеносных сосудов нарушается. Учитывая, что жизнь учащейся молодежи связана со стресс-факторами в виде зачетов, экзаменов, контрольных и, как правило, протекает на фоне predisposing болезнетворных факторов (нарушение пищевого и водного режимов, нарушение формулы сна и т. д.), а зачастую сопровождается вредными привычками (курение, алкоголь, крепкий чай и кофе, избыток соли в пище), становится понятным значение развивающегося статического перенапряжения в формировании пограничных поражений сердечно-сосудистой системы [8].

2. Цель исследования

Целью нашего исследования явилось изучение закономерностей формирования статического мышечного перенапряжения как одного из главных predisposing факторов развития пограничных гипертонических состояний у учащейся молодежи и использование новых перспективных методов нечеткой кластеризации для обработки полученных данных.

Работа проводилась на базе Харьковского областного врачебно-физкультурного диспансера, кафедры внутренних болезней, спортивной медицины и лечебной физкультуры Харьковского государственного медицинского университета и кафедры биомедицинских электронных устройств и систем Харьковского национального университета радиоэлектроники. Объектами наблюдения являлись 368 студентов младших (1–2) и старших (4–5) курсов Харьковского государственного медицинского университета, Харьковской государственной академии физического воспитания, Харьковского национального университета им. Каразина и Харьковского национального университета радиоэлектроники. Обследуемые студенты, возраст которых составлял от 16 до 26 лет, были распределены на три группы.

3. Сжатие данных с помощью трехслойной нейронной сети и кластеризация данных с помощью нечеткой BSB-модели

В настоящее время задачи кластеризации данных различной природы успешно выполняются с помощью широкого класса нейронных сетей. Решение данной задачи основывается на принципах самообучения, которые заложены в основу конструкции автоассоциативной памяти как одного из видов искусственных нейронных сетей, отличающихся простотой архитектуры и наличием достаточно эффективных алгоритмов обучения.

Предварительный анализ данных показал, что исходная размерность пространства избыточна для решения задачи кластеризации, поскольку существуют неинформативные признаки. Для сокращения пространства исходных данных, т. е. уменьшения количества признаков, была реализована процедура сжатия информации с помощью автоассоциативной трехслойной нейронной сети типа «Bottle neck». Данная сеть характеризуется тем, что количество нейронов первого скрытого и третьего (выходного) слоев принято равным количеству признаков классифицируемых объектов (в нашем случае 19), а количество нейронов второго скрытого

слоя равно сокращенному количеству признаков. В таком случае выходы второго скрытого слоя представляют собой сжатые данные, а выходного слоя — восстановленные. Описанная нейронная сеть настраивается с помощью алгоритма на основе полиномиальной функции активации [9]

$$\psi_j(u_j) = 1.5 \left(u_j(k) - \frac{u_j^3(k)}{3} \right);$$

$$-1 < \psi_j(u_j) < 1; -1 < u_j < 1,$$

обеспечивающей более высокую скорость настройки синаптических весов сети по сравнению с традиционными процедурами.

В результате процедуры сжатия количество признаков сокращено до 7. Для обработки полученных данных была использована искусственная нейронная сеть автоассоциативной памяти специального вида, известная под названием «Brain-State-In-A-Box Model» [10, 11]. BSB-модель представляет собой нейродинамическую нелинейную замкнутую систему с амплитудными ограничениями, охваченную положительной обратной связью. Динамика этой системы определяется парой уравнений в пространстве состояний

$$\begin{cases} y(k, \tau) = x(k, \tau) + \alpha Wx(k, \tau), \\ x(k, \tau + 1) = \psi(y(k, \tau)), \end{cases}$$

где $x(k, 0) \equiv x(k)$ — $(n \times 1)$ -вектор-образ, подаваемый в систему; $k = 1, 2, \dots, l$ — номер конкретного образа в множестве фундаментальной памяти; $\tau = 0, 1, 2, \dots, T$ — итерации машинного времени; $x(k, T)$ — вектор состояний в установившемся режиме; α — малый положительный параметр обратной связи; W — $(n \times n)$ -матрица синаптических весов линейной корреляционной автоассоциативной памяти, представляющей собой по сути однослойную нейронную сеть, образованную адаптивными линейными ассоциаторами; $\psi(\bullet)$ — активационная кусочно-линейная функция с насыщением, действующая покомпонентно на элементы вектора $y(k, \tau)$ так, что

$$x_i(k, \tau + 1) = \psi(y_i(k, \tau)) = \begin{cases} +1, & \text{если } y_i(k, \tau) > +1; \\ y_i(k, \tau), & \text{если } -1 \leq y_i(k, \tau) < +1; \\ -1, & \text{если } y_i(k, \tau) < -1; \end{cases}$$

$$i = 1, 2, \dots, n.$$

Фазовое пространство BSB-нейромодели ограничено n -мерным гиперкубом с центром в начале координат и ребром длиной 2. Присутствие в модели положительной обратной связи приводит к тому, что подаваемые в модель сигналы, предварительно закодированные для работы в гиперкубе, усиливаются до тех пор, пока все состояния $x_i(k, \tau)$ не войдут в насыщение, при этом аналоговые входные сигналы преобразуются в дискретную форму, соответствующую одной из вершин гиперкуба.

BSB-модель решает задачу кластеризации заданного массива данных $x(k), k = 1, 2, \dots, l$ благодаря тому, что вершины гиперкуба в процессе обработки действуют как точечные аттракторы с некоторыми областями притяжения, которые делят n -мерное пространство признаков. В связи с тем, что количество аттракторов, т. е. вершин гиперкуба, равно $2^n \gg n$, а емкость линейной автоассоциативной памяти W не может превышать n (абсолютная емкость $l/n \leq 1$), возникает ситуация при которой большинство вершин будут оставаться пустыми. В то же время данные могут принадлежать нескольким или всем кластерам одновременно, так как многие реальные задачи не могут быть решены путем однозначного разделения всех объектов на группы.

Таким образом, целесообразно ввести некоторую функцию, которая будет описывать уровень принадлежности объекта каждому из классов [12]. Для этого можно воспользоваться идеями нечеткой кластеризации, используя простую и эффективную треугольную функцию принадлежности:

$$\mu_q(x(k, T)) = 1 - \frac{d(x(k, T), x_q)}{2n},$$

где $d(x(k, T), x_q) = \sum_{i=1}^n |x_i(k, T) - x_{q,i}|$ — хэммингово расстояние.

В случае гиперкуба уровень принадлежности описывает близость каждой вершины объекта $x(k, T)$ к q -й вершине, в которой находится явный представитель группы (известный априори). Такая вершина определяется как центр кластера.

Качество работы BSB-модели определяется в значительной мере емкостью автоассоциативной линейной памяти, включенной в контур обратной связи, а эта емкость, в свою очередь, существенно зависит от принятой процедуры настройки n^2 синаптических весов адаптивных линейных ассоциаторов. Для настройки синаптических весов данной нейросетевой модели был использован алгоритм Уидроу—Хоффа [13], который обеспечивает достаточное быстроедействие и обладает необходимыми следящими свойствами для решения задачи:

$$W(k + 1) = W(k) + \eta(x(k + 1) - W(k)x(k + 1))x^T(k + 1),$$

где η — параметр шага обучения.

В результате обработки данных пониженной размерности с помощью нечеткой BSB-нейромодели было выявлено шесть устойчивых кластеров, соответствующих различным состояниям организма.

4. Понижение размерности входного пространства с помощью метода главных компонент и кластеризация данных fcm-методом

Поскольку, как было замечено выше, исходная размерность пространства избыточна для решения задачи кластеризации из-за наличия неинформативных признаков, то для сокращения входного пространства был использован метод главных компонент

(PCA — Principal component analysis) [14, 15]. Этот метод применяется в тех случаях, когда приходится сталкиваться с ситуациями, в которых общее число признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ очень велико и, тем не менее,

многомерные наблюдения $X_i = \begin{pmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(p)} \end{pmatrix}, i = 1,$

2, ..., n следует подвергнуть обработке.

Каждое из наблюдений представляется в виде вектора Z некоторых вспомогательных показателей $z^{(1)}, z^{(2)}, \dots, z^{(p')}$ ($p' \ll p$). Формально задача перехода к новому набору признаков $\tilde{z}^{(1)}, \tilde{z}^{(2)}, \dots, \tilde{z}^{(p')}$ заключается в определении такого набора признаков \tilde{Z} , найденного в классе F допустимых преобразований исходных показателей $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, чтобы

$$I_{p'}(\tilde{Z}(X)) = \max_{Z \in F} \{I_{p'}(Z(X))\},$$

где $Z(X) = (z^{(1)}(X), \dots, z^{(p')}(X))$.

В качестве класса допустимых преобразований F определяются всевозможные линейные ортогональные нормированные комбинации исходных показателей, т. е.

$$z^{(j)}(X) = c_{j1}x^{(1)} - \mu^{(1)} + \dots + c_{jp}x^{(p)} - \mu^{(p)};$$

$$\sum_{v=1}^p c_{jv}^2 = 1, j = 1, 2, \dots, p';$$

$$\sum_{v=1}^p c_{jv}c_{kv} = 0, j, k = 1, 2, \dots, p'; j \neq k;$$

(здесь $\mu^{(v)} = Ex^{(v)}$ — математическое ожидание

$x^{(v)}$); $I_{p'}(\tilde{Z}(X)) = \frac{Dz^{(1)} + \dots + Dz^{(p')}}{Dx^{(1)} + \dots + Dx^{(p)}}$ — мера информативности p' -мерной системы признаков

$Z(X) = (z^{(1)}(X), \dots, z^{(p')}(X))$ (здесь D — знак операции вычисления дисперсии соответствующей случайной величины).

Для обработки данных было предложено использовать FCM-алгоритм (Fuzzy C-Means Algorithm). Это один из алгоритмов нечеткой кластеризации данных [14, 16].

Исходной информацией является выборка наблюдений, сформированная из Nn -мерных векторов признаков $X = \{(x(1), x(2), \dots, x(N))\}, x(k) \in X, k = 1, 2, \dots, N$. Результат работы алгоритма представляет собой разбиение исходного массива данных на m классов с некоторым уровнем $w_j(k)$ принадлежности k -го вектора признаков j -му кластеру.

Поступающие на обработку данные предварительно центрируются и стандартизируются по всем признакам так, чтобы все наблюдения принадлежали гиперкубу $[-1, 1]^n$. Центрирование осуществляется либо относительно среднего, вычисляемого с помощью соотношения

$$m_i(k) = m_i(k-1) + \frac{1}{k}(x_i(k) - m_i(k-1)),$$

либо с целью придания процедуре центрирования робастных свойств (защита от аномальных наблюдений) оно осуществляется относительно медианы, которая вычисляется согласно рекуррентному соотношению

$$me_i(k) = me_i(k-1) + \eta_m \operatorname{sign}(x_i(k) - me_i(k-1)),$$

$$i = 1, 2, \dots, n,$$

где η_m — параметр шага поиска, выбираемый в стационарном случае в соответствии с условиями Дворецкого.

Алгоритмы, основанные на целевых функциях [17], предназначены для решения задачи кластеризации путем оптимизации некоторого наперед заданного критерия качества кластеризации и являются наиболее строгими с математической точки зрения.

Целевая функция, подлежащая минимизации, имеет вид

$$E(w_j(k), c_j) = \sum_{k=1}^N \sum_{j=1}^m w_j^\beta(k) d^2(x(k), c_j)$$

при ограничениях

$$\sum_{j=1}^m w_j(k) = 1, k = 1, \dots, N;$$

$$0 < \sum_{k=1}^N w_j(k) < N, j = 1, \dots, m.$$

Здесь $w_j(k) \in [0, 1]$ — уровень принадлежности вектора $x(k)$ к j -му кластеру, c_j — прототип (центр) j -го кластера. β — неотрицательный параметр, именуемый «фазификатором» (обычно принимается равным 2), $d^2(x(k), c_j)$ — расстояние между $x(k)$ и c_j в принятой метрике.

В результате кластеризации получаем $N \times m$ матрицу $W = \{w_j(k)\}$, называемую *матрицей нечеткого разбиения*.

Поскольку элементы матрицы W могут рассматриваться как вероятности принадлежности векторов данных определенным кластерам, то такие процедуры называются *вероятностными алгоритмами кластеризации*.

Введем функцию Лагранжа

$$L(w_j(k), c_j, \lambda(k)) = \sum_{k=1}^N \sum_{j=1}^m w_j^\beta(k) d^2(x(k), c_j) +$$

$$+ \sum_{k=1}^N \lambda(k) \left(\sum_{j=1}^m w_j(k) - 1 \right) =$$

$$= \sum_{k=1}^N \left(\sum_{j=1}^m w_j^\beta(k) d^2(x(k), c_j) + \lambda(k) \left(\sum_{j=1}^m w_j(k) - 1 \right) \right),$$

где $\lambda(k)$ — неопределенный множитель Лагранжа.

Решая систему уравнений Куна—Таккера

$$\begin{cases} \frac{\partial L(w_j(k), c_j, \lambda(k))}{\partial w_j(k)} = 0, \\ \nabla_{c_j} L(w_j(k), c_j, \lambda(k)) = 0, \\ \frac{\partial L(w_j(k), c_j, \lambda(k))}{\partial \lambda(k)} = 0, \end{cases}$$

получим решение в виде

$$w_j^{pr}(k) = \frac{(d^2(x(k), c_j))^{1-\beta}}{\sum_{l=1}^m (d^2(x(k), c_l))^{1-\beta}},$$

$$c_j^{pr} = \frac{\sum_{k=1}^N w_j^\beta(k) x(k)}{\sum_{k=1}^N w_j^\beta(k)},$$

$$\lambda(k) = - \left(\sum_{l=1}^m (\beta d^2(x(k), c_l))^{1-\beta} \right)^{\frac{1}{1-\beta}}.$$

Выбирая $\beta = 2$ и принимая евклидово расстояние $d^2(x(k), c_j) = \|x(k) - c_j\|^2$, получаем алгоритм нечеткой кластеризации Бездека [17], называемый еще FCM-алгоритмом (Fuzzy C-Means Algorithm) [14], который и использовался в нашей работе:

$$w_j^{pr}(k) = \frac{\|x(k) - c_j\|^{-2}}{\sum_{l=1}^m \|x(k) - c_l\|^{-2}};$$

$$c_j^{pr} = \frac{\sum_{k=1}^N w_j^2(k) x(k)}{\sum_{k=1}^N w_j^2(k)};$$

$$\lambda(k) = - \sum_{l=1}^m \left(\frac{\|x(k) - c_l\|^{-2}}{2} \right)^{-1}.$$

При работе алгоритма на заранее сжатой выборке было получено 6 устойчивых классов.

5. Выводы

Результатом кластеризации сжатых данных с помощью рассмотренных выше моделей является разделение объектов на группы таким образом, что можно четко определить принадлежность каждого объекта только к одному классу.

Использование новых перспективных методов при обработке полученных данных показали высокую степень достоверности полученных результатов, что дает основание использовать такие методы в дальнейших исследованиях.

Список литературы: 1. Бурханов А. И., Муценко Т. А. Характеристика функции внешнего дыхания у студентов // Гигиена и санитария. – 1997. – № 2. – С. 32–34. 2. Залимов Р. Ю. Специфика адаптации студентов к условиям образовательного процесса и результативность их учебной деятельности в зависимости от состояния физиологических функций и личностных способностей. Физиологические основы здоровья студентов // Тр. МНС по экспериментальной и прикладной физиологии / Под редакцией К. В. Судакова. – М.: НИИИФ им. П. К. Анохина РАМН, 2001. – 10. – С. 69–83. 3. Кашалев М. А., Лисицин В. И., Возженкина Р. В. Состояние здоровья студентов медицинского института // Здоровоохранение Казахстана. – 1992. – № 1. – С. 65–67. 4. Самохвалов В. Г., Самохвалов А. В. Динамика психологической и физиологической адаптации студентов к учебным нагрузкам. Физиологические основы здоровья студентов // Тр. МНС по экспериментальной и прикладной физиологии / Под редакцией К. В. Судакова. – М.: НИИИФ им. П. К. Анохина РАМН, 2001. – 10. – С. 84–106. 5. Филатов О. М., Шедрина А. Г. Роль индивидуальной изменчивости организма в формировании здоровья студентов // Гигиена и санитария. – 1996. – № 6. – С. 29–32. 6. Аль Табак М. Особенности влияния сподученої електростимуляції скелетних м'язів на функціональний стан організму спортсменів // Практична медицина. – 1999. – № 1–2. – С. 92–96. 7. Босенко А. И., Белинова А. Г., Цонева Т. Н., Годына О. В. Оценка резервных возможностей дыхания, кардио- и гемодинамики юных спортсменов // Гигиена и санитария. – 1995. – № 2. – С. 20–22. 8. Антоненко П. А., Литвинов В. И. Автоматизированная система врачебного контроля функционального состояния организма спортсмена // Теория и практика физической культуры. – 1983. – № 1. – С. 42–44. 9. Тесленко Н. А. Алгоритм обучения автоассоциативной искусственной многослойной нейронной сети // Бионика интеллекта. – 2004. – Вып. 61(1). – С. 103–106. 10. Anderson J. A. Cognitive and psychological computation with neural models // IEEE Trans. on Systems, Man, and Cybernetics. – 1983. – 13. – P. 799–815. 11. Anderson J. A., Silverstein J. W., Ritz S. A., Jones R. S. Distinctive features. Categorical perception and probability learning: Some applications of a neural model // Neurocomputing: Foundations of Research / Ed. by J. A. Anderson, E. Rosenfeld. – Cambridge, MA: MIT Press, 1988. – P. 413–451. 12. Боянский Е. В., Тесленко Н. А., Чурюмова И. Г. Нечеткая BSB-нейромодель для обработки биомедицинской информации. Интеллектуальные системы принятия решений и прикладные аспекты информационных технологий // М-лы научно-практической конференции: Т. 2. – Херсон: Изд-во Херсонского морского института, 2006. – С. 184–187. 13. Боянский Е. В., Колодажный В. В., Тесленко Н. А. Алгоритм обучения модифицированной BSB-модели в задачах кластеризации // М-лы 12-й Международной конференции по автоматическому управлению «Автоматика-2005». – Харьков: НТУ «ХПИ». – 2005. – Т. 1. – С. 49–50. 14. Nelles O. Nonlinear System Identification: from classical approaches to neural networks and fuzzy models. – Springer, 2001. 15. Айвазян С. А., Бухитабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 607 с. 16. Боянский Е. В., Горшков Е. В. и др. Об адаптивном алгоритме нечеткой кластеризации данных // АСАУ. – 2002. – 5(25). 17. Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms. – N. Y.: Plenum Press, 1981. – 272 p.

Поступила в редакцию 16.03.2006

УДК 681.51:007.5

В. А. Гороховатский, С. В. Кузьмин

ИССЛЕДОВАНИЕ ДЕТЕКТОРОВ ТОЧЕК ИНТЕРЕСА ПРИ ПОСТРОЕНИИ ЛОКАЛЬНЫХ ПРИЗНАКОВ ИЗОБРАЖЕНИЙ

1. Введение

Построение эффективных систем локальных признаков при распознавании изображений является важнейшей проблемой современного компьютерного зрения [1-4]. Локальные признаки используются в структурных методах распознавания, которые обладают универсальностью и гибкостью применения в затрудненных условиях при действии помех и сложного фона [5, 6]. Распознавание изображений при этом сводится к сопоставлению их представлений в пространстве локальных признаков.

Под «точками интереса» в системах компьютерного зрения понимают существенные для процесса распознавания двумерные изменения сигнала [1, 4]. Реально на изображениях это соответствует углам, контрастным и изолированным точкам, концам кривых, линий и т. д. Эти локальные признаки при геометрических преобразованиях изображений, как правило, почти не претерпевают изменений. Кроме того, множество точек интереса по мощности значительно меньше часто применяемого множества краевых (граничных) точек, поэтому представляет интерес при построении эффективных методов структурного анализа и распознавания изображений [2]. Как правило, точки интереса используются в качестве отправного момента для построения в их окрестности более сложной векторной системы признаков [6], хотя имеют и самостоятельное применение [2, 4].

Целью работы является изучение свойств, особенностей, формализация, выбор параметров и компьютерное моделирование детекторов точек интереса для реальных изображений.

Задачи исследования — формализация описания признаков точек интереса, анализ вариантов их построения, применение их для решения проблемы представления изображений в признаковых пространствах, исследование статистических, инвариантных и вычислительных свойств этих признаков.

2. Формальное описание и пути построения детекторов точек интереса

Одним из путей построения признаков точек интереса является анализ автокорреляционной функции изображения, которая в значительной мере отражает локальные особенности сигнала. Автокорреляционная функция в дискретном варианте имеет следующий вид:

$$f(x, y, \Delta x, \Delta y) = \sum_{(x_k, y_k) \in W(x, y)} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2, \quad (1)$$

где $I(x, y)$ — функция яркости изображения, (x_k, y_k) — пиксели окна W с центром в точке (x, y) . Обычно окно $W(x, y)$ — это окрестность 5×5 с центром в точке (x, y) , по множеству пикселей которой вычисляется значение $f(x, y)$. С увеличением окрестности в целом растут помехозащищенность и вместе с тем вычислительные затраты, поэтому наиболее популярны на практике маски размеров 5×5 , 7×7 .

В литературе с общих позиций анализируют три варианта классификации области $W(x, y)$ фиксированных размеров. Решение принимается на основе значений автокорреляционной функции $f(x, y)$ от функции яркости $I(x, y)$, $(x, y) \in W$, заданной на этом множестве точек.

Пусть

$$\delta_1(x, y) = \min_{\Delta x, \Delta y} f(x, y, \Delta x, \Delta y),$$

$$\delta_2(x, y) = \max_{\Delta x, \Delta y} f(x, y, \Delta x, \Delta y), \quad \delta_{\max} = \max_{x, y} \delta_2(x, y).$$

Тогда область $W(x, y)$ можно классифицировать следующим образом [4]:

- Область приблизительно однородна — минимальное и максимальное по всем $(\Delta x, \Delta y)$ значения функции $f(x, y)$ незначительны (с точностью до некоторого порога δ_0) относительно глобального максимума δ_{\max} функции, определенного по всему изображению, т. е. когда выполнены условия

$$\delta_1 < \delta_0, \quad \delta_2 < \delta_0, \quad \delta_0 \ll \delta_{\max}.$$

Здесь символ \ll понимается как «существенно меньше» (что в реальности обычно соответствует доли в 5...10 %).

- Краевая (граничная) точка — минимальный отклик не велик, а максимальный отклик значителен относительно глобального максимума, т. е.

$$\delta_1 < \delta_0, \quad \delta_2 \approx \delta_{\max}, \quad \delta_0 \ll \delta_{\max}.$$

При этом направления минимума и максимума, задаваемые параметрами Δx , Δy , должны быть перпендикулярны между собой в пределах дискретной сетки пикселей, хотя отдельные исследователи считают это условие необязательным.

- Точка интереса — минимальный и максимальный отклики по направлениям, определяемым параметрами Δx , Δy , значительны относительно величины глобального максимума, т. е.

$$\delta_1 \approx \delta_{\max}, \delta_2 \approx \delta_{\max}. \quad (2)$$

Таким образом, точками интереса считаются пиксели изображения, полученные путем анализа множества откликов автокорреляционной функции в окрестности $W(x, y)$ по всем направлениям (в смысле 4- или 8-связности), причем оба оптимума превышают заданный порог $\delta \approx \delta_{\max}$.

В основе классификации на три типа лежит фундаментальное понятие кривизны поверхности, которую формирует функция яркости двух переменных в окрестности заданной точки [1].

Заметим, что классификация точек изображения на три указанных типа полностью определяется глобальными порогами $\delta, \delta_0, \delta_{\max}$. Реально все множество значений функции $f(x, y)$ разбивается на три области: $[0, \delta_0], (\delta_0, \delta], (\delta, \delta_{\max}]$. Пороги определяются, в основном, экспериментальным путем и зависят от контраста, освещенности и действия шумов, хотя при некоторых условиях существуют и оптимальные пути выбора порогов.

В результате анализа в соответствии с условиями (2) формируется множество точек интереса в виде

$$T = \{x, y, t \mid t = f(x, y, \Delta x, \Delta y), f(x, y, \Delta x, \Delta y) > \delta.$$

Затем сформированное множество T традиционно подвергается действию процедуры U отсеивания точек с целью сокращения количества пространственно группирующихся однотипных признаков [1]. Дело в том, что точки интереса на реальных изображениях имеют свойство группироваться в окрестности выделенного признака. Это связано со свойством гладкости автокорреляционной функции. В результате обработки имеем множество

$$T^* = U(T, \Omega),$$

где Ω — пространственная область, которой принадлежат точки из T , на основе анализа координат которых формируются точки множества T^* . Стандартная обработка сводится к анализу прямоугольной области изображения $\Omega = 7 \times 7$ точек с выбором максимальной по значению отклика из попавших в эту область точек. Увеличение размеров области Ω уменьшает количество точек интереса. Сформированный таким образом максимальный отклик приписывается координатам центра области Ω , а остальные точки окрестности выбрасываются из дальнейшего рассмотрения. Такая обработка получила название «немаксимального подавления значений» [4].

На базе анализа автокорреляционной функции построен ряд схем обработки [2, 3]. Например, детектор Моравица основан на поиске локального минимума в каждой точке среди откликов автокорреляционной функции по всем направлениям, опре-

деляемым вектором сдвига $(\Delta x, \Delta y)$. В дальнейшем минимумы используются для принятия решения о точке интереса с использованием заданного глобального порога. Таким образом, здесь множество точек интереса формируется в виде

$$M = \{x, y, t \mid t = \delta_1(x, y), \delta_1 > \delta.$$

Обычно порог δ задается как величина, пропорциональная значению максимального отклика детектора на всем множестве точек изображения. Диапазон изменения откликов функции (1) зависит от контраста изображения и для области $W(x, y)$ размером 5×5 изменяется в пределах от 0 до $5 * 255$.

По сути, автокорреляционная функция аппроксимирует производную функции яркости по направлению, задаваемому параметрами $(\Delta x, \Delta y)$. Поэтому в детекторе Харриса [1] анализ автокорреляционной функции в точке (x, y) сводят к рассмотрению матрицы вида:

$$A(x, y) = \begin{bmatrix} \sum_{(x_k, y_k) \in W} (I_x(x_k, y_k))^2 & \sum_{(x_k, y_k) \in W} I_x(x_k, y_k) I_y(x_k, y_k) \\ \sum_{(x_k, y_k) \in W} I_x(x_k, y_k) I_y(x_k, y_k) & \sum_{(x_k, y_k) \in W} (I_y(x_k, y_k))^2 \end{bmatrix},$$

где I_x, I_y — частные производные функции изображения I в точке (x_k, y_k) .

С целью сглаживания помех суммирование значений частных производных в матрице $A(x, y)$ реализуется через свертку с матрицей гауссиана. Один из вариантов такой матрицы для сглаживания имеет вид

$$\begin{bmatrix} 0,779 & 0,882 & 0,779 \\ 0,882 & 1 & 0,882 \\ 0,769 & 0,882 & 0,779 \end{bmatrix}.$$

Кривизна автокорреляционной функции изображения определяется собственными значениями матрицы $A(x, y)$, которые в общем случае отражают понятия главной кривизны поверхности, что позволяет учитывать направления изменения значений кривизны [2].

Реализация детектора Харриса в результате сводится к анализу величины R , связанной с собственными значениями и построенной на основе взаимосвязи следа и определителя матрицы $A(x, y)$:

$$R = \text{Det} -\beta \text{Tr}^2,$$

где Tr — след (сумма элементов главной диагонали); Det — детерминант (определитель) матрицы; β — параметр, задающий свойства (классификацию) выделенных точек и фактически определяющий их количество в связи с пороговой обработкой величины R , $\beta \in [0, 1]$.

Разработано большое количество методов выделения краев (границ), которые также опираются на информацию о производных от изображений. Основные принципы их построения сводятся к анализу связности пикселей или построению параметрических моделей для аппроксимации участков кривых. Для выделения краев, например, в настоящее время популярен детектор Кэнни [4], использующий на первом этапе значение и направление градиента, вычисленные с помощью оператора Робертса. Точки интереса относятся детектором Кэнни ко множеству краевых точек путем сопоставления значения градиента с двумя порогами — на минимум возможного отклика и на обнаружение дополнительных точек-кандидатов, включаемых во множество пикселей формируемого края. На базе этих порогов и направлений градиентов соседних связанных точек можно формировать также и точки интереса. Однако, как показали эксперименты [1], краевые методы несколько проигрывают детектору Харриса по критерию инвариантности к группе движений.

Другое направление построения точек интереса связано с оценкой параметров аналитически заданных прямых и кривых на сформированных бинарных сигналах, полученных при обработке изображения. Его ограничением является привязка к какому-либо специфическому типу точек, например, L -, T -, X -, Y -, \uparrow - углам [1].

3. Результаты компьютерных экспериментов

Целью моделирования был анализ целесообразности и эффективности применения детекторов точек интереса к решению задач распознавания реальных изображений. Распознавание в такой постановке сводится к сопоставлению множеств точек интереса некоторого объекта с соответствующими множествами для эталонов.

Одно из направлений исследования состоит в оценке влияния параметра β детектора Харриса на качество и устойчивость выделения локальных признаков.

На рис. 1 приведены исходные бинарные изображения, включающие все типы классифицируемых точек, а также величины откликов при различных β , нормализованные в целях отображения к диапазону яркостей 0...255.

Проведенные эксперименты показывают, что при изменении коэффициента β в пределах от 0 до 0,1 увеличивается величина отклика R для однородных областей, резко выделяются краевые точки, уменьшается количество выделенных точек интереса, вплоть до их полного слияния с откликами однородных областей при $\beta=0,1$. Инверсия исходных изображений приводит к аналогичному результату.

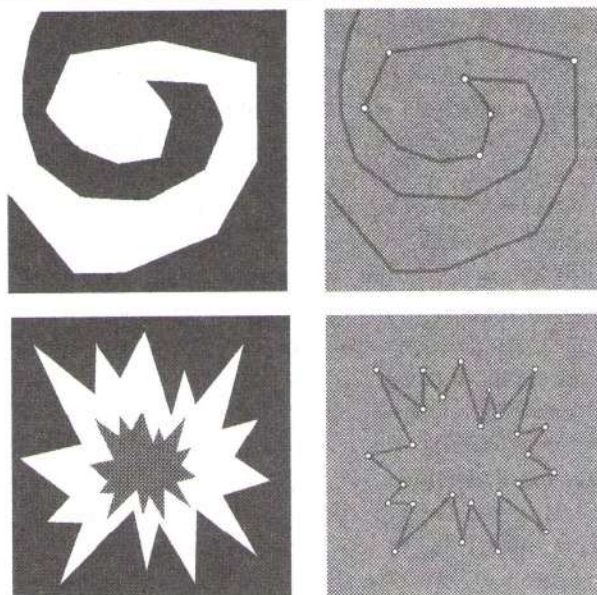


Рис. 1. Тестовые изображения и отклики на них детектора Харриса при $\beta = 0,06$

Известно, что детектор Харриса, как, впрочем, и другие детекторы, обладает инвариантностью к преобразованиям смещения (так как реализуется путем сканирования) и к повороту (поскольку анализ направлений изменения автокорреляционной функции учитывает ее повороты на 45° вместе с изображением).

С целью анализа влияния величины выделяемых углов на значение отклика R для бинарных изображений в качестве тестового использовалось изображение, которое представляет собой сектор с фиксированным значением угла (см. рис. 2). Результаты экспериментов для разных значений углов приведены в таблице 1. Согласно результатам экспериментов, наибольший отклик дают углы сектора в $30...90^\circ$, существенно меньший — $0...15^\circ$, а углы $120...165^\circ$ дают отрицательный отклик и поэтому относятся детектором к краевым точкам, а не к точкам интереса.

Приведем теперь результаты в виде гистограммы значений, где по горизонтальной оси отложены диапазоны принимаемых детектором Харриса значений, а по вертикали — количество точек в каждом

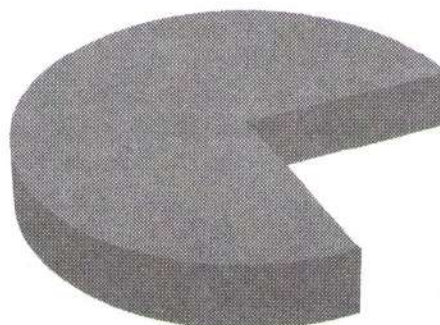


Рис. 2. Тестовое изображение сектора с фиксированным углом

Таблица 1

Значения откликов детектора Харриса для изображений углов (диапазон 15...120°, горизонталь) от величины угла поворота сектора рис. 2 (вертикаль)

| Углы | 15 | 30 | 45 | 60 | 75 | 90 | 105 | 120 |
|-----------|------|-------|-------|-------|-------|-------|-------|-------|
| 0 | 2,98 | 12,33 | 12,33 | 13,37 | 11,69 | 9,55 | 3,94 | -0,03 |
| 20 | 3,56 | 12,73 | 12,73 | 14,86 | 10,31 | 10,31 | 8,23 | -1,54 |
| 40 | 6,31 | 12,30 | 12,30 | 13,44 | 9,02 | 7,57 | 3,89 | -0,29 |
| 60 | 2,41 | 14,85 | 14,85 | 15,51 | 10,82 | 10,46 | 4,3 | -1,09 |
| 80 | 1,19 | 12,25 | 12,25 | 13,21 | 9,25 | 6,08 | 4,3 | -2,14 |
| 100 | 4,08 | 14,20 | 14,21 | 13,41 | 15,53 | 7,22 | 10,59 | -1,5 |
| 120 | 6,65 | 17,42 | 17,42 | 11,65 | 10,89 | 5,73 | 3,7 | 3,69 |
| 140 | 1,48 | 13,93 | 13,93 | 15,64 | 8,56 | 8,23 | 3,88 | 5,73 |
| 160 | 2,49 | 7,96 | 7,96 | 10,08 | 12,55 | 15,58 | 10,94 | -1,48 |
| 180 | 2,99 | 12,30 | 12,30 | 13,33 | 13,22 | 9,55 | 4,87 | 1,41 |
| Среднее | 3,41 | 13,02 | 13,02 | 13,22 | 11,18 | 9,03 | 5,86 | 0,27 |
| Дисперсия | 3,36 | 5,86 | 5,86 | 2,62 | 4,60 | 8,08 | 8,42 | 6,70 |

Таблица 2

Отклики лапласиана гауссоиды и детектора Харриса для угла 45° при разных углах поворота сектора

| Углы поворота | 0 | 20 | 40 | 60 | 80 | 100 | Среднее | Дисперсия |
|---------------------|--------|--------|--------|-------|--------|-------|----------|-----------|
| Лапласиан гауссоиды | -0,607 | -3,213 | -1,213 | -1,82 | -1,213 | -1,82 | -1,64767 | 0,79 |
| Детектор Харриса | 12,33 | 12,73 | 12,30 | 14,85 | 12,25 | 14,21 | 13,02 | 5,86 |

из них. Гистограммы (рис. 3, а) показывают, что инвариантность детектора Харриса к повороту понимается как стабильность (в некотором диапазоне) множества наибольших значений отклика. Гистограмма после удаления откликов краевых точек, соответствующих отрицательным значениям R , приобретает вид рис. 3, б. Основная масса значений здесь относится к максимальному диапазону (3...5). Поэтому обнаружение и распознавание объектов по точкам интереса можно построить путем анализа множества значений детектора, близких к максимальному значению.

Разновидностью точек интереса являются углы типа L, T, X, Y, ↑ [2, 4]. Эксперименты показали, что все они устойчиво определяются описанными детекторами.

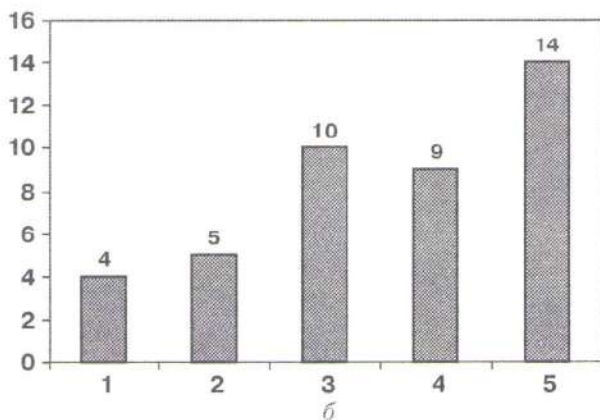
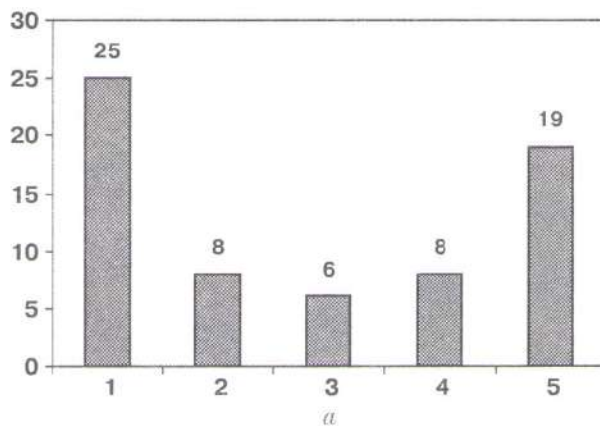


Рис. 3. Гистограммы значений детектора Харриса: а — полная, б — после удаления откликов краевых точек

Из-за особенности построения многоэтапной обработки при реализации детектора Харриса возникают «перескоки» откликов между собой в пределах группы углов, поэтому при необходимости распознавания углов разного собой нужно использовать дополнительные признаки, позволяющие различать углы разных размеров в пределах уже выделенных групп точек. Одним из популярных признаков такого рода является лапласиан гауссоиды. Для сравнения в таблице 2 приведены значения откликов, а также среднее и дисперсия лапласиана гауссоиды и детектора Харриса для угла 45° градусов. Разброс значений для лапласиана гауссоиды, как видно из таблицы 2, несколько меньше, чем у детектора Харриса.

Эксперименты на полутоновых изображениях показали следующие результаты. На рис. 4 приведены исходное изображение (рис. 4, а), а также изображение с преобразованиями смещения и поворота на 20° на сложном фоне (рис. 4, б). Доля правильно перешедших в процессе преобразований точек интереса составляет около 94 %. Как видим из рисунков, на фоне возникает некоторое дополнительное количество выделенных признаков.

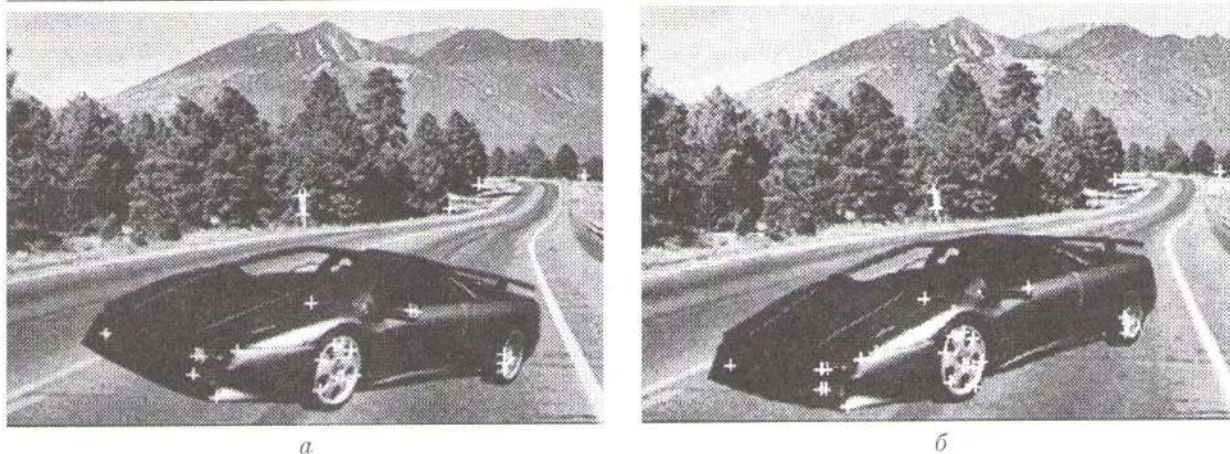


Рис. 4. Преобразование полутонового изображения: а — выделена 31 точка; б — выделены 28 точек

4. Выводы

Новизна предложенного подхода состоит в построении, формальном описании и оценке свойств эффективной системы признаков — точек интереса для целей распознавания объектов на изображениях.

Важной особенностью детектора Харриса является то, что он может успешно справляться как с детектированием углов, так и краев (в зависимости от задаваемого параметра β). Второй особенностью при детектировании углов является чувствительность детектора к их величине — изображения углов в пределах от 30° ... 90° дают существенно больший отклик, нежели остальные.

Функция R детектора Харриса принимает положительные значения в точках интереса, отрицательные — в краевых точках и близкие к нулю значения в областях изображения, которые можно приближенно считать однородными. При выделении краев принимают $\beta \geq 1$, а для точек интереса β принимает значения в диапазоне $\beta = 0,04$... $0,06$.

Практическая значимость полученных результатов заключается в экспериментальном подтверждении таких свойств детекторов, как инвариантность, устойчивость в пределах некоторого диапазона выделяемых углов, связь с классическими детекторами краев, универсальность применения для разных видов локальных признаков.

Список литературы: 1. Schmid C., Mohr R., Bauckhage C. Evaluation of interest point detectors // International Journal of Computer Vision.— 2000.— 37, No. 2.— P. 151–172.
2. Harris C. Stephens M. A combined edge and corner detector // Alvey Vision Conf.— 1988.— 4.— P. 189–192.
3. Schmid C., Mohr R. Local Greyvalue Invariants for Image Retrieval // IEEE Transactions on Pattern Analysis and Machine Intelligence.— 1997.— 19, No. 5.— P. 530–534.
4. Шапиро Л., Стокман Дж. Компьютерное зрение.— Бинном. Лаборатория знаний, 2006.— 752 с.
5. Гороховатский В. А. Распознавание изображений в условиях неполной информации.— Харьков: ХНУРЭ, 2003.— 112 с.
6. Пуплатин С. П., Гороховатский В. О., Маташ О. О. Методы та алгоритми комп'ютерного зору: Навч. посібник. — Харків: ТОВ «Компанія СМІТ», 2006.— 236 с.

Поступила в редакцию 27.08.2006

УДК 004.93.673: 61

Н. В. Белоус, Е. В. Высоцкая, Т. В. Жемчужкина, А. Н. Ковалев, Г. А. Кобзарь, А. П. Порван

ОБ ОДНОМ ПОДХОДЕ К РАСПОЗНАВАНИЮ ЭЛЕКТРОКАРДИОГРАФИЧЕСКОГО СИГНАЛА В РЕАЛЬНОМ МАСШТАБЕ ВРЕМЕНИ

1. Введение

Исследование, обнаружение и диагностика заболеваний сердечно-сосудистой системы и, в частности, заболеваний сердца являются одной из важных проблем современной медицины. Одним из ведущих методов изучения биоэлектрической активности сердца является электрокардиография. Она на сегодняшний день незаменима в диагностике инфарктов миокарда, ишемической болезни сердца, нарушений ритма и проводимости, гипертрофий предсердий и желудочков и других заболеваний сердца. Использование традиционно применяемых методов не всегда оказывается эффективным, о чем свидетельствует тот факт, что число неточных диагностических заключений до сих пор достаточно велико. Одной из сложных проблем является большой объем трудозатрат медперсонала на получение точного диагноза, который иногда ставится лишь после смерти больного.

Поток электрокардиограмм (ЭКГ), в большинстве случаев не содержащих признаков сердечно-сосудистых патологий, перегружает врачей-кардиологов, что ведет к ослаблению их внимания, переутомлению и, как следствие, к пропуску пограничных случаев или неточной дифференциальной диагностике.

Важно определить оптимальный набор структурных параметров электрокардиографического сигнала, обеспечивающих наиболее достоверную диагностику, а также разработать методы выделения и обработки характерных структур регистрируемого сигнала [1].

При визуальном анализе ЭКГ часто не принимаются во внимание или не замечаются его незначительные изменения, в которых может содержаться информация о важных нюансах в работе сердца. Эти незаметные на глаз отклонения от нормы могут развиваться в патологии, поэтому автоматизированный анализ ЭКГ является основным направлением в развитии кардиологических систем. Такие системы предназначены для выделения электрокардиосигналов (ЭКС) в норме и подсказки врачу-кардиологу наиболее вероятных кардиологических синдромов, содержащихся в пограничной и патологических группах ЭКС.

В настоящее время известно большое число алгоритмов автоматического анализа ЭКГ, разнообразие которых определяется:

- спецификой задачи (диагностика, мониторинг, слежение за состоянием пациента, оценка состояния оператора, профилактическое обследование и т. п.);

- спецификой анализа ЭКГ, регистрируемой в покое или при нагрузке;
- использованием различных систем отведений и методов получения вторичного описания ЭКГ и его диагностической интерпретации;
- способами реализации алгоритма (программой для ЭВМ или в виде специализированного вычислительного устройства).

Специфика задач, решаемых при анализе ЭКГ, накладывает свой отпечаток прежде всего на сложность возможных алгоритмов обработки.

При мониторингном слежении важна оперативность получения информации, что ведет к упрощению методов анализа для повышения их быстродействия. Для функциональной диагностики главное — точность и достоверность результатов, поэтому здесь возможно применение очень сложных алгоритмов анализа.

Анализаторы ЭКС кардиологических систем строятся на базе различных методов автоматического анализа ЭКГ. Большинство таких методов являются производными мануальных методов, которые используют кардиологи во врачебной практике в отсутствие автоматизированных компьютерных кардиологических систем (структурные методы). Врачебные признаки характеризуют амплитуду и длительность зубцов ЭКС, полярность, скорость изменения, расщепление, интервалы времени между зубцами. Такие методы обладают рядом достоинств: простота реализации, быстродействие, наглядность. Основным недостатком этих методов — неадекватное использование вычислительной мощности современной компьютерной техники.

В данной статье предлагается подход к анализу ЭКС в целом с использованием современных методов быстрого распознавания сигналов по форме. Этот подход предусматривает архивацию эталонов ЭКС в определенном формате, а также быструю идентификацию участков ЭКС по созданной базе эталонов.

2. Анализ существующих методов

Основой для построения алгоритмического и программного обеспечения большинства автоматизированных систем служит следующая последовательность этапов обработки ЭКГ-сигнала: ввод ЭКГ; предварительная обработка сигнала; распознавание характерных элементов; определение информативных параметров и их анализ; интерпретация результатов анализа.

Ввод ЭКГ подразумевает дискретизацию и квантование сигнала, предобработка — цифровую фильтрацию для устранения шумов и помех. Наиболее важным этапом обработки сигнала ЭКГ является распознавание элементов сигнала (QRS-комплекса) и выявление его характерных точек (вершин и границ зубцов), что дает возможность определить важные характеристики ЭКГ-сигнала — амплитуды и длительности всех зубцов и сегментов. При анализе ЭКГ первостепенный интерес для врача представляют временная эволюция и статистика распределения структурных параметров: временных интервалов и амплитуд. Поэтому важную группу составляют методы, основанные на автоматическом определении параметров, используемых врачом при расшифровке ЭКГ. К ним относятся значения амплитуд и длительностей зубцов ЭКГ, а также интервалов между ними.

На сегодняшний день разработано большое количество разнообразных методов автоматизированного анализа ЭКГ [2]. Большое разнообразие существующих методов обусловлено как различием решаемых задач, так и спецификой исследуемых параметров сигнала. Можно выделить три основные группы методов анализа ЭКС:

- методы, основанные на анализе производной сигнала, использующие пороговые значения производной для каждого из зубцов ЭКГ, а также временные маски (временной интервал, в пределах которого ожидается появление зубца);
- структурные методы, основанные на предварительной сегментации сигнала с представлением его в виде последовательности простейших элементов и последующим грамматическим разбором получаемой цепочки символов;
- корреляционные методы, использующие анализ корреляционной функции между ЭКГ-сигналом и образцами желудочкового комплекса.

Достоинством первых двух групп методов является простота реализации, недостатком является разбиение кардиосигнала на отдельные участки.

Корреляционные методы анализируют кардиосигнал в целом, что является достоинством этого метода, однако при таком подходе количество вычислительных операций при сравнении анализируемого кардиосигнала с набором эталонных форм кардиоциклов значительно возрастает. Таким образом, актуальными являются задачи оптимизации размера эталонов и разработки методов идентификации входного сигнала по базе эталонов с целью построения системы распознавания ЭКС в реальном времени.

3. Распознавание ЭКС как единого целого

Предлагается использовать непосредственное распознавание участков ЭКС с помощью методов распознавания по форме. Такой подход предполагает рассмотрение сигнала как единого целого, но при этом позволит оценить его локальные характеристи-

ки, избегая проблем зашумления в большинстве случаев.

Одной из основных задач при реализации предложенного подхода непосредственного распознавания ЭКГ сигнала по форме является разработка информативного представления формы (или дескриптора) сигнала. К настоящему времени разработаны три дескриптора формы, которые в наибольшей степени удовлетворяют требованиям устойчивости, компактности, вычислительной простоте извлечения: дескрипторы Фурье (FD — Fourier descriptors), дескрипторы скалярного пространства кривизны (CSSD — curvature scale space descriptors), а также сеточные дескрипторы (GD — grid descriptors).

В целом, существуют только два типа дескрипторов формы: контурные и областные. FD и CSSD являются контурными представлениями, а GD — областные. Областные дескрипторы не подходят для распознавания сигнала, так как подразумевают оценку замкнутой формы. Однако GD метод можно применить к сигналу, рассматривая область, ограниченную сигналом сверху, как объект для распознавания.

Таким образом, для обеспечения распознавания интервалов ЭКГ сигналов необходимо создать базу эталонов, представленных в соответствии с одним из существующих дескрипторов. Базу эталонов можно сформировать, используя существующие таблицы ЭКГ, открытые базы ЭКГ в сети Интернет (Германия, Япония, США) [3].

Рассмотрим и проведем анализ перечисленных дескрипторов на предмет их применения в качестве базы для распознавания ЭКС.

4. Разработка эффективного подхода представления и распознавания ЭКС

Дескрипторы Фурье вычисляются с помощью преобразования Фурье, применяемого к преобразованному сигналу (первая и вторая производная, кривизна и др.; в качестве входных данных можно использовать и исходный сигнал). Полученные коэффициенты Фурье называют дескрипторами Фурье. Чтобы обеспечить применение преобразования Фурье, необходимо нормировать количество точек дискретного сигнала. Дискретное преобразование Фурье последовательности точек будет иметь следующий вид:

$$u_n = \frac{1}{N} \sum_{i=0}^{N-1} r_i \exp\left(\frac{-j2\pi ni}{N}\right), \quad n=0, 1, \dots, N-1, \quad (1)$$

где N — количество точек последовательности, r_i — значения в точках последовательности.

Коэффициенты u_n , $n=0, 1, \dots, N-1$, обычно называют дескрипторами Фурье (ДФ) формы и записывают в виде FD_n , $n=0, 1, \dots, N-1$ [4].

Для оценки сходства анализируемой формы и целевой формы из базы данных используется простое евклидово расстояние между векторами признаков анализируемого и эталонного сигнала. Преобразо-

вание Фурье позволяет ориентироваться как на мелкие детали формы, так и структуру формы в целом, в зависимости от того, какого порядка коэффициенты используются. Для дескрипторов Фурье не существует проблемы шумов — их воздействие отражается только на высоких частотах, которые чаще всего отбрасываются. Небольшие изменения контуров формы незначительно отражаются на виде конечного представления. Благодаря применению быстрого преобразования Фурье (FFT), процесс извлечения дескрипторов очень эффективен. Недостатком дескрипторов Фурье является невозможность оценки локальных характеристик формы, так как известны только абсолютные величины частот, а их положение не известно.

5. Дескриптор CSS

Дескрипторы CSS (curvature scale space — скалярное пространство кривизны) предназначены для представления локальных ключевых характеристик формы. Представление сигнала в скалярном пространстве позволяет определять не только положения вогнутостей (или выпуклостей) линии сигнала, но и степень искривления. Первым шагом извлечения дескриптора CSS является построение карты CSS. Для построения карты CSS вычисляется кривизна точек формы $(x_i, y_i), i = 1, 2, \dots, L$:

$$k_i = (\dot{x}_i \ddot{y}_i - \ddot{x}_i \dot{y}_i) / (\dot{x}_i^2 + \dot{y}_i^2)^{3/2}, \quad (2)$$

где \dot{x}_i, \dot{y}_i и \ddot{x}_i, \ddot{y}_i — первые и вторые производные в точках сигнала с индексом i соответственно.

Затем определяются точки перегиба. После этого сигнал подвергается гауссовому сглаживанию для перехода к следующему измерению (итерации). Этот процесс продолжается до тех пор, пока можно найти точки перегиба. Результирующая карта CSS представляет собой композицию всех точек перегиба $z_c(i, \sigma)$, где i — одномерная координата точки сигнала, σ — итерация, на которой z_c была найдена. Затем извлекаются пики полученной карты (не ниже некоторого порогового значения), которые записываются в убывающем порядке и таким образом формируют дескриптор CSS, позволяющий идентифицировать сигнал по форме [5].

Сходство двух форм измеряется с помощью суммы разностей всех совпадающих пиков и значений несовпадающих пиков. Достоинством дескрипторов CSS является способность оценки локальных ключевых признаков, таких как положение и степень вогнутости (выпуклости) сегментов линии сигнала. Эти признаки играют важную роль в человеческом восприятии при оценке сходства двух форм. Размер дескрипторов CSS невелик, и представление является устойчивым благодаря применению гауссового сглаживания. Однако дескрипторы CSS описывают только локальные ключевые признаки. Глобальные признаки, которые также очень важны для объек-

тивного представления форм, не учитываются. Поэтому такие глобальные признаки, как изменение среднего значения сигнала, должны быть включены в набор CSS для получения более объективного представления.

6. Сетчатые дескрипторы

Для получения сетчатого представления сигнал проектируется на квадратичную сетку фиксированного размера. Ячейки сетки, лежащие под линией сигнала, получают значение 1, 0 — если лежат выше. Индекс сигнала, представленный бинарной последовательностью, генерируется последовательным сканированием сетки слева направо и сверху вниз. Эта бинарная последовательность используется в качестве дескриптора формы.

Расстояние между двумя наборами сетчатых дескрипторов измеряется просто количеством отличающихся элементов. Это наиболее простой с вычислительной стороны метод распознавания, однако наименее устойчивый к помехам. Кроме того, наложение низкочастотных колебаний большой амплитуды на сигнал ЭКГ (что нередко происходит) влечет за собой кардинальное изменение дескриптора.

7. Построение кардиологической системы на основе предложенного подхода.

Ограничения подхода и их устранение

Из приведенного выше анализа достоинств и недостатков дескрипторов формы применительно к задаче распознавания сигнала ЭКГ по форме следует, что наиболее помехоустойчивым является метод распознавания, основанный на дескрипторе Фурье. Однако данный дескриптор не отражает сугубо локальные изменения сигнала, которые имеют критическое значение при анализе ЭКГ, так что дескриптор Фурье может быть использован на начальной стадии при распознавании сигнала ЭКГ для определения группы (широкого класса) сигнала. Выделенный недостаток устраняется последующим уточнением классификации сигнала с помощью CSS дескриптора. Применение сеточного дескриптора для анализа ЭКГ не представляется эффективным.

Недостатком предлагаемого подхода является необходимость точного выделения RR интервалов ЭКС на предварительном этапе, что не всегда возможно. Неточное выделение участка ЭКС, ограниченного R зубцами, повлечет за собой ошибочную идентификацию и некорректный диагноз. Этот недостаток можно устранить, включив в этап предварительной обработки контроль вариабельности сердечного ритма (к настоящему времени разработано множество методов оценки), а также усреднение сигнала по N участкам [6]. Число N необходимо в дальнейшем определить эмпирическим путем.

На рис. 1 представлена предлагаемая схема поэтапной идентификации ЭКС.

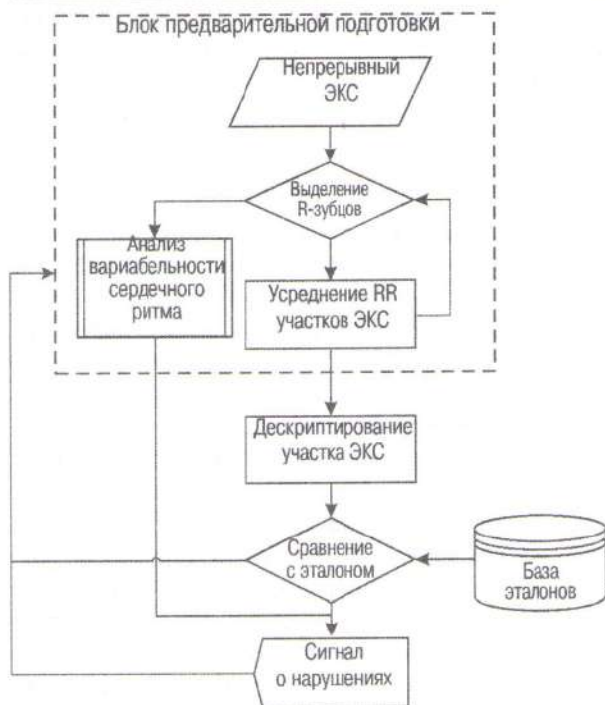


Рис. 1. Схема идентификации ЭКС

Локализация *R* зубцов осуществляется стандартными средствами (анализ производной сигнала, оценка кривизны). Для анализа variability сердечного ритма также предлагается использование новейших методов. Описание этих двух этапов выходит за рамки тематики данной статьи.

Усреднение участков ЭКС осуществляется с целью устранения шумов. Внештатные ситуации (экстрасистолы, нарушения сердечного ритма, связанные с выпадением QRS комплекса, и др.) рассматриваются отдельно на этапе анализа variability сердечного ритма.

Под дескрипированием подразумевается формирование вектора признаков входного участка ЭКС согласно выбранным дескрипторам CSS и Фурье. На этапе сравнения происходит сопоставление полученного вектора признаков с эталонами базы ЭКС для распознавания ЭКС.

По результатам анализа variability сердечного ритма и распознавания сигнала с помощью ба-

зы эталонов выдается сигнал о нормальном состоянии пациента или каких-либо нарушениях с указанием этих нарушений.

8. Выводы

В данной статье рассмотрена проблема распознавания ЭКС в реальном масштабе времени. Большинство существующих кардиологических систем основаны на подходе к анализу ЭКГ, который применяют кардиологи в повседневной практике. Этот подход не в полной мере использует вычислительную мощность и аналитические возможности современных компьютерных технологий. Предлагается подход к распознаванию участков ЭКС, ограниченного RR интервалом, как единого целого без какого-либо дополнительного деления сигнала. Этот подход отличается от корреляционного тем, что эталоны базы распознавания представляют собой векторы признаков эталонных ЭКС, извлеченных автоматически в соответствии с определенными дескрипторами формы сигнала (CSS и Фурье), что позволяет значительно уменьшить объем базы эталонов и, соответственно, ускорить процесс распознавания.

Список литературы: 1. Горбунова Е. В. Исследование и разработка методов оценивания контролируемых параметров сигналов в автоматизированных кардиологических комплексах. — Барнаул: АлтГТУ, 1999. 2. Garvey J. Lee. ECG Techniques and Technologies // Emerg. Med. Clin. — 2006. — No. Am 24. — P. 209–225. 3. PhysioNet the research resource for complex physiologic signals. — <http://physionet.ph.biu.ac.il/>, 2006. 4. Zhang D. S., Lu G. J. Shape Retrieval Using Fourier Descriptors // Int. Conf. on Multimedia and Distance Education, Fargo, ND, USA, June 2001. 5. Mokhtarian F., Abbasi S., Kittler J. Efficient and Robust Retrieval by Shape Content through Curvature Scale Space // Int. Workshop on Image Databases and Multimedia Search. — Amsterdam, 1996. — P. 35–42. 6. Москаленко А. В., Кукушкин Н. И., Стармер Ч. Ф. и др. Количественный анализ variability электрокардиограмм, типичных для полиморфных аритмий // Биофизика. — 2001. — Т. 46, вып. 2. — С. 319–329.

Поступила в редколлегию 16.08.2006

УДК 681.513:519.7

Н. Е. Кулишова

ИНТЕРПОЛЯЦИЯ КОЭФФИЦИЕНТОВ ОТРАЖЕНИЯ КРАСОК С ПОМОЩЬЮ РАДИАЛЬНО-БАЗИСНОЙ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ

1. Введение

Синтез многокрасочной полиграфической печати использует три базовые краски: голубую, пурпурную и желтую (ГПЖ, в англоязычной литературе — СМУ). При наложении на бумагу они последовательно или одновременно поглощают из внешнего освещения соответственно красную, зеленую и синюю спектральные составляющие. При этом базовые краски составляют линейную комбинацию. Количество каждой базовой краски определяется в ходе цветоделения изображения и влияет на формируемый цвет оттиска.

Подобный принцип позволяет получать на оттисках большинство цветов из всех доступных для восприятия человека, определение этого базового набора является основной проблемой при синтезе цвета. Реальный полиграфический синтез цвета использует не три, а четыре базовые составляющие — к желтой, пурпурной и голубой добавляют черную краску, количество которой линейно связано с количеством трех остальных. Этот факт противоречит линейной теории цветовоспроизведения [1] и является основным источником различий между теоретически прогнозируемыми и реальными цветами автотипного синтеза.

Процедура цветоделения является ключевой при печати, так как результат цветоделения — готовый оттиск — можно получить лишь после создания фотоформ, печатных форм и проведения печати всеми красками. Дефекты цветоделения приводят к неустраняемому браку печатной продукции и, следовательно, к существенным дополнительным затратам.

Важной и актуальной является задача определения количества базовых красок для каждого цвета с учетом нелинейной зависимости спектральных характеристик печатных красок от длины волны излучения и от относительной запечатываемой площади растровой ячейки S . В общем случае спектральные характеристики не имеют аналитического выражения, их получают экспериментально путем измерений и восстанавливают в виде совокупности численных значений. Интерполяция двумерных спектральных характеристик красок является первым этапом решения оптимизационной задачи разложения заданных цветов на базовые составляющие полиграфического синтеза.

2. Постановка задачи

Пусть видимый человеком диапазон излучения разбит на N интервалов. Интенсивность излучения в зависимости от длины его волны — спектр $I(\lambda)$ — можно представить N -мерным вектором I . Фотоны

этого светового потока, падающие на некоторую поверхность, частично поглощаются молекулами этой поверхности (рис. 1), частично отражаются. Отношение отраженного потока к падающему световому потоку, в зависимости от длины волны излучения, характеризуется спектральным коэффициентом отражения поверхности $\rho(\lambda)$. В видимом диапазоне излучения его можно представить N -мерным вектором r .

Таким образом, спектр излучения, отраженного объектом со спектральным коэффициентом отражения r под действием освещения I , можно выразить в виде Lr , где L — диагональная матрица ($N \times N$), образованная элементами вектора I . Для описания цвета используем стандартное цветовое пространство CIEXYZ, определяемое кривыми сложения, которые заданы матрицей A ($N \times 3$) [2], столбцы которой содержат значения стандартных кривых сложения x , следовательно, цветовые CIEXYZ координаты спектра Lr могут быть записаны в виде выражения

$$t = A^T Lr. \quad (1)$$

Пусть имеется некоторое цифровое цветное изображение. Этому изображению соответствует трехмерный массив $I(M_x \times M_y \times 3)$, где M_x — количество пикселей в изображении вдоль одной границы, M_y — количество пикселей вдоль другой границы, 3 — количество координат цвета t каждого из пикселей изображения. Таким образом, информация о цветовом содержании изображения представлена совокупностью CIEXYZ координат t_T .

Пусть необходимо создать точную копию данного изображения с помощью офсетной печати. Цветовые CIEXYZ координаты i -го пиксела в офсетном оттиске определяются выражением:

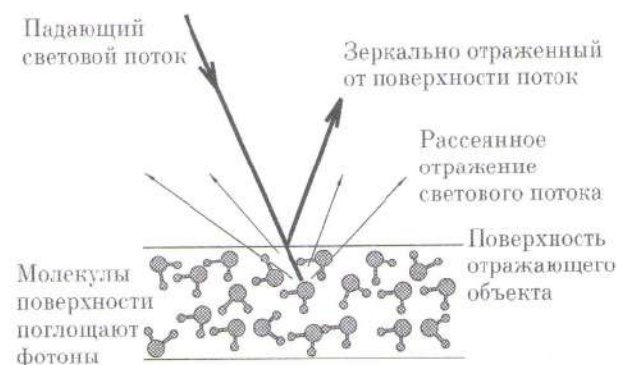


Рис. 1. Ощущение цвета возникает в результате восприятия отраженного светового потока, окрашенного в результирующий цвет вычитанием спектральных составляющих освещения

$$t_{\text{офсет}} = A^T L_1 r_i. \quad (2)$$

Здесь L_1 — матрица, описывающая освещение при наблюдении оттиска; r_i — спектральный коэффициент отражения поверхности оттиска на участке, соответствующем i -му пикселу исходного файла.

В соответствии с линейной теорией цветовоспроизведения, спектральный коэффициент отражения определяется как

$$r_{i \text{ отг}} = b_1 r_C + b_2 r_M + b_3 r_Y, \quad (3)$$

где r_C, r_M, r_Y — спектральные характеристики голубой (*cyan*), пурпурной (*magenta*) и желтой (*yellow*) красок синтеза. Они удерживаются в достаточно узких пределах допусков и считаются стабильными.

Кроме этих трех красок, в печати используется еще и черная (*black*). Технология ввода этой составляющей основана на положении, что ее количество соответствует некоторой линейной комбинации трех базовых цветов [3–7]. Следовательно, формируемый спектр оттиска:

$$r_{i \text{ отг}} = b_1 r_C + b_2 r_M + b_3 r_Y + b_4 r_K, \quad (4)$$

где индекс K соответствует черной краске синтеза, при этом

$$b_4 = \alpha(b_1 + b_2 + b_3). \quad (5)$$

Такая модель предполагает, что краски прозрачны, а функции поглощения красителей не перекрываются и имеют прямоугольную форму. На практике эта теория не обеспечивает приемлемого качества восстановления цвета и не учитывает взаимодействие красок и подложки. Реальные краски не полностью прозрачны, их спектральные характеристики не имеют аналитической записи и в реальности являются линейно зависимыми если не во всем диапазоне видимого спектра, то в отдельных его интервалах.

Полиграфический синтез использует автотипный принцип формирования изображения с помощью растровой точки переменной площади, коэффициент отражения оттиска зависит от запечатываемой площади растровой ячейки S и меняется с изменением длины волны видимого спектра λ . Значения спектрального коэффициента отражения известны лишь в узловых точках, образованных конечными множествами значений переменных S и λ .

Таким образом, формирование офсетного оттиска сводится к определению значений относительных площадей растровых точек основных цветов S_C, S_M, S_Y, S_K , которые могут быть получены в результате решения оптимизационной задачи:

$$S_j = \arg \left(\min_{\lambda} \sum_{j=1}^{M_p} \|t_{T_j} - A^T L_1 r_i(S_i, \lambda)\|^2 \right). \quad (6)$$

где i — номер пиксела или соответствующей ему растровой ячейки оттиска; M_p — количество пикселов в изображении.

Получение искомого решения затруднено, поскольку целевая функция (6) не может быть выражена аналитически, так как спектральный коэффициент отражения $r_i(S_i, \lambda)$ задан табличным способом. Следовательно, поиск решения задачи оптимизации непосредственно связан с интерполяцией поверхности $r_i(S_i, \lambda)$, при этом интерполирующую функцию можно использовать для аналитической записи целевой функции (6).

3. Радиально-базисная искусственная нейронная сеть

Для решения задачи многомерной интерполяции предлагается использовать радиально-базисную искусственную нейронную сеть [8], архитектура которой показана на рис. 2.

Исходной информацией для обучения радиально-базисной сети является набор значений функции двух переменных $f(x)$, заданной в N узлах решетки $f(1), f(2), \dots, f(N)$. Результатом обучения является восстановление отображения $F: R^2 \rightarrow R^1$, удовлетворяющее условиям интерполяции:

$$F(x(k)) = f(k), \quad k = 1, 2, \dots, N. \quad (7)$$

Выход радиально-базисной сети в данном случае будет вычисляться в соответствии с выражением:

$$F(x) = \sum_{k=1}^N w_k \Phi_k(\|x - c_k\|, \sigma_k) = \sum_{k=1}^N w_k \Phi_k(\|x - x(k)\|, \sigma_k), \quad (8)$$

где $\Phi(\bullet)$ — радиально-базисная функция, центрами c_k которой приняты узлы интерполяции $x(k)$; $\sigma_k > 0$ — параметр ширины, определяющий локальную область входного пространства, на которую «реагирует» данная функция. Чаще всего $\Phi(\bullet)$ — это гауссиан вида

$$\begin{aligned} \Phi_k(\|x - c_k\|, \sigma_k) &= e^{-\frac{\|x - c_k\|^2}{\sigma_k^2}} = \\ &= \exp\left(-\sigma_k^{-2} \|x - c_k\|^2\right). \end{aligned} \quad (9)$$

где σ_k выбирается из эмпирических соображений. Для определения синаптических весов w_k необходимо решить систему линейных уравнений

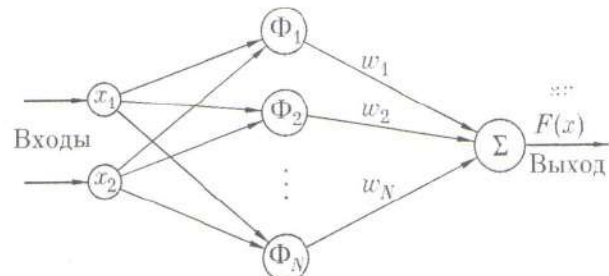


Рис. 2. Архитектура радиально-базисной искусственной нейронной сети

$$\begin{pmatrix} \Phi_{11} & \Phi_{12} & \dots & \Phi_{1N} \\ \Phi_{21} & \Phi_{22} & & \Phi_{2N} \\ \vdots & & & \\ \Phi_{N1} & \Phi_{N2} & & \Phi_{NN} \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} = \begin{pmatrix} f(1) \\ f(2) \\ \vdots \\ f(N) \end{pmatrix}, \quad (10)$$

где $\Phi_{kj}(\|x(j)-x(k)\|, \sigma_k)$, $k=1, \dots, N$; $j=1, \dots, N$. Введя в рассмотрение векторы $f=(f(1), f(2), \dots, f(N))^T$, $w=(w_1, w_2, \dots, w_N)^T$ и $(N \times N)$ матрицу $\Phi=\{\Phi_{kj}\}$ — $(N \times N)$, можно записать систему (8) в матричной форме:

$$\Phi w = f, \quad (11)$$

откуда

$$w = \Phi^+ f, \quad (12)$$

где Φ^+ — матрица, псевдообратная по Муру—Пенроузу к Φ . Таким образом, при использовании радиально-базисной нейронной сети в соответствии с (8) получаем значения функции $F(x)$, интерполирующей неизвестную функцию $f(x)$ в заданных узлах.

4. Экспериментальное исследование

Для исследования интерполирующих свойств радиально-базисной искусственной нейронной сети в задаче интерполяции спектральных коэффициентов отражения красок использованы результаты спектрофотометрических измерений коэффициента отражения полиграфической тестовой шкалы. Эта шкала содержит четыре ряда полей, отпечатанных четырьмя красками синтеза, каждое из полей запечатано растровыми точками разных размеров — от 1 до 100 % заполнения относительной площади ячейки. Полям соответствуют следующие значения относительной запечатанной площади: 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 92%, 94%, 96%, 98%, 100%. Измерения проводятся прибором автоматически в видимом диапазоне излучений, что соответствует длинам волн от 380 до

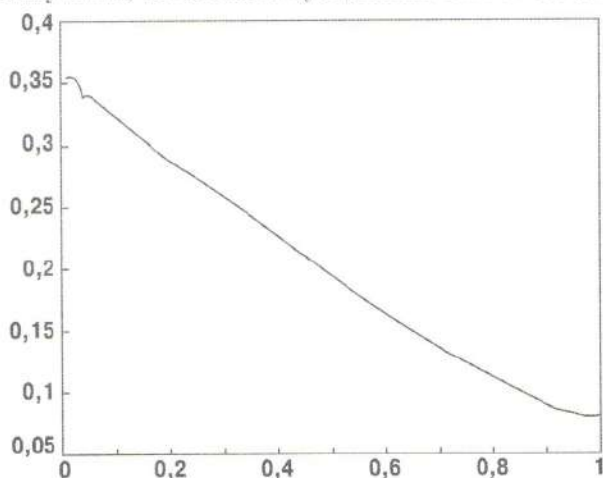


Рис. 3. Интерполяция коэффициента отражения голубой краски в зависимости от относительной площади растровой точки (сплошная линия — исходная функция, пунктирная линия — результат интерполяции) при $\lambda = 380$ нм

730 нм. Этот диапазон разбит на интервалы по 10 нм. Таким образом, решетка интерполяции является прямоугольной, но неравномерной по одной из осей.

Проведена одномерная и двумерная интерполяция коэффициента отражения с помощью радиально-базисной сети. В качестве примера рассмотрены коэффициенты отражения полей шкалы в интервале длин волн 380...390 нм. В качестве радиально-базисной функции был выбран гауссиан (9) с $\sigma = 0,8$. На рис. 3–6 показаны результаты интерполяции коэффициента отражения для разных красок синтеза в узловых точках. Для удобства восприятия результаты отображены в виде одномерных срезов двумерных функций при фиксированных значениях длины волны $\lambda = 380$ нм.

Точность интерполяции в узловых точках решетки оценивалась с помощью среднеквадратичного отклонения. Результаты приведены в таблице.

Точность интерполяции коэффициента отражения красок синтеза

| Краска | Точность интерполяции (среднеквадратичное отклонение в узловых точках) |
|-----------|--|
| Голубая | $3,67 \cdot 10^{-15}$ |
| Пурпурная | $3,09 \cdot 10^{-15}$ |
| Желтая | $3,35 \cdot 10^{-15}$ |
| Черная | $3,46 \cdot 10^{-15}$ |

5. Заключение

Научная новизна данной работы заключается в применении радиально-базисной искусственной нейронной сети для интерполяции двумерной функции спектрального коэффициента отражения при решении насущной задачи подготовки цветных изображений к полиграфической печати — цветоделения. Практическое значение имеет высокая точность

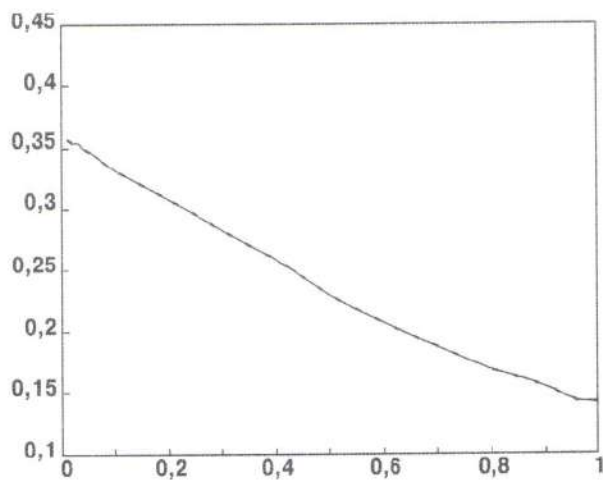


Рис. 4. Интерполяция коэффициента отражения пурпурной краски в зависимости от относительной площади растровой точки (сплошная линия — исходная функция, пунктирная линия — результат интерполяции) при $\lambda = 380$ нм

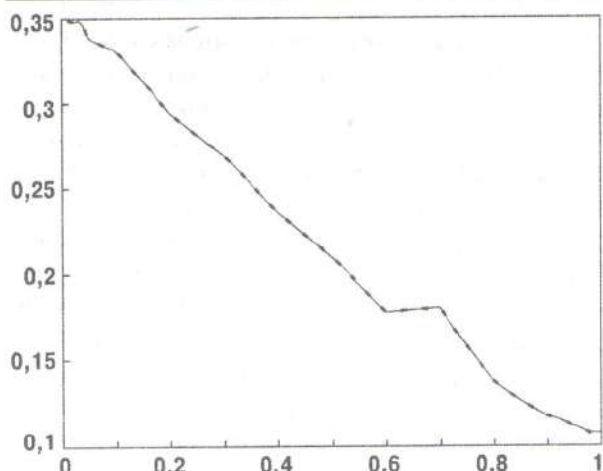


Рис. 5. Интерполяция коэффициента отражения желтой краски в зависимости от относительной площади растровой точки (сплошная линия — исходная функция, пунктирная линия — результат интерполяции) при $\lambda = 380$ нм

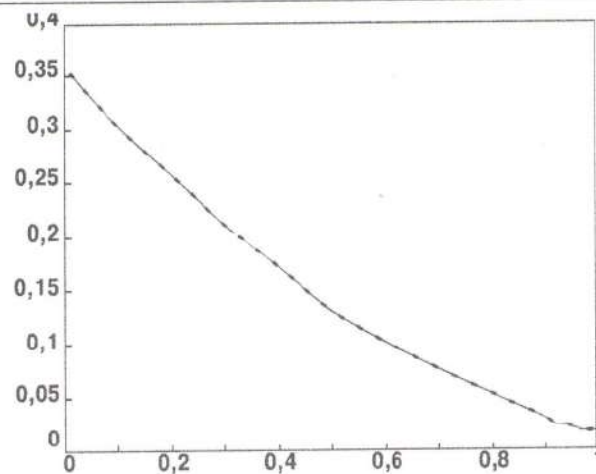


Рис. 6. Интерполяция коэффициента отражения черной краски в зависимости от относительной площади растровой точки (сплошная линия — исходная функция, пунктирная линия — результат интерполяции) при $\lambda = 380$ нм

интерполяции в узловых точках функции, что позволяет в дальнейшем проводить поиск решения задачи оптимизации количества базовых красок синтеза.

Список литературы: 1. Шашлов Б. А. Цвет и цветовоспроизведение.— М.: Книга, 1986.— 208 с. 2. Sharma G. Digital Color Imaging Handbook.— Boca Raton, FL: CRC, 2003. 3. Барановський І., Шаблій І. Про сучасні технології кольороподілу і кастрування // Палітра друку.— 1995.— № 4.— С. 52–53. 4. Барановський І., Шаблій І. Методи виготовлення кольороподілених растрових зображень // Палітра друку.— 1994.— № 1.— С. 19. 5. Berns R. S.,

Billmeyer F. W., Sacher R. S. Methods for generating spectral reflectance functions leading to color-constant properties // Col. Res. Appl.—1985.— Vol. 10.— P. 73–83. 6. Heuberger K. J., Jing Z. M., Persiev S. Color transformations and lookup tables // Proc. TAGA/ISCC.— 1992.— Vol. 2.— P. 863–881. 7. ISO 13655, Graphic technology — Spectral measurements and colorimetric computation for graphic art images. International Organization for Standardization, Geneva, Switzerland. 7. Cichocki A., Unbehauen R. Neural Networks for Optimization and Signal Processing.— Stuttgart: Teubner, 1993.— 526 p.

Поступила в редакцію 26.08.2006

УДК 519.7

О бионике интеллекта / М. Ф. Бондаренко, Ю. П. Шабанов-Кушнаренко // Бионика интеллекта: Научн.-техн. журнал. – 2006. – № 1 (64). – С. 3–5.

В статье рассматривается общая характеристика научного направления «бионика интеллекта». Приведена также характеристика алгебры предикатов как эффективного средства формализации объектов. Исследован подход к переходу на логические сети.

Библиогр.: 3 назв.

УДК 519.7

Про біоніку інтелекту / М. Ф. Бондаренко, Ю. П. Шабанов-Кушнаренко // Біоніка інтелекту: Наук.-техн. журнал. – 2006. – № 1 (64). – С. 3–5.

У статті розглянуто загальну характеристику наукового напрямку «біоніка інтелекту». Наведено також характеристику алгебри предикатів як ефективного засобу формалізації об'єктів. Досліджено підхід до переходу на логічні мережі.

Бібліогр.: 3 найм.

UDC 519.7

About Bionics of Intelligence / M. F. Bondarenko, Yu. P. Shabanov-Kushnarenko // Bionica intellecta: Sci. Mag. – 2006. – No. 1 (64). – P. 3–5.

Scientific direction general characteristic as bionics of intelligence is considered in this article. The algebra of predicates characteristic as effective means of object formalization is given. The transition to logical network is studied.

Ref.: 3 items.

УДК [519.95 + 518.5]: 622.692.4

Интеллектуальная система обнаружения криминальных отборов из продуктопроводов / А. Д. Тевяшев, К. В. Кобылинский // Бионика интеллекта: Научн.-техн. журнал. – 2006. – № 1 (64). – С. 6–18.

Рассматривается интеллектуальная система обнаружения криминальных отборов из продуктопроводов, в основу которой положены интеллектуальные алгоритмы автоматической обработки, анализа и принятия решений по оперативным данным, поступающим из замерных узлов. Описывается математическая модель квазистационарных и нестационарных неизотермических режимов транспорта жидких продуктов по магистральным трубопроводам. Предложен комплекс методов по обнаружению криминальных отборов и оцениванию их параметров в режиме реального времени.

Табл.: 1. Ил.: 4. Библиогр.: 6 назв.

УДК [519.95 + 518.5]: 622.692.4

Интеллектуальна система виявлення кримінальних відборів із продуктопроводів / А. Д. Тевяшев, К. В. Кобылинський // Біоніка інтелекту: Наук.-техн. журнал. – 2006. – № 1 (64). – С. 6–18.

Розглянуто інтелектуальну систему виявлення кримінальних відборів із продуктопроводів, в основу якої покладені інтелектуальні алгоритми автоматичної обробки, аналізу й прийняття рішень за оперативним даними, що надходять із вимірювальних вузлів. Описано математичну модель квазістаціонарних і нестаціонарних неизотермічних режимів транспорту рідких продуктів по магістральних трубопроводах. Запропоновано комплекс методів щодо виявлення кримінальних відборів й оцінювання їхніх параметрів у режимі реального часу.

Табл.: 1. Іл.: 4. Бібліогр.: 6 найм.

UDC [519.95 + 518.5]: 622.692.4

Intelligence system for detection of criminal selections on product pipelines / A. D. Tevjashev, K. V. Kobylinskij // Bionica intellecta: Sci. Mag. – 2006. – No. 1 (64). – P. 6–18.

The work describes an intelligence system for detection of criminal selections on product pipelines. The intelligent algorithms of automatic processing, analysis and decision-making with on-line data coming from measurement points are in base of the system. The mathematical model of the stationary and unstationary unisothermal modes of liquid products transport on main pipelines is described. Several methods of criminal selections detection and evaluation of their parameters in the real-time mode is offered.

Tab.: 1. Fig.: 4. Ref.: 6.

УДК 519.6

Индукционная согласованность отношений в задачах грануляции информации / В. П. Машталир, В. В. Шляхов // Бионика интеллекта: Научн.-техн. журнал. – 2006. – № 1 (64). – С. 19–26.

Обсуждаются свойства мультиотношений, индуцированных на классах эквивалентности произвольными n -арными отношениями. Обоснована целесообразность использования мультиотношений как инструмента моделирования для грануляции информации. Доказаны необходимые и достаточные условия равенства носителей мультиотношений.

Табл.: 2. Ил.: 2. Библиогр.: 14 назв.

УДК 519.6

Индукowana uzгодженість відношень у задачах грануляції інформації / В. П. Машталір, В. В. Шляхов // Біоніка інтелекту: Наук.-техн. журнал. – 2006. – № 1 (64). – С. 19–26.

Обговорюються властивості мультивідношень, що індуються на класах еквівалентності довільними n -арними відношеннями. Обґрунтовано доцільність використання мультивідношень як інструменту моделювання для грануляції інформації. Доведено необхідні та достатні умови рівності носіїв мультивідношень.

Табл.: 2. Іл.: 2. Бібліогр.: 14 найм.

UDC 519.6

An induced agreement of relations in problems of information granulation / V. P. Mashtalir, V. V. Shlyakhov // Bionica intellecta: Sci. Mag. – 2006. – No. 1 (64). – P. 19–26.

Properties of multirelations that are induced on equivalence classes by arbitrary n -arity relations are discussed. Multirelation appropriateness as a simulation tool has been vindicated for granular computing. Necessary and sufficient conditions of multirelation carriers equality have been proved.

Tab.: 2. Fig.: 2. Ref.: 14 items.

УДК 519.7

Модели цветового зрения / М. Ф. Бондаренко, Ю. П. Шабанов-Кушнаренко // Бионика интеллекта: Научн.-техн. журнал. – 2006. – № 1 (64). – С. 27–41.

В статье рассматривается задача моделирования зрительных процессов. Получено обобщение модели цветового зрения на случай произвольных зрительных картин. Исследованы законы Грассмана.

Библиогр.: 30 назв.

УДК 519.7

Моделі кольорового бачення / М. Ф. Бондаренко, Ю. П. Шабанов-Кушнаренко // Біоніка інтелекту: Наук.-техн. журнал. – 2006. – № 1 (64). – С. 27–41.

У статті розглянуто задачу моделювання процесів бачення. Отримано узагальнення моделі кольорового бачення на випадок довільних зорових картин. Досліджено закони Грассмана.

Бібліогр.: 30 найм.

UDC 519.7

Models of chromatic vision / M. F. Bondarenko, Yu. P. Shabanov-Kushnarenko // Bionica intellecta: Sci. Mag. – 2006. – No. 1 (64). – P. 27–41.

Modelling task of stationary and homogeneous process of chromatic vision modeling is considered in this article. The generalization of the chromatic vision model is given. The Grassman laws are studied.

Ref.: 30 items.

УДК 519.7:007.52

Идентификация фаз сна в реальном времени на основе гибридных нейронных сетей / Е. В. Бодянский, Е. А. Винокурова, А. Н. Слипченко, А. В. Шило // Бионика интеллекта: Научн.-техн. журнал. – 2006. – № 1 (64). – С. 42–45.

В статье предложен подход для идентификации фаз сна в автоматическом режиме в реальном времени. Отслеживание изменения свойств исследуемого процесса проводилось на основе нейронной сети с ортогональными функциями активации, а для классификации временного ряда была разработана модель на основе байесовского подхода.

Ил.: 1. Библиогр.: 13 назв.

УДК 519.7:007.52

Идентифікація фаз сну в реальному часі на основі гібридних нейронних мереж / Є. В. Бодяньський, О. А. Винокурова, О. М. Слипченко, О. В. Шило // Біоніка інтелекту: Наук.-техн. журнал. – 2006. – № 1 (64). – С. 42–45.

У статті запропоновано підхід до ідентифікації фаз сну в автоматичному режимі в реальному часі. Відстеження зміни властивостей процесу, що досліджується, проводилося на основі нейронної мережі з ортогональними функціями активації, а для класифікації часового ряду була розроблена модель на основі байєсівського підходу.

Іл.: 1. Бібліогр.: 13 найм.

UDC 519.7:007.52

Sleep phases identification in real time based on the hybrid neural networks / Ye. V. Bodyanskiy, O. A. Vynokurova, O. M. Slipchenko, O. V. Shilo // *Bionica intellecta: Sci. Mag.* – 2006. – No. 1 (64). – P. 42–45.

In this paper the approach to the sleep phases identification in the automatic mode in real time is proposed. Tracing the changes in properties of the process is implemented using neural network with orthogonal activation functions. The model for time series classification based on Bayesian approach is developed.

Fig.: 1. Ref.: 13 items.

УДК 681.518:004.912

Идентификация нештатных ситуаций в информационных сетях / А. Л. Ерохин, А. П. Турута // *Бионика интеллекта: Научн.-техн. журнал.* – 2006. – № 1 (64). – С. 46–55.

Рассмотрено решение задачи создания моделей нештатных ситуаций, связанных с отказами в обслуживании в информационных сетях *web*. Предложенные модели позволяют улучшить качество идентификации нештатных ситуаций путем составления и решения предикатных уравнений.

Табл.: 1. Ил.: 3. Библиогр.: 15 назв.

УДК 681.518:004.912

Идентифікація позаштатних ситуацій в інформаційних мережах / А. Л. Єрохін, О. П. Турута // *Біоніка інтелекту: Наук.-техн. журнал.* – 2006. – № 1 (64). – С. 46–55.

Розглянуто розв'язання задачі створення моделей позаштатних ситуацій, пов'язаних з відмовами в обслуговуванні в інформаційних мережах *web*. Запропоновані моделі дозволяють покращити якість ідентифікації позаштатних ситуацій шляхом складання та розв'язання предикатних рівнянь.

Табл.: 1. Іл.: 3. Бібліогр.: 15 найм.

UDC 681.518:004.912

Identification of supernumerary situations in information networks / A. L. Yerokhin, A. P. Turuta // *Bionica intellecta: Sci. Mag.* – 2006. – No. 1 (64). – P. 46–55.

The decision of models creation problem for the supernumerary situations connected to refusals in service in information networks *web* is considered. The offered models allow to improve quality of supernumerary situations identification by decisions of the predicate equations.

Fig.: 4. Tab.: 1. Ref.: 15 items.

УДК 004.89

Автоматизация принятия решений при разработке межцеховых технологических маршрутов на приборостроительных предприятиях / П. Б. Сироджа, Г. А. Фролова // *Бионика интеллекта: Научн.-техн. журнал.* – 2006. – № 1 (64). – С. 56–58.

Рассмотрены проблемы внедрения информационных технологий на современных приборостроительных предприятиях, выделены требующие автоматизации задачи технологической подготовки производства. Предлагается использовать знаниеориентированные методы для решения слабоструктурированных задач принятия решений в системах автоматизации разработки межцеховых технологических маршрутов.

Ил.: 1, Библиогр.: 14 назв.

УДК 004.89

Автоматизація прийняття рішень при розробці міжцехових технологічних маршрутів на приладобудівних підприємствах / П. Б. Сіроджа, Г. О. Фролова // *Біоніка інтелекту: Наук.-техн. журнал.* – 2006. – № 1 (64). – С. 56–58.

Розглянуто проблеми впровадження інформаційних технологій на сучасних приладобудівних підприємствах, задачі технологічної підготовки виробництва, які потребують автоматизації. Запропоновано використати знанняорієнтовані методи для прийняття рішень у системах автоматизації розробки міжцехових технологічних маршрутів.

Іл.: 1. Бібліогр.: 14 найм.

UDC 004.89

Automation of decisions making at development of intershop technological routes at the instrument-making plants / I. Sirodzha, G. Frolova // *Bionica intellecta: Sci. Mag.* – 2006. – No. 1 (64). – P. 56–58.

In paper the problem implantation of information technologies in modern instrument-making plants. The tasks of technological productions are selected which require automation. It is offered to use knowledge-oriented methods for decisions making of semistructured tasks in systems of automation of development of intershop technological routes.

Fig.: 1. Ref.: 14 items.

УДК 681.82:519.81

Использование алгебры предикатов и предикатных операций для формализации декларативной и процедурной составляющих знаний / З. А. Алисейко, В. И. Булкин, О. В. Канищева, Н. В. Шаронова // Бионика интеллекта: Научн.-техн. журнал.– 2006.– № 1 (64).– С. 59–63.

В статье рассмотрены различные формы представления знаний, показано использование алгебры предикатов и предикатных операций для формализации декларативной и процедурной составляющих знаний. Рассмотрены современные формы представления знаний, проводится их сравнительный анализ, а также подробно рассматриваются возможности алгебры предикатов и предикатных операций для представления знаний.

Ил.: 2. Библиогр.: 5 назв.

УДК 681.82:519.81

Використовування алгебри предикатів та предикатних операцій для формалізації декларативної та процедурної складових знань / З. А. Алісейко, В. І. Булкін, О. В. Канищева, Н. В. Шаронова // Біоніка інтелекту: Наук.-техн. журнал.– 2006.– № 1 (64).– С. 59–63.

У статті розглянуто різні форми подання знань, показано використання алгебри предикатів та предикатних операцій для формалізації декларативної та процедурної складових знань. Розглянуто сучасні форми подання знань, наведено їх порівняльний аналіз, а також докладно розглянуто можливості алгебри предикатів та предикатних операцій для подання знань.

Ил.: 2. Библиогр.: 5 найм.

UDC 681.82:519.81

Use of algebra predicate and predicate operations for formalization declarative and procedure knowledge / Z. A. Aliseyko, V. I. Bulkin, O. V. Kanisheva, N. V. Sharonova // Bionica intellecta: Sci. Mag.– 2006.– No. 1 (64).– P. 59–63.

Different forms for produce knowledge were considered and algebra predicate and predicate operations for formalization declarative and procedure knowledge showed use also in this article. Modern forms for produce knowledge were considered, their comparative analysis cited and possibility predicate and predicate operations for produce knowledge detail were considered.

Fig.: 2. Ref.: 5 items.

УДК 519.62

Обоснование выбора и принципы построения многозначных обратимых логических элементов / Г. Г. Четвериков, М. А. Дмитриева // Бионика интеллекта: Научн.-техн. журнал.– 2006.– № 1 (64).– С. 64–67.

В статье рассматриваются теоретические основы построения многозначных элементов, модулей и структур. Модифицированы методы построения многозначных обратимых неоднородных логических элементов и модулей (АКП-структур первого и второго рода) для языковых систем искусственного интеллекта с обратимыми свойствами. Проведенные исследования аппаратных и программных средств реализации полученных моделей естественного языка и методов синтеза АКП-структур позволили определить рациональные пути для их создания, а также сформулировать критерии выбора АКП-структур для реализации отношений, которые составляют основу действия языковых систем.

Табл.: 3. Ил.: 1. Библиогр.: 5 назв.

УДК 519.62

Обґрунтування вибору та принципи побудовання багатозначних оборотних логічних елементів / Г. Г. Четвериков, М. А. Дмитрієва // Біоніка інтелекту: Наук.-техн. журнал.– 2006.– № 1 (64).– С. 64–67.

У статті розглянуто теоретичні засади побудови багатозначних елементів, модулів та структур. Модифіковано методи побудови багатозначних оборотних неоднорідних логічних елементів та модулів (АСП-структур першого та другого роду) для мовних систем штучного інтелекту з оборотними властивостями. Дослідження апаратних та програмних засобів реалізації отриманих моделей природної мови та методів синтезу АСП-структур дозволили визначити раціональні шляхи їх створення, а також сформулювати критерії вибору АСП-структур для реалізації відношень, що складають основу дії мовних систем.

Табл.: 3. Ил.: 1. Библиогр.: 5 найм.

UDC 519.62

Choosing and principles of multiple-valued invertible logical elements building / G. G. Chetverikov, M. A. Dmitrieva // Bionica intellecta: Sci. Mag.– 2006.– No. 1 (64).– P. 64–67.

In article the theoretical construction principles of multiple-valued elements and structures have been developed. The analysis of their practical application in information system with k-valued coding has been tested. The comparative analysis of parallel methods of linguistics equation system solving which enables to study and compare proposed software and hardware implementation of obtained language models has been carried out. On the basis of general principles and methods of universality hybridism and parallelism of k-valued spatial structures a new class of universal converters –

AFV-structures of the third kind which allows to implement analysis, normalization and synthesis of Ukrainian language morphology problems has been created.

Tab.: 3, Fig.: 1, Ref.: 5 items

УДК 519.7

Разработка правил пополнения иерархической базы знаний / В. М. Левыкин, Т. М. Неофитная // Бионика интеллекта: Научн.-техн. журнал.— 2006.— № 1 (64).— С. 68–71.

В работе рассмотрены семантические закономерности иерархических структур систем знаний предметных областей и предложены правила пополнения иерархических баз знаний автоматизированных систем. Использование данных правил в рамках процедуры создания базы знаний в диалоге с пользователем позволит проводить проверку корректности базы знаний и упростит процесс ее уточнения.

Ил.: 3. Библиогр.: 4 назв.

УДК 519.7

Розробка правил поповнення ієрархічної бази знань / В. М. Левикін, Т. М. Неофітна // Біоніка інтелекту: Наук.-техн. журнал.— 2006.— № 1 (64).— С. 68–71.

У роботі розглянуто семантичні закономірності ієрархічних структур систем знань предметних галузей і запропоновано правила поповнення ієрархічних баз знань автоматизованих систем. Використання цих правил у рамках процедури створення бази знань у діалозі з користувачем дозволить проводити перевірку коректності бази знань і спростить процес її уточнення.

Іл.: 3. Бібліогр.: 4 найм.

UDC 519.7

Development of rules for acquisition of hierarchical knowledge base / V. M. Levikin, T. M. Neophitnaya // Bionica intellecta: Sci. Mag.— 2006.— No. 1 (64).— P. 68–71.

The paper describes semantic regularities typical for hierarchical structures of domain knowledge systems. For expressing these regularities the rules for acquisition of hierarchical knowledge base of automated systems are suggested. Using of these rules in procedure of knowledge base development permits to check of knowledge base correctness and to simplify process of its elaboration.

Figs. 3, Ref.: 4 items.

УДК 658.012.56

Морфологическая модель словоизменения флективного языка и электронный грамматический словарь / Т. П. Любченко // Бионика интеллекта: Научн.-техн. журнал.— 2006.— № 1 (64).— С. 72–77.

Представлена морфологическая модель словоизменения флективного языка, представляющая концептуальную основу для компьютерного моделирования и реализации функции парадигматических отношений. Материал изложен на примере русского языка.

Табл.: 1, Библиогр.: 4 назв.

УДК 658.012.56

Морфологічна модель словозміни флективної мови та електронний граматичний словник / Т. П. Любченко // Біоніка інтелекту: Наук.-техн. журнал.— 2006.— № 1 (64).— С. 72–77.

Подано морфологічну модель словозміни флективної мови, яка є концептуальною основою для комп'ютерного моделювання та реалізації функцій парадигматичних відношень. Матеріал викладено на прикладі російської мови.

Табл.: 1, Бібліогр.: 4 найм.

UDC 658.012.56

The morphological model of word-inflection of flective languages and electronic grammatical dictionary / T. P. Lyubchenko // Bionica intellecta: Sci. Mag.— 2006.— No. 1 (64).— P. 72–77.

The morphological model of word-inflection of inflecting languages is represented. The model is a conceptual basis for computational modelling and realization of the function of paradigmatic relations. The material is developed by the example of Russian.

Tab.: 1, Ref.: 4 items.

УДК 004.934

Некоторые проблемы распознавания речевых образов / М. Ф. Бондаренко, А. В. Работягов, С. В. Щепковский // Бионика интеллекта: Научн.-техн. журнал.— 2006.— № 1 (64).— С. 78–85.

В статье приведены результаты анализа современного состояния исследований в области речевых технологий. Указаны некоторые проблемы спектрального метода распознавания речевых образов.

Библиогр.: 17 назв.

УДК 004.934

Деякі проблеми розпізнавання мовних образів / М. Ф. Бондаренко, А. В. Работягов, С. В. Щепковський // Біоніка інтелекту: Наук.-техн. журнал.– 2006.– № 1 (64).– С. 78–85.

У статті наведено результати аналізу сучасного стану досліджень у галузі мовних технологій. Зазначено деякі проблеми спектрального методу розпізнавання мовних образів.

Бібліогр.: 17 найм.

UDC 004.934

Some problems of vocal patterns recognition / M. Bondarenko, A. Robotyagov, S. Schepkovsky // Bionica intellecta: Sci. Mag.– 2006.– No. 1 (64).– P. 78–85.

In the article the results of analysis of the modern state of researches are resulted in area of vocal technologies. Some problems of spectral method vocal patterns recognition are indicated.

Ref.: 17 items.

УДК 681.327.12

Модифицированный метод потенциальных функций // Е. В. Волченко // Бионика интеллекта: Научн.-техн. журнал.– 2006.– № 1 (64).– С. 86–91.

В статье предложен новый метод уменьшения размера обучающей выборки путем объединения групп объектов в мета-объекты. Описывается модифицированный метод потенциальных функций, использующий обучающую выборку мета-объектов. Приведено численное сравнение результатов работы простого и модифицированного методов потенциальных функций, оценивающее эффективность использования мета-объектов для нахождения решающей функции обучающейся системы распознавания.

Табл.: 8. Ил.: 6. Библиогр.: 11 назв.

УДК 681.327.12

Модифікований метод потенційних функцій // О. В. Волченко // Біоніка інтелекту: Наук.-техн. журнал.– 2006.– № 1 (64).– С. 86–91.

У статті запропоновано новий метод зменшення розміру навчальної вибірки за рахунок об'єднання груп об'єктів у мета-об'єкти. Описується модифікований метод потенційних функцій, який використовує навчальну вибірку мета-об'єктів. Наведено чисельні порівняння результатів роботи звичайного та модифікованого методів потенційних функцій, яке оцінює ефективність використання мета-об'єктів для знаходження розв'язувальної функції системи розпізнавання, що навчається.

Табл.: 8. Іл.: 6. Бібліогр.: 11 найм.

UDC 681.327.12

Modified method of potential functions / E. V. Volchenko // Bionica intellecta: Sci. Mag.– 2006.– No. 1 (64).– P. 86–91.

In given article the new method of reduction of the training sample size is offered by association of objects groups in meta-objects. The modified method of potential functions using training sample of meta-objects is described. Numerical comparison of work results of the simple and modified method of the potential functions, estimating efficiency of use of meta-objects for a finding of decision function of training system of recognition is resulted.

Tab.: 8. Fig.: 6. Ref.: 11 items.

УДК [615.471:616-071]:681.513

Нечеткая кластеризация массивов биомедицинских данных в условиях избыточности информации / Н. А. Тесленко, И. Г. Чурюмова // Бионика интеллекта: Научн.-техн. журнал.– 2006.– № 1 (64).– С. 92–95.

В статье исследованы закономерности формирования мышечного перенапряжения как одного из главных факторов развития пограничных гипертонических состояний. Для решения задачи кластеризации данных использованы два метода: нечеткая искусственная нейронная сеть автоассоциативной памяти «Brain-State-In-A-Box Model» и алгоритм нечеткой кластеризации данных «Fuzzy C-Means Algorithm». Размерность пространства исходных данных была предварительно понижена с помощью автоассоциативной трехслойной нейронной сети типа «Bottle neck» в первом случае и с помощью метода главных компонент — во втором.

Библиогр.: 17 назв.

УДК [615.471:616-071]:681.513

Нечітка кластеризація масивів біомедичних даних в умовах надлишкової інформації / Н. О. Тесленко, І. Г. Чурюмова // Біоніка інтелекту: Наук.-техн. журнал.– 2006.– № 1 (64).– С. 92–95.

У статті досліджено закономірності формування м'язової перенапруги як одного з головних факторів розвитку приміжєвих гіпертонічних станів. Для розв'язання завдання кластеризації даних було використано два методи:

нечітку штучну нейронну мережу автоасоціативної пам'яті «Brain-State-In-A-Box Model» і алгоритм нечіткої кластеризації даних «Fuzzy C-Means Algorithm». Розмірність простору вихідних даних була попередньо знижена за допомогою автоасоціативної тришарової нейронної мережі типу «Bottle neck» у першому випадку та за допомогою методу головних компонент — у другому.

Бібліогр.: 17 найм.

UDC [615.471:616-071]:681.513

Fuzzy clustering of data biomedical sets in conditions of redundant information / N. A. Teslenko, I. G. Churyumova // *Bionika intellecta: Sci. Mag.*— 2006.— No. 1 (64).— P. 92–95.

The rules of muscular overstrain formation as one of the main factors of border-line hypertensive states development were researched in the article. Fuzzy artificial neural network of autoassociative memory «Brain-State-In-A-Box Model» and «Fuzzy C-Means Algorithm» was used for solving the problem of data clusterization. Space dimensionality of original data was reduced previously using autoassociative three-layer neural network «Bottle neck» in the first case and with the help of a principal component analysis in second.

Ref.: 17 items

УДК 681.51:007.5

Исследование детекторов точек интереса при построении локальных признаков изображений / В. А. Гороховатский, С. В. Кузьмин // *Бионика интеллекта: Научн.-техн. журнал.*— 2006.— № 1 (64).— С. 96–100.

Приведены результаты исследования признаков точек интереса в задачах построения локальных характеристик изображений. Анализируются их статистические, инвариантные и вычислительные свойства, а также особенности применения для реальных полутоновых изображений.

Табл.: 2. Ил.: 4. Библиогр.: 6 назв.

УДК 681.51:007.5

Дослідження детекторів точок інтересу при побудові локальних ознак зображень / В. О. Гороховатський, С. В. Кузьмін // *Біоніка інтелекту: Наук.-техн. журнал.*— 2006.— № 1 (64).— С. 96–100.

Наведено результати досліджень ознак точок інтересу в задачах побудови локальних характеристик зображень. Аналізуються їх статистичні, інваріантні та обчислювальні властивості, а також особливості застосування для реальних напівтонових зображень.

Табл.: 2. Іл.: 4. Бібліогр.: 6 найм.

UDC 681.51:007.5

Research of detectors of interest points at construction of local attributes of images / V. A. Gorohovatsky, S. V. Kuzmin // *Bionika intellecta: Sci. Mag.*— 2006.— No. 1 (64).— P. 96–100.

Results of research of attributes of interest points tasks of construction of local characteristics of images are resulted. Their statistical, invariant and computing properties, and also features of application for real half-tone pictures are analyzed.

Tab.: 2. Fig.: 4. Ref.: 6 items.

УДК 004.93.673:61

Об одном подходе к распознаванию ЭКС в реальном масштабе времени / Н. В. Белоус, Е. В. Высоцкая, Т. В. Жемчужкина, А. Н. Ковалев, Г. А. Кобзарь, А. П. Порван // *Бионика интеллекта: Научн.-техн. журнал.*— 2006.— № 1 (64).— С. 101–104.

Рассматривается проблема распознавания ЭКС в реальном масштабе времени. Предлагается подход к распознаванию участков ЭКС, ограниченных RR-интервалом, как единого целого без какого-либо дополнительного деления сигнала. Этот подход отличается от корреляционного тем, что эталоны базы распознавания представляют собой векторы признаков эталонных ЭКС, извлеченных автоматически в соответствии с определенными дескрипторами формы сигнала (CSS и Фурье), что позволяет значительно уменьшить объем базы эталонов и, соответственно, ускорить процесс распознавания.

Ил.: 1. Библиогр.: 6 назв.

УДК 004.93.673:61

Про один підхід до розпізнавання ЕКС у реальному масштабі часу / Н. В. Білоус, О. В. Висоцька, Т. В. Жемчужкіна, А. М. Ковальов, Г. А. Кобзарь, А. П. Порван // *Біоніка інтелекту: Наук.-техн. журнал.*— 2006.— № 1 (64).— С. 101–104.

Розглянуто проблему розпізнавання ЕКС у реальному масштабі часу. Запропоновано підхід до розпізнавання ділянок ЕКС, обмежених RR-інтервалом, як єдиного цілого без будь-якого додаткового ділення сигналу. Цей підхід відрізняється від кореляційного тим, що еталонні бази розпізнавання являють собою вектори ознак еталонних ЕКС.

видлучених автоматично відповідно до певних дескрипторів форми сигналу (CSS і Фур'є), що дозволяє значно зменшити обсяг бази еталонів і, відповідно, прискорити процес розпізнавання.

Л.: 1. Бібліогр.: 6 найм.

UDC 004.93.673:61

About one approach recognition of ECS in the real time / N. V. Belous, E. V. Visotskaya, T. V. Gemchugkina, A. N. Kovalev, G. A. Kobzar, A. P. Porvan // *Bionika intellecta: Sci. Mag.* – 2006. – No. 1 (64). – P. 101–104.

There is considered the problem of recognition ECS in the real time. Approach of recognition of areas ECS limited by the RR interval as a single whole without some additional division signal is offered. This approach differs from correlation to those, that the standards of base of recognition are vector of signs of standard ECS extracted automatically in accordance with the definite descriptors of signal form (CSS and Fure) that allows considerably to decrease the volume of base of standards and accordingly accelerate the process of recognition.

Fig.: 1. Ref.: 6 items.

УДК 681.513: 519.7

Интерполяция коэффициентов отражения красок с помощью радиально-базисной искусственной нейронной сети / Н. Е. Кулишова // *Бионика интеллекта: Науч.-техн. журнал.* – 2006. – № 1 (64). – С. 105–108.

В статье рассмотрена задача интерполяции двумерных спектральных коэффициентов отражения полиграфических красок для повышения точности цветовоспроизведения при печати. Задача решена с помощью радиально-базисной искусственной нейронной сети. Коэффициенты отражения интерполированы в фиксированных точках неравномерной двумерной решетки с заданной точностью.

Табл.: 1. Ил.: 6. Библиогр.: 7 назв.

УДК 681.513: 519.7

Интерполяция коефіцієнтів відбивання фарб за допомогою радіально-базисної штучної нейронної мережі / Н. Є. Кулішова // *Біоніка інтелекту: Наук.-техн. журнал.* – 2006. – № 1 (64). – С. 105–108.

У статті розглянуто задачу інтерполяції двовимірних спектральних коефіцієнтів відбивання поліграфічних фарб для підвищення точності кольоровідтворення під час друку. Задачу розв'язано за допомогою радіально-базисної штучної нейронної мережі. Коефіцієнти відбивання інтерпольовано у фіксованих точках нерівномірної двовимірної решітки із заданою точністю.

Табл.: 1. Іл.: 6. Бібліогр.: 7 найм.

UDC 681.513: 519.7

Interpolation of printing paints reflections by using of radial-basis artificial neural net / N. Ye. Kulishova // *Bionika intellecta: Sci. Mag.* – 2006. – No. 1 (64). – P. 109–112.

The interpolation problem of twodimensional spectral reflections of printing paints for increasing of color reproduction accuracy in printing is considered in this article. The problem is solved by using of radial-basis artificial neural net. Spectral reflections are interpolated in fixed points of irregular two-dimensional grid with given accuracy.

Tab.: 1. Fig.: 6. Ref.: 7 items.

СВЕДЕНИЯ ОБ АВТОРАХ

| | |
|-----------------------------------|--|
| Алисейко Зоя Анатольевна | аспирант кафедры автоматизированных систем управления, НТУ «ХПИ» |
| Белоус Наталья Валентиновна | канд. техн. наук, профессор кафедры программного обеспечения ЭВМ, ХНУРЭ |
| Бодянский Евгений Владимирович | д-р техн. наук, профессор кафедры искусственного интеллекта, научн. руководитель ПНИЛ АСУ, ХНУРЭ |
| Бондаренко Михаил Федорович | д-р техн. наук, профессор, ректор ХНУРЭ |
| Булкин Виталий Иванович | канд. техн. наук, доцент кафедры прикладной математики и информационных технологий Макеевского экономико-гуманитарного института |
| Винокурова Елена Анатольевна | канд. техн. наук, ст. науч. сотр. ПНИЛ АСУ, ХНУРЭ |
| Волченко Елена Владимировна | ассистент кафедры программного обеспечения интеллектуальных систем ДГИИИ |
| Высоцкая Елена Владимировна | канд. техн. наук, доцент кафедры биомедицинских электронных устройств и систем ХНУРЭ |
| Гореховатский Владимир Алексеевич | канд. техн. наук, доцент кафедры информатики ХНУРЭ |
| Дмитриева Мария Александровна | студентка гр. ПОАС-01-1, ХНУРЭ |
| Ерохин Андрей Леопольдович | канд. техн. наук, доцент, начальник кафедры информатики Харьковского национального университета внутренних дел |
| Жемчужкина Татьяна Владимировна | канд. техн. наук, ассистент кафедры биомедицинских электронных устройств и систем ХНУРЭ |
| Кавищева Ольга Валерьевна | аспирант кафедры автоматизированных систем управления, НТУ «ХПИ» |
| Кобзарь Глеб Адольфович | аспирант кафедры программного обеспечения ЭВМ, ХНУРЭ |
| Кобылинский Константин Валерьевич | мл. науч. сотр. кафедры прикладной математики, ХНУРЭ |
| Ковалев Андрей Николаевич | аспирант кафедры программного обеспечения ЭВМ, ХНУРЭ |
| Кузьмин Сергей Владимирович | аспирант кафедры информатики, ХНУРЭ |
| Кулишова Нонна Евгеньевна | канд. техн. наук, доцент кафедры инженерной и компьютерной графики, ХНУРЭ |
| Левыкин Виктор Макарович | д-р техн. наук, профессор кафедры информационно управляющих систем, ХНУРЭ |
| Любченко Татьяна Петровна | мл. науч. сотр., Украинский языково-информационный фонд НАН Украины |
| Машталир Владимир Петрович | д-р техн. наук, профессор, декан факультета компьютерных наук, ХНУРЭ |
| Неофитная Татьяна Михайловна | науч. сотр. НИИ лазерных технологий |
| Порван Андрей Павлович | инженер кафедры биомедицинских электронных устройств и систем, ХНУРЭ |
| Работягов Андрей Валентинович | канд. техн. наук, науч. сотр. кафедры программного обеспечения ЭВМ, ХНУРЭ |
| Сиродка Игорь Борисович | д-р техн. наук, профессор, Харьковский Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ» |
| Слипченко Александр Николаевич | канд. техн. наук, ст. науч. сотр. ПНИЛ АСУ, ХНУРЭ |
| Тевяшев Андрей Дмитриевич | д-р техн. наук, профессор, заведующий кафедрой прикладной математики, ХНУРЭ |
| Тесленко Наталья Александровна | аспирант кафедры искусственного интеллекта, ХНУРЭ |
| Турута Алексей Петрович | преподаватель кафедры информатики Харьковского национального университета внутренних дел |
| Фролова Галина Александровна | студентка гр. 667-п, Харьковский национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ» |
| Четвериков Григорий Григорьевич | д-р техн. наук, профессор кафедры программного обеспечения ЭВМ, ХНУРЭ |
| Чурюмова Ирина Геннадиевна | аспирант кафедры биомедицинских электронных устройств и систем ХНУРЭ |
| Шабанов-Кушнарченко Юрий Петрович | д-р техн. наук, профессор кафедры программного обеспечения ЭВМ, ХНУРЭ |
| Шаронова Наталья Валериевна | д-р техн. наук, профессор кафедры автоматизированных систем управления, НТУ «ХПИ» |
| Шило Александр Владимирович | канд. биол. наук, ст. науч. сотр. Института проблем криобиологии и криомедицины НАН Украины |
| Шляхов Владислав Викторович | канд. техн. наук, доцент, вед. науч. сотр. кафедры информатики, ХНУРЭ |
| Щепковский Сергей Вадимович | ведущий инженер кафедры программного обеспечения ЭВМ, ХНУРЭ |

ПРАВИЛА оформления рукописей для авторов научно-технического журнала «БИОНИКА ИНТЕЛЛЕКТА»

Статьи оформляются в редакторе Microsoft Word. Формат страницы — А4 (210x297 мм), поля: верхнее — 25 мм, нижнее — 20 мм, левое, правое — 17 мм. Количество колонок — 2, интервал между ними — 5 мм, основной шрифт Times New Roman, кегль основного текста — 10 пунктов, междустрочный интервал — множитель (1,1), абзацный отступ — 6 мм. Объем рукописи — от 4 до 12 листов (языки: русский, украинский, английский).

Согласно Постановлению ВАК Украины от 15.01.2003 № 7-05/1 (Бюллетень ВАК, № 1, 2003, с. 2) статья должна иметь следующие необходимые элементы: постановку проблемы в общем виде и ее связь с важными научными или практическими заданиями; анализ последних исследований, публикаций и выделение нерешенных ранее частей общей проблемы в данной области; формулирование целей и задач исследования; изложение основного материала исследований с полным обоснованием полученных научных результатов; выводы по данному исследованию и перспективы дальнейших исследований в данном направлении.

Структура рукописи: УДК, инициалы и фамилии авторов, заголовок, основной текст, список литературы.

Заголовок статьи — кегль 12.

ОБРАЗЕЦ ОФОРМЛЕНИЯ

УДК 621.396.961.4

И. О. Фамилия, И. О. Фамилия

НАЗВАНИЕ РУКОПИСИ

Статья должна начинаться с введения и заканчиваться заключением.

Основной текст разделяется на 2 и более подразделов с подзаголовками, пронумерованными арабскими цифрами, выделенными жирным шрифтом, отступы: сверху — 6 пт, снизу — 3 пт.

1. Название раздела

Определения выделяются в тексте курсивом.

Рисунки и таблицы (черно-белые, контрастные) помещаются в текст после первой ссылки в виде встроенных объектов и отдельно нумеруются арабскими цифрами сквозной нумерацией, при наличии более одного объекта. Все надписи в рисунках и таблицах должны быть выполнены шрифтом Times New Roman, кегль в рисунках — 10, в таблицах — 9.

Рисунок содержит подрисуючную центрированную подпись (вне рисунка), кегль 9, по центру, отступы сверху и снизу по 6 пт. Ширина рисунка должна соответствовать ширине колонки (или ширине страницы).

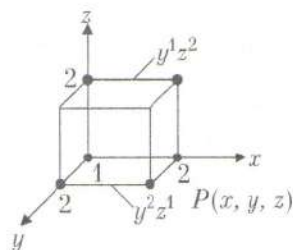


Рис. 1. График трехместного бинарного предиката

Табличный заголовок располагается справа над таблицей (кегль 9).

Формулы, символы, переменные, должны быть набраны в редакторе формул MathType (Equation). Формулы располагаются по

центру и нумеруются при наличии ссылок на них в рукописи. Шрифт — Times New Roman. Высота переменной — 10 пунктов, индексов и подиндексов — 8 пт, основной мат. символ — 12 (10) пт. Курсив в стиле формул включают только для текста и переменных. Сокращения русских слов выполняются прямым начертанием. Переменные, содержащиеся в тексте, также набираются в редакторе формул.

$$F(X_1, \dots, X_n) = \vee F(\delta_1, \dots, \delta_n) x_1^{\delta_1} \dots x_n^{\delta_n}. \quad (1)$$

Список литературы включает опубликованные источники, на которые имеются ссылки в тексте, заключенные в квадратные скобки [1], печатается без абзацного отступа, кегль 9 пунктов, отступ сверху 6 пт.

Список литературы: 1. Фамилия И. О. Название книги. — Город: Издательство, 1900. — 000 с. 2. Фамилия И. О. Название статьи // Название журнала. Название серии. — 1997. — Т. 00, № 00. — С. 00–00.

Поступила в редакцию 00.00.2005

Рефераты на трех языках: русском, украинском и английском (Times New Roman, кегль — 9 пунктов).

УДК 000.000.00

Название статьи / И. О. Фамилия, И. О. Фамилия // Бionika интеллекта: Научн.-техн. журнал. — 2006. — № 00. — С. 00–00.

Текст реферата (3–4 предложения).

Табл. 00. Ил. 00. Библиогр.: 00 назв.

УДК 000.000.00

Назва статті / І. Б. Прізвище, І. Б. Прізвище // Біоніка інтелекту: Наук.-техн. журнал. — 2006. — № 00. — С. 00–00.

Текст реферату.

Табл. 00. Іл. 00. Бібліогр.: 00 найм.

UDC 000.000.00

Title of paper / Initials, Surname // Bionika intellecta: Sci. Mag. — 2006. — No. 00. — P. 000-000.

Text.

Tab. 00. Fig. 00. Ref.: 00 items.

Сведения об авторах: фамилия, имя, отчество, ученая степень, звание, должность, место работы, адрес, контактные телефоны.

Подача материалов: рукопись, рефераты, сведения об авторах — на новой дискете 3,5 дюйма. Твердая копия материалов в одном экземпляре: рукопись, рефераты, сведения об авторах, рецензия, подписанная доктором наук, акт экспертизы, заявление на имя главного редактора со сведениями об авторах.

Материалы, не соответствующие требованиям, редакцией не рассматриваются.

Адрес редакции: Украина, 61726, Харьков, пр. Ленина 14, ХНУРЭ, к. 139, тел. 702-13-07, e-mail: innov@kture.kharkov.ua

ТЕМАТИКА ЖУРНАЛА:

- алгебро-логические системы
- базы знаний
- генная инженерия
- игры и машинное творчество
- иерархические и динамические системы
- интеллектуальные агенты
- интеллектуальные роботы
- интеллектуальные сети
- лингвистические технологии
- математическая психология
- машинное обучение и эволюция
- методология информатизации
- мозгоподобные ЭВМ
- нейроматематика
- объектное моделирование
- онтологические системы
- параллельная обработка информации
- поиск информации
- понимание
- принятие решений
- распознавание и синтез речи
- распознавание образов
- теория интеллекта
- философия информатизации
- чувственное восприятие
- экспертные системы
- электронная нервная система

Адрес редакции:

Украина, 61166, Харьков-166, пр. Ленина, 14, к. 139
Харьковский национальный университет радиоэлектроники (ХНУРЭ),
тел. 702-13-07,
e-mail: innov@kture.kharkov.ua

Наукове видання

БІОНІКА ІНТЕЛЕКТУ

Науково-технічний журнал

Російською мовою

Науковий редактор — *Ю. П. Шабанов-Кушмаренко*

Редактор — *Л. В. Ситнік*

Відповідальний за випуск — *Г. Г. Четвериков*

Комп'ютерна верстка — *І. Г. П'ятигорська*

Рекомендовано Вченою Радою
Харківського національного університету радіоелектроніки
(протокол № 27 від 24.11.06)

Підписано до друку 24.11.2006. Формат 60 × 84 $\frac{1}{8}$. Друк ризографічний.
Папір офсетний. Гарнітура New Standard. Умов. друк. арк. 13,95. Обл.-вид. арк. 16,7. Тираж 300 прим. Зам. № 52

ТОВ «Компанія СМІТ»
61166, м. Харків, просп. Леніна, 14
Тел.: 8 (057) 717-54-94
Факс: 8 (057) 702-13-07
E-mail: book@smit.com.ua

Свідоцтво про внесення суб'єкта видавничої справи
до Державного реєстру видавців, виготівників
і розповсюджувачів видавничої продукції
ДК № 26.04.2001

Друк — ФОП Васильєва Н. В.
м. Харків, просп. Леніна, 14
тел. 8 (057) 702-13-07